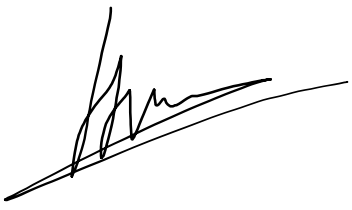
Genetics in the diagnosis and management of thyroid cancer

Submitted by Jonathan Fussey to the University of Exeter as a thesis for the degree of Doctor of Medicine in August 2022

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Signature: 

ABSTRACT

Background

Thyroid cancer is the commonest endocrine malignancy, constituting around 1% of all cancers. Medullary thyroid carcinoma (MTC) is usually sporadic, but may occur in patients with germline mutations in the rearranged during transfection (*RET)* gene. The sporadic form lacks germline mutations but may harbour somatic *RET* mutations. The risk factors for differentiated thyroid carcinoma (DTC) are not well understood, and genetic predisposition accounts for a smaller proportion of DTC than MTC.

Methods

Presenting features of patients undergoing germline *RET* testing were analysed in order to explore the prevalence of germline *RET* mutations in patients with different clinical presentations in the UK population, and a systematic review of the literature was undertaken to investigate the clinical usefulness of somatic mutations in patients with sporadic MTC. Mendelian randomisation was utilised to investigate the causal roles of several proposed modifiable risk factors for thyroid cancer.

Results

By analysing results of 1,058 patients' germline *RET* analysis, I found that 8.5% of patients with presumed sporadic MTC in fact have hereditary disease. The

systematic review on molecular genetics in sporadic MTC highlights the emerging role of somatic mutations and epigenetic markers in prognostication in sporadic MTC. Regarding risk factors for DTC, my Mendelian randomisation studies identified a causal association between lower thyroid stimulating hormone (TSH) levels and DTC, and a possible causal relationship between type 2 diabetes mellitus (T2DM) and DTC. I was however unable to provide any evidence using genetic epidemiological techniques for a causal role for obesity, smoking, alcohol consumption or physical exercise.

Conclusions

This thesis helps to illustrate patterns of clinical presentations of patients with germline *RET* mutations in the United Kingdom, and reports the novel use of mendelian randomisation to investigate the role of a range of lifestyle risk factors on the risk of thyroid cancer, raising some questions about the validity of previously reported observational findings, and paving the way for further research with the vital aim of understanding the modifiable risk factors for thyroid cancer.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

metabolic disease can only be tested in this context with genetic variants. 1-sample MR was not performed for T2DM due to the difficulty interpreting results using binary variables).

Chapter 4 Figure 3: Forest plot showing observed, 1 sample MR and 2 sample MR (instrumental variable analysis) associations between BMI, WHR, adjWHR, favourable adiposity, T2DM and thyroid cancer. Odds ratios and 95% confidence intervals are shown. (There are no observational estimates for favourable adiposity because this phenotype of higher adiposity and lower risk of metabolic disease can only be tested in this context with genetic variants. 1-sample MR was not performed for T2DM due to the difficulty interpreting results using binary variables).

Chapter 5 Figure 1: Basic concept of Mendelian randomisation: if high TSH truly causes thyroid cancer, then the genetic variants for TSH should also be associated with thyroid cancer. Unlike measured TSH, these genetic variants are not susceptible to confounding by other risk factors or reverse causation (dashed lines) as they are assigned at conception. Horizontal pleiotropy (dotted line), whereby the variants of interest also affect the outcome via their association with another trait is a possible limitation of MR.

Chapter 5 Figure 2: Forest plot showing two-sample MR (instrumental variable analysis) associations between genetically instrumented TSH and thyroid cancer and benign nodular thyroid disease. Odds ratios of outcome per standard deviation higher genetically instrumented TSH, and 95% confidence intervals are shown.

Chapter 5 Figure 3: Forest plot showing associations between genetically instrumented TSH and thyroid cancer and benign nodular thyroid disease. Odds ratios of outcome per standard deviation higher genetically instrumented TSH, and 95% confidence intervals are shown.

Chapter 5 Figure 4: Forest plot showing associations between quartiles for TSH genetic risk and thyroid cancer and benign nodular thyroid disease. Odds ratios and 95% confidence intervals are shown for each quartile as compared to those in the lowest quartile.

Chapter 6 Figure 1: Forest plot showing the odds of benign nodular thyroid disease per 1 standard deviation increase in physical activity on observational

analysis, and using weak instrument MR with inverse variance weighted (IVW), Egger and MR RAPS analyses

AUTHOR'S DECLARATION

I confirm that the work contained in this thesis is my own, but acknowledge the great support of my supervisors and co-authors throughout my research. Chapters 2, 3, 4 and 5 represent the findings of work which has undergone scientific peer-review during the course of my studies and has been published. Chapter 6 represents work which is currently undergoing peer review for publication. At the beginning of each chapter, I have listed my co-authors for the work contained in that chapter. I am first and corresponding author for every piece of work contained herein.

ACKNOWLEDGMENTS

Jessica Tyrrell	*Supervisor, and senior lecturer in genetics, University of Exeter*
Bijay Vaidya	*Supervisor, and consultant endocrinologist & honorary clinical professor, Royal Devon & Exeter Hospital and University of Exeter*
Joel Smith	*Supervisor, and consultant head and neck surgeon, Royal Devon & Exeter Hospital*
Sian Ellard	*Supervisor, and Professor of Genomic Medicine, University of Exeter*
Robin Beaumont	*Research fellow, genetics of complex traits, University of Exeter*
Andrew Wood	*Lecturer in genetics, University of Exeter*
Samuel Jones	*Research Fellow in molecular medicine, University of Helsinki*
Martina Owens	*Clinical Scientist, department of molecular genetics, Royal Devon & Exeter Hospital*
Ruth Cleaver	Consultant *Clinical Geneticist, Royal Devon & Exeter Hospital*
Christopher Bowles	*Clinical Scientist, department of molecular genetics Royal Devon & Exeter Hospital*

LIST OF ABBREVIATIONS

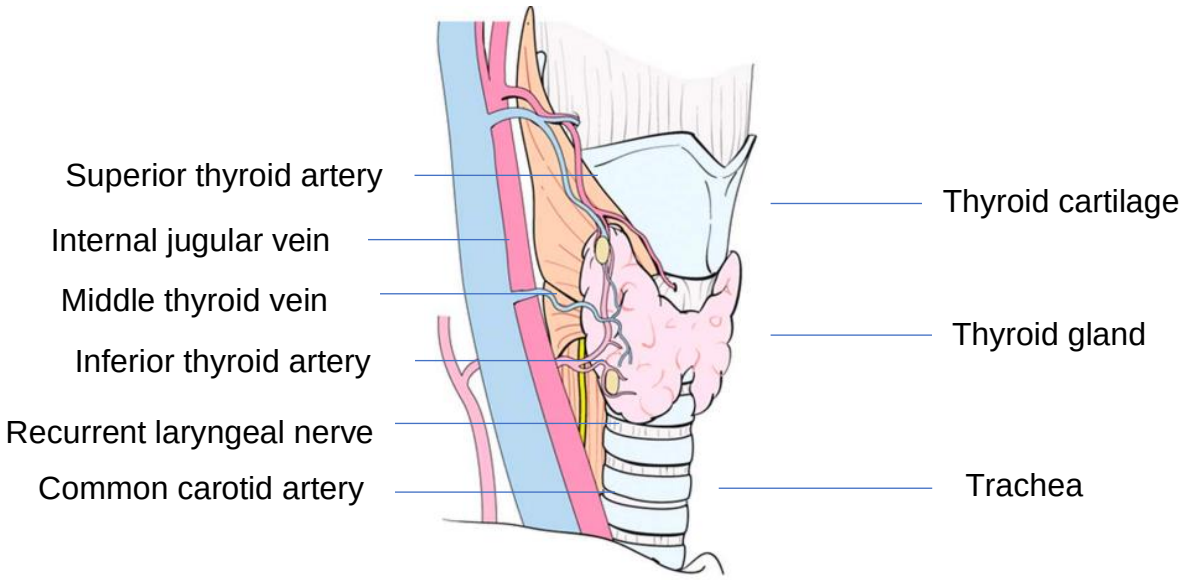
AUS	Atypia of uncertain significance
CCND	Central compartment neck dissection
DTC	Differentiated thyroid carcinoma
FDA	Food and Drug Administration
FLUS	Follicular lesion of uncertain significance
FMTC	Familial medullary thyroid carcinoma
FNAB	Fine needle aspiration biopsy
FTC	Follicular thyroid carcinoma
hMTC	Hereditary medullary thyroid carcinoma
IGF-1	Insulin like growth factor 1
LD	Linkage disequilibrium
MTC	Medullary thyroid carcinoma
sMTC	Sporadic medullary thyroid carcinoma
MEN2A	Multiple endocrine neoplasia 2A
MEN2B	Multiple endocrine neoplasia 2B
PRS	Polygenic risk score
PTC	Papillary thyroid carcinoma
SNP	Single nucleotide polymorphism
T2DM	Type 2 diabetes mellitus
T3	Triiodothyronine
T4	Tetraiodothyronine (thyroxine)
TG	Thyroglobulin
TKI	Tyrosine Kinase inhibitor
TPO	Thyroperoxidase
TSH	Thyroid stimulating hormone
UK	United Kingdom
US	Ultrasound

CHAPTER 1:

INTRODUCTION

1.1 THYROID STRUCTURE AND FUNCTION

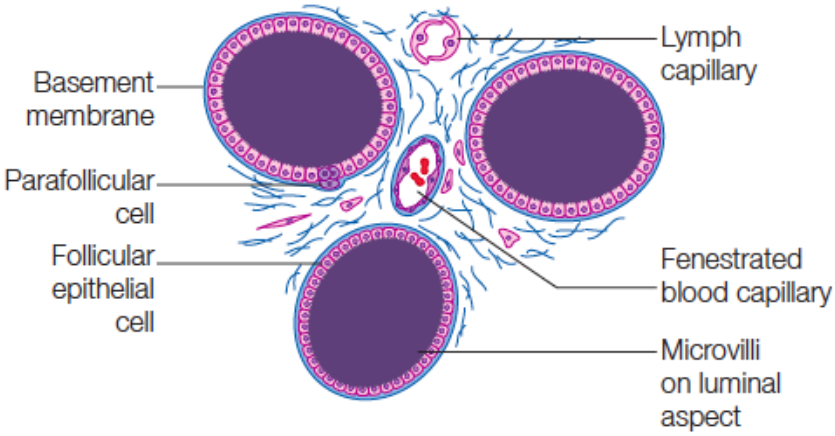
The thyroid gland is an endocrine organ located in the anterior neck. It is formed of two lobes either side of the trachea, linked by an isthmus overlying the second to fourth tracheal rings. As well as the trachea, it is closely related to the oesophagus, parathyroid glands, common carotid arteries, internal jugular veins and recurrent laryngeal nerves, making it a challenging site for surgery (McGlashan 2018; Mohebati & Shaha 2012) (Figure 1).



Chapter 1 Figure1: Thyroid gland and its anatomical relations (Mohebati & Shaha 2012)

Embryologically, the thyroid develops from between the first and second pharyngeal pouches at around the 3rd week of gestation, and over the following 4 weeks, it descends from the foramen caecum at the base of the tongue to its final position in the neck. During this process, the ultimobranchial bodies from the 4th and 5th pharyngeal pouches contribute the calcitonin-secreting parafollicular C cells (Policeni *et al.* 2012).

Microscopically, the thyroid gland is formed of multiple lobules, themselves composed of follicles. These follicles are formed of follicular epithelial cells arranged in a spherical pattern around colloid (Figure 2). Scattered throughout the stoma separating these follicles are parafollicular C cells, which are stimulated to secrete calcitonin by elevated serum calcium levels, or gastrin (McGlashan 2018).



Chapter 1 Figure 2: Microscopic structure of the thyroid gland showing follicular cells arranged around central colloid (McGlashan 2018).

The primary function of the thyroid gland is the production and secretion of triiodothyronine (T3) and tetraiodothyronine (T4), in response to thyroid stimulating hormone (TSH) secreted by the anterior pituitary. This process relies upon the absorption of dietary iodine, which is transported to the thyroid follicles as iodide. This is then combined with thyroglobulin via thyroperoxidase to form mono-iodothyronine and di-iodothyronine which can be combined to form T3 and T4. These are then secreted into the circulation where they are bound to binding proteins. Deiodination in peripheral tissues converts T4 to the more physiologically potent T3. The process of thyroid hormone production and release is dependent on sufficient dietary iodine, as well as the hypothalamo-pituitary-thyroid axis, whereby thyrotropin releasing hormone (TRH) is released

from the hypothalamus and triggers the release of TSH from the pituitary which then stimulates T3 and T4 production and release, with negative feedback control (Weickert 2018). Thyroid hormones have wide ranging effects on almost all tissues in the body, including increasing the metabolic rate and tissue oxygen consumption, increasing carbohydrate absorption and metabolism, increasing heart rate and contractility, and increasing fat breakdown (Weickert 2018).

Calcitonin is secreted by the parafollicular C cells of the thyroid in response to high serum calcium levels. The physiological role of calcitonin has long been debated, with some authors questioning its importance based on the fact that patients who have undergone thyroidectomy and thus have low calcitonin seem to have no ill-effects. However, there is evidence that calcitonin regulates bone turnover as well as calcium homeostasis (Davey & Finlay 2013). Importantly in the context of thyroid cancer, calcitonin is secreted by medullary thyroid carcinoma cells and is therefore a reliable biomarker for diagnosis and monitoring following treatment (Perros *et al.* 2014).

1.2 THYROID NODULES AND CANCER

Thyroid nodules, defined as “discrete lesions within the thyroid gland, radiologically distinct from surrounding thyroid parenchyma” (Cooper *et al.* 2009) are a common clinical entity. Palpable nodules are present in 2-6% of the population (Dean & Gharib 2008), but can be identified in up to 68% using ultrasound (US) imaging (Guth *et al.* 2009). Thyroid nodules are four times commoner in women than in men (Mazzaferri 1993), and their prevalence

increases with increasing age, and iodine deficiency (Popoveniuc & Jonklaas 2012). The sex hormones oestrogen and progesterone are thought to be responsible for the sex discrepancy, as evidenced by increased new nodule formation and growth of existing nodules during pregnancy (Kung *et al.* 2002). In men, there is no evidence that oestrogen or testosterone levels affect nodule formation or growth, but thyroid nodules have been shown to be associated with lower levels of sex hormone binding globulin (Chen *et al.* 2017). Ionising radiation is also a well-recognised risk factor for thyroid nodules, conferring a 2% increased risk per year (Popveniuc & Jonklaas 2012).

The majority of thyroid nodules are benign and represent dominant nodules within a multinodular goitre (Mazzaferri 1993). These require no management unless associated with deranged thyroid function, cosmetic concerns or compression symptoms in the case of very large nodules. However, 5-10% of thyroid nodules represent malignant tumours (Hegedüs 2004; Perros *et al.* 2014). Thyroid cancer can broadly be categorised as differentiated thyroid carcinoma (DTC) which arises in the follicular cells and includes papillary and follicular subtypes, and medullary thyroid carcinoma (MTC) which arises in the parafollicular C cells. Anaplastic thyroid carcinoma makes up only around 2% of thyroid cancers, and is characterised by rapid growth and local invasion with dismal prognosis. The median survival is 3-10 months and 1-year survival rate is 20% (Smallrige & Copeland 2010). Rarer forms of thyroid tumour include thyroid lymphoma and metastasis from distant sites. The focus of this thesis will be DTC and MTC.

1.3 DIFFERENTIATED THYROID CARCINOMA

Differentiated thyroid carcinoma (DTC) makes up the majority of all thyroid cancers, and is itself divided into papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC). The presenting features normally consist of a neck lump, either representing a thyroid mass if in the central neck, or a lymph node metastasis in the lateral neck. Concerning features at presentation include the presence of lymphadenopathy, and voice change suggesting infiltration of the recurrent laryngeal nerve. PTC has a good prognosis, with a 10-year overall survival of 97% (Ito *et al*. 2018). Age has a significant impact on survival, with 10-year survival approaching 99% in those under 45 years of age (Shah 2015). PTC spreads via lymphatics with cervical lymph node metastasis in up to 40-90% of cases (Lundgren *et al.* 2006). Management depends on stage of disease at presentation but usually consists of surgery in the form of partial or total thyroidectomy with or without post-operative radioactive iodine ablation and long-term TSH suppression with thyroxine supplementation (Perros *et al.* 2014). Targeted therapy with tyrosine kinase inhibitors (TKI) is reserved for the treatment of advanced metastatic DTC (Haugen 2017). The term papillary thyroid microcarcinoma is used to describe PTCs less than 1cm in diameter. They have a very good prognosis indeed and there is some evidence that those identified incidentally on histological examination of thyroidectomy specimens behave more favourably than those identified radiologically (Mehanna *et al.* 2014). In the case of the former, no further treatment is required (Perros *et al*. 2014). In the latter group individual risk factors and adverse features are taken into account and the management options include observation with active surveillance, or surgery. The rate of tumour growth over a 10-year period of active

surveillance has been reported as 8%, and the rate of new lymph node metastasis over the same period 3.8% (Ito *et al.* 2014).

1.4 MEDULLARY THYROID CARCINOMA

Medullary thyroid carcinoma (MTC) is a neural crest-derived malignant tumour arising from the parafollicular C cells of the thyroid. Although it represents a distinct clinical entity, the initial presentation is often indistinguishable from DTC, with a new neck mass either in the thyroid or a cervical lymph node. It is much rarer than DTC, making up only around 2% of all thyroid cancers (Wells *et al.* 2015). MTC is sporadic in the majority of cases, but in 20-25% it occurs in a hereditary form either alone when it is known as familial MTC (FMTC), or in association with other extrathyroidal manifestations of the multiple endocrine neoplasia syndromes types 2A and 2B (MEN2A and MEN2B) (Pelizzo *et al.* 2007). MEN2A comprises of MTC with or without phaeochromocytoma and primary hyperparathyroidism, whereas MEN2B is characterised by aggressive form of MTC (often diagnosed in childhood) with or without phaeochromocytoma along with typical physical features such as marfanoid body habitus, mucosal neuromata, corneal nerve thickening and gastrointestinal ganglioneuromata.

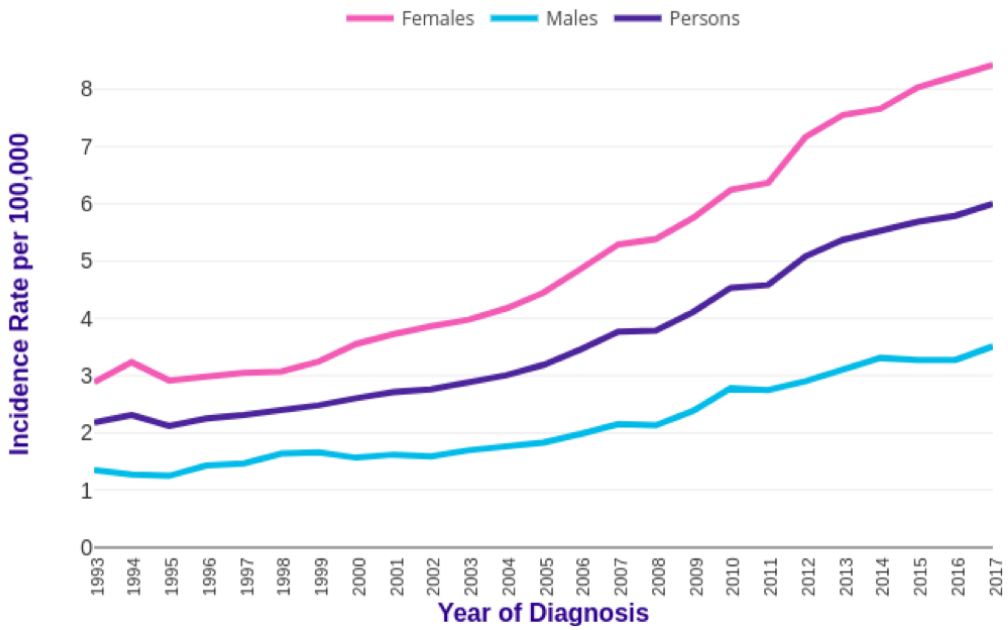
The REarranged during Transfection (*RET)* gene was first described in 1985 (Takahashi *et al.* 1985), and soon after its discovery, *RET* pathogenic variants were identified in almost all patients with hereditary MTC (Donis-Keller *et al.* 1993). Located on chromosome 10q11.2, *RET* encodes a tyrosine kinase involved in the development of the parathyroid glands, adrenal medulla, parafollicular C cells and enteric ganglia, thus accounting for the clinical

manifestations of the MEN2 syndromes. Over recent decades, germline analysis of the *RET* gene has become a routine investigation in the work-up of newly diagnosed MTC and is recommended in all patients with MTC, in order to identify those patients at risk of developing other MEN2-associated conditions as well as those whose family members are also at risk (Wells *et al.* 2015; Perros *et al.* 2014; Elisei *et al.* 2012). By definition, patients with sporadic MTC lack germline pathogenic variants in *RET*, although up to 66% of sporadic tumours are found to have somatic *RET* variants (Romei *et al.* 1996; Marsh *et al.* 1996). There is increasing evidence that the presence of somatic *RET* mutations in sporadic MTC predicts poorer prognosis (Ciampi *et al.* 2019), and this information may be useful in risk-stratifying patients and guiding targeted therapy, although somatic *RET* analysis is not currently recommended by international guidelines.

The management of both hereditary and sporadic MTC is surgical, in the form of thyroidectomy with or without cervical lymph node dissection. Iodine is not taken up by the parafollicular C cells of the thyroid, so radioactive iodine ablation is not effective as in DTC, and radiotherapy and TSH suppression are similarly ineffective (Wells *et al.* 2015). Novel targeted therapies such as tyrosine kinase inhibitors have been shown to increase progression-free survival, but are not effective as curative treatments (Wells *et al.* 2012; Elisei *et al.* 2013), and selective *RET* inhibitors such as selpercatinib are in the relatively early stages of clinical validation (Wirth *et al.* 2020). In the case of sporadic MTC, there is significant variation in opinion as to the recommended extent of first surgery (Fussey *et al.* 2020), and this is at least in part due to the difficulty in accurately stratifying the risk of aggressive disease.

1.5 Changing epidemiology of thyroid cancer

Recent decades have seen a significant increase in the incidence of thyroid cancer (Wiltshire *et al.* 2016), and the most recent Cancer Research UK data reports an incidence of 8.42 per 100,000 women and 3.51 per 100,000 men (Cancer Research UK 2017, Figure 3).



Chapter 1 figure 3: Incidence of thyroid cancer in the UK 1993-2017 (Cancer Research UK 2017)

The incidence is expected to continue to rise, by up to 74% by 2035 (Smittenaar *et al.* 2016), and this change primarily affects differentiated thyroid cancer, with rates of MTC remaining relatively stable. There are several proposed mechanisms for this increase in incidence, however it is largely explained by the increasing accuracy of imaging techniques, and the increasing use of cross-sectional imaging for other reasons, which identify incidental thyroid nodules that would otherwise have gone undiagnosed. This is supported by the fact that

mortality rates from thyroid cancer are stable (Davies & Welch 2014), and the findings of several authors that the detection rate of smaller tumours is rising disproportionately (Davies *et al.* 2010; Husson *et al.* 2013).

However, other authors argue that the indolent nature of DTC means that a true increase in incidence would not be reflected in mortality rates for several decades (Vigneri *et al.* 2020), and point to recent data suggesting that in fact, incidence-based mortality is increasing, and the increase in incidence is not limited to small tumours that previously would have gone undetected (Yan *et al.* 2020). This suggests the possibility of a true increase in both thyroid cancer incidence and aggressiveness, contrary to the case in most other cancers, and raises the question of what may be driving such a change.

1.6 Risk factors for thyroid cancer

Genetic risk factors for both DTC and MTC exist, with a more important effect in the latter, due to the prevalence of germline pathogenic variants in the *RET* gene in those with MTC as discussed above. A hereditary predisposition to DTC may be associated with rare familial cancer syndromes, for example familial adenomatous polyposis coli (Plail *et al.* 1987), Cowden Syndrome (Nelen *et al.* 1997) and Werner Syndrome (Goto *et al.* 1996). Where there is no syndromic diagnosis, epidemiological studies have demonstrated the role of family history in the development of DTC (Oakley *et al.* 2013; Fallah *et al.* 2013) with a three-to-five-fold increased risk for first degree relatives of a family member affected with DTC, and an increased rate of PTC in consanguineous families (Zayed *et al.* 2021). it is estimated that non-syndromic familial DTC accounts for up to 5% of

all DTC (Vriens *et al.* 2009), and large genome wide association studies (GWAS) have identified susceptibility loci for DTC in European populations (Köhler *et al.* 2013; Gudmundsson *et al.* 2017). There is however relatively little known about modifiable, or lifestyle-associated risk factors for thyroid cancer.

A well-established link exists between ionizing radiation and DTC, with evidence form large population-based studies following the use of the atomic bombs in Japan, and the nuclear disaster at Chernobyl (Nikiforov *et al.* 2006; Furukawa *et al.* 2013), and the effect being greatest in those exposed as children or adolescents (Ron *et al.* 1995).

1.6.1 The effect of sex on thyroid cancer risk

Thyroid cancer in general, as well as benign nodular thyroid disease is commoner in women than men, although this difference is less marked in medullary and anaplastic carcinoma (Gilliland *et al.* 1997). Papillary carcinoma is three times more common in women than men, with the maximum disparity being attributable to women of child-bearing age (Rahbari *et al.* 2010). Furthermore, although women tend to be younger at diagnosis of thyroid cancer, men have a poorer prognosis with a higher mortality and lower disease-free survival (Kilfoy *et al.* 2009). The risk of malignancy in thyroid nodules is higher in men (Boelaert *et al.* 2006; Belfiore *et al.* 1992; Fighera *et al.* 2015), with an odds ratio for malignancy in men with a nodule reported to be up to 3.45 (Witczak *et al.* 2016). It is known that oestrogen can alter thyroid cancer cell proliferation via oestrogen receptors (ER) alpha and beta, which are expressed by thyroid cells (Chen *et al.* 2008). It has therefore been proposed that the sex disparity in DTC incidence is due to the effects of female sex hormones on thyroid cells. The evidence for this however is

limited, and whilst sex hormones may explain the fact that nodules are commoner in women, the explanation for the higher rate of malignancy in men is not well understood.

The link between sex hormone exposure and thyroid cancer risk has been extensively studied. Weak associations between late menarche, menstrual cycle longer than 30 days, older age at menopause, prolonged breastfeeding, parity and advanced age at first pregnancy, and DTC have been described (Negri *et al.* 1999; Horn-Ross *et al.* 2011; Cao *et al.* 2015; Yi *et al.* 2016; Zhu *et al.* 2016). Furthermore, an increased risk of DTC in the first 5 years after a pregnancy has been reported by prospective cohort studies and meta-analyses (Horn-Ross *et al.* 2011). The evidence on oral contraceptive use is controversial however, with conflicting findings from meta-analyses (Cao *et al.* 2015; Wu *et al.* 2015).

Although most research into hormonal risk factors for DTC has focussed on oestrogen, there is some evidence that women treated for infertility with progesterone or clomiphene (an oestrogen receptor inhibitor) have an increased risk of DTC (Hannibal *et al.* 2007; Yu *et al.* 2018). This suggests that if hormonal factors are responsible for the sex disparity in DTC incidence, they may be more complex than simply exposure to oestrogen. The relationship between sex hormones and DTC in men has been less well studied, however an increased prevalence of nodules in men with lower sex hormone binding globulin levels, but no significant association with testosterone or oestrogen levels has been reported (Chen *et al.* 2017).

1.6.2 OBESITY AS A RISK FACTOR FOR THYROID CANCER

Obesity, which is closely associated with the components of the metabolic syndrome (reduced HDL-cholesterol, raised triglycerides, blood pressure and fasting plasma glucose) has long been known to be a risk factor for cardiovascular disease and type 2 diabetes mellitus (T2DM) (Han 2016). It has also been linked with many cancer types (Basen-Engquist & Chang 2011), with some estimates suggesting that it is responsible for 20% of all cancers (Wolin *et al.* 2010).

Case control studies have shown a positive association between obesity and DTC in women, although the link was initially less clear in men (Dal Maso *et al.* 2000; Guignard *et al.* 2007). However, a pooled analysis of prospective studies containing over 800,000 participants has provided support for the hypothesis that obesity is an independent risk factor for DTC in both men and women (Kitahara *et al.* 2011). Two meta-analyses investigating obesity and thyroid cancer have both demonstrated significant positive associations in both men and women for differentiated and anaplastic cancers, but an inverse association with medullary thyroid carcinoma (Ma *et al.* 2015; Schmid *et al.* 2015). In addition to the positive association between body weight and thyroid cancer, there is some evidence from retrospective studies that high BMI is associated with more aggressive features of PTC including larger tumour size, extrathyroidal extension, more advanced tumour stage and persistent disease following treatment (Kim *et al.* 2013; Trésallet *et al.* 2014; Choi *et al.* 2015). One prospective study measuring BMI at the time of cytological analysis of thyroid nodules found an inverse relationship between obesity and malignant features on cytology, and no

relationship when considering a smaller subset of patients with surgical histology available (Rotondi *et al.* 2016).

There is some evidence that obese patients who undergo bariatric surgery have a lower overall cancer risk than obese controls, although this effect has only been demonstrated in female patients (Tee *et al.* 2013; Bruno & Berger 2020). This data comes primarily from retrospective population-based studies and thus it is difficult to comment on the mechanisms for the observed reduced cancer risk. In the case of thyroid cancer, it is relevant that in addition to reducing the incidence of T2DM, bariatric surgery also appears to result in a reduction in serum TSH levels (Juiz-Valiña *et al.* 2019), both of which have been implicated in DTC risk.

The biological mechanisms underlying this link are still being elucidated, however it is now clear that the role of human adipose tissue is far more complex than energy storage and insulation. Many hormones, adipokines and cytokines are released from adipose tissue, which play important roles in metabolism, angiogenesis and inflammation. Obesity dysregulates these processes and leads to insulin resistance, glucose intolerance, oxidative stress and chronic systemic inflammation, all of which have been implicated in carcinogenesis (Booth *et al.* 2015).

The systemic inflammation seen in obesity points to a possibility that inflammatory mediators are responsible for the observed increased risk of thyroid cancer in obese patients. Of particular interest amongst the adipose-tissue proteins is adiponectin, which has insulin-sensitising and anti-inflammatory properties, and which is paradoxically inversely associated with obesity

(Dalamaga *et al.* 2012). Adiponectin is inversely associated with other cancers including endometrial and post-menopausal breast cancer (Cust *et al.* 2007; Tworoger *et al.* 2007), and a large recent European prospective study has also identified an inverse relationship between serum levels and thyroid cancer risk in women (Dossus *et al.* 2018).

1.6.3 Lifestyle factors and thyroid cancer

Unlike many other cancers, there is little known about lifestyle-related risk factors for thyroid cancer. Smoking is an important risk factor for many cancers, and many researchers have investigated the link between smoking and thyroid cancer. However, large studies and meta-analyses have consistently shown an inverse relationship between smoking status and thyroid cancer risk (Cho & Kim 2014; Yeo *et al.* 2020). Proposed mechanisms for this finding include the suppressive effect of cigarette smoking on TSH and oestrogen, and the fact that smokers have a lower BMI than non-smokers (Piirtola *et al.* 2018).

The effect of alcohol consumption on thyroid cancer has also been well investigated by large observational studies, and once again the relationship has been found to be an inverse association (Yeo *et al.* 2020). In fact, there appears to be a linear dose-response relationship, with reducing risk of thyroid cancer as frequency and volume of alcohol consumed increases (Yeo *et al.* 2020; Kitihara *et al.* 2012). Once again, alcohol affects TSH levels by depressing the hypothalamic-pituitary-thyroid axis, and this has been proposed as a possible explanation for the findings.

Diet and exercise are difficult risk factors to investigate using observational study designs due to the wide variation in definitions of what constitutes sufficient exercise and a healthy diet, and the difficulty in objectively measuring them. A meta-analysis investigating the role of physical activity on thyroid cancer risk found no evidence of an association (Schmid 2013), however many dietary factors have been reported to influence thyroid cancer risk. Diets rich in cruciferous vegetables, seafood, dairy products and iodine have been found to reduce the risk of thyroid cancer (Choi & Kim 2014; Dal Maso *et al. 2009*; Fiore *et al.* 2020), whilst starch-rich diets have been reported to increase the risk (Choi & Kim 2014). The effect of iodine intake on thyroid cancer risk is by no means clear, in contrast with the well-established link between iodine deficiency and benign nodular thyroid disease (Zimmerman & Boelaert 2015). Animal studies have demonstrated a weak initiator effect and a strong promoter effect of low iodine on thyroid cancer development, suggesting that it may potentiate the effects of other carcinogens or promote carcinogenesis by increasing TSH levels (Kanno *et al.* 1992). However, the same authors also reported a weaker promoter effect of excess iodine on thyroid cancer development. Population studies comparing thyroid cancer incidence before and after salt iodination have reported a reduction in thyroid cancer incidence mainly affecting the anaplastic and follicular subtypes following salt iodination (Zimmerman & Galetti 2015). Conflicting findings have been reported for the effect of caffeine consumption on thyroid cancer risk, although a meta-analysis found no overall effect. (Han & Kim 2017).

The circadian rhythm that controls the sleep/wake cycle is known to also affect thyroid hormone production (Ikegami *et al.* 2019). Disruption of circadian rhythm,

for example by shift-working has been found to be associated with many health conditions including obesity, cardiovascular disease, T2DM and several cancers (Buxton *et al.* 2012; Kettner *et al.* 2014). Altered expression of the genes that control circadian rhythm has been demonstrated in thyroid cancer cells, but not in normal thyroid tissue or benign nodules (Mannic *et al.* 2013). The causal association between sleep cycle disturbance and thyroid cancer is not clear, and is likely to be further confounded by the other health conditions discussed in this section which are associated with both altered circadian rhythm and thyroid cancer.

1.6.4 Type 2 diabetes and thyroid cancer

There is a clear association between T2DM and overall cancer risk, which cannot be explained simply by the fact that both conditions are common (Oberaigner *et al.* 2014; Vigneri *et al.* 2009). The reasons for this association are poorly understood but some authors have proposed common risk factors for both cancer and diabetes including obesity, and sex hormone dysregulation (Giovannucci *et al.* 2010; Oberaigner *et al.* 2014). The incidence of T2DM is increasing globally in parallel with the incidence of thyroid cancer, which has led many consider a possible association between the two diseases. Indeed, there is evidence from meta-analyses for a positive association, which is stronger in women (Li *et al.* 2017; Yeo *et al.* 2014). The authors suggest an important role for the insulin-like growth factor pathway. Insulin like growth factor 1 (IGF-1) receptors are overexpressed on thyroid cancer cells and may be activated by chronically elevated levels of circulating insulin leading to cell proliferation (Yeo *et al.* 2014). This theory is supported by the finding that thyroid nodules are associated with

insulin resistance (Tang *et al.* 2017). Another possible mechanism is direct cellular oxidative stress caused by increased glucose concentration.

Somewhat complicating the link between T2DM and thyroid cancer is the finding that metformin has been found to inhibit thyroid cancer cell proliferation and trigger apoptosis *in vitro* and in animal models (Cho *et al.* 2014). Furthermore, a large cohort study in humans has suggested a protective role of metformin on thyroid cancer incidence (Tseng 2014), and a small study found a protective effect of metformin on thyroid cancer in diabetic patients in terms of complete response and progression-free survival (Klubo-Gwiezdzinska *et al.* 2013). If this represents a true association, then the widespread use of metformin in the management of T2DM may mask the true relationship between diabetes and thyroid cancer in cohort and case-control studies.

1.6.5 The effect of thyroid stimulating hormone on thyroid cancer risk

The association between TSH and the risk of thyroid cancer remains controversial, with some animal evidence of a role in tumorigenesis (Franco *et al.* 2011). A positive association has been reported in case control and cross-sectional studies (Ye *et al.* 2013; Chiu *et al.* 2012), with some authors identifying raised TSH as a predictor of thyroid cancer in patients with nodular thyroid disease (Golbert *et al.* 2017; Zafon *et al.* 2015; Boelaert *et al.* 2006; Fighera *et al.* 2015; Krátký *et al.* 2018). A meta-analysis of 28 studies concluded that high TSH is associated with an increased risk of thyroid cancer (McCleod *et al*. 2012), however, several case-control studies have found an inverse relationship between TSH levels and thyroid cancer risk (Rinaldi *et al*. 2014; Huang *et al.* 2017), and thyroid cancer is known to occur in a wide range of TSH levels,

including in patients with completely suppressed TSH (Satta *et al.* 1993), which casts doubt on a causative relationship. Furthermore, an important genome-wide association study of 3,001 thyroid cancer patients revealed five genetic variants associated with both low TSH levels and thyroid cancer risk, which lends some weight to the hypothesis of an inverse association (Gudmundsson *et al.* 2012).

The role of thyroid hormones is equally controversial. There is some evidence from *in vitro* studies of a tumour-promoting effect of thyroxine on cancer cells from a number of different tissues (Moeller & Führer 2013; Moriggi *et al.* 2011), and epidemiological studies have linked raised tri-iodothyronine (T3) levels to prostate cancer (Lehrer *et al.* 2002). However, the epidemiological evidence linking thyroid hormone levels to thyroid cancer risk is conflicting, with some authors reporting that lower hormone levels are associated with a higher risk of thyroid cancer (Huang *et al.* 2017; Gul *et al.* 2010; Jonklaas *et al.* 2008), and others reporting no association (Rinaldi 2014).

There is a large body of observational evidence linking autoimmune thyroid diseases to the development of DTC, with reported increased risk of DTC in the presence of both Hashimoto's thyroiditis and Graves' disease (Ye *et al. 2013;* Fiore *et al.* 2011; Staniforth *et al.* 2016). It appears that the relationship is more complex than simply the deranged TSH levels in patients with these conditions, as there is also evidence of a link between thyroid autoantibodies and DTC (Lun *et al.* 2013). However, the immunological mechanisms underpinning these associations are not well understood, and there is conflicting evidence on the relationship between autoimmune thyroid disease and DTC prognosis. Patients with Hashimoto's thyroiditis and DTC have been found to have a better prognosis

than patients without autoimmune thyroid disease (Lun *et al.* 2013; Jeong *et al.* 2012), possibly due to anti-cancer activity of Th17 cells which play a key role in autoimmunity (Zhou *et al.* 2010). The prognosis of patients with DTC and Graves' disease on the other hand, appears to be poorer than that of euthyroid patients with DTC, although the extent to which this is due to thyroid hormone derangements as opposed to autoimmune factors is not clear (Pellegriti *et al.* 2013). There is a well-documented link between both anti-thyroperoxidase (anti-TPO) and anti-thyroglobulin (anti-TG) antibodies and DTC incidence, with some evidence for their effect on prognosis in the form of an increased risk of cervical nodal metastasis but a reduced rate of distant metastasis (Shen *et al.* 2017; Wu *et al.* 2014).

Despite the evidence from case-control and cohort studies for a link between obesity, T2DM, TSH and thyroid cancer, there is significant potential for bias with these study designs due to the overlap between the three risk factors: obese patients are often screened for thyroid function more than the general population and tend to have higher TSH levels; obese patients are more likely to have T2DM; and patients with T2DM are more likely to have hypothyroidism (Tamez-Perez *et al.* 2012; Chaker *et al.* 2016). Finally, leptin is known to stimulate TSH release (Radwanska 2014), and leptin is inhibited by metformin (Nar & Gedik 2009). Thus, investigating the causal effects of each of these risk factors on thyroid cancer using epidemiological or observational studies is fraught with risk of confounding.

1.7 Diagnosis and diagnostic challenges in thyroid cancer

Thyroid cancer usually presents as a thyroid nodule, and the diagnostic work-up includes ultrasonography to characterise the nodule itself and identify any abnormal cervical lymph nodes, plus fine needle aspiration biopsy (FNAB) for cytological assessment (Perros *et al.* 2014). Nodules can be classified into five groups according to their ultrasonographic appearances (U1-U5) ranging from normal thyroid parenchyma at U1, benign nodules at U2, indeterminate nodules at U3, suspicious nodules at U4, to malignant nodules at U5, with good inter-observer agreement and sensitivity (Weller *et al.* 2020). Current guidelines in the UK recommend FNAB for nodules classified as U3 or higher on ultrasound (Perros *et al.* 2014) and the cytological findings can also be classified according to the THY system, used in the UK, or the Bethesda system, used in the US as outlined in Table 1 (Cross *et al.* 2016; Cibas & Ali 2009). The Thy3 category is divided into Thy3a denoting samples with nuclear or architectural atypia (Bethesda diagnostic category III), and Thy3f suggesting evidence of follicular neoplasm (Bethesda diagnostic category IV). The usefulness of FNAB in the diagnosis of MTC has been reported to be less than 50% (Trimboli *et al.* 2015), although this can be significantly improved by performing immunohistochemical analysis and calcitonin measurement on FNAB samples (Trimboli *et al.* 2014).

Thy classification	Bethesda classification	Description
Thy1	I – Non-diagnostic	Non-diagnostic
Thy2	II – Benign	Benign
Thy3a	III – AUS or FLUS	Features of nuclear or architectural atypia
Thy3f	IV – Follicular or Hurthle cell neoplasm	Follicular neoplasm
Thy4	V – Suspect for malignancy	Suspicious for malignancy
Thy5	VI - Malignant	Malignant

Chapter 1 Table 1: Cytological classification of thyroid nodules (AUS = Atypia of uncertain significance; FLUS = Follicular lesion of uncertain significance)

1.7.1 DTC

The indeterminate thyroid nodule poses a significant diagnostic challenge, and accounts for almost 20% of nodules (Bongiovanni *et al.* 2012). Nodules falling into the Thy3 (Bethesda diagnostic categories III and IV) category cannot be classified as benign or malignant without diagnostic surgery in the form of thyroid lobectomy. This carries not only the risks of surgery under general anaesthesia but also the risk of recurrent laryngeal nerve injury resulting in vocal cord paralysis. Between 15-30% of nodules with indeterminate cytology are in fact malignant neoplasms (Bongiovanni *et al.* 2012) which means that the majority of these operations serve no therapeutic purpose whatsoever. Furthermore, some patients who undergo diagnostic thyroid lobectomy for indeterminate nodules are found to have thyroid cancers with adverse features which would have been better treated by a total thyroidectomy, and therefore require completion surgery. In addition to these clinical costs of the indeterminate nodule, are the financial and resource costs placed upon the healthcare system of repeat scans and biopsies and in some cases, multiple operations, in the context of a steady rise in incidence of thyroid nodules. There is therefore a real need for a better system for risk stratifying indeterminate thyroid nodules to determine which ones are

likely to be benign and require surveillance only, and which ones are likely to require surgical treatment. This is an area of interest for researchers using molecular genetic markers which can be identified on cytological specimens (Livhits *et al.* 2020) and in the peripheral blood, in the form of cell-free DNA albeit with somewhat conflicting reports of efficacy. (Thakur *et al.* 2019)

Micropapillary carcinoma, defined as a PTC measuring 10mm or less in diameter, pose a further challenge in the workup and management of DTC. Their incidence is increasing significantly, partly due to increased sensitivity of modern imaging techniques, but also due to an increase in identification on pathological assessment of excised thyroid tissue for benign diseases (Zalzali *et al.* 2019). The outcomes for patients with micropapillary PTC are generally very good, with overall mortality, and distant metastasis rates of less than 1%, and locoregional recurrence rates of around 2.5% (Roti *et al.* 2008). Because of these favourable outcomes, less aggressive management has been proposed namely thyroid lobectomy without postoperative radioactive iodine ablation or TSH suppression (Perros *et al.* 2014), with some authors advocating active surveillance rather than surgery (Brito *et al.* 2016). However, a significant proportion of micropapillary carcinomas are known to be higher risk, with a propensity for local and distant metastasis, and recurrence following surgery (Page *et al.* 2009). Those advocating more conservative management emphasise the importance of assessing pre-operative risk factors such as previous neck irradiation, multifocal disease and family history of thyroid cancer, however a significant proportion of those identified as low risk pre-operatively will be found to have high risk features on post-operative pathological assessment (Gao *et al.* 2019). It is therefore

difficult with currently available approaches to risk-stratify micropapillary carcinomas prior to treatment.

1.7.2 MTC

Like DTC, MTC commonly presents as a thyroid nodule or neck mass and the initial work-up is as for any thyroid nodule, including ultrasonography of the thyroid and neck, and FNAB. If there is a suspicion of MTC based on cytological findings, then calcitonin assays can be performed on FNAB samples, however the diagnosis is often not confirmed until definitive pathology is available following diagnostic hemithyroidectomy or total thyroidectomy. The use of serum calcitonin assays to screen patients with thyroid nodules is a controversial issue, with the British Thyroid Association advising against it, and the American Thyroid Association neither advocating it or advising against it (Perros *et al.* 2014; Wells *et al.* 2015). Regardless of international guidelines, it is performed in many centres and has been shown to be highly sensitive and specific with a sensitivity and specificity of 99% (Vardarli *et al.* 2021). However, the low prevalence of MTC amongst patients with thyroid nodules means that the positive predictive value of serum calcitonin has been reported to be as low as 7.7%, and thus its routine use may result in unnecessary thyroid surgery for many patients with false positive results (Verbeek *et al.* 2021). Once the diagnosis is confirmed, genetic analysis of the *RET* proto-oncogene is indicated in order to identify familial cases (Wells *et al.* 2015). This is important not only in order to direct investigations to identify other manifestations of the MEN2 syndromes, but also to offer genetic counselling and predictive testing for unaffected family members, and to provide information on the likely disease phenotype, which is useful in treatment planning (Wells *et al.* 2015). Even patients with apparently sporadic MTC and no family

history should be offered *RET* testing, as up to 14.9% of patients with presumed sporadic MTC are found to harbour pathogenic variants (Kihara *et al.* 2016), and up to 75% of MEN2B cases arise in patients with *de novo* p.M918T variants (Wells *et al.* 2015).

Genetic analysis of the *RET* proto-oncogene is performed on germline DNA extracted from peripheral blood leukocytes and is targeted to the exons in which pathogenic variants are known to occur. The extent of screening of the *RET* proto-oncogene has increased in recent decades, as new variants have been identified in patients with MTC (Fazioli *et al.* 2008; Castellone *et al.* 2010; Latteyer *et al.* 2016). Despite this, there remains a subgroup of patients with MTC who have either a family history of MEN2-associated conditions and/or a personal history of other MEN2-associated conditions and yet in whom no germline *RET* pathogenic variant is identified (Smith *et al.* 2016). This raises the possibility of as yet undiscovered *RET* variants, or variants in other genes for example *ESR2* being responsible for familial MTC (Smith *et al.* 2016).

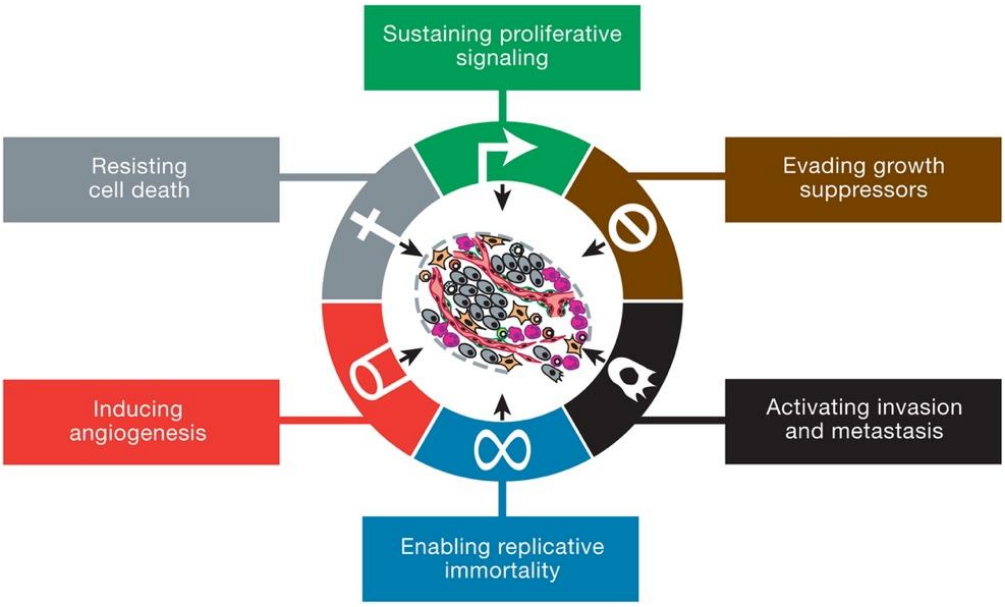
Somatic *RET* variants are identified in the tumour DNA of up to 66% of patients with sporadic MTC who by definition lack germline *RET* pathogenic variants (Romei *et al.* 1996; Marsh *et al.* 1996). The usefulness of these somatic variants to predict disease behaviour has been extensively studied, with some authors suggesting that the somatic p.M918T variant which is associated with MEN2B when present in the germline DNA, predicts more aggressive disease and a worse outcome in sporadic disease (Moura *et al.* 2009; Schilling *et al.* 2001; Elisei *et al.* 2008; Ciampi *et al.* 2019). Furthermore, the somatic mutational profile of sporadic MTC may have an effect on the efficacy of multi-kinase inhibitor therapy

in patients with advanced disease, with randomised controlled trial evidence suggesting that those with somatic *RET* M918T mutations benefit from longer progression-free survival than those without when treated with carbozantinib (Sherman *et al.* 2016). Despite these potential uses of somatic *RET* testing in sporadic MTC, there is no consensus between international guidelines on its use in clinical practice, and is not routinely used in the UK outside of the research setting (Wells *et al.* 2015; Elisei *et al.* 2012).

The lack of a readily available and evidence-based risk stratification tool for sporadic MTC results in significant difficulty in planning surgical management. Although tumour size, preoperative serum calcitonin levels and certain somatic mutations have been reported to predict more aggressive disease and a higher likelihood of lymph node metastasis (Machens & Dralle 2010; Fan *et al.* 2018), rates of occult nodal metastasis of over 25% have been reported for patients with low preoperative serum calcitonin and tumours less than 2cm in diameter (Ito *et al.* 2018), suggesting that these factors alone are not sufficient to accurately risk stratify patients.

1.8 Cancer genetics

Cancer is a genetic disorder, in that all cancers result from genetic abnormalities resulting in altered function of the proteins which control the cell cycle and other important cellular processes, as summarised by the six hallmarks of cancer (Hanahan & Weinberg 2011, Figure 4).



Chapter 1 Figure 4: The six hallmarks of cancer (Hanahan & Weinberg *2011) showing the essential qualities which must be acquired by cancer cells to enable successful growth and metastatic spread.*

The process of quality control and genome maintenance during DNA replication is usually highly effective (Hanahan & Weinberg 2011), however acquired DNA sequence aberrations (somatic mutations) can arise as a result of chance or by the action of a mutagenic exogenous factor, for example ionising radiation, chemical carcinogens, tobacco smoke and viruses such as the human papilloma virus (HPV) and Epstein Barr virus (EBV) (Stratton *et al.* 2009). These aberrations may take the form of single base substitutions, insertions or deletions of segments of DNA, rearrangements, copy number increases or copy number reductions (Stratton *et al.* 2009).

The result of these aberrations is a clonal expansion of abnormal cells, resulting in the clinical manifestations of cancer depending on the tissue in which it arises and the behaviour of the abnormal cells themselves. The advent of cancer genome sequencing has resulted in the discovery of very high levels of

heterogeneity in tumour DNA between different regions of the same tumour and between primary and recurrent tumours in the same individual (McGranahan & Swanton 2017; Gerlinger *et al.* 2012). This suggests a process of stepwise mutation with branching 'evolution' of different clonal populations as each cancer genome becomes more unstable and susceptible to further mutations. These cell populations may have different properties and compete with each other for resources within the tumour microenvironment, resulting in the spectrum of aggressiveness we can observe amongst different tumours of the same histological type.

The rate of expansion in the understanding of the role of genetics in the development of cancer since the first identification of so-called 'cancer susceptibility genes' (Harris *et al.* 1969) has been immense. Tumour suppressor genes, as they are now referred to, are involved in cell cycle control and DNA repair. They account for most heritable forms of cancer, as they rely on a germline mutation on one copy of the gene followed by a 'second hit' acquired somatic mutation in the other copy, resulting in the observation that they are usually recessive in nature, as first described by Knudson in his study of retinoblastoma (Knudson 1971). Contrary to tumour suppressor genes, which cause cancer following loss of function mutations, oncogenes are genes which can initiate cancer by triggering cellular processes following activating or gain of function mutations (Vogelstein & Kinzler 2004).

The *RET* proto-oncogene, on chromosome 10q11.2 was first described in 1985 (Takahashi *et al.* 1985) and encodes a tyrosine kinase which is expressed in the developing urogenital system, and in neuroendocrine tissues such as the thyroid

C cells and adrenal chromaffin cells (Mulligan 2014). Inherited *RET* mutations are generally point mutations which result in tyrosine kinase activation and are associated with familial MTC, either alone or as part of the MEN2A and MEN2B syndromes. There is a high level of correlation between genotype and phenotype in these conditions, with certain mutations such as M918T being strongly associated with the early onset and aggressive MTC typically seen in MEN2B (Raue & Frank-Raue 2012). Somatic mutations are also identified in over 60% of sporadic MTC (Romei *et al.* 1996), and rearrangements of the *RET* locus are associated with DTC as well as lung adenocarcinoma (Ferrara *et al.* 2018; Prescott *et al.* 2015).

The most commonly identified somatic mutation in DTC is the activating V600E *BRAF* proto-oncogene, which is almost exclusively found in PTC, and has been shown to promote cancer progression and predict a more aggressive phenotype (Kebebew *et al.* 2007). For this reason, it has been the focus of much research into its use as an aid to diagnosis in thyroid nodules, and for risk stratification and management planning in PTC (Trimboli *et al.* 2020; Huang *et al.* 2018). Somatic gain of function mutations in *RAS* are also commonly identified in patients with DTC, and are primarily found in follicular carcinoma (FTC) or follicular variant PTC (FVPTC), although their clinical significance and exact role in carcinogenesis are less well understood (Howell *et al.* 2013).

Recent years have seen an advance in the understanding of cancer genetics beyond the relatively simplistic model of DNA sequence aberrations resulting in loss of function in tumour suppressor genes or gain of function of oncogenes, to incorporate the more nuanced field of epigenetics. Epigenetics refers to the

inherited modulation in the expression of genes without altering the DNA sequence (Peschansky & Wahlestedt 2014), via a number of mechanisms including DNA methylation, histone modification, and the action of microRNAs on gene expression. A detailed exploration of the vast and expanding field of cancer epigenetics is beyond the scope of this chapter, although there is increasing evidence that epigenetics and genetics play interconnected roles in carcinogenesis, and epigenetic factors can be used to determine prognosis and target treatment (You & Jones 2012; Cheng *et al.* 2019).

1.9 Genetic epidemiology

The molecular structure of DNA was first described in 1953 (Watson & Crick 1953), and as early as 1954, researchers were investigating the role of genetics in the pathogenesis of complex disease (Neel & Schull 1954). Initially this process consisted of identifying diseases that clustered in families and estimating heritability. With the advent of genome sequencing came the ability to test these hypotheses, and indeed many genes were mapped for specific diseases. Single nucleotide polymorphisms (SNPs), defined as single bases with varying alleles at a frequency of at least 1% in a random set of individuals in the population, were identified and SNPs within segments of the human genome with linkage disequilibrium (LD) were found to correlate closely with one another (Gabriel *et al.* 2002). However, it was clear that for complex diseases the estimated heritable component was not completely explained by the observed single nucleotide polymorphisms (SNP) identified in GWAS (Duggal *et al.* 2018). The model of multifactorial, or polygenic, susceptibility to complex diseases, where multiple

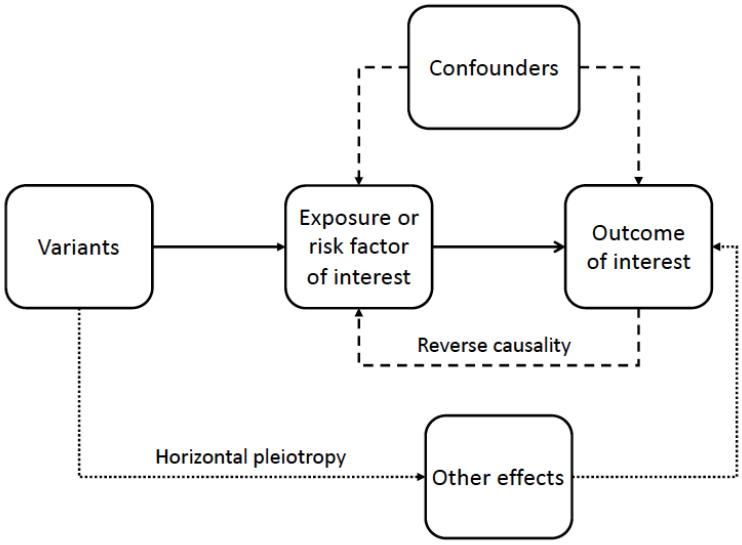
SNPs each with a small effect size contribute to the risk of complex diseases is now generally accepted (Lange 1997; Bodmer & Bonilla 2008).

In the case of cancer, whilst some germline mutations in oncogenes or tumour suppressor genes may result in inherited cancer syndromes as discussed above, some cancer susceptibility can be ascribed to relatively common SNPs. These tend to have a low effect on cancer risk in isolation but may result in a significant risk in combination, and have an effect on the action of environmental carcinogens (Lang & Pelkonen 1999). However, genetic epidemiology techniques such as Mendelian randomisation (MR) can also be used to investigate the effect of a wide range of modifiable risk factors on the risk of cancer. Traditional observational studies have significant limitations when used to examine risk factors for cancer, as they are susceptible to bias from confounding factors and reverse causality (Davey Smith & Phillips 1992; Sattar & Preiss 2017). Thus, whilst observational epidemiological studies may identify an association between a risk factor and a disease, it is difficult for them to irrefutably demonstrate a causal relationship. Mendelian randomisation uses genetic variants as instruments for exposures, and measures their causal effects on an outcome of interest, for example a cancer diagnosis. This has been made possible by GWAS data demonstrating the role of numerous SNPs in predisposing to various traits, from type 2 diabetes to educational attainment (Mahajan *et al.* 2018; Okbay *et al.* 2016). The advantage of MR over observational studies in the investigation of causal relationships between exposures and outcomes, is its ability to exploit the natural randomisation of alleles at conception due to the assortment of variants from both parents during meiosis. This random allocation of variants and the fact that they are set at

conception means that they are not subject to reverse causality, and at a population level, behavioural and environmental confounding factors should be equally distributed between groups with and without the variants of interest. This has led many to compare MR to randomised controlled trials, as the individuals in a population are randomly assigned at conception to an 'intervention' or control group based on their genetic variants relevant to a particular modifiable exposure (Yarmolinsky *et al.* 2018). MR is described as one-sample MR if the associations between the genetic variants and the risk factor of interest, and between the genetic variants and the outcome of interest are both measured in the same population. This is not always feasible and is often limited by the number of participants with available genetic data, risk factor exposure data and outcome data. Two-sample MR describes a technique whereby the variant-exposure association and variant-outcome association are measured in different groups of patients. This allows for the use of readily available GWAS summary data, and obviates the need for complete data on risk factor exposure in the population in which the outcome of interest is measured.

The validity of findings from MR studies does rely upon three assumptions: first, that the genetic variants used as the instrumental variable are reliably associated with the risk factor of interest; second, that the association between the instrumental variable and the outcome of interest is not confounded by other factors; and third, that the only association between the instrumental variable and the outcome of interest is by its effect on the risk factor of interest. If these criteria are not met, there is a risk of bias affecting the outcome of the analysis. These principles are illustrated in Figure 5. Since its development as a technique in genetic epidemiology, MR has successfully been used to identify many causal

associations, for example between obesity and depression (Tyrrell *et al.* 2018), apolipoprotein B and cardiovascular disease (Richardson *et al.* 2020), and educational attainment and Alzheimer's disease (Larsson *et al.* 2017).



Chapter 1 figure 5: General principles of Mendelian randomisation

Mendelian randomisation is not without its limitations. For example, the low proportion of variance in most traits explained by genetic variants means that very large sample sizes are required in order to demonstrate a difference between test and control groups. Another possible limitation is the undetected action of a variant on the outcome of interest via a pathway other than through its effect on the exposure of interest – known as horizontal pleiotropy (Figure 5). This results from the fact that some variants have multiple functions, and can be counteracted by using multiple genetic variants as the instrumental variable, and by ensuring that the function of each variant is fully understood (Lawlor *et al.* 2008). Linkage disequilibrium also poses a potential problem for MR studies, as if the variant being used as the instrumental variable is closely correlated to another variant (in LD with it) and that second variant has an effect on the outcome of interest by any other means than via an effect on the risk factor of

interest, then confounding will arise (Lawlor *et al.* 2008). This confounding can be tempered by a thorough understanding of the function of the variants used as instrumental variables, and the variants in LD with them. The use of multiple genetic variants can also help to mitigate the effects of LD. Developmental compensation for variants inherited at conception and expressed in early life poses a potential limitation of MR of uncertain significance (Davey Smith & Ebrahim 2003).

In the investigation of causal roles of risk factors for cancer, MR has several advantages over conventional observational studies. The long latency usually seen between risk factor exposure and cancer diagnosis means that it takes a long time to investigate the effect of a given exposure on the development of cancer using conventional epidemiological techniques. Using genetic variants that are fixed at conception generally allows estimation of the effect of life-long exposure to the risk factor of interest, unless the variants predispose to an exposure at a certain age, for example early menarche or late menopause. Thus, MR may detect causal relationships between temporally distant exposures and outcomes, for example the finding that whilst observational studies have identified a positive association between body mass index (BMI) and breast cancer, genetic variants associated with higher BMI in early life and middle age have found inverse associations (Guo *et al.* 2016). This suggests that the findings of observational studies may be confounded by other environmental risk factors that cause both obesity in later life, and breast cancer.

Thyroid cancer is a good example of a condition for which traditional observational epidemiological methodology has struggled to unpick the causal

roles for various proposed modifiable risk factors. The reported associations between obesity, diabetes, TSH, smoking and alcohol, and thyroid cancer discussed above are closely inter-connected and thus difficult to fully control for in observational population studies. Furthermore, the interplay between thyroid function, body mass index, and blood sugar control means that in a study group containing individuals who have had treatment for thyroid cancer (which often involves suppression of TSH levels) there is a significant risk of bias from reverse causality when retrospectively examining observational associations between the risk factors and outcome of interest. Thus, using known genetic variants (randomly assigned at conception) associated with a particular risk factor as instrumental variables in effect randomises the population and should result in an even distribution of other confounding factors, as well as effectively eliminating the risk of reverse causality.

1.10 Aims, objectives and results

We are entering a new era of genomic medicine, where our understanding of the human genome is allowing ever more detailed analysis of the genetic processes contributing to cancer development, its behaviour, and possible novel therapies. In the case of MTC, there is a well-established role for germline testing of the *RET* gene in diagnosis as well as counselling of unaffected family members. The role of somatic mutation analysis in risk stratification and management of sporadic MTC is less well understood. In the case of DTC, there is relatively little known about modifiable risk factors, and the development of genetic epidemiological techniques such as Mendelian randomisation and large datasets

such as the UK Biobank including genotyping data has allowed the use of genomics to study possible risk factors.

Thus, the aims of this thesis are to:

1. Explore the role of genetics in the diagnosis and management of MTC
2. Add to current knowledge on the risk factors for DTC, using genetic variants and Mendelian randomisation techniques.

The objectives are:

1. To generate and analyse a database of the largest UK series of *RET* testing in patients with MTC and other MEN2-associated diseases.
2. To perform a systematic review of the literature on the use of genetic analysis in the management of patients with *RET*-negative sporadic MTC.
3. To use Mendelian randomisation to investigate causative effects of various modifiable risk factors on benign nodular thyroid disease, and DTC.

Results will be presented in each chapter individually, in the form of discrete pieces of work which can be grouped together towards the overall aim of improving understanding of genetics in the diagnosis and management of thyroid cancer.

CHAPTER 2:

DIAGNOSTIC *RET* GENETIC TESTING IN 1,058 INDEX PATIENTS: A UK CENTRE PERSPECTIVE

This chapter investigates the role of germline *RET* proto-oncogene analysis in medullary thyroid carcinoma (MTC) in the clinical setting in the United Kingdom. As discussed in chapter 1, although the majority of cases of MTC are sporadic and thus by definition lack germline *RET* mutations, genetic testing is indicated in order to identify hereditary cases (Wells *et al.* 2015). This is important not only in order to offer surveillance for other manifestations of the multiple endocrine neoplasia 2 (MEN2) syndromes, but also to allow screening of unaffected family members. Much of the evidence on the correlation between clinical manifestations and germline *RET* mutations comes from continental Europe (Elisei *et al.* 2019; Romei *et al* 2011; Machens *et al.* 2001), with very few studies in the UK setting. The aim of this work was to report the use of genetic testing in MTC and MEN2 in the UK, to establish the rate of mutations in those with and without family history and/or other clinical manifestations of the MEN2 syndromes. We also investigated the rates and types of mutation according to clinical presentation in order to examine genotype – phenotype correlation in a UK population.

The work reported in this chapter has been published in Clinical Endocrinology (2020: 00: 1-8**,** DOI: 10.1111/cen.143950). The author list and their contributions are as follows:

Jonathan Mark Fussey	– data collection, analysis, and manuscript preparation
Joel Anthony Smith	– manuscript preparation
Ruth Cleaver	– manuscript preparation
Christopher Bowles	– data collection and manuscript preparation
Sian Ellard	– project inception and oversight, manuscript preparation
Bijay Vaidya	– manuscript preparation
Martina Owens	– project inception and oversight, data collection, analysis and manuscript preparation

2.1 – Introduction

Medullary thyroid carcinoma (MTC) is a neural crest derived malignant tumour accounting for up to 2% of thyroid cancers (Wells *et al.* 2015). Approximately 20-25% of cases are hereditary (hMTC), either as part of the multiple endocrine neoplasia types 2A and 2B (MEN2A and MEN2B) or in isolation (FMTC), however in the majority of cases the disease is sporadic (sMTC) (Pelizzo *et al.* 2007). Whilst the aetiology of sMTC is not fully understood, the underlying cause for hMTC is well established. The REarranged during Transfection (*RET*) gene on chromosome 10q11.2 was first described in 1985 (Takahashi 1985). It encodes a tyrosine kinase involved in the development of the parathyroid glands, adrenal medulla, parafollicular C cells and enteric ganglia. It is primarily these structures that are affected in patients with MEN2, with MEN2A comprising hereditary medullary thyroid carcinoma with or without phaeochromocytoma and primary hyperparathyroidism, and MEN2B comprising hereditary MTC with or without phaeochromocytoma and other characteristic phenotypic features such as mucosal neuromata, marfanoid habitus, corneal nerve thickening, and gastrointestinal ganglioneuromata. Shortly after its discovery, germline *RET* pathogenic variants were found to be present in almost all patients with hMTC (Donis-Keller *et al.* 1993). Despite the lack of heritable germline *RET* pathogenic variants in the sporadic form of MTC, somatic *RET* variants are identified in up to 66% of cases (Romei *et al.* 1996; Marsh *et al.* 1996).

A genetic diagnosis of MEN2 is important not only in order to direct investigation for other manifestations of the syndromes, but also to enable genetic counselling and

predictive testing for family members. It is particularly important to offer *RET* testing to all patients with MTC regardless of family history, as around 7% of those with apparently sporadic MTC do indeed harbour germline pathogenic variants (Romei *et al.* 2011), and 75% of MEN2B cases result from *de novo RET* pathogenic variants (Wells 2015). For these reasons, the American Thyroid Association, British Thyroid Association and European Thyroid Association guidelines all recommend *RET* analysis in any patient with confirmed MTC regardless of family history (Wells *et al.* 2015; Perros *et al.* 2014; Elisei *et al.* 2012).

In a research setting, germline *RET* pathogenic variants have been identified in exons 5, 8, 10, 11, 13-16 in over 95% of patients with MEN2A and FMTC, and a pathogenic variant in codon 918 (M918T) has been identified in over 95% of patients with MEN2B (Eng *et al.* 1996). However, *RET* pathogenic variant detection rates in the routine clinical setting are less studied. Furthermore, the last two decades have seen an increase in reports of novel germline *RET* pathogenic variants in hMTC, and this has resulted in a shift in the extent of *RET* analysis in clinical practice. The objective of the present study is to report the results of diagnostic *RET* analysis performed in a clinical setting at a single centre over a 21-year period.

2.2 – METHODS

2.2.1 – PATIENTS

Between 1997 and 2018, 1,058 index patients with MTC and other MEN2-related clinical features were referred from across the United Kingdom to the Exeter Genomics Laboratory at the Royal Devon and Exeter NHS Foundation Trust for diagnostic germline *RET* testing. An online request form was used to collect clinical data including age at diagnosis, MTC and other MEN2-related clinical features and information on family history. In addition, 551 clinically unaffected family members of patients found to have germline *RET* pathogenic variants underwent predictive testing. Informed consent was obtained by the treating clinician from all patients prior to genetic testing.

2.2.2 – GENETIC ANALYSIS

Genomic DNA was analysed for variants in selected exons of the *RET* gene using sequence specific primers (sequences available on request). PCR products were sequenced using Big Dye Terminator chemistry on an ABI DNA sequencer (Applied Biosystems, Warrington, UK), and the sequences were compared to the published sequence (NM_020975).

During the period of this study, the extent of *RET* sequencing has evolved in line with reports of new causative pathogenic variants. When diagnostic *RET* sequencing was introduced at our institution in 1997, only exons 10 – 11 were included, with subsequent expansion to include exons 13, 14, 15 and 16 in the same year. Testing

for exon 8 was introduced in 2007, exon 5 in 2010 and exon 7 in 2016 in response to reports of novel pathogenic variants in those regions (Fazioli *et al.* 2008; Castellone *et al.* 2010; Latteyer *et al.* 2016). Patients with a possible diagnosis of MEN2B are tested for exons 15 and 16 only. Unaffected relatives of a proband with a germline *RET* pathogenic variant are routinely offered predictive testing.

2.3 – RESULTS

The clinical indications for germline *RET* analysis in the 1,058 index patients are shown in table 1. Overall, pathogenic variants were identified in 108 index patients (10.2%). The detection rate was higher amongst the 514 patients referred after 1st December 2010 (n=63, 11.9%) than amongst the 544 patients referred before this date (n=45, 8.3%); however, the indication for testing also varied according to date, with 65.1% of those referred prior to December 2010 having MTC compared with 80.2% of those referred after December 2010. The prevalence of individual variants according to clinical features is shown in supplementary table 1. Of the 551 family members who underwent predictive testing, 160 (29%) were found to carry the familial variant.

Clinical features	N	Germline pathogenic variants (%)
All index patients with MTC[a]	766	92 (12.0)
Index patients with isolated MTC	690	68 (9.9)
Isolated MTC with no known family history of endocrine tumours	657	56 (8.5)
Isolated MTC with family history of MEN2 components	33	12 (36.4)
Index patients with multiple MEN2 related tumours	91	21 (23.1)
MTC and phaeochromocytoma or PHTPT	69	20 (29.0)
Phaeochromocytoma and PHPT	22	1 (4.5)
Index patients with phaeochromocytoma only	165	6 (3.6)
Index patients with PHPT only	56	0 (0)
Index patients with MEN2B phenotype	44	13 (29.5)
With MTC only	6	3 (50)
With phaeochromocytoma only	4	2 (50)
With MTC and phaeochromocytoma	1	1 (100)
With no endocrine tumour	33	7 (21.2)
Index patients with C-cell hyperplasia	5	0 (0)
Index patients with unspecified paraganglioma	4	0 (0)
Index patients with carcinoid	1	0 (0)
Index patients with pituitary adenoma	1	0 (0)
Index patients with lichen amyloidosis	1	0 (0)
Unaffected family members undergoing predictive testing	551	160 (29)

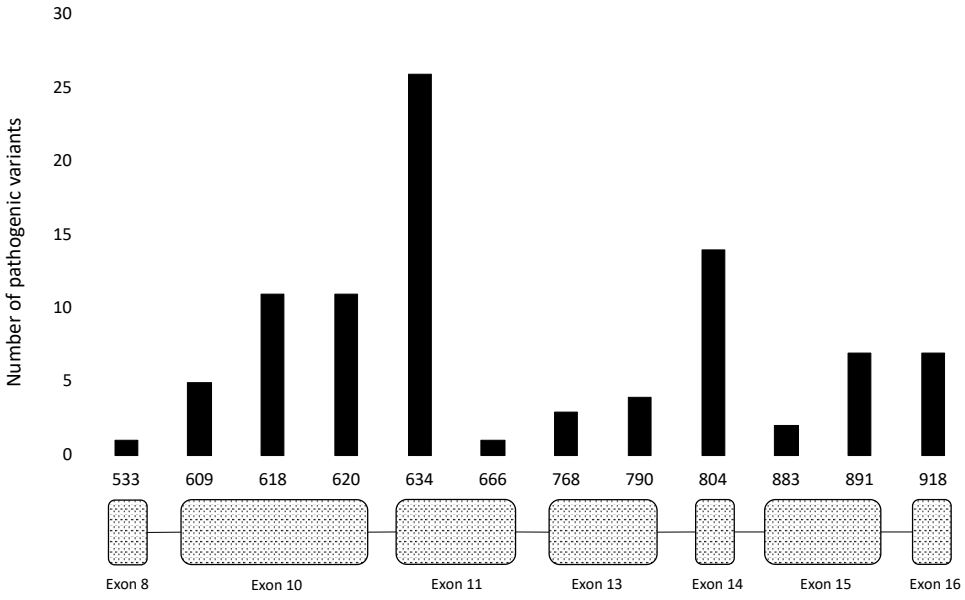
Chapter 2 Table 1: Pathogenic variant rates in UK patients according to clinical features and family history (MTC = medullary thyroid carcinoma; PHPT = primary hyperparathyroidism; MEN2B = multiple endocrine neoplasia type 2B)

[a]All index patients with MTC include index patients with isolated MTC (n=690), MTC and phaeochromocytoma or PHTPT (n=69), MTC with MEN2B phenotype (n=6), and MTC with phaeochromocytoma and MEN2B phenotype (n=1)

2.3.1 – INDEX PATIENTS WITH MTC

Of the 766 UK patients with MTC, 433 were female (56.5%). The median age at referral was 51 years (Range 1-94 years). In total, 92 patients (12%) were found to

harbour a germline *RET* pathogenic variant. Variants in exons 10, 11 and 14 of the *RET* gene made up the majority, with the most commonly affected codons overall being 634 followed by 804 (figure 1). As expected, pathogenic variants affecting the cysteine residues were the most common (58.2%, n=53). The p.(Val804Met) variant was identified in 14 of the 91 patients (15.4%). Of the 766 index patients with MTC, 690 had isolated MTC with no other features of MEN2A or MEN2B.



Chapter 2, Figure 1: Locations of the 92 germline RET *pathogenic variants identified in index patients with medullary thyroid carcinoma from the UK.*

2.3.1.1 – PATIENTS WITH ISOLATED MTC AND NO KNOWN FAMILY HISTORY OF ENDOCRINE TUMOURS

Of the 657 patients with confirmed isolated MTC and no family history (therefore presumed to have sporadic disease), 56 were found to harbour germline *RET* pathogenic variants (8.5%, table 1). The median age of these patients was 58 years (age range: 2-94 years), and there was a 1.3:1 female to male ratio. One patient in this group who was diagnosed with isolated MTC at the age of 30 was found to harbour

a pathogenic variant at codon 883 - p.(Ala883Phe) which is associated with MEN2B. She had no known family or personal history of endocrine tumours and no phenotypic features of MEN2B. In addition, three patients in this group aged seven, 13 and 19 were found to have the MEN2B-associated p.(Met918Thr) variant. According to their referrals, their family history was unknown and they had no personal history of phenotypic features of MEN2B or phaeochromocytoma. A variant of uncertain significance (VUS), p.(Val292Met), was identified in a female with isolated MTC. This variant has previously been described in 44-year-old male with phaeochromocytoma and MTC (Castellone *et al.* 2010). *In vitro* assays indicated a low-grade transforming potential, and the authors proposed that other genetic determinants may also have contributed to tumorigenesis.

2.3.1.2 – Patients with isolated MTC and a family history of MTC or MEN2 components

This subgroup was formed of patients with a diagnosis of MTC and a positive family history of MTC, C-cell hyperplasia, phaeochromocytoma or primary hyperparathyroidism, and included 33 patients. There were 12 pathogenic variants identified in this group (36.4%, supplementary table 1).

2.3.2 – Patients with multiple MEN2-associated conditions

A clinical diagnosis of a MEN2 syndrome was suspected based on the presence of 2 or more of MTC, phaeochromocytoma and primary hyperparathyroidism. This group included 91 patients, of whom 6 were reported to have all three diagnoses, 25 had

MTC and primary hyperparathyroidism, 38 had MTC and phaeochromocytoma and 22 had phaeochromocytoma and primary hyperparathyroidism.

Amongst patients with multiple MEN2-associated conditions, 21 (23.1%) were found to harbour a germline *RET* pathogenic variant, and one was found to have a variant of uncertain significance (supplementary table 1).

2.3.2.1 – PATIENTS WITH MTC PLUS PHAEOCHROMOCYTOMA OR PRIMARY HYPERPARATHYROIDISM

In those with MTC and Phaeochromocytoma the pathogenic variant detection rate was 44.7% (17 out of 38), compared to 8% (2 out of 25) in those with MTC and primary hyperparathyroidism. We identified a previously reported variant, p.(Glu818Lys), in a patient in this group. This variant has been described in a patient with MTC (Paszko *et al.* 2007), however further functional analysis is required to determine the significance of this variant on the transforming potential of RET and it was therefore classified as being of uncertain significance. A pathogenic *RET* variant was identified in only one of the six patients reported by the referring clinician to have all three diagnoses. However, of the five patients with no variant identified, one had undergone testing of exons 10, 11, 13, 14 and 15 elsewhere and only underwent testing of exons 1-7, 9, 12 and 16-20 at our laboratory so the possibility of a missed variant cannot be excluded. The remaining four patients were referred as having MTC, phaeochromocytoma and primary hyperparathyroidism but on further inquiry none had confirmed MTC, but rather raised calcitonin along with confirmed phaeochromocytoma and primary hyperparathyroidism.

Amongst patients with MTC and phaeochromocytoma, variants at codon 634 were the commonest, being present in 76.5% of those with a pathogenic variant. This is in contrast to patients with only MTC, amongst whom, p.(Val804Met) was the most common variant. Furthermore, when the whole series was grouped according to pathogenic variants by codon, 20 of the 31 patients (64.5%) with pathogenic variants at codon 634 had phaeochromocytoma compared to only 7 of 75 patients (9.3%) with pathogenic variants at all other codons. One patient in this group who was diagnosed with MTC and phaeochromocytoma at the age of 13 years, with uncertain family history was heterozygous for the MEN2B associated germline p.(Met918Thr) pathogenic variant.

2.3.2.2 – INDEX PATIENTS WITH PHAEOCHROMOCYTOMA AND PRIMARY HYPERPARATHYROIDISM

Of the 22 index patients with a diagnosis of both phaeochromocytoma and primary hyperparathyroidism, only 1 (4.5%) was found to have a germline *RET* pathogenic variant (p.Cys634Arg). None of these patients had a family history of MEN2-related conditions.

2.3.3 – INDEX PATIENTS WITH ISOLATED PHAEOCHROMOCYTOMA

165 patients were referred for germline *RET* testing with a diagnosis of isolated phaeochromocytoma, with a median age of 39 years (range: 1-88 years), and only 6 were found to harbour pathogenic variants (3.6%). Of these, 4 were found to have

pathogenic variants at codon 634, and one had a family history of MEN2-associated conditions.

2.3.4 – Index patients with isolated primary hyperparathyroidism

There were 56 patients with isolated primary hyperparathyroidism. The median age at referral was 36 years (range: 12 – 73 years), and 11 patients (19.6%) had a family history of MEN2-associated conditions. No patients with isolated primary hyperparathyroidism were found to harbour a pathogenic *RET* variant. One patient had a previously reported *RET* variant, p.(Asn783Ser), which is currently of uncertain clinical significance. This variant has previously been reported in a patient with MTC (Lebault *et al.* 2015).

2.3.5 – Index patients with the MEN2B phenotype

The final group comprised 44 patients with a clinical suspicion of MEN2B based on phenotypic features such as mucosal neuromata, intestinal ganglioneuromata, corneal nerve thickening and marfanoid body habitus, with or without MTC or phaeochromocytoma. Six of these patients had MTC only, four had phaeochromocytoma only, and one had both MTC and phaeochromocytoma; however, 33 patients did not have MTC or phaeochromocytoma at the time of referral (table 1). The median age of this group without MTC or phaeochromocytoma was 12 years, and they were referred purely on the basis that their phenotype was suspicious for MEN2B. Germline *RET* pathogenic variants were identified in six of the 11 patients with phenotypic features of MEN2B plus MTC and/or phaeochromocytoma (54.5%), and seven of the 33 patients with phenotypic features of MEN2B but no diagnosis of

MTC of phaeochromocytoma at the time of referral (21.2%). The p.(Met918Thr) variant accounted for every pathogenic variant in this latter group, and the median age at testing was 9 years. There was one patient with an isolated phaeochromocytoma and marfanoid features included in this group who was heterozygous for the p.(Cys634Tyr) variant, which is not associated with MEN2B.

2.4 – DISCUSSION

This study represents the largest series of germline *RET* testing in a clinical diagnostic setting in the United Kingdom. 1,058 index patients were referred over a 21-year period, and a further 551 unaffected family members underwent predictive testing. Amongst the 690 index patients with a diagnosis of isolated MTC, the principal finding was of a lower than expected overall pathogenic variant rate (9.9%). This may be due to less stringent testing criteria in a clinical setting than in the research setting, however the rates of pathogenic variant detection in those with other MEN2-associated conditions and with a positive family history were also lower than expected (29% and 36.4% respectively). We did find an increase in detection rate amongst those referred since 2010 compared with those referred prior to 2010, and this may reflect evolution in testing over recent decades, although the indications for tests before and after 2010 were not comparable so this information should be interpreted with caution. Our finding that 29% of unaffected family members were found to have pathogenic variants on predictive testing suggests that this is a valuable use of resources.

Amongst patients with presumed sporadic MTC, there was an 8.5% rate of germline *RET* pathogenic variants, which is in keeping with figures from other studies, which range from 6.2 – 14.9% (Romei *et al.* 2011; Elisei *et al.* 2019; Kihara *et al.* 2016; Sarika *et al.* 2015), and justifies international guidelines to offer germline *RET* testing to all patients diagnosed with MTC (Wells *et al.* 2015; Perros *et al.* 2014). The use of testing somatic *RET* mutations in presumed sporadic MTC in order to inform prognosis has been discussed extensively elsewhere, and there is evidence that certain somatic *RET* mutations predict a poorer outcome (Moura *et al.* 2009; Schilling *et al.* 2001; Elisei *et al.* 2008; Fussey *et al.* 2019). Despite this increasing body of evidence, guidelines on the management of sporadic MTC disagree on the use of somatic *RET* analysis in the routine clinical management of sporadic MTC (Wells *et al.* 2015; Ciampi *et al.* 2019). Another use for somatic *RET* analysis is to confirm the truly sporadic nature of the disease in cases where a somatic variant is identified that was not present on germline analysis, thus providing reassurance to family members regarding their risk of hereditary disease. In addition, it can guide the use of targeted therapy in advanced disease.

The rate of detection of germline pathogenic variants in patients with confirmed hereditary MTC has been reported as high as 98.3% (Elisei *et al.* 2019). In our series, 39.4% of index patients with a family history of either MTC or another MEN2-related condition were found to harbour pathogenic variants in *RET.* The rate was 29% in those with MTC plus phaeochromocytoma and/or primary hyperparathyroidism, with 80% of these being in codon *634.* These rates are lower than expected given the high likelihood of a MEN2 syndrome in patients with both MTC and phaeochromocytoma, and raises the possibility that either the clinical information provided by referring

clinicians was not accurate, or that a *RET* pathogenic variant was indeed present and not detected in a proportion of cases. Patients with no variant identified were referred from across the UK with no evidence of geographical clustering, and it is notable that despite the introduction of screening for exons 5 and 7, no additional pathogenic variants were identified. In the case of familial MTC it is possible that either hitherto unidentified variants in *RET,* or variants in genes other than *RET* have a role in tumorigenesis. For example, Smith *et al.* identified a novel germline frameshift deletion in the *ESR2* gene using exome sequencing in a patient with MTC and 3 family members with C-cell hyperplasia but no identifiable germline *RET* pathogenic variant (Smith *et al.* 2016). They showed that the effect was loss of *ER*β and indirect upregulation of *RET*, which raises the possibility of alternative pathways in the development of MTC.

Overall, the most commonly affected codons amongst the UK patients with MTC were 634 in exon 11, followed by 804 in exon 14. The largest studies on *RET* analysis in MTC are from Italy, and they consistently find codon 804 to be most commonly mutated in MTC (Elisei *et al.* 2019; Romei *et al.* 2016), however codon 634 has been reported as the most commonly affected in both Germany and Brazil (Machens *et al.* 2003; Maciel *et al.* 2019). Interestingly, a large Danish study found that variants at codon 611 were most prevalent (Mathiesen *et al.* 2017). In keeping with reports from Italy, we also found that cysteine pathogenic variants at codon 634 were more common in those with MTC with a phaeochromocytoma (Elisei *et al.* 2019).

The lowest pathogenic variant detection rates in the current series were in the subgroups of index patients with isolated primary hyperparathyroidism (0%), and

isolated phaeochromocytoma (3.6%). A previous series of 271 patients with apparently sporadic phaeochromocytoma reported a germline *RET* pathogenic variant rate of 4.8% (Neumann *et al.* 2002). Furthermore, our finding that 21 out of 32 patients (65.6%) with pathogenic variants at codon 634 had phaeochromocytoma compared to only 7 of 76 patients (9.2%) with pathogenic variants at all other codons supports findings by other authors that pathogenic variants at codon 634 are associated with phaeochromocytoma in patients with MEN2 (Frank-Raue *et al.* 1996; Machens *et al.* 2001). Primary hyperparathyroidism is a relatively common condition, with an incidence of 6.72 per 1000 in the UK population (Yu *et al.* 2009). This may account for the low rates of pathogenic variant detection in those with primary hyperparathyroidism, even in addition to MTC or phaeochromocytoma. These patients may in fact have sporadic MTC or sporadic phaeochromocytoma with incidental co-existing primary hyperparathyroidism, rather than MEN2A. Our finding that none of the 56 patients with isolated primary hyperparathyroidism had a pathogenic variant in the *RET* gene (which is included in the gene panel testing for primary hyperparathyroidism in the UK) is consistent with a recent study showing that primary hyperparathyroidism is a rare presenting manifestation of MEN2 (Larsen *et al.* 2020). These low rates in patients with isolated phaeochromocytoma and primary hyperparathyroidism should be borne in mind when counselling these patients for genetic testing.

In our series, patients with phenotypic features of MEN2B plus either MTC or phaeochromocytoma were the most likely to harbour a *RET* pathogenic variant, with a detection rate of 54.5%. A single feature of MEN2B was sufficient for inclusion in this group, and some features such as a marfanoid habitus are somewhat subjective,

meaning that it is inevitable that some patients in this group in fact had sporadic MTC or phaeochromocytoma, or MEN2A in the case of the individual with a pathogenic variant at codon 634. Furthermore, it is well recognised that some patients can present with the characteristic physical phenotypes of MEN2B but do not carry *RET* pathogenic variants or develop endocrinopathies (such as MTC or phaeochromocytoma), the condition termed 'pure mucosal neuroma syndrome' (Spyer *et al.* 2006). We have recently shown that several of these patients carry pathogenic variants in the *SOS1* gene, providing evidence that this condition is a distinct clinical condition unrelated to MEN2B despite similar physical phenotype (Owens *et al.* 2016). On the other hand, seven index patients with phenotypic features of MEN2B but not known to have MTC or phaeochromocytoma at the time of referral were found to carry the MEN2B-associated pathogenic variant p.(Met918Thr) in this study, suggesting that the physical phenotype (such as, mucosal neuromas, corneal nerve thickening and Marfanoid body habitus) can be the first presenting feature of MEN2B.

2.4.1 – Strengths and limitations

This study represents the largest UK series of germline *RET* analysis to date, and gives a perspective on the realities of molecular genetic diagnostic testing in a clinical setting. The main limitations are firstly, the limited clinical information regarding index patients and their family members on referral forms in some cases. This makes it difficult to confirm the clinical diagnosis and therefore correlate this with identified pathogenic variants. To try and capture as much clinical information as possible, including family history, ethnicity and age at diagnosis, the laboratory has a request form that can be downloaded from our website by the clinician for completion.

Secondly, although patients referred from outside the UK were excluded in the analysis of specific pathogenic variants, we were not able to ascertain the ethnicity of the included patients. It is known that *RET* variants in MTC and MEN2 differ according to geographical population (Hedayati *et al.* 2016), and therefore the ethnicity of patients in our series may have had an effect on the pathogenic variants identified. Finally, although samples from across the UK were tested, the applicability of our findings to other populations is limited, and we were unable to obtain follow-up data to identify metachronous MEN2-associated conditions diagnosed since referral for *RET* analysis.

2.4.2 – Conclusions

In a UK clinical setting as in other settings, the rate of germline *RET* pathogenic variants in presumed sporadic MTC is significant, and all patients with a diagnosis of MTC should be screened in order to identify familial cases. In patients with clinical manifestations of MEN2A, the majority did not receive a confirmatory genetic diagnosis. The rate of detection in patients with isolated phaeochromocytoma is not insignificant and therefore it seems this is an appropriate indication for germline *RET* analysis, however the absence of a single positive result amongst patients with isolated primary hyperparathyroidism suggests that this may not be an effective use of resources. Finally, the use of somatic *RET* analysis to confirm the diagnosis of sporadic MTC in patients with no identified germline *RET* variants may be a useful adjunct both in terms of reassuring family members about the lack of a heritable pathogenic germline variant, and risk-stratifying sporadic tumours based on somatic variants.

Exon	Variant (NM_020975.4)	Total number of patients	Isolated MTC	MTC with family history of MEN2-related condition	≥2 MEN2-associated conditions[a]	Isolated phaeochromocytoma	MEN2B features[b]	Classification (Richards *et al.* 2015)
5	c.874G>A p.(Val292Met)	1	1	0	0	0	0	Uncertain significance
8	c.1597G>T p.(Gly533Cys)	1	0	1	0	0	0	Pathogenic
10	c.1826G>A p.(Cys609Tyr)	5	5	0	0	0	0	Pathogenic
10	c.1852T>A p.(Cys618Ser)	1	0	0	0	1	0	Pathogenic
10	c.1852T>C p.(Cys618Arg)	5	3	2	0	0	0	Pathogenic
10	c.1852T>G p.(Cys618Gly)	1	1	0	0	0	0	Pathogenic
10	c.1853G>C p.(Cys618Ser)	5	3	2	0	0	0	Pathogenic
10	c.1858T>C p.(Cys620Arg)	6	4	1	1	0	0	Pathogenic
10	c.1858T>G p.(Cys620Gly)	1	1	0	0	0	0	Pathogenic
10	c.1858T>A p.(Cys620Ser)	3	1	2	0	0	0	Pathogenic
10	c.1859G>T p.(Cys620Phe)	1	1	0	0	0	0	Pathogenic
11	c.1900T>A p.(Cys634Ser)	2	2	0	0	0	0	Pathogenic
11	c.1900T>C p.(Cys634Arg)	11	2	0	7	2	0	Pathogenic
11	c.1900T>G p.(Cys634Gly)	6	1*	1	3	1**	0	Pathogenic
11	c.1901G>A p.(Cys634Tyr)	11	3	1	5	1	1	Pathogenic
11	c.1901G>T p.(Cys634Phe)	1	0	0	1	0	0	Pathogenic
11	c.1902C>G p.(Cys634Trp)	1	0	0	1	0	0	Pathogenic
11	c.1996A>G p.(Lys666Glu)	1	0	0	1	0	0	Likely pathogenic
13	c.2304G>C p.(Glu768Asp)	3	3	0	0	0	0	Pathogenic
13	c.2370G>T p.(Leu790Phe)	4	4	0	0	0	0	Pathogenic
14	c.2410G>A p.(Val804Met)	14	13	1	0	0	0	Pathogenic
14	c.2452G>A p.(Glu818Lys)	1	0	0	1	0	0	Uncertain significance
15	c.2647_2648delinsTT p.(Ala883Phe)	2	1	0	0	0	1	Pathogenic
15	c.2671T>G p.(Ser891Ala)	8	5	1	1	1	0	Pathogenic
16	c.2753T>C p.(Met918Thr)	15	3	0	1	0	11	Pathogenic
	Totals	110	57	12	22	6	13	

Chapter 2 Supplementary table 1: Prevalence of variants according to presenting features (MTC = medullary thyroid carcinoma; MEN = multiple endocrine neoplasia)
[a] MTC, primary hyperparathyroidism, phaeochromocytoma. [b] Mucosal neuromata, intestinal ganglioneuromata, corneal nerve thickening, marfanoid body habitus
** This patient underwent variant confirmation rather than full RET analysis ** This patient had a thyroid lesion, not confirmed MTC, therefore presumed isolated phaeochromocytoma*

CHAPTER 3:

THE ROLE OF MOLECULAR GENETICS IN THE CLINICAL MANAGEMENT OF SPORADIC MEDULLARY THYROID CARCINOMA: *A SYSTEMATIC REVIEW*

Following my work on germline *RET* proto-oncogene analysis in familial medullary thyroid carcinoma (MTC), I wanted to focus on the application of molecular genetics in the investigation and management of sporadic MTC. The clinical behaviour of familial MTC can be predicted based on the specific germline *RET* mutation identified (Wells *et al.* 2015), and this information can be used to tailor management to the individual patient. Sporadic MTC however has a variable clinical course and natural history, with stage at presentation only partly predicting tumour behaviour. Around 60% of sporadic MTC is found to harbour a somatic *RET* mutation (Romei *et al.* 1996; Marsh *et al.* 1996), and although there is some evidence that these somatic mutations have a negative impact on prognosis, there is also an emerging body of evidence that other somatic genetic and epigenetic markers may be used to predict tumour behaviour. The aim of this systematic review of the literature was to thoroughly examine the evidence base for the use of various somatic genetic and epigenetic variants in the risk stratification and prognostication of patients with sporadic MTC.

The systematic review forming this chapter has been published in *Clinical Endocrinology* (2019; 91(6):697-707; DOI: 10.1111/cen.14060). The author list and their contributions are as follows:

Jonathan Mark Fussey	– literature search, data extraction, analysis and manuscript preparation
Bijay Vaidya	– project oversight and manuscript preparation
Dae Kim	– manuscript preparation
Jonathan Clark	– manuscript preparation
Sian Elard	– manuscript preparation
Joel Anthony Smith	– data extraction, project oversight and manuscript preparation

3.1 INTRODUCTION

3.1.1 RATIONALE

Medullary thyroid carcinoma (MTC) is a malignant tumour of the neural crest derived parafollicular C cells. Due to the recent increase in incidence of papillary thyroid carcinoma, MTC now comprises only 1-2% of all thyroid cancers (Wells *et al.* 2015), but accounts for a significant proportion of thyroid cancer morbidity and mortality. The rate of regional and distant metastasis at presentation is up to 35% and 13%, respectively (Roman 2006), and there has been no trend towards earlier stage at diagnosis or improved overall survival in recent decades (Kebebew *et al.* 2005).

Approximately 20-25% of patients have hereditary MTC (hMTC) either as part of the multiple endocrine neoplasia syndromes (MEN2A and MEN2B) or in isolation as familial medullary thyroid carcinoma (FMTC), however in the majority of cases the disease is sporadic (sMTC) (Pelizzo *et al.* 2007). Although around 7% of those with apparently sporadic MTC do indeed harbour germline mutations (Romei *et al.* 2011), this group is a minority and is not included in the definition of sMTC used in this review. sMTC often follows an unpredictable course, with some patients suffering rapidly progressive and ultimately fatal disease, some surviving for many decades with incurable but stable disease, and others with relatively indolent disease amenable to cure following surgery.

The gold standard treatment for patients with biopsy proven, node negative sMTC is total thyroidectomy and central compartment neck dissection (CCND) (Wells *et*

al. 2015; Perros *et al.* 2014). The role of prophylactic lateral compartment neck dissection is less clear, with subtle differences in the guidance offered by the American Thyroid Association (ATA) and British Thyroid Association (BTA) (Wells *et al.* 2015; Perros *et al.* 2014). There is also a lack of unanimity on the subject of limited surgery for small tumours, with the ATA suggesting that hemithyroidectomy is sufficient if the post-operative serum calcitonin is less than 1000pg/ml (Wells *et al.* 2015), and the BTA using a tumour diameter of less than 5mm as a cut-off (Perros *et al.* 2014). Other authors have advocated thyroidectomy alone without central compartment neck dissection in patients with tumours smaller than 2cm (Esfandiari *et al.* 2014).

These subtle but important differences suggest that there is a lack of good evidence to define the best approach when considering the extent of initial surgery in MTC. In reality, many patients receive multiple operations, either to address the lateral compartment lymph nodes armed with information on central compartment occult nodal metastases and post-operative calcitonin levels, or as a result of a conservative approach with subsequent disease recurrence in cervical lymph nodes.

3.1.2 Objective

The difficulty in predicting the clinical course and prognosis at the time of diagnosis makes management planning particularly challenging in sMTC. Currently the role of somatic genetic and epigenetic profiles in the management of sMTC is unknown. Their use in risk stratification may help to guide treatment and allow improved prognostication, as well as more individualised follow-up. It

would allow better planning of the initial operation, for example to include a more aggressive neck dissection in cases with a high risk of cervical lymph node metastasis. Furthermore, it would rationalise inclusion criteria for clinical trials in the current era of personalised medicine and a rapidly increasing selection of novel anti-cancer drugs. The objective of this review is to systematically evaluate the current evidence for value of somatic molecular genetic and epigenetic markers in the risk stratification of patients with sMTC.

3.2 Materials and methods

3.2.1 Study design and eligibility criteria

The search strategy and inclusion criteria were outlined in advance in a protocol and registered with the PROSPERO international prospective register of systematic reviews. The protocol can be accessed at https://www.crd.york.ac.uk/prospero/ (registration number: CRD42019131092). All studies reporting the use of genetic or epigenetic markers in risk stratification and prognostication of patients with sMTC were considered. Articles were excluded if they did not focus primarily on patients with sMTC, or contained less than 10 subjects. Articles in languages other than English were excluded, as were articles not published in the last 25 years. Case reports, correspondence, commentaries and reviews were not eligible for inclusion.

3.2.2 Search strategy

The Cochrane library and PROSPERO international prospective register of systematic reviews were searched using a single search term: “medullary thyroid carcinoma” to identify relevant previous reviews on the topic, however none were identified. A search of Medline and Embase using the Ovid platform was then performed using the MeSH search terms “MEDULLARY CARCINOMA”, “EPIGENETICS”, “MOLECULAR GENETICS”, “MICRORNAS”; and free text terms “MEDULLARY CARCINOMA”, “SPORADIC MEDULLARY THYROID CANCER”, “SPORADIC MEDULLARY THYROID CARCIMONA”, “RET”, “RAS” and “miR”. Finally, a manual search of bibliographies of included studies was performed. The last search was performed on 3rd April 2019. The search log and Boolean operators used are outlined in Appendix 1.

3.2.3 Study selection and data extraction

Titles and abstracts of identified articles were screened for inclusion independently by two assessors based on the criteria outlined in section 3.2.1. If relevance was not evident after reading the title and abstract then the full text was studied before making a decision. Any disagreement between assessors was settled by a third assessor if consensus could not be reached. A data capture sheet was used to record study design; number of subjects; specific markers investigated; outcome measures; and key findings from included articles.

3.2.4 Data analysis

Due to the variation in study design and outcome reporting, formal statistical meta-analysis was not possible. Therefore, following article identification, key findings were grouped according to the genetic marker used and presented as a narrative synthesis.

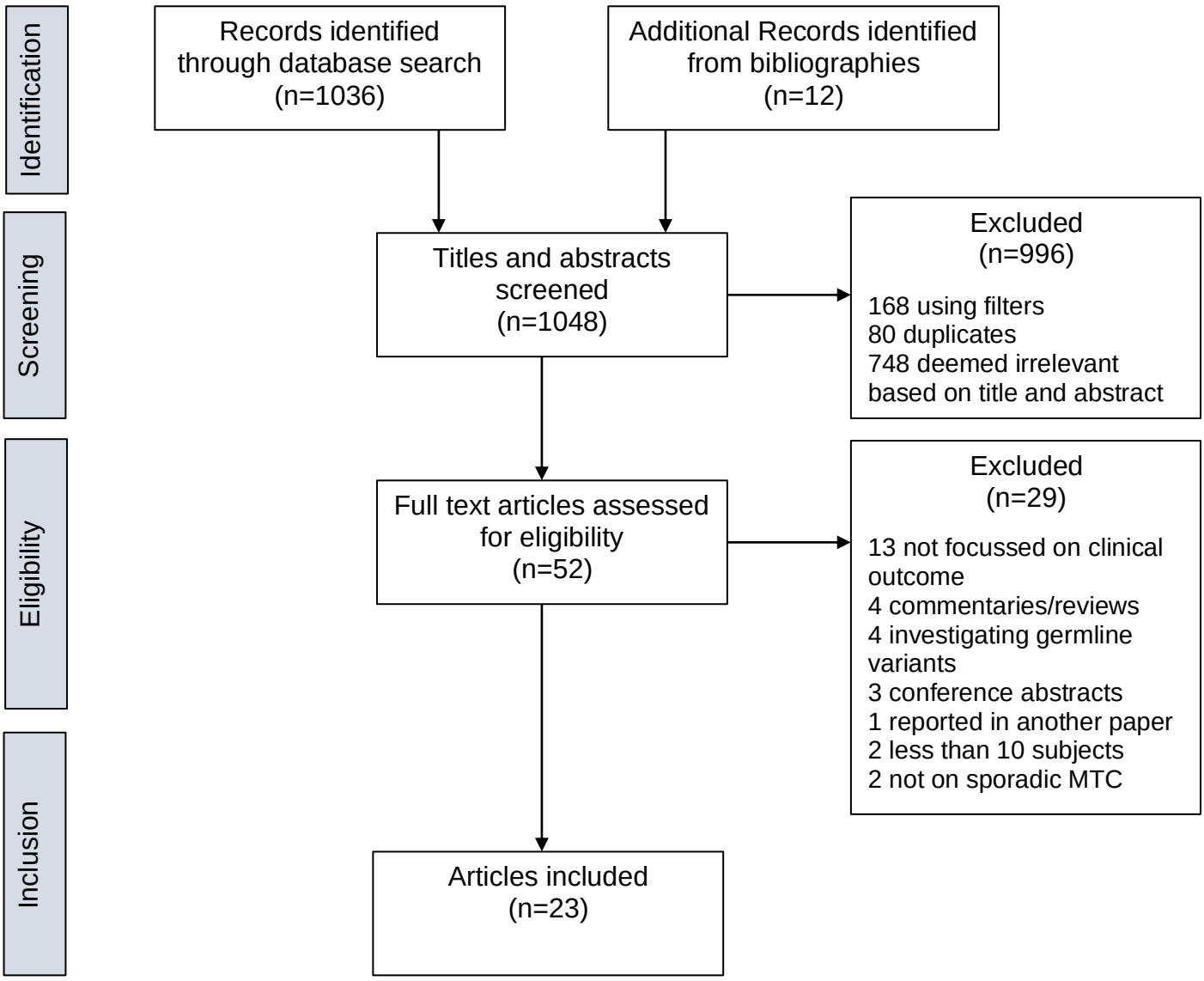
3.2.5 Risk of bias assessment

The studies included in this review were not randomised and were not testing an intervention, so tools for assessing the risk of bias in clinical trials were not applicable. A modification of the Newcastle Ottawa scale (Appendix 2) was therefore used to estimate the risk of bias in included studies.

3.3 Results

3.3.1 Search results and study selection

The initial search outlined in section 3.2 identified 1,036 articles. A manual search of bibliographies yielded a further 12 articles giving a total of 1,048. After screening articles for eligibility, and filtering for articles published in the last 25 years and with human subjects 1,025 articles could be excluded. This left 23 articles meeting the inclusion criteria (Figure 1).



Chapter 3 Figure 1: Preferred Reporting Items for Systematic reviews and Meta-Analyses 2009 (PRISMA) diagram

3.3.2 Characteristics of included studies

The twenty-three included studies were published between 1996 and 2018, and apart from one randomised controlled trial (Wells 2012) and one prospective

cohort study (Romeo *et al.* 2018), all were retrospective in design. They contained a combined total of 1,713 patients. Although all studies focused on the use of genetic and epigenetic markers in the risk stratification or prognostication of sMTC, there were a variety of specific markers used. The majority investigated the clinical role of somatic *RET* mutations (Wells *et al.* 2012; Cote *et al.* 2017; Romei *et al.* 2016; Simbolo *et al.* 2014; Ciampi *et al.* 2013; Romei *et al.* 2012; Mian *et al.* 2011; Moura *et al.* 2009; Dvorakova *et al.* 2008; Elisei *et al.* 2008; Schilling *et al.* 2001; Romei *et al.* 1996), however four included somatic *RAS* mutations (Simbolo *et al.* 2014; Ciampi *et al.* 2013; Cavedon *et al.* 2017; Moura *et al.* 2011) and one focussed on somatic *CDKN2C* mutations (Grubbs *et al.* 2016). Six studies utilised tumour miRNA expression (Romeo *et al.* 2018; Cavedon *et al.* 2017; Aubert *et al.* 2018; Galuppini *et al.* 2017; Mian *et al.* 2012; Abraham *et al.* 2011), two focussed on tumour methylation levels (Ceolin *et al.* 2018; Wang *et al.* 2016), and one on tumour mTOR expression Lyra *et al.* 2014). The general characteristics of included studies are represented in table 1. The risk of bias in the included studies varied significantly, with modified Newcastle Ottawa scores ranging from 3 to 7 out of 8 (8 being highest quality). Sources of potential bias arose from non-consecutive patient selection and inadequate follow-up duration in most cases, although none were deemed unworthy of inclusion solely based on their risk of bias.

Author	Year	Design	N=	Markers	Main findings
Ceolin	2018	Retrospective	24	Global DNA methylation levels	Higher global DNA methylation in sMTC than hMTC, but no correlation with tumour characteristics and clinical outcome
Aubert	2018	Retrospective	54	Tumour miR-183 and miR-21 expression	Both miR-21 and miR-183 were associated with lymph node involvement, and miR-21 was identified as an independent prognostic factor for lymph node involvement.
Romeo	2018	Prospective	33	Circulating miR-375	Circulating miR-375 levels were able to distinguish between patients with persistent sMTC and those in remission. Distant metastasis was higher and overall survival lower in those with higher levels of circulating miR-375.
Cote	2017	Retrospective	75	Circulating *RET* M918T	Circulating *RET* M918T mutated DNA found in 32% of those with the mutation identified in tissue biopsy, and strongly correlated with worse overall survival
Cavedon	2017	Retrospective	107	Somatic miR-224 and *RAS*	Positive association between miR-224 and *RAS.* miR-224 was a positive prognostic marker, inversely associated with persistent/progressive disease, calcitonin and disease-related death.
Galuppini	2017	Retrospective	104	Tumour miR-375 expression	miR-375 expression was associated with tumour size, capsule invasion, lymph node involvement and stage at diagnosis, although 26 hMTCs were included in the analysis.
Romei	2016	Retrospective	70	Somatic *RET*	Positive correlation between poor prognosis and number of somatic *RET* mutations identified
Grubbs	2016	Retrospective	62	Somatic *CDKN2C* copy number loss	Increased distant metastasis and reduced overall survival in the presence of Somatic *CDKN2C* copy number loss
Wang	2016	Retrospective	39	Somatic *TERT* promoter methylation	Lower overall and disease-specific survival in patients with high *TERT* promoter methylation
Simbolo	2014	Retrospective	20	Somatic *RET* and *RAS* mutations	No significant association between somatic mutation and clinical outcomes.
Lyra	2014	Retrospective	77	Tumour mTOR activation	Increased mTOR activation (as measured by p-S6 expression) associated with lymph node metastasis and invasive tumours, although 10 hMTCs were included in the analysis.
Ciampi	2013	Retrospective	175	Somatic *RET and* RAS mutations	Non-significant correlation between somatic *RAS* mutations and better outcomes. Somatic *RET* mutations predicted worse biochemical cure rates and higher disease progression rates than both *RAS* negative *RET* negative and *RAS* positive *RET* negative cases.
Mian	2012	Retrospective	34	Somatic expression of miR-21, miR-127, miR-154, miR-224, miR-323, miR-370,	miR-224 expression levels inversely associated with nodal metastasis and disease stage, and positively associated with biochemical cure.

				miR-9*, miR-183, and miR-375	
Romei	2012	Retrospective	160	Somatic *RET* M918T mutations	Somatic *RET* M918T mutations were associated with larger primary tumours.
Wells	2012	RCT	298	Somatic *RET* M918T	Phase III clinical trial on vandetanib, which found a higher response rate to the drug in sMTC patients with the somatic *RET* M918T mutation.
Abraham	2011	Retrospective	12	Tumour miR-183 and miR-375 expression	Increased expression of miR-183 and miR-375 correlated with lateral compartment nodal metastasis, distant metastasis and mortality, although 7 hMTCs were included in the analysis.
Mian	2011	Retrospective	60	Somatic *RET* and *Ki-67* expression	Both somatic *RET* and tumour Ki-67 expression correlated positively with tumour size, nodal metastasis, distant metastasis and low overall survival.
Moura	2011	Retrospective	65	Somatic *RET* and *RAS* mutations	No statistically significant differences in clinicopathological parameters between *RAS* positive and *RAS* negative cases
Moura	2009	Retrospective	51	Somatic *RET* mutations	Tumours with somatic *RET* mutations affecting exons 15 and 16 were associated with higher rates of lymph node metastasis, redisual disease, advanced disease and persistently raised calcitonin compared with tumours harbouring other *RET* mutations or no *RET* mutation.
Dvorakova	2008	Retrospective	48	Somatic *RET* mutation	Somatic *RET* mutations were associated with more advanced stage at presentation, but not with other clinical and pathological characteristics.
Elisei	2008	Retrospective	100	Somatic *RET* mutation	Somatic *RET* correlated with advanced stage a presentation and worse overall survival (10 years follow-up). M918T somatic *RET* mutations correlated with larger tumours, nodal metastases and distant metastases.
Schilling	2001	Retrospective	34	Somatic *RET* mutation	Somatic M918T *RET* mutations were associated with higher rates of distant metastasis, lower metastasis-free survival and lower overall survival.
Romei	1996	Retrospective	18	Somatic *RET* mutations	Somatic *RET* mutations were associated with higher rates of post-treatment recurrence or high serum calcitonin.

Chapter 3 Table 1: General study characteristics

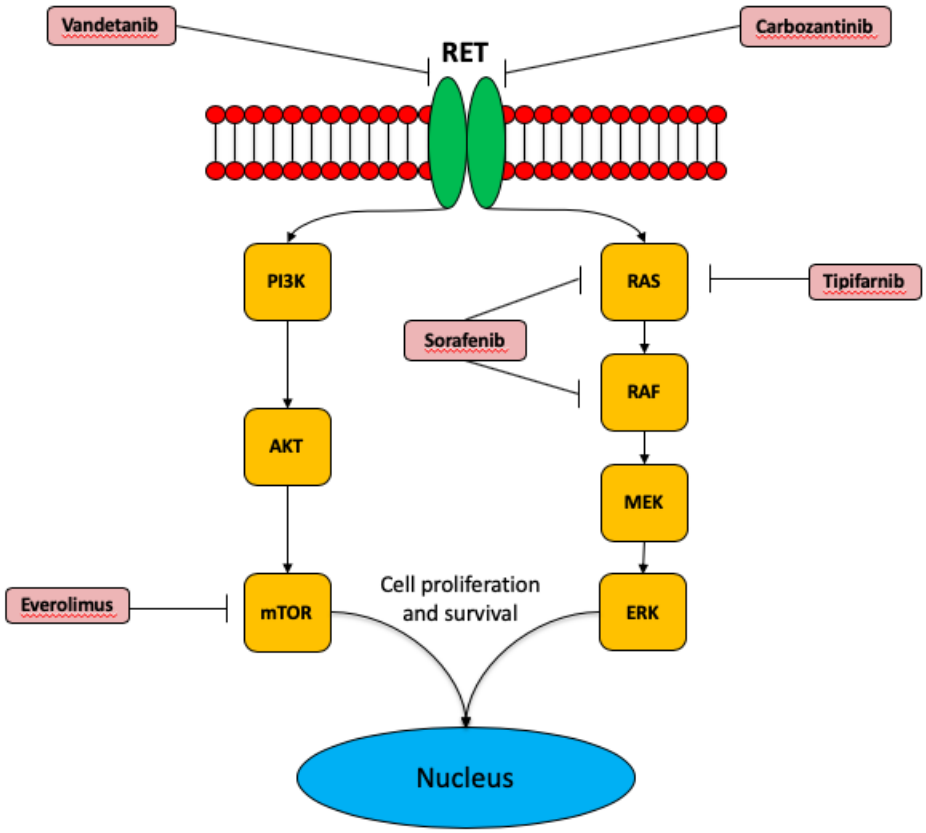
3.3.3 SOMATIC *RET* MUTATIONS

Chance heterozygous mutations in somatic tissue will be retained in the clonal progeny of the cell in which they occur, and can occur either as a cause or effect of cancer. Somatic RET mutations are present in up to 66% of sMTC (Dvorakova *et al.* 2008; Romei *et al.* 1996), and there is significant evidence to suggest that they are associated with worse outcomes. This was first demonstrated in 1996 (Romei *et al.* 1996), and later by Elisei *et al* in their study of 100 patients with 10-year follow-up (Elisei *et al.* 2008). The authors found a correlation between somatic *RET* mutations and lymph node metastasis at presentation, lower biochemical cure rate following surgery, and worse long-term survival. Other authors have also demonstrated more advanced tumour stage and worse outcomes in patients with somatic *RET* mutations (Moura 2009; Dvorakova *et al.* 2008), and improved specificity when combining the use of somatic *RET* mutations and tumour Ki-67 expression levels (Mian *et al.* 2011).

Although there are a large number of described somatic *RET* mutations in sMTC, the commonest is the p.Met918Thr (M918T) mutation, which is also the germline mutation responsible for around 95% of cases of MEN2B (Wells *et al.* 2015). The somatic M918T mutation has been identified in up to 68% of sMTC (Marsh *et al.* 1996). It is considered to predict the worst outcome (Schilling *et al.* 2001), and along with the p.Ala883Phe (A883F) mutation is associated with higher rates of lymph node metastasis, multifocality and persistent disease at last follow-up when compared with patients with other somatic *RE*T mutations or none (Moura *et al.* 2009). Double *RET* mutations have also been associated with worse outcomes (Romei *et al.* 2016). A meta-analysis of 23 studies found that somatic *RET*

mutations are associated with a higher rate of lymph node metastasis, distant metastasis, advanced stage at diagnosis, recurrence and mortality (Vuong *et al.* 2018). The M918T mutated tumour DNA can also be identified in peripheral blood, and there is some evidence that circulating M918T mutated DNA portends a worse overall survival and more accurately predicts outcome than calcitonin doubling time (Cote *et al.* 2017).

Given the established role of RET in hMTC and that almost half of sMTCs harbour somatic *RET* mutations, most efforts at novel drug development have focussed on inhibiting RET and RET-related pathways (Figure 2). This has resulted in the licencing of tyrosine kinase inhibitors (TKI) with specificity for RET, vandetanib and cabozantinib. Both drugs have been found to be superior to placebo in double blind randomised controlled trials in terms of progression free survival in populations consisting of both hMTC and sMTC (Wells *et al.* 2012; Elisei *et al.* 2013). In the original vandetanib trial, an M918T *RET* mutation was identified in 142 of 298 patients with sMTC, and subgroup analysis revealed a higher response rate and superior progression-free survival in these patients (Wells *et al.* 2012). However, a recent meta-analysis of studies on vandetanib found minimal evidence of significant benefit, with no evidence of improved overall survival and a significant risk of side effects, which may significantly impact quality of life (Trimboli *et al.* 2018). It is notable that the studies included in this meta-analysis consisted of patients with both hMTC and sMTC, and subgroup analysis according to somatic mutational profile was not possible. It is conceivable that stratification by genetic profile may have altered the results given the evidence of improved response rates of M918T RET sMTC.



Chapter 3 Figure 2: Schematic representation of selected downstream signalling pathways activated by RET, with sites of possible novel therapy action. Vandetanib and cabozantinib directly inhibit RET. Tipifarnib is a farnesyltransferase inhibitor which ultimately inactivates RAS. Sorafenib is a tyrosine kinase inhibitor, which deactivates the RAS/RAF/MEK/ERK pathway. Everolimus binds to a protein receptor, which directly inhibits mTOR. (PI3K = phosphoinositide 3-kinase; AKT = protein kinase B; mTOR = mammalian target of rapamycin; RAF = rapidly accelerated fibrosarcoma kinase; MEK = MAPK/ERK kinase; ERK = Extracellular signal kinase)

More recently, two selective *RET* inhibitors have been developed which are still being evaluated in clinical trials. LOXO-292 is under investigation in a phase 2 study, however preliminary results have demonstrated preclinical activity against various *RET* alterations. Furthermore, its use in a patient with advanced sMTC with a confirmed somatic *RET* M918T mutation resulted in tumour regression, a

drop in serum calcitonin and symptom improvement (Subbiah *et al.* 2018(a)). The second drug, BLU-667 is also a selective RET inhibitor still in phase 2 trials, however once again preliminary data has shown a good response in a patient with sMTC and multiple *RET* mutations (Subbiah *et al.* 2018(b)). Clearly it is too early to draw conclusions on the efficacy of these novel drugs or the relative effects in patients with different RET mutations, however these preliminary results are promising.

3.3.4 SOMATIC *RAS* MUTATIONS

Somatic *RAS* mutations are present in 30% of all human cancers (Forbes 2010). In sMTC, somatic *RAS* mutations have been identified in up to 81% of *RE*T negative tumours, and almost never co-exist in *RET* mutated tumours (Simbolo *et al.* 2014; Ciampi *et al.* 2013; Moura *et al.* 2011; Lyra *et al.* 2014). In contrast to follicular cell thyroid cancers, MTC tends to harbour *HRAS* mutations most commonly and *NRAS* mutations rarely, with the commonest being *HRAS* p.Gln61Arg (Moura *et al.* 2015).

The clinical significance of somatic *RAS* mutations has been less thoroughly investigated, and although some data suggests that they predict a more favourable outcome compared to somatic *RET* mutations in sMTC, none of the studies identified in this review were able to demonstrate this with statistical significance (Simbolo *et al.* 2014; Ciampi *et al.* 2013; Moura *et al.* 2011). Moura *et al* divided their 2011 cohort of sMTC patients into four groups depending on the presence of high-risk somatic *RET* mutations (M918T and A883F), other *RET* mutations, *RAS* mutations and no mutations (Moura *et al.* 2015). They found that

tumours with somatic *RAS* mutations behaved less aggressively than those with high risk *RET* mutations, but more aggressively than those with other *RET* mutations. Although Cavedon *et al* were unable to identify a significant association between somatic *RAS* mutation and favourable outcome; they did find that tumour miR-224 under-expression correlated with shorter overall survival, and that somatic *RAS* mutations correlated with miR-224 over-expression (Cavedon *et al.* 2017). They therefore concluded that somatic *RAS* mutations in sMTC predict a less aggressive phenotype.

3.3.5 Other genetic pathways

In up to 45% of sMTC, no somatic *RET* or *RAS* mutation is identified. This has led some authors to investigate the role of other related pathways, as well as epigenetic regulation of gene expression. In general, the quest for novel genes regulating tumorigenesis in MTC has focused on oncogenes in other cancer types, and been relatively fruitless (Cerrato *et al.* 2009). However, exome sequencing on a patient with MTC and no identifiable *RET* mutation has identified a germline frameshift c.948delT mutation in the *ESR2* gene (Smith *et al.* 2016). This mutation was also present in 3 family members with C-cell hyperplasia, and immunohistochemical studies confirmed its effect being the loss of ERβ and overexpression of RET. This finding of indirect up-regulation of RET expression leading to MTC, provides an interesting insight into alternate pathways for developing MTC, although it has yet to be replicated by other researchers, and its role in sMTC is as yet unknown (Ruiz-Ferrer *et al.* 2017).

The *CDKN* gene family, encoding cyclin-dependent kinase inhibitors has attracted some interest in the setting of MTC. These genes play a vital role in cell cycle control and have been implicated in many cancers, suggesting a tumour suppressor role (El Naofal *et al.* 2017). Somatic copy number loss in the *CDKN2C* gene has been identified in 19% of a cohort of 62 sMTC cases, and associated with higher rates of distant metastasis and worse overall survival (Grubbs 2016). Genotyping of the *CDKN* genes revealed a significant association between single nucleotide polymorphisms (SNPs) in the CDKN1B and CDKN2A genes and susceptibility to sMTC (Barbieri *et al.* 2014). Both of these genes encode cyclin dependent kinases with cell cycle control roles, with the latter being an important stabiliser of the p53 tumour suppressor protein. More recently, fluorescence in situ hybridisation has been used to identify loss of heterozygosity in *CDKN2C* and *CDKN2D* in formalin fixed paraffin embedded MTC samples, raising the possibility of a clinically useful prognostic marker (El Naofal *et al.* 2017).

Based on evidence that the *CDK/RB* pathway may be an alternative target for treatment of MTC, *in vitro* research has demonstrated that dinaciclib (a CDK1/2/5/9 inhibitor) reduced mRNA levels of CD7 and *RET* in MTC cells (Valenciaga *et al.* 2018). The effect was synergistic when dinaciclib was used in combination with a TKI, and the authors postulate the possibility of its use to improve therapeutic *RET* targeting in MTC.

The mTOR pathway is activated in both hMTC and sMTC (Tamburrino *et al.* 2012), and there is some evidence that high mTOR activity as measured by tumour p-S6 expression is associated with more invasive sMTC and higher rates

of lymph node metastasis (Lyra *et al.* 2014). Inhibitors of the mTOR pathway have therefore been investigated in advanced MTC, with a phase II trial of Everolimus showing some benefit (Schneide *et al.*r 2015).

3.3.6 EPIGENETIC MARKERS IN SMTC

Epigenetics can be defined as stably inherited modulations in the expression of genes without altering DNA sequence (Peschansky *et al.* 2014). The epigenetic control of gene expression is known to influence oncogenesis and tumour behaviour in many cancer types, mainly via aberrations in the histone acetylation and methylation pathways (Dawson *et al.* 2012). Although there has been little research on methylation profiles in MTC, a recent study by Ceolin *et al* found that global DNA methylation levels in peripheral blood leucocytes were higher in sMTC patients than hMTC patients (Ceolin *et al.* 2018). The authors were unable to identify any correlation between DNA methylation levels and tumour characteristics or clinical outcome.

A significant increase in gene expression of the histone methyltransferases EZH2 and SMYD3 has been found in samples from MTC patients with local and distant metastasis, irrespective of germline and somatic *RET* and *RAS* mutation status (Sponziello *et al.* 2014). This adds further weight to the suggestion that MTC tumour progression may be controlled by epigenetic factors independent of *RET* and *RA*S pathways. Telomerase activity is controlled by the *TER*T gene, which is usually suppressed in normal tissue, but activated in many human cancers to allow cell proliferation. Although *TERT* gene promoter mutations have not been identified in MTC (Liu *et al.* 2014), *TERT* copy number gain has been

demonstrated in a subset of patients with sMTC, as well as increased methylation of the *TERT* promoter region in both hMTC and sMTC patients (Wang *et al.* 2016). The same group found a statistically significant association between higher TERT methylation index, and poorer disease-free and overall survival. They postulated that the methylation of the TERT promoter region stimulates TERT expression and telomerase activation thereby contributing to more aggressive disease.

MicroRNAs (miRNAs) are molecules capable of down-regulating gene expression, and have thus been implicated in both carcinogenesis via down-regulation of tumour suppression genes, and tumour progression. Overexpression of miR-183 and miR-375 have been identified in sMTC compared with hMTC, and are associated with increased rates of lateral cervical lymph node metastases, distant metastases and mortality (Abraham *et al.* 2011). Other investigators have also identified a positive association between miR-375 levels and tumour stage at diagnosis, lymph node metastasis, calcitonin level at diagnosis and disease progression, independent of *RET* and *RAS* status (Galuppini *et al.* 2017).

miR-183 has been found to down-regulate the expression of the pro-apoptotic gene *PDCD4 (Li et al. 2010)*, and has been implicated in the oncogenesis of other cancers (Lin *et al.* 2008; Motoyama *et al.* 2009). Down regulation of *PDCD4* in association with overexpression of miR-29 is a possible carcinogenic factor in both sMTC and hMTC (Pennelli *et al.* 2015). More recently, the expression of six miRNAs was measured in a cohort of 54 sMTCs, and miR-183 was again found to correlate strongly with lymph node metastasis. In addition to miR-183, miR-21

expression was also found to correlate with lymph node metastasis as well as tumour size, baseline serum calcitonin and T3/T4 tumours at presentation (Aubert *et al.* 2018). The authors postulated that routinely testing this marker could guide surgeons in the extent of cervical lymph node dissection. Table 2 summarises the prognostic effects of various somatic genetic and epigenetic changes.

Biomarker	Outcome							References
	Stage at presentation	Calcitonin at presentation	Lymph node metastasis	Distant metastasis	Persistent disease at follow up	DFS	OS	
Somatic *RET* M918T or A883F versus *RET* wild-type			↑	↑	↑		↓	Moura 2009; Schilling 2001
Somatic *RET* mutation versus *RET* wild-type	↑		↑		↑		↓	Elisei 2008; Dvorakova 2008; Ciampi 2013
Tumour miR-375 over-expression		↑	↑	↑			↓	Abraham 2011; Galuppini 2017
Tumour miR-183 over-expression			↑	↑			↓	Abraham 2011
Tumour miR-21 over-expression	↑	↑	↑					Aubert 2018
Tumour miR-224 under-expression	↑	↑			↑	↑		Cavedon 2017
Tumour *TERT* hypermethylation						↓	↓	Wang 2016
Somatic *CDKN2C* copy number loss							↓	Grubbs 2016
Somatic *RAS* mutation versus *RAS* wild-type			↓ (NS)	↓ (NS)	↓ (NS)			Ciampi 2013

Chapter 3 Table 2. Summary of prognostic biomarkers in sporadic medullary thyroid carcinoma (WT=wild-type; DMFS = distant metastasis free survival; DFS = disease free survival; OS = overall survival; NS = not statistically significant)

The recent interest in epigenetic factors in sMTC has led some to investigate the feasibility of targeting epigenetic regulators with the aim of developing novel therapeutic agents. As early as 2005, *in vitro* experiments revealed the efficacy of histone deacetylase inhibitors in supressing proliferation and inducing apoptosis in thyroid cancer cell lines (Mitsiades *et al.* 2005). miRNA therapy has now been trialled in other forms of human cancer, with the first molecule to be studied being MRX34, an miR-34a mimic which supresses oncogenes targeted by miR-34a (Beg *et al.* 2017). Initial phase I studies showed promising anti-tumour efficacy, but significant adverse effect profiles. Interestingly, miR-34a expression has been found to be raised in MTC, rather than reduced as in other cancers (Shabani *et al.* 2018). Furthermore, Lassalle et al have demonstrated *in vitro* that overexpression of miR-375, one of the most studied miRNAs in MTC, results in improved efficacy of vandetanib on cancer cells (Lassalle *et al.* 2016).

3.4 Discussion

3.4.1 Summary of key findings

This review demonstrates the emerging complexity of the molecular genetic and epigenetic landscape in sMTC. There is now a significant body of evidence supporting the use of various genetic and epigenetic markers in risk stratification of patients with sMTC, however few authors have used this information to personalise management plans. Currently the clinical stage and serum calcitonin levels dictate the extent of initial and subsequent surgery in these patients, although with increasing availability of genetic profiling of tumours, the

incorporation of tumour-specific genetic and epigenetic markers into the decision-making process is becoming more feasible.

There is substantial evidence that somatic *RET* mutations predict more aggressive tumour behaviour and a worse outcome, however it is not clear whether these mutations actually drive tumorigenesis in sMTC. Some patients demonstrate heterogeneity of *RET* mutation in cell subpopulations both within different areas of the same primary tumour, and within different metastases from the same patient (Esfandiari *et al.* 2014; Eng *et al.* 1996). This suggests that the mutation is not an early event in tumorigenesis, and incidentally may also explain the wide variation in reported frequency, as its detection depends on which part of a tumour is tested. Furthermore, the presence of somatic *RET* mutations is not unique to MTC and may represent a marker of poor prognosis in many cancers, with various mutations, including M918T, having been identified in breast, lung and liver cancers, correlating with advanced disease stage and poor survival (Griseri *et al.* 2016; Dalbir *et al.* 2014; Ye *et al.* 2017). This suggests that the somatic *RET* mutation may play more of a role in tumour progression than tumorigenesis.

The role of somatic *RAS* mutations in predicting clinical course is still not well understood, and the findings of some authors that their presence predicts more favourable outcomes may simply reflect the fact that these tumours almost always lack somatic *RET* mutations. Up to 45% of sMTC express neither *RET* or *RAS* somatic mutations, which suggests a role for other mechanisms. The field of epigenetics is growing rapidly, and there is increasing evidence for a role for several epigenetic markers in risk stratification of sMTC. The most promising of

these are miR-21, miR-183 and miR-375, which have been shown by several authors to be associated with increased risk of lymph node metastasis in sMTC (Aubert *et al.* 2018; Galuppini *et al.* 2017; Abraham *et al.* 2011). These markers could complement serum calcitonin levels and stage of disease in helping to guide decision-making when considering the need for lateral compartment neck dissections in the absence of clinical or radiological evidence of nodal metastases.

A significant barrier to the use of these tumour specific markers in risk stratification and management planning is the fact that they are frequently not available at the time of initial diagnosis as their detection relies upon analysis of excised tumour tissue. The work of Cote *et al* on the detection of circulating *RET*-mutated tumour DNA (Cote *et al.* 2017), and Romeo *et al* on circulating miR-375 (Romeo *et al.* 2018) is therefore particularly interesting. In addition to their potential use in prognosticating and planning surgical management, genetic and epigenetic characteristics of sMTC tumours may also be valuable in predicting response to novel therapies and developing new therapeutic agents. The improved response rate to vandetanib amongst patients with somatic *RET* M918T mutations demonstrated by Wells *et al,* and the suggestion by Lassalle *et al* that tumour miR-375 overexpression increases susceptibility to vandetanib raise the possibility of using somatic mutation profiles to guide medical therapy (Wells *et al.* 2012; Lassalle *et al.* 2016).

3.4.2 LIMITATIONS

The field of molecular genetics in MTC is rapidly evolving and there is substantial evidence in the form of retrospective studies to support associations between

various tumour genetic and epigenetic markers and clinico-pathological features. However, the number of studies is still relatively limited, and with each using different molecular genetic characteristics and slightly different outcome measures, comparison of findings is difficult. Furthermore, there is a dearth of prospective evidence on the use of these markers to inform management of patients with sMTC. The majority of research to date has focussed on somatic *RET* mutations, and whilst the evidence for their role in predicting tumour behaviour is strong, it is clear that other molecular genetic mechanisms play a role. Further research is required to confidently base management decisions on these genetic and epigenetic characteristics.

3.4.3 Conclusions

Currently clinical stage and calcitonin levels dictate the extent of initial and subsequent surgery for sMTC, with novel therapies being reserved for local or distant disease progression. The evidence presented in this review suggests that somatic *RET* mutations and tumour overexpression of certain miRNAs may predict aggressive tumour behaviour and lymph node metastasis. There appears therefore to be scope to improve risk stratification in sMTC. This could allow for more individualised decision-making, with the extent of surgery being tailored to patients. Due to the rarity of these cancers, and as analysis of somatic mutations and epigenetic expression is not routinely analysed in most centres, a subspecialist or even nationalised approach to such cases may be warranted. Further research should aim to further clarify molecular genetic and epigenetic drivers of sMTC in order to provide more accurate prognostic information, inform management strategies and provide potential targets for novel therapeutic agents.

CHAPTER 4:

DOES OBESITY CAUSE THYROID CANCER? A MENDELIAN RANDOMIZATION STUDY

Obesity is an important public health issue and has been linked with many cancers (Fang *et al.* 2021). Although much has been written on observed associations between obesity and differentiated thyroid carcinoma (Dal Maso *et al.* 2000; Kitahara *et al.* 2011), it is difficult to tease apart the causal effect of obesity due to the many confounding factors acting on both obesity and thyroid cancer risk. As discussed in Chapter 1, Mendelian randomisation (MR) is a genetic epidemiological technique used to examine causal effects of proposed risk factors on various disease traits. It is well suited to investigating risk factors for cancer and has been used with some success in other cancers (Fang *et al.* 2021; Guo *et al.* 2016). In the case of thyroid cancer, it has so far been less widely used. This chapter reports the first use of MR to investigate the causal association between obesity and thyroid cancer, using genetic and observational data from the UK Biobank, a large longitudinal study of 500,000 participants with extensive demographic, health and genetic data.

The study reported in this chapter was published in *Journal of Clinical Endocrinology and Metabolism* (2020; 105(7): e2398-e2407; DOI: 10.1210/clinem/dgaa250). The list of authors and their contributions to the study is as follows:

Jonathan Mark Fussey	– Data collection, analysis, and manuscript preparation
Robin N Beaumont	– Data analysis
Andrew R Wood	– Data analysis and manuscript preparation
Bijay Vaidya	– Project oversight, manuscript preparation
Joel Smith	– Project oversight, manuscript preparation
Jessica Tyrrell	– Project oversight, data analysis, manuscript preparation

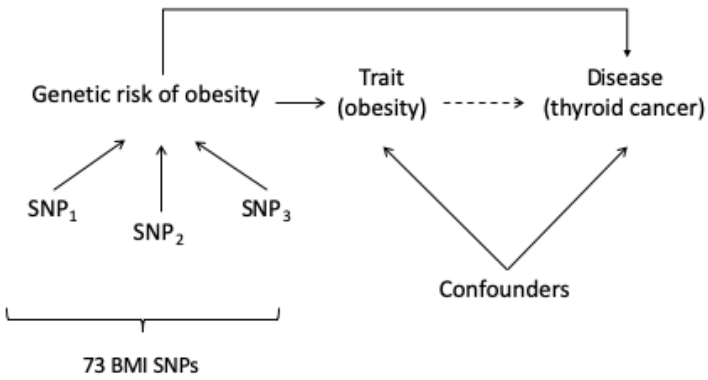
4.1 INTRODUCTION

Thyroid cancer is the commonest endocrine malignancy, and its incidence is increasing, with rates expected to rise by 74% by 2035 in the United Kingdom (Smittenaar *et al.* 2016). Along with childhood radiation exposure, a history of benign nodular thyroid disease is an important risk factor for differentiated thyroid cancer, suggesting that rather than representing distinct entities, benign nodules and thyroid cancers may fall on a spectrum of thyroid neoplasia (Franceschi *et al.* 1999; Kitahara *et al.* 2018).

Amongst the frequently cited risk factors for malignancy in a patient presenting with a thyroid nodule is obesity (Perros *et al.* 2014), which is itself a growing public health concern. Large pooled analyses of case control studies (6,796 and 848,932 participants respectively) have demonstrated an association between body mass index (BMI) and thyroid cancer risk in men and women (Dal Maso *et al.* 2000; Kitahara *et al.* 2011), and two subsequent meta-analyses have corroborated these findings (Ma *et al.* 2015; Schmid *et al.* 2015). In addition, there is also some evidence from retrospective studies that obesity is associated with more aggressive features of thyroid cancers, including larger tumour size, extrathyroidal extension, more advanced tumour stage and persistent disease following treatment (Choi *et al.* 2015; Kim *et al.* 2013; Trésallet *et al.* 2014). However, these studies tend to be observational, which carries an inherent risk of bias due to confounding factors and reverse causality. For example, obese patients may be more likely to undergo thyroid function screening than the general population, and tend to have higher thyroid stimulating hormone (TSH) levels, which may be an independent risk factor for thyroid cancer (Boelaert *et al.*

2016). Efforts have been made to remove this bias by measuring BMI at the time of cytological analysis in previously undiagnosed patients with thyroid nodules, and found no association between BMI and thyroid cancer (Rotondi *et al.* 2016). Further complicating the link between obesity and thyroid cancer is the finding that type 2 diabetes mellitus (T2DM), a disease strongly associated with obesity, has been identified as a risk factor both for increased TSH levels (Tamez-Perez *et al.* 2012), and thyroid cancer (Li & Qian 2014; Yeo *et al.* 2014).

The use of genetic epidemiology to unravel environmental determinants of disease relies upon the fact that inheritance of genetic variants at conception is random and cannot be confounded by other risk factors. Mendelian Randomization (MR; Figure 1) uses genetic determinants of a trait such as obesity to test the hypothesis that the trait increases the risk of a disease (such as thyroid cancer) in the absence of bias from reverse causality and confounding (Davey Smith & Ebrahim 2003).



Chapter 4 Figure 1: If a risk factor e.g. BMI truly causes an outcome e.g. thyroid cancer, then the genetic variants for the risk factor should also be associated with the outcome. Unlike the observed risk factor, the genetic variants are not susceptible to confounding by other risk factors as they are assigned at conception.

Genome-wide association studies have identified many polymorphisms associated with obesity (Yengo *et al.* 2018), which can be used to construct individual genetic risk scores and perform MR. Here, we test the hypothesis that obesity and other adiposity-related factors increase the risk of benign nodular thyroid disease and differentiated thyroid cancer. We used cancer registry data and clinical and genetic information from 379,708 unrelated participants of European ancestry in the UK Biobank to perform one-sample MR, and 451,025 participants (without exclusion of related individuals) to perform two-sample MR.

4.2 Materials and Methods

4.2.1 Participants

The UK Biobank is a longitudinal study of 500,000 participants between the ages of 40 and 69 recruited between 2006 and 2010 (Collins 2012). Demographic and health-related information was obtained via questionnaires and interviews, and anthropometric measurements, blood pressure readings, blood urine and saliva samples were taken at enrolment (Sudlow *et al.* 2015). Extensive health data is available via linkage with national cancer and death registries and the Hospital Episode Statistics database.

Genotyping was performed by Affymetrix (Santa Clara, Ca, US) using DNA extracted from whole blood samples. Two specially designed single nucleotide polymorphism (SNP) arrays with over 95% content overlap were used: the UKBiobank Axiom array® was used for ~440,000 participants, and the UK BiLEVE Axiom array® for 50,000 participants (Welsh *et a*l. *2017*). Sample quality

control was performed by removing duplicated individuals, those identified as sex mismatches, of non-European descent, or outliers of heterozygosity, with an overall proportion of samples identified as poor quality of 0.2% (Bycroft *et al.* 2018).

4.2.2 Exposure and outcome measures

In order to investigate the effect of obesity on the risk of developing benign nodular thyroid disease and thyroid cancer, a range of obesity-related exposure measures were used, including BMI, waist-hip ratio (WHR) and WHR adjusted for BMI. These were defined from the baseline measures, with BMI calculated from measured weight and height and WHR from measured waist and hip circumference. Type 2 diabetes mellitus was also included as an exposure measure. Type 2 diabetes was identified from the baseline questionnaire as described previously (Tyrrell *et al. 2013*). In brief, participants who had ever been told they had diabetes by a doctor were identified, then excluded if they were diagnosed under the age of 35, had received insulin within the first year of diagnosis, or were diagnosed less than a year before the date of enrolment in order to reduce the risk of inadvertently including participants with type 1 diabetes mellitus. Serum lipid levels, blood glucose and glycated haemoglobin were also used as biomarkers for metabolic health and their associations with benign thyroid disease and thyroid cancer were investigated.

Thyroid cancer cases were identified by the ICD-10 code “C73” in the linked Cancer Registry data. Histology codes were utilised to refine the case definition. Specifically, participants were excluded if they had a diagnosis of medullary or

anaplastic carcinoma, thyroid lymphoma or unspecified histological subtypes, due to the different aetiology and pathophysiology of these subtypes. Finally, the date of first diagnosis was used to ensure that duplicated records and recurrences were excluded. This resulted in 638 cases including related individuals (and 425 after exclusion of related individuals). Benign nodular thyroid disease was defined using the ICD-10 code “D34” from Hospital Episode Statistics data. Patients with toxic nodules were excluded. This resulted in 2149 cases including related individuals (and 1812 after exclusion of related individuals). Date of diagnosis was used to exclude individuals diagnosed with both benign nodular thyroid disease and thyroid cancer within 12 months, in order to exclude cases of diagnostic uncertainty. Separate control groups were used for the two cohorts in order to maximise comparability and reduce confounding, so that controls had no history of any cancer (n=310,176), and no history of nodular thyroid disease (n=377,896) respectively.

4.2.3 Observational analysis

Initial analysis was performed to ascertain demographic differences between cases and controls. We investigated associations between clinical measures of obesity and related metabolic parameters (including serum HDL, LDL, triglycerides, glucose and glycated haemoglobin), and thyroid cancer and benign nodular thyroid disease using age and sex adjusted logistic regression models. Only factors observationally associated with either benign nodular thyroid disease or thyroid cancer were taken forward for Mendelian randomization.

4.2.4 GENETIC DATA

Genetic variants for each potential risk factor were taken from published genome wide association studies (GWAS) that excluded the UK Biobank. Seventy-three genetic variants associated with BMI were identified from the GIANT consortium study of up to 338,224 individuals (Locke *et al.* 2015), after exclusion of those not associated with BMI in all people of European ancestry, those reaching genome-wide levels of statistical significance in only one sex, and those known to be classified as a secondary signal within a locus. In a similar fashion, variants for WHR, waist hip ratio adjusted for BMI (adjWHR) (Pulit *et al.* 2018) and type two diabetes risk (Mahajan *et al.* 2018) were identified. Finally, 14 common genetic variants associated a high body fat percentage but low metabolic disease risk, as identified by Yaghotkar *et al* (Yaghootkar *et al.* 2016) were used as favourable adiposity (FA) variants. Weighted genetic risk scores (GRS) for each potential risk factor trait were produced using the size of the effect of each variant as reported in the primary GWAS, and rescaled according to the number of trait-causing alleles. All genetic risk scores strongly associated with their corresponding traits, with all f statistics greater than 10.

4.2.5 GENETIC ANALYSIS

Mendelian randomization (MR, Figure 1) utilises genetic variants as instrumental variables, which are associated with an outcome only through their association with a particular risk factor (for example measured BMI) (Lawlor *et al.* 2008). MR relies upon three assumptions: first, that the instrumental variable is associated with the risk factor of interest; second, that the instrumental variable is not affected by the confounding factors acting upon the association between the risk

factor and outcome of interest; and finally, that the instrumental variable is associated with the outcome of interest *only* via its effect on the modifiable risk factor.

One-sample MR was performed in two stages using the GRS. First, the strength of the instrumental variables was tested by regressing against the corresponding observed risk factors, using linear regression for continuous predictors and logistic regression for binary predictors, with adjustment for age, sex, assessment centre, ancestral principal components and genotyping platform. Predicted values and residuals from this regression model were saved (representing unconfounded estimates of variation in observed variants), and used in the second stage as the independent variable, with disease status as the dependent variable. Robust standard errors were used to allow for uncertainty in the estimate. Sensitivity analysis was performed by excluding patients with a background of hypothyroidism or hyperthyroidism (table 1).

Analysis	Exposure	OR thyroid cancer	95% CI	OR Benign nodular thyroid disease	95% CI
Primary analysis	BMI	1.18	0.55 – 2.52	0.87	0.61 – 1.25
	WHR	1.45	0.55 – 3.85	1.52	0.93 – 2.50
	adjWHR	0.80	0.53 – 1.21	1.07	0.86 – 1.32
	T2DM	1.22	0.89 – 1.69	0.99	0.84 – 1.16
	FA	0.99	0.95 – 1.03	1.01	0.98 – 1.08
Sensitivity analysis	BMI	1.18	0.55 – 2.52	0.86	0.60 – 1.22
	WHR	1.45	0.55 – 3.85	1.55	0.94 – 2.54
	adjWHR	0.79	0.52 – 1.19	1.06	0.86 – 1.32
	T2DM	1.23	0.98 – 1.70	0.99	0.84 – 1.16
	FA	0.99	0.95 – 1.09	1.01	0.99 – 1.04

Chapter 4 Table 1: Sensitivity analysis for 1-sample MR with exclusion of patients with a diagnosis of hyperthyroidism or hypothyroidism. (BMI=body mass index; WHR=waist-to-hip ratio; adjWHR=waist-to-hip ratio adjusted for BMI; T2DM=type two diabetes mellitus; FA=favourable adiposity)

Two-sample MR was also performed in a larger related subsample (n=451,025). Here, GWAS of the thyroid cancer and benign nodular thyroid disease variables were performed using BOLT-LMM software to correct for inter-relatedness (Loh *et al.* 2015). Known genetic variants for our predictor traits of interest were then extracted and three different methods of 2-sample MR performed. Firstly, inverse variant weighted (IVW) instrumental variable analysis. IVW assumes that all genetic instruments are valid and is therefore susceptible to horizontal pleiotropy whereby variants have an effect on the outcome via a route other than the risk factor of interest. To reduce this potential source of bias, we also used the MR-Egger and Median MR techniques (Bowden *et al.* 2016*a;* Bowden *et al.* 2016*b*) which are more robust to pleiotropy. In MR-Egger analysis the intercept is unconstrained to remove the assumption that all variants are valid instrumental variables and allow a weighted regression. This reduces the possibility of variants

having a stronger effect on the outcome than the exposure trait. Median-MR uses the median instrumental variable from all included variants, and allows for up to 50% of the variants to be invalid, and thus is also more resistant to pleiotropy.

4.3 Results

Basic demographic information on included participants and controls is displayed in table 2. Participants in both the benign nodular thyroid disease and thyroid cancer groups were more likely to be female and older at recruitment than controls. Other demographics including smoking status, Townsend Deprivation Index, alcohol intake, mean BMI and waist hip ratio (WHR) were not different between the thyroid cancer and control groups. There were however differences in the benign nodular thyroid disease group, with cases more likely to be obese, more likely to have T2DM, less likely to be current smokers, have a lower alcohol intake, be more deprived (based on the Townsend Deprivation Index) and have a lower WHR. Of the 425 patients with a diagnosis of differentiated thyroid cancer, 117 had a previous history of benign nodular thyroid disease.

Characteristic	Benign nodular thyroid disease	Controls	p-value	Thyroid cancer	Cancer-free Controls	p-value
N	1,812	377,896		425	310,176	
Female sex	1,492 (82.34)	203,244 (53.78)	$<1x10^{-15}$	316 (74.35)	165,537 (53.37)	$5.6x10^{-17}$
Mean age at recruitment (SD), years	58.92 (7.39)	57.23 (8.01)	$<1x10^{-15}$	57.72 (7.29)	56.54 (8.04)	0.0022
Smoking status:						
Never	976 (53.86)	203,242 (53.78)		233 (54.82)	169,785 (54.74)	
Former	649 (35.82)	133,744 (35.40)	0.04	152 (35.76)	107,027 (34.51)	0.40
Current	166 (9.16)	35,774 (9.47)		32 (7.53)	29,261 (9.43)	
Missing	21 (1.16)	5,136 (1.36)		8 (1.88)	4,103 (1.32)	
Mean Townsend Deprivation index (SD)	-1.16 (3.09)	-1.48 (2.99)	$4.8x10^{-8}$	-1.63 (2.94)	-1.45 (2.99)	0.42
Inverse normalised mean units of alcohol per week (SD)	0.20 (1.07)	0.57 (1.04)	$2.5x10^{-6}$	0.37 (1.00)	0.57 (1.04)	0.48
Mean BMI (SD), kg/m^2	27.96 (5.23)	27.38 (4.78)	$4.1x10^{-12}$	27.55 (5.34)	27.38 (4.77)	0.16
Obese: BMI 30-40 kg/m^2	471 (25.99)	83,503 (22.10)	$1.6x10^{-9}$	111 (26.12)	68,563 (22.10)	0.068
Mean waist hip ratio (SD)	0.85 (0.087)	0.87 (0.090)	$1.4x10^{-9}$	0.85 (0.90)	0.87 (0.09)	0.083
Type 2 Diabetes	70 (3.86)	11,996 (3.17)	0.0037	16 (3.76)	9,290 (3.0)	0.16

Chapter 4 Table 2: Summary of demographic characteristics of included participants. Values stated are numbers (percentages), unless otherwise stated. P-values calculated using logistic regression adjusted for age and sex.

4.3.1 Observational associations

4.3.1.1 Benign nodular thyroid disease

Higher adiposity was associated with a higher odds of benign nodular thyroid disease, for example obese patients (defined by a BMI between 30-40kg/m^2) had a 1.47 higher odds (95% CI 1.30-1.67) than those with a normal BMI, and each 1 standard deviation increase in BMI (4.8 kg/m^2) resulted in a 1.16 higher odds of benign nodular thyroid disease (95% CI 1.11-1.22). T2DM was also associated with benign nodular thyroid disease (Table 3). WHR and adjWHR were

associated with benign nodular thyroid disease, although some sex differences were observed, with stronger associations in men than in women (Table 4).

Measured serum high-density lipoprotein (HDL) and low-density lipoprotein (LDL) levels were both associated with a lower odds of benign nodular thyroid disease, and higher serum triglyceride levels were associated with a marginally higher odds (Table 3). There was no significant association between serum glucose levels and benign nodular thyroid disease (Table 3).

Trait	Odds ratio benign nodular disease	95% Confidence interval	p-value	Odds ratio thyroid cancer	95% Confidence interval	p-value
Obesity (BMI 30-40kg/m^2)	1.47	1.30 – 1.67	$1.6x10^{-9}$	1.26	0.98 – 1.62	$6.8x10^{-2}$
BMI	1.16	1.11 – 1.22	$4.1x10^{-10}$	1.04	0.95 – 1.15	0.39
WHR	1.15	1.10 – 1.21	$1.4x10^{-9}$	1.09	0.99 – 1.20	$8.3x10^{-2}$
AdjWHR	1.09	1.04 – 1.15	$1.4x10^{-4}$	1.07	0.98 – 1.81	0.14
T2DM	1.40	1.04 – 1.90	$2.9x10^{-2}$	1.22	0.62 – 2.41	0.57
Serum HDL	0.60	0.52 – 0.69	$3.1x10^{-12}$	0.59	0.44 – 0.80	$7.2x10^{-4}$
Serum LDL	0.89	0.84 – 0.94	$3.6x10^{-5}$	0.93	0.83 – 1.04	0.19
Serum triglycerides	1.05	1.00 – 1.11	$3.7x10^{-2}$	1.04	0.94 – 1.15	0.50
Serum glucose	1.01	0.98 – 1.06	0.42	1.06	0.98 – 1.14	0.15

Chapter 4 Table 3: Observational associations between measured traits and benign nodular thyroid disease and thyroid cancer. P-values calculated using logistic regression adjusted for age and sex. Continuous variables are reported as odds ratio per 1 standard deviation higher predictor.

Characteristic	Odds ratio benign nodular disease	95% Confidence interval	p-value	Odds ratio thyroid cancer	95% Confidence interval	p-value
Female BMI						
Measured	1.18	1.12 – 1.24	$3.5x10^{-10}$	1.06	0.95 – 1.19	0.28
1-sample MR	0.93	0.63 – 1.38	0.71	0.87	.036 – 2.09	0.75
Male BMI						
Measured	1.08	0.96 – 1.20	0.19	0.98	0.81 – 1.19	0.85
1-sample MR	0.80	0.34 – 1.88	0.6	3.13	0.73 – 13.43	0.12
Female WHR						
Measured	1.14	1.08 – 1.20	$9.0x10^{7}$	1.05	0.94 – 1.17	0.39
1-sample MR	1.45	0.85 – 2.50	0.17	1.05	0.33 – 3.30	0.94
Male WHR						
Measured	1.24	1.11 – 1.38	$1.4x10^{-4}$	1.21	1.00 – 1.46	$4.5x10^{-2}$
1-sample MR	1.69	0.50 – 5.66	0.4	3.67	0.56 – 24.12	0.18
Female WHR adjusted for BMI						
Measured	1.07	1.02 – 1.12	$1.1x10^{-2}$	1.03	0.91 – 1.13	0.81
1-sample MR	1.00	0.79 – 1.26	0.97	0.74	0.45 – 1.20	0.22
Male WHR adjusted for BMI						
Measured	1.22	1.09 – 1.36	$4.0x10^{-4}$	1.27	1.06 – 1.53	$1.2x10^{-2}$
1-sample MR	1.36	0.80 – 2.29	0.25	1.04	0.46 – 2.33	0.93

Chapter 4 Table 4: Observational and 1-sample MR associations between BMI, WHR and type 2 diabetes and benign nodular thyroid disease and thyroid cancer in men and women. P-values calculated using logistic regression adjusted for age and sex.

4.3.1.2 THYROID CANCER

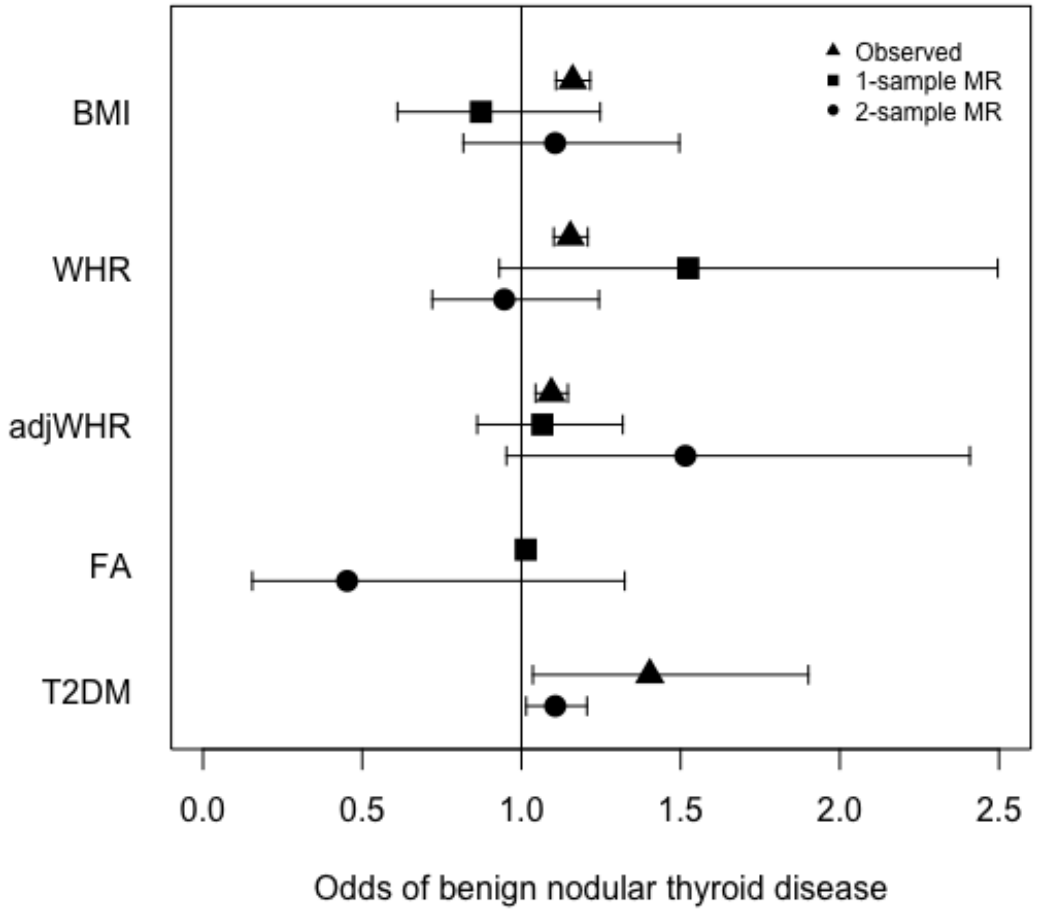
Obesity was trending towards higher odds of thyroid cancer (OR 1.26; 95% CI 0.98-1.62; Table 3), but in all individuals a per unit higher BMI was not associated with thyroid cancer. Similarly, there was no association between measured WHR, adjWHR and T2DM and thyroid cancer.

As in benign nodular thyroid disease, a higher serum HDL level was associated with a lower odds of thyroid cancer, with each 1 standard deviation (0.38mmol/L) higher HDL associating with 0.59 lower odds of thyroid cancer (95% CI 0.44-0.80). LDL and triglyceride levels were not associated with thyroid cancer (Table 3).

4.3.2 Mendelian randomization

4.3.2.1 – Benign nodular thyroid disease

There was some evidence that a higher genetic liability to T2DM caused benign thyroid disease (OR 1.11; 95% CI 1.01-1.21; Figure 2 and Table 5). The effect estimate was consistent across the more pleiotropy-resistant methods, but the confidence intervals crossed the null. 1-sample MR was not performed for T2DM due to the difficulty in interpreting MR using binary variables. There was no evidence that BMI, WHR, or favourable adiposity cause benign nodular thyroid disease, however in all cases the confidence intervals overlapped the observational estimates (Figure 2). There was tentative evidence for a causal relationship between genetically instrumented higher WHR adjusted for BMI and benign nodular thyroid disease (OR 1.51; 95% CI 0.95-2.41; p=0.085). One-sample MR did not reveal evidence of a causative role for HDL or LDL in benign nodular thyroid disease, however it revealed a protective effect of serum TG (OR 0.69; 95% CI 0.53 – 0.91). Two-sample MR did not demonstrate any associations between biomarkers and benign thyroid disease.



Chapter 4 Figure 2: Forest plot showing observed, 1 sample MR and 2 sample MR (instrumental variable analysis) associations between BMI, WHR, adjWHR, favourable adiposity, T2DM and benign nodular thyroid disease. Odds ratios and 95% confidence intervals are shown (there are no observational estimates for favourable adiposity because this phenotype of higher adiposity and lower risk of metabolic disease can only be tested in this context with genetic variants. 1-sample MR was not performed for T2DM due to the difficulty interpreting results using binary variables).

	IVW		Egger			WM		PWM	
Trait	OR (95% CI)	p-value	OR (95% CI)	p-value	Int P	OR (95% CI)	p-value	OR (95% CI)	p-value
BMI	1.10 (0.82-1.50)	0.51	0.59 (0.31-1.12)	0.11	0.03	0.67 (0.43-1.01)	0.07	0.67 (0.43-1.04)	0.08
WHR	0.95 (0.72-1.24)	0.69	1.06 (0.48-2.33)	0.89	0.76	0.90 (0.61-1.33)	0.59	0.90 (0.60-1.34)	0.60
WHR adjusted for BMI	1.52 (0.95-2.41)	0.08	0.54 (0.11-2.60)	0.45	0.18	1.06 (0.62-1.83)	0.83	1.06 (0.63-1.80)	0.82
Favourable adiposity	0.45 (0.15-1.32)	0.17	0.02 (0.00-0.32)	0.02	0.04	0.43 (0.13-1.46)	0.18	0.42 (0.12-1.52)	0.19
T2DM	1.11 (1.01-2.21)	0.02	1.12 (0.95-1.33)	0.17	0.84	1.11 (0.97-1.26)	0.12	1.11 (0.98-1.27)	0.11

Chapter 4 Table 5: Results of two-sample MR analysis for benign nodular thyroid disease. Odds ratios, 95% confidence intervals and p-values for the inverse variant weighted (IVW), Egger, weighted median (WM) and penalised weighted median (PWM) analyses are displayed. Int P represents the p-intercept of Egger analysis, which is a measure of horizontal pleiotropy.

4.3.2.2 THYROID CANCER

MR analysis provided no evidence for a causal role for BMI, WHR, WHR adjusted for BMI or favourable adiposity, according to neither raw GRS (Figure 3; Table 6), nor when analysed by quartile for GRS. Although conventional MR approaches did not provide evidence that a higher genetic liability for T2DM caused thyroid cancer, individuals in the top 25% of the GRS for T2DM were at higher odds of thyroid cancer (OR 1.45; CI 1.11-1.90; Table 7). Using 1-sample and 2-sample MR, no causative role was found for LDL, HDL or TG levels.

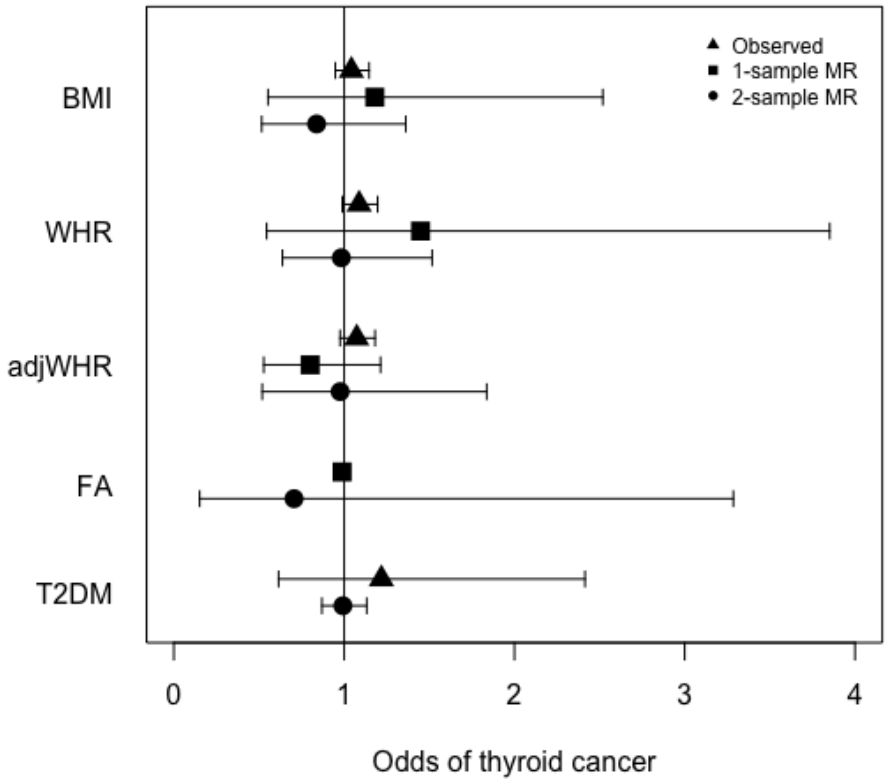


Figure 3: Forest plot showing observed, 1 sample MR and 2 sample MR (instrumental variable analysis) associations between BMI, WHR, adjWHR, favourable adiposity, T2DM and thyroid cancer. Odds ratios and 95% confidence intervals are shown. (There are no observational estimates for favourable adiposity because this phenotype of higher adiposity and lower risk of

metabolic disease can only be tested in this context with genetic variants. 1-sample MR was not performed for T2DM due to the difficulty interpreting results using binary variables).

	IVW		Egger			WM		PWM	
Trait	OR (95% CI)	p-value	OR (95% CI)	p-value	Int P	OR (95% CI)	p-value	OR (95% CI)	p-value
BMI	0.84 (0.52-1.36)	0.48	0.70 (0.24-2.01)	0.51	0.71	1.04 (0.46-2.34)	0.93	1.04 (0.46-2.32)	0.93
WHR	0.98 (0.64-1.52)	0.94	1.46 (0.42-5.11)	0.55	0.51	1.04 (0.53-2.04)	0.91	1.07 (0.56-2.04)	0.84
WHR adjusted for BMI	0.98 (0.52-1.84)	0.94	0.51 (0.06-4.34)	0.54	0.54	1.27 (0.51-3.17)	0.61	1.32 (0.54-3.20)	0.54
Favourable adiposity	0.71 (0.15-3.29)	0.66	1.89 (0.02-162.9)	0.78	0.63	0.76 (0.09-6.53)	0.80	0.76 (0.09-6.65)	0.80
T2DM	0.99 (0.87-1.13)	0.91	0.92 (0.71-1.81)	0.50	0.43	0.94 (0.76-1.17)	0.58	0.94 (0.76-1.71)	0.59

Chapter 4 Table 6: Results of two-sample MR analysis for thyroid cancer. Odds ratios, 95% confidence intervals and p-values for the inverse variant weighted (IVW), Egger, weighted median (WM) and penalised weighted median (PWM) analyses are displayed. Int P represents the p-intercept of Egger analysis, which is a measure of horizontal pleiotropy.

	Number (%)			
Quartiles	Cases	Controls	OR (95% CI)	P
1	89 (20.9)	77,358 (24.9)	1.00	
2	101 (23.8)	77,529 (24.9)	1.14 (0.85-1.51)	0.38
3	106 (24.9)	77,701 (25.1)	1.19 (0.90-1.57)	0.23
4	129 (30.4)	77,588 (25.0)	1.45 (1.11-1.90)	0.0068

Chapter 4 Table 7: Odds ratios for thyroid cancer by quartiles of genetic risk of T2DM. Odds ratios and P values adjusted for age and sex with a logistic regression model.

4.4 Discussion

4.4.1 Principal findings

This study examined the causal role of adiposity and related traits in both benign nodular thyroid disease and thyroid cancer. There was no evidence of an association between observed BMI and thyroid cancer amongst UK Biobank participants, although there was a positive association between both observed BMI and WHR adjusted for BMI and benign nodular thyroid disease.

Other published observational studies have reported somewhat inconsistent results regarding obesity and thyroid cancer. Positive associations have been reported by pooled analysis of retrospective cohort studies (Perros *et al.* 2014) and meta-analyses (Ma *et al.* 2015; Schmid *et al.* 2015). However, these analyses relied upon a heterogeneous group of studies with varying levels of adjustment for confounding factors such as a previous diagnosis of benign nodular thyroid disease which is in itself a known risk factor for the development of thyroid cancer (Kitahara *et al.* 2018). Furthermore, a prospective study focussing on a group of patients known to have thyroid nodules measured their

BMI at the time of cytological analysis and found an inverse relationship between obesity and malignant features on cytology, and no relationship when considering a smaller subset of patients with surgical histology available (Rotondi *et al.* 2016).

Using MR, we did not find evidence of a causal link between obesity and benign nodular thyroid disease to support our observational findings. This could be due to confounding factors such as T2DM or TSH levels that increase both the risk of benign nodular thyroid disease and obesity resulting in false associations when using clinical observations, which are limited by using genetic variants. In addition to this, the MR findings do not support a causal role for obesity in thyroid cancer.

MR provided some evidence for a causal link between T2DM and benign nodular thyroid disease, which supports the observational association in the UK Biobank. Although there was no observational association between T2DM and thyroid cancer, the patients in the top quartile for genetic risk of T2DM did have a significantly higher odds of thyroid cancer, suggesting a possible causative role.

Previous observational studies investigating the link between T2DM and thyroid cancer have identified a positive association (Li & Qian 2017; Yeo *et al.* 2014). This along with the negative findings for obesity-related risk factors suggests that type 2 diabetes may be a risk factor for thyroid cancer independent of obesity. The mechanisms for this association are not clear, however insulin like growth factor 1 (IGF-1) receptors are overexpressed on thyroid cancer cells and may be activated by chronically elevated levels of circulating insulin leading to cell proliferation (Yeo *et al.* 2014). This theory is supported by the finding that thyroid nodules are associated with insulin resistance (Tang *et al.* 2017). Further work is

required to confirm the causal role of T2DM, however if confirmed it could have implications for risk stratification and management of indeterminate nodules in diabetic patients.

We also investigated the association between serum lipid levels and both benign nodular thyroid disease and thyroid cancer, using observed levels as well as genetically instrumented levels. We found no association between measured HDL or LDL and either benign nodular thyroid disease or thyroid cancer, and this was supported by the findings of the MR analysis which were negative. MR did however suggest an unexpected protective role for higher TG levels in benign nodular thyroid disease but not thyroid cancer. This association needs further validation.

4.4.2 Strengths and limitations

This represents the first Mendelian randomization study on the effect of adiposity and related traits on nodular thyroid disease and thyroid cancer. However, we acknowledge some limitations to our approach. First, we only had 425 thyroid cancer cases which limits our power for both observational and MR analyses. Second, the UK Biobank is not population representative, as participants were limited to adults between the ages of 40 and 69 living in the UK, with some evidence of healthy volunteer bias, resulting in lower rates of obesity and cancer incidence than age and sex matched members of the general population (Fry *et al.* 2017). Third, our definition of T2DM was based on self-reported type 2 diabetes, which is susceptible to recall bias, although patients recently started on insulin, or diagnosed under the age of 35 years were excluded in order to remove

those patients with type 1 diabetes. Fourth, whilst we performed sensitivity analysis excluding patients with a diagnosis of hypothyroidism or hyperthyroidism, the diagnosis of benign nodular thyroid disease covers a wide range of conditions from single non-toxic nodule to multinodular goitre. Fifth, identification of diagnoses in biobanks relies upon accurate coding. Cancer registry coding is robust in the UK, however coding of benign thyroid nodules is likely to be more sporadic, so our results are likely to under-represent benign nodular thyroid disease within the UK Biobank population. Similarly, identification of controls relied upon the lack of a record of the outcomes of interest in the cancer registry or HES registry rather than full clinical and radiological examination excluding them. As in all biobank research this results in a source of detection bias for control groups. Finally, the BMI GRS we used has been shown previously to be associated not only with BMI at enrolment in the UK Biobank, but also with self-reported obesity at the age of 10 (Yaghootkar *et al.* 2016). Other authors have suggested that body shape in middle age is a more important risk factor for thyroid cancer than body shape in early adulthood (Clavel-Chapelon *et al.* 2010). The advantage of our BMI GRS is that it predicts high BMI over an individual's lifetime, as opposed to the one-time measurements used in many observational studies, however the disadvantage is the fact that it cannot discriminate causal effects of obesity at a particular time in life.

4.4.3 Conclusions

The incidence of thyroid cancer is increasing and there is limited evidence for modifiable risk factors. Obesity and T2DM are both reported by observational studies as being risk factors, and whilst we were able to demonstrate evidence

of a causative role for T2DM using Mendelian randomization, we were unable to find a causative link between obesity and either benign nodular thyroid disease or thyroid cancer. This suggests the possibility of other factors confounding the reported observational associations in the case of obesity. More work is needed to improve our understanding of the underlying mechanisms for the associations between thyroid nodules, thyroid cancer and obesity.

CHAPTER 5:

MENDELIAN RANDOMIZATION TO INVESTIGATE THE LINK BETWEEN TSH AND THYROID CANCER

Following on from my work using genetic variants to investigate the causal effect of obesity and type 2 diabetes (T2DM) on thyroid cancer risk, I decided to focus on another risk factor proposed in the observational literature: serum thyroid stimulating hormone (TSH) level. The causal association between TSH levels and thyroid cancer is far from clear based on observational studies alone, as many the other proposed metabolic risk factors for thyroid cancer (such as, obesity and T2DM) are also closely related to TSH levels. Furthermore, TSH is often only tested when a diagnosis of thyroid cancer has been made or is suspected, which introduces the possibility of reverse causality. There are well-described genetic variants associated with TSH levels (Porcu *et al.* 2013), so I decided to use these with MR analysis to explore the causal association between TSH and thyroid cancer. .At the time of data analysis and manuscript preparation for this project there were no previously published studies using MR to investigate the effect of TSH on thyroid cancer risk, however two such studies have recently been published (Zhou *et al.* 2020; Yuan *et al.* 2020a), allowing us to validate our findings.

This chapter is a modified version of a report published in *Endocrine Related Cancer* as a research letter (2021, DOI: 10.1530/ERC-21-0156). The author list and contributions is as follows:

Jonathan Mark Fussey	– Data collection, analysis and manuscript preparation
Robin N Beaumont	– Data analysis
Andrew R Wood	– Data analysis
Bijay Vaidya	– Project oversight, manuscript preparation
Joel Smith	– Project oversight, manuscript preparation
Jessica Tyrrell	– Project oversight, data collection, analysis and manuscript preparation

5.1 – INTRODUCTION

Thyroid cancer is the commonest endocrine malignancy, with recent years seeing a significant increase in incidence (Kitahara & Sosa 2016). The majority of patients present with a thyroid nodule; however, nodules are common with a prevalence of up to 34% on ultrasound (Moon *et al.* 2018), and only 5-10% of nodules are malignant (Hegedüs 2004; Perros *et al.* 2014). Following clinical assessment, ultrasonography and cytological analysis of fine needle aspiration biopsy the true nature of the nodule remains indeterminate in up to 25% of cases, and up to one third of surgically excised nodules are in fact benign (Bongiovanni *et al.* 2012). There is therefore a need for better stratification of thyroid nodules according to risk of malignancy, and whilst recent years have seen the development of genomic tests to aid diagnosis (Nikiforov & Baloch 2019), these are yet to be widely used in clinical practice.

In addition to ultrasonography and fine needle aspiration biopsy, thyroid function tests including serum thyroid stimulating hormone (TSH) level are recommended in the investigation of a thyroid nodule (Perros *et al.* 2014; Haugen *et al.* 2016). The role of serum TSH testing is primarily to identify hyperthyroidism caused by a toxic nodule, however several authors have reported an association between raised serum TSH and thyroid cancer (Boelaert *et al.* 2006; Zafón *et al.* 2015; Golbert *et al.* 2017; Fighera *et al.* 2015) as well as a correlation between serum TSH and aggressiveness of thyroid cancer (Haymart *et al.* 2008; Haymart *et al.* 2009; McLeod *et al.* 2014).

Despite the strong observational basis for an association between serum TSH and thyroid cancer in patients with thyroid nodules, there is significant difficulty in

establishing a causal link due to the difficulty in adequately adjusting for confounding factors such as iodine deficiency, autoimmune thyroid disease and cigarette smoking (Zimmermann & Galetti 2015; Singh *et al.* 1999; Cho *et al.* 2018), and the possibility of reverse causality. This difficulty can be overcome by using genetic epidemiological techniques such as Mendelian Randomization (MR), which relies upon the fact that genetic variants predisposing to a certain trait are inherited randomly at conception and are not affected by many confounders acting upon observational associations. The aim of this study was to test the hypothesis that elevated serum TSH has a causal role in the development of benign thyroid nodules and differentiated thyroid cancer, using genetic data from 451,025 participants in the UK Biobank.

5.2 – METHODS

5.2.1 – PARTICIPANTS

We used data from the UK Biobank, a longitudinal study of 500,000 participants aged between 40 and 69 years who were recruited between 2006 and 2010 (Collins 2012; Sudlow *et al.* 2015). After giving informed written consent, participants gave blood, urine and saliva samples for biomarker measurement and genotyping, as well as having anthropometric measurements at the time of enrolment. In addition, patients filled out demographic and health questionnaires and consented to record linkage with national cancer and death registries as well as the Hospital Episode Statistics (HES) database. UK Biobank has received ethics approval from the National Health Service National Research Ethics Service (ref 11/NW/0382).

Genotyping was performed on DNA extracted from whole blood samples by Affymetrix (Santa Clara, Ca, US), using two specially designed single nucleotide polymorphism (SNP) arrays with over 95% content overlap. The UKBiobank Axiom array® was used for ~450,000 participants, and the UK BiLEVE Axiom array® for 50,000 participants (Welsh *et al*. 2017). Extensive central quality control was carried out by the UK Biobank (Bycroft *et al.* 2018). Genotyping data was subjected to quality control, with exclusion of sex mismatches (genetic sex different from reported sex), duplicated individuals and outliers of heterozygosity. After these exclusions, as well as those individuals who have subsequently withdrawn from the UK Biobank Study, 451,025 individuals of White European descent were identified using principal components computed from the 1000 Genomes Project. Kinship coefficients provided by UK Biobank were then used to remove related individuals up to three degrees of separation, resulting in a maximal subset of 379,708 unrelated individuals for sensitivity analysis. Principal components were generated amongst this group of individuals to adjust for genetic ancestry.

5.2.2 – EXPOSURE AND OUTCOME MEASURES

We identified two outcome groups of interest within the UK Biobank: those with a diagnosis of differentiated thyroid cancer and those with a diagnosis of benign nodular thyroid disease. In order avoid reporting bias, we discounted self-reported diagnoses in the health questionnaire and used Cancer Registry and HES data only. Thyroid cancer cases were identified by the ICD-10 code “C73”. Histology codes were then used to exclude patients with medullary thyroid carcinoma, anaplastic carcinoma, thyroid lymphoma and unclassified subtypes, leaving only patients with differentiated thyroid carcinoma. This resulted in 462

cases. Patients with a diagnosis of benign nodular thyroid disease were identified by the ICD-10 code “D34”. Patients diagnosed with both nodular thyroid disease and thyroid cancer were excluded to avoid bias caused by misdiagnosis (n=147). This left 2031 cases in total. Control groups with no previous diagnosis of cancer, and no previous diagnosis of benign thyroid disease were also created, numbering 391,458 and 448,532 respectively.

The exposure of interest in this study was a genetic predisposition to higher TSH levels. In order to quantify this, we utilised 20 SNPs associated with TSH levels at genome wide significance (p-value threshold of $5x10^{-8}$) in a meta-analysis of 26,420 individuals of European ancestry across 18 cohorts (Porcu *et al.* 2013) (Table 1). Linkage disequilibrium (LD) was calculated using the LDlink SNPClip application, with an R^2 threshold of 0.1 and a minor allele frequency (MAF) threshold of 0.01 (Machiela & Chanock 2015). No SNPs were removed due to LD. The 20 SNPs have been reported to account for 5.64% of variance in serum TSH levels (Porcu *et al.* 2013). The effect of the included SNPs on thyroid-specific gene expression was investigated using the Genotype-Tissue Expression Consortium tool (www.gtexportal.org; last accessed 8th June 2020) and is shown in table 1.

Chromosome	Position	Effect/Other allele	SNP	Effect size	Standard error	Thyroid-specific gene expression
5	76566105	A/G	rs6885099	-0.141	0.009	No
6	165966473	C/G	rs753760	0.100	0.010	Yes
1	19713761	A/G	rs10799824	-0.113	0.012	Yes
16	78306854	T/C	rs3813582	0.082	0.010	Yes
6	43919740	T/C	rs9472138	-0.079	0.010	Yes
6	44012758	T/C	rs11755845	-0.065	0.010	Yes
4	149888956	T/C	rs10032216	0.087	0.011	Yes
2	217333768	A/G	rs13015993	0.078	0.010	Yes
17	67639131	T/C	rs9915657	-0.064	0.009	Yes
1	61393084	A/G	rs334699	-0.141	0.021	No
15	47533656	A/T	rs10519227	-0.072	0.011	Yes
11	45184143	T/C	rs17723470	-0.065	0.010	Yes
15	86920108	A/G	rs17776563	-0.060	0.010	Yes
19	7174848	T/G	rs4804416	-0.057	0.009	No
9	135129086	A/C	rs657152	0.058	0.009	No
14	92665344	A/C	rs11624776	-0.064	0.011	Yes
8	32535816	A/G	rs7825175	-0.066	0.001	Yes
14	35643769	T/C	rs1537424	-0.052	0.009	Yes
6	148562985	T/C	rs9497965	0.051	0.009	No
9	4257209	A/G	rs1571583	0.057	0.010	No

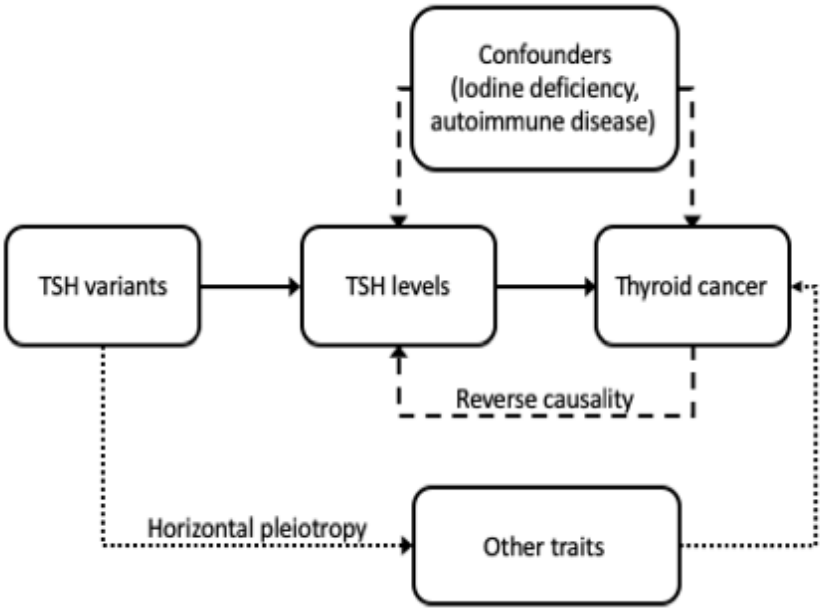
Chapter 5 Table 1: TSH SNPs, their effect sizes on TSH (modified from Porcu et al 2013) and their effect on thyroid specific gene expression.
Positions given are in build 36. Tissue specific expression information from the Genotype-Tissue Expression Consortium (www.gtexportal.org)

5.2.3 – DATA ANALYSIS

In order to identify the causal effect of TSH on thyroid cancer and benign nodular thyroid disease we performed two-sample Mendelian randomization (MR) (Bowden *et al.*2016 *a*) using the 20 known genetic variants associated with TSH (Porcu *et al.* 2013). We corrected for inter-relatedness by performing a GWAS of the thyroid cancer and benign nodular thyroid disease variables using a linear mixed model in the BOLT-LMM software (Loh *et al.* 2015). The 20 known TSH

variants were then extracted from these GWAS analyses for use in the two-sample MR. The SNP-TSH association was taken from the primary GWAS by Porcu *et al.*, whilst the SNP-outcome association was taken from the BOLT-LMM GWAS of our UK Biobank outcome traits. Using the median standard deviation (0.8mIU/L) from the meta-analysis of studies used to identify genetic variants associated with TSH (Porcu *et al.* 2013), the coefficients from the two-sample MR were adjusted to approximate the odds ratios per unit increase in TSH.

MR relies upon the fact that the instrumental variable being used is associated with the risk factor of interest, that the instrumental variable is not subject to the confounding factors acting upon the association between the risk factor and outcome of interest, and that the instrumental variable is associated with the outcome of interest *only* via its effect on the risk factor of interest. Figure 1 illustrates the principles of MR.



Chapter 5 Figure 1: Basic concept of Mendelian randomisation: if high TSH truly causes thyroid cancer, then the genetic variants for TSH should also be associated with thyroid cancer. Unlike measured TSH, these genetic variants are not susceptible to confounding by other risk factors or reverse causation (dashed lines) as they are assigned at conception. Horizontal pleiotropy (dotted line), whereby the variants of interest also affect the outcome via their association with another trait is a possible limitation of MR.

Three different methods of two-sample MR were performed. Firstly, inverse variant weighted (IVW) instrumental variable analysis. IVW assumes that all genetic instruments are valid and is therefore susceptible to horizontal pleiotropy (variants affecting the outcome via a route other than the risk factor of interest). To reduce this potential source of bias, we also used the MR-Egger and Median MR techniques (Bowden *et al.* 2016*a*; Bowden *et al.* 2016*b*) which are more robust to pleiotropy. In MR-Egger analysis the intercept is unconstrained to allow a weighted regression, thus removing the assumption that all variants are valid instrumental variables. This reduces the possibility of variants having a stronger effect on the outcome than the exposure itself (actual TSH levels). Median-MR is

also more resistant to pleiotropy, as it uses the median instrumental variable from all included variants, and allows for up to 50% of the variants to be invalid.

In order to allow for sensitivity analyses, the effect sizes of the 20 included TSH-associated SNPs were used to produce a genetic risk score (GRS) for TSH in the smaller subset of 379,708 unrelated UK Biobank participants. This group contained 425 cases of differentiated thyroid cancer and 1,812 cases of benign nodular thyroid disease. The association between these GRSs and the outcome traits of benign nodular thyroid disease and thyroid cancer was then tested using a logistic regression model with adjustment for age, sex, ancestral principal component and recruitment centre. Robust standard errors were used to allow for uncertainty in the estimate. In addition, the 379,708 unrelated participants were divided into quartiles based on their GRS for TSH, and logistic regression analysis was used to test the association between GRS quartile and both disease outcomes. Finally, sensitivity analysis was performed to account for patients with a diagnosis of either hyper- or hypothyroidism as identified in the self-reported health questionnaire or the Hospital Episode Statistics data. A *post hoc* power calculation using a Mendelian randomisation power calculator (Brion *et al.* 2013) was performed to calculate the power of the study to detect a true odds ratio of 0.2 for the binary outcome of benign nodular thyroid disease per one standard deviation increase in genetically determined TSH (Power=1.00; α=0.05; non-centrality parameter =73.67; F statistic=26959.26); and a true odds ratio of 0.5 for the binary outcome of thyroid cancer per one standard deviation increase in genetically determined TSH (Power 0.71; α=0.05; non-centrality parameter=6.36; F statistic 26959.26).

5.3 – RESULTS

Table 2 displays basic information on the unrelated subgroup of participants and controls. Participants with a diagnosis of benign nodular thyroid disease were more likely than controls to be female, have a higher BMI and have co-existent type two diabetes mellitus (T2DM), hyper- or hypothyroidism. Those with a diagnosis of thyroid cancer were also more likely than controls to be female, and marginally more likely to be obese, with a higher proportion than controls having co-existent hyper- or hypothyroidism.

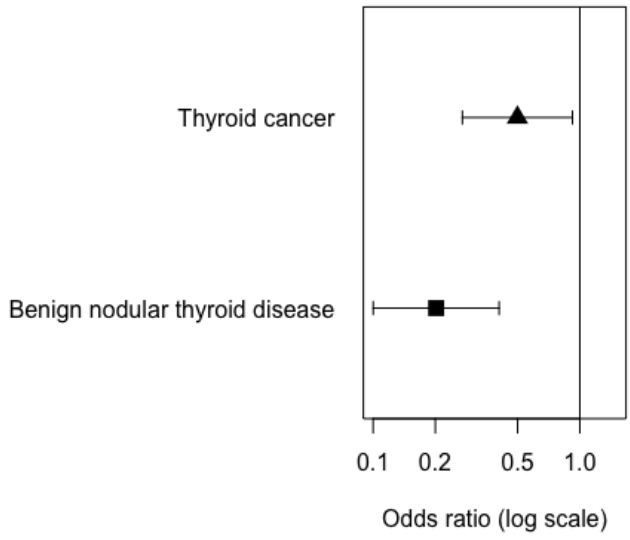
Characteristic	Benign nodular thyroid disease	Controls	p-value	Thyroid cancer	Cancer-free Controls	p-value
N	1,812	377,896		425	310,176	
Female sex	1,492 (82.34)	203,244 (53.78)	$<1x10^{-15}$	316 (74.35)	165,537 (53.37)	$5.6x10^{-17}$
Mean age at recruitment (SD), years	58.92 (7.39)	57.23 (8.01)	$<1x10^{-15}$	57.72 (7.29)	56.54 (8.04)	0.0022
Smoking status:						
Never	976 (53.86)	203,242 (53.78)	0.04	233 (54.82)	169,785 (54.74)	0.40
Former	649 (35.82)	133,744 (35.40)		152 (35.76)	107,027 (34.51)	
Current	166 (9.16)	35,774 (9.47)		32 (7.53)	29,261 (9.43)	
Missing	21 (1.16)	5,136 (1.36)		8 (1.88)	4,103 (1.32)	
Mean Townsend Deprivation index (SD)	-1.16 (3.09)	-1.48 (2.99)	$4.8x10^{-8}$	-1.63 (2.94)	-1.45 (2.99)	0.42
Mean units of alcohol (SD)	0.20 (1.07)	0.57 (1.04)	$2.5x10^{-6}$	0.37 (1.00)	0.57 (1.04)	0.48
Mean BMI (SD), kg/m^2	27.96 (5.23)	27.38 (4.78)	$4.1x10^{-12}$	27.55 (5.34)	27.38 (4.77)	0.16
Obese: BMI 30-40 kg/m^2	471 (25.99)	83,503 (22.10)	$1.6x10^{-9}$	111 (26.12)	68,563 (22.10)	0.068
Type 2 Diabetes	70 (3.86)	11,996 (3.17)	0.0037	16 (3.76)	9,290 (3.0)	0.16
Hypothyroidism	182 (10.04)	18,600 (4.92)	$4.8x10^{-9}$	28 (6.59)	15,238 (4.91)	0.12
Hyperthyroidism	59 (3.26)	2,897 (0.77)	$1.6x10^{-4}$	7 (1.65)	2,377 (0.77)	0.35

Chapter 5 Table 2: Summary of demographic characteristics of participants
Table represents data after exclusion of related individuals due to the difficulty in adjusting observational analyses for relatedness. Values stated are numbers (percentages), unless otherwise stated. P-values calculated using logistic regression adjusted for age and sex.

5.3.1 – Two-sample MR

The two-sample MR analysis provided evidence that higher TSH was associated with a lower risk of both benign nodular thyroid disease and thyroid cancer. A one standard deviation (approximately 0.8 mIU/L) higher genetically instrumented

TSH level was associated with an odds ratio of 0.2 for benign nodular thyroid disease (95% CI 0.10 – 0.41) and 0.5 for thyroid cancer (95% CI 0.27 – 0.92; Figure 2 & Table 3. This approximates to an 86% reduction in the risk of benign nodular thyroid disease (OR 0.14; 95% CI 0.07 – 0.27) and a 58% reduction in the risk of thyroid cancer (OR 0.42; 95% CI 0.23 – 0.77) per 1 mlU/L higher TSH. The more pleiotropy-resistant two-sample analyses were directionally consistent, and MR-Egger intercept provided no evidence of horizontal pleiotropy for either outcome trait (Table 3).



Chapter 5 Figure 2: Forest plot showing two-sample MR (instrumental variable analysis) associations between genetically instrumented TSH and thyroid cancer and benign nodular thyroid disease. Odds ratios of outcome per standard deviation higher genetically instrumented TSH, and 95% confidence intervals are shown.

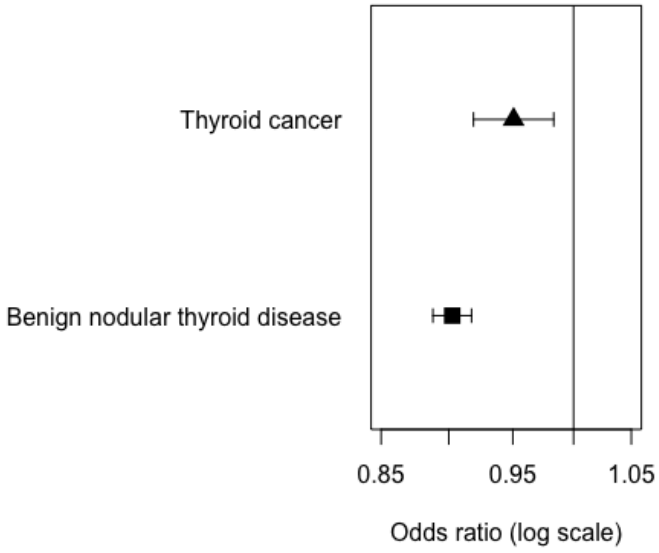
	IVW		Egger			WM	
Trait	OR (95% CI)	p-value	OR (95% CI)	p-value	Int P	OR (95% CI)	p-value
Thyroid cancer	0.50 (0.27-0.92)	0.039	0.58 (0.08-4.33)	0.60	0.87	0.53 (0.30-0.95)	0.034
Benign nodular thyroid disease	0.20 (0.10-0.41)	2.59×10^{-4}	0.68 (0.07-6.26)	0.74	0.27	0.26 (0.15-0.44)	5.41×10^{-7}

Chapter 5 Table 3: Results of two-sample MR analysis for benign nodular thyroid disease and thyroid cancer.
Odds ratios, 95% confidence intervals and P values for the inverse variance weighted (IVW), Egger, weighted median (WM) and penalised weighted median (PWM) analyses are displayed. Int P represents the p-intercept of Egger analysis, which is a measure of horizontal pleiotropy. Standard errors are displayed in parentheses.

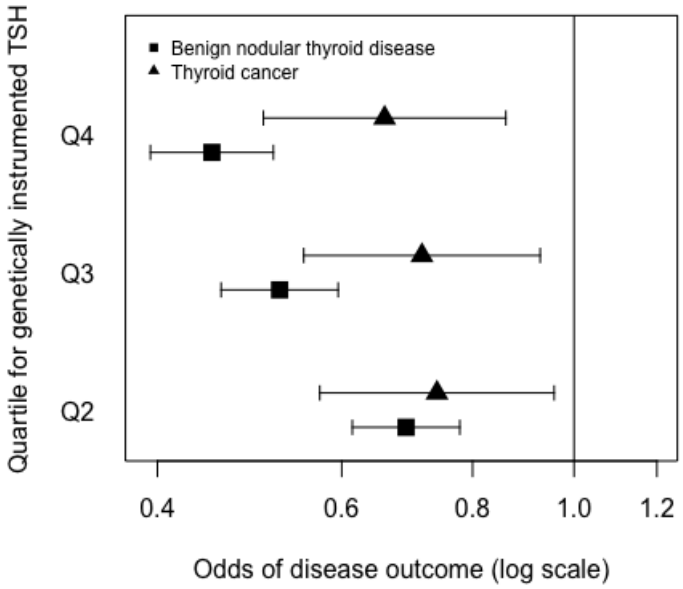
5.3.2 – Sensitivity analyses

The GRS for TSH was inversely associated with both benign nodular thyroid disease and thyroid cancer in the subgroup of unrelated participants. Using logistic regression adjusted for age, sex, ancestral principal component and assessment centre with the GRS for TSH as the dependent variable, we found a negative association with benign nodular thyroid disease, with each weighted allele increase in the TSH GRS resulting in odds ratio of 0.90 for benign nodular thyroid disease (95% CI 0.89 – 0.92; Figure 3). Similarly, there was a negative association with thyroid cancer, with an odds ratio of 0.95 (95% CI 0.92 – 0.98; Figure 3). Furthermore, individuals in the top 25% of the TSH GRS were at a lower risk of thyroid cancer (OR 0.66; 95% CI 0.50 – 0.86; Figure 4) and benign nodular thyroid disease (OR 0.45; 95% CI 0.39 – 0.52) than those in the bottom 25%. Results of the sensitivity analysis after excluding participants with hypothyroidism or hyperthyroidism, which did not affect the associations, are shown in Table 4. Finally, the analysis using TSH GRS in the unrelated subgroup

was repeated separately in male and female participants (Table 5) with the associations with both benign nodular disease and thyroid cancer remaining stable in females, but the inverse association with thyroid cancer weakening in males.



Chapter 5 Figure 3: Forest plot showing associations between genetically instrumented TSH and thyroid cancer and benign nodular thyroid disease. Odds ratios of outcome per standard deviation higher genetically instrumented TSH, and 95% confidence intervals are shown.



Chapter 5 Figure 4: Forest plot showing associations between quartiles for TSH genetic risk and thyroid cancer and benign nodular thyroid disease. Odds ratios and 95% confidence intervals are shown for each quartile as compared to those in the lowest quartile.

Analysis	Exposure	OR thyroid cancer	N= cases (controls)	95% CI	OR Benign nodular thyroid disease	N= cases (controls)	95% CI
Primary analysis	TSH	0.95	425 (310,176)	0.92 – 0.98	0.90	1,812 (377,896)	0.89 – 0.92
Sensitivity analysis	TSH	0.95	423 (309,526)	0.92 – 0.99	0.89	1,803 (377,110)	0.87 – 0.90

Chapter 5 Table 4: Results of sensitivity analysis for both thyroid cancer and benign nodular thyroid disease after exclusion of participants with co-existing diagnoses of hypothyroidism or hyperthyroidism.

Analysis	Exposure	OR thyroid cancer	N= cases (controls)	95% CI	OR Benign nodular thyroid disease	N= cases (controls)	95% CI
Primary analysis	TSH	0.95	462 (391,458)	0.92 – 0.98	0.90	1,812 (377,896)	0.89 – 0.92
Male sex	TSH	0.99	109 (144,639)	0.93 – 1.06	0.89	320 (174,652)	0.86 – 0.93
Female sex	TSH	0.94	316 (165,537)	0.90 – 0.97	0.89	1,492 (203,244)	0.87 – 0.90

Chapter 5 Table 5: Results of sensitivity analysis for both thyroid cancer and benign nodular thyroid disease according to sex.

5.4 – Discussion

5.4.1 – Principal findings

This study investigated the causal relationship between serum TSH levels and thyroid cancer risk, and benign nodular thyroid disease risk. Our results reveal a significant inverse association between genetic variants predicting higher serum TSH, and both diagnoses. Furthermore, these results were consistent when using different approaches.

The first study to identify an association between serum TSH levels and risk of thyroid cancer was published by Boelaert *et al.* in 2006 (Boelaert *et al.* 2006), and since then many authors have reported similar findings, with two meta-analyses reporting positive associations (McLeod *et al.* 2012; Zheng *et al.* 2016). However, the majority of included studies were retrospective, resulting in bias and difficulty in adjusting for all confounders. There is also some controversy on the subject, with the EPIC case-control study reporting a negative association amongst its 357 cases and 767 matched controls (Rinaldi *et al.* 2014), and a more recent case-control study on papillary thyroid carcinoma identifying a positive association in men and a negative association in women (Huang *et al.* 2017). Furthermore, differentiated thyroid cancer has been reported even in patients with completely suppressed TSH levels (Satta *et al.* 1993), although the authors note that tumours may have arisen prior to development of thyrotoxicosis. Here, we demonstrated stronger evidence of a role for TSH in thyroid cancer in women than in men. However, these analyses were limited by the low numbers of male participants with thyroid cancer in the UK Biobank study. When testing the role of TSH on benign thyroid nodules, results were similar in men and women.

The widely proposed mechanism explaining the positive association between serum TSH levels and both benign thyroid nodules and thyroid cancer reported by most studies is the thyrotropic effect of TSH. TSH is known to induce proliferation of thyrocytes, and both benign and malignant thyroid tumours have long been known to express TSH receptors (Ichikawa *et al.* 1976). Furthermore, suppression of TSH is a well-established adjunct to the management of differentiated thyroid carcinoma (Perros *et al.* 2014), and has in the past been used in the treatment of benign nodular thyroid disease although has now largely been abandoned in this setting due to long-term adverse effects (Wémeau *et al* 2002). Animal studies have demonstrated a convincing role for TSH in thyroid tumorigenesis (Lu *et al.* 2010) however the cause-effect relationship between high TSH and thyroid cancer in humans is unclear due to the difficulty in accounting for known and unknown confounders.

Despite this, a recent study of participants from Iceland, Michigan and the UK Biobank reported two-sample MR findings for 94 TSH-associated variants, and identified a protective effect of genetically instrumented TSH upon thyroid cancer risk (OR 0.55; 95% CI 0.40-0.74) and benign goitre risk (OR 0.28; 95% CI 0.20 – 0.41) (Zhou *et al.* 2020). Furthermore, they demonstrated an increased risk of thyroid cancer and benign goitre with low TSH. The authors point to the fact that most TSH-lowering variants were also associated with higher free thyroxine levels and suggest that their findings are likely to be a result of the effect of TSH-associated variants on the thyroid gland itself. Another recent Mendelian randomisation study investigating the link between TSH and all cancer sites identified a positive association between genetically instrumented TSH and

overall cancer risk, but an inverse association with thyroid cancer risk (OR 0.47; 95% CI 0.30-0.73; p = 0.001) (Yuan *et al.* 2020a). Our two-sample MR findings of inverse associations between genetically instrumented TSH and both thyroid cancer and benign nodular thyroid disease closely match the results of these authors, despite using a more select instrument of only 20 TSH-associated SNPs. These surprising results in the light of existing observational evidence may be explained by a temporal difference in the way that observational and genetic studies estimate the link between TSH and thyroid cancer risk. MR techniques use genetic predisposition to raised TSH, and therefore provide an estimate of the association between lifetime exposure to TSH and thyroid cancer, whereas observational studies measure TSH at the time of diagnosis. It may therefore be the case that TSH has differing effects at different points in life, for example suppressing nodule and cancer development in those with normal thyroid parenchyma, but promoting nodule growth and tumour progression in those with existing thyroid nodules or cancers.

The aim of this study was to test the hypothesis that TSH has a causal role in the development of thyroid cancer, using genetic techniques. Whilst our findings support those of Zhou *et al.*, they are not sufficient to change clinical practice. They do however add to the debate on the topic, and represent further evidence using a novel technique that higher TSH reduces the risk of both benign nodular thyroid disease and thyroid cancer. It is vital that the causal relationship between TSH and thyroid cancer is understood when considering previously published observational and clinical studies on the subject. Furthermore, our findings help pave the way for future research on the ability of genetic instruments to distinguish between benign and malignant thyroid nodules, with the advantage

that they are not confounded by environmental factors. This information may eventually be used along with other genetic instruments to formulate pre-diagnostic screening or stratification tools to aid clinical decision making in patients with thyroid nodules.

5.4.2 – Strengths and limitations

The advantage of the present study is the ability to limit the effect of environmental confounders by using genetic predictors of TSH levels, which are fixed at conception. The main limitation of the study is the lack of an observational measure of TSH in the UK Biobank cohort. This problem is partially overcome by using data from large genome wide association studies to confirm robust associations between the TSH SNPs we used and TSH levels in European subjects, however the UK Biobank participants are inevitably not completely comparable to those cohorts. Secondly, the power of our MR analyses is somewhat limited by the small number (462) of thyroid cancers amongst the UK Biobank participants. Finally, the UK Biobank is not representative of the wider population, including only UK residents aged between 40-69 at enrolment. There is some evidence of healthy volunteer bias within the UK Biobank, with lower rates of cancer incidence than the age and sex matched general population (Fry *et al*. 2017).

5.4.3 - Conclusions

Using MR, we have demonstrated an inverse association between genetically instrumented TSH and both benign nodular thyroid disease and differentiated thyroid cancer. This supports the findings of the only other published MR studies

on this topic, but contradicts the positive associations reported by many observational studies. Further research is required to investigate the physiological mechanism for these findings, and to investigate their clinical applicability in patients with thyroid nodules.

CHAPTER 6:

USING MENDELIAN RANDOMISATION TO EXPLORE LIFESTYLE RISK FACTORS FOR THYROID CANCER

The final part of my research using genetics to investigate risk factors for thyroid cancer focussed on lifestyle-related risk factors. Various authors have described observational inverse associations between smoking, alcohol consumption and physical exercise and thyroid cancer (Yeo *et al.* 2020; Xhaard *et al.* 2016), however the causal basis for these associations is not understood. Once again, this appeared to be a question that MR could help to answer, by using genetic variants associated with these behaviours as instrumental variables. The techniques used in this study differed from those employed in Chapters 4 and 5 due to the nature of the risk factors being investigated. Genetically instrumented smoking and alcohol consumption is only valid in people who smoke or drink alcohol, thus 2-sample MR (whereby the association between the genetic variants and the risk factor are taken from a different population to the association between the genetic variants and the outcome) could not be used. In addition, in the case of physical activity, weak instrument MR was required to account for the fact that genetic variants account for only a small proportion of variance in the trait.

The work reported in this chapter is being prepared for peer review. The author list and contributions is as follows:

Jonathan Mark Fussey	– Data collection, analysis and manuscript preparation
Francesco Cassanova	– Data analysis and manuscript preparation
Robin N Beaumont	– Data analysis
Andrew R Wood	– Data analysis
Joel Smith	– Project oversight, manuscript preparation
Bijay Vaidya	– Project oversight, manuscript preparation
Jessica Tyrrell	– Project oversight, manuscript preparation

6.1 - INTRODUCTION

The incidence of thyroid cancer is rising (Miranda-Filho *et al.* 2021), and is expected to continue to rise (Smittenaar *et al.* 2016). Differentiated thyroid carcinoma (DTC), arising from the follicular cells of the thyroid gland, is the commonest form of thyroid cancer and accounts for over 90% of all thyroid cancers (Lim *et al.* 2017). It has a very low mortality rate, with a 5-year overall survival rate of almost 100% (Howlader *et al.* 2020). However, its treatment, consisting primarily of surgery with or without post-operative radioactive iodine ablation and thyroid stimulating hormone (TSH) suppression, carries a significant risk of complications and long-term side-effects. Furthermore, the fact that it is often diagnosed at a relatively young age and carries such a favourable prognosis means that there is a high and increasing prevalence of thyroid cancer survivors requiring long-term surveillance and treatment (Aschebrook-Kilfoy *et al.* 2013)

Genetic predisposition accounts for only a small proportion of cases of differentiated thyroid carcinoma (DTC) (Yang *et al.* 2016), and whilst there are several other well-known risk factors such as ionising radiation and female sex (Furukawa *et al.* 2013; Rahbari *et al.* 2010), relatively little is known about modifiable risk factors for DTC. Many authors have reported observational associations between risk factors such as obesity and thyroid cancer (Dal Maso *et al.* 2000; Kitahara *et al.* 2011; Ma *et al.* 2015; Schmid *et al.* 2015) but the causal nature of these associations is not entirely clear (Fussey *et al.* 2020). Similarly, much of the evidence on the impact of lifestyle factors on the risk of thyroid cancer comes from observational studies, with their inherent risk of confounding and reverse causality. For example, a reduced risk of thyroid cancer has been

reported in cigarette smokers and alcohol drinkers (Yeo *et al.* 2020). Positive associations have been reported between both sleep deprivation and physical inactivity, and thyroid cancer (Luo *et al.* 2013; Xhaard *et al.* 2016). A meta-analysis investigating the association between coffee consumption and thyroid cancer risk found no association overall, but an inverse association amongst hospital-based case control studies (Han *et al.* 2017; Riza *et al.* 2015; Takezaki *et al.* 1996).

Mendelian randomisation is a genetic epidemiological technique which can be used to explore causal associations between risk factors and diseases. It relies upon the fact that many physical or behavioural traits are associated with genetic variants, which are assigned randomly at conception. Using these variants as surrogate markers for the exposure of interest means that in a large enough population, confounding factors should be evenly distributed, allowing unbiased assessment of the causal effect of the exposure on the disease of interest (Davey Smith & Ebrahim 2013). Here we use genetic and clinical data from 379,708 unrelated individuals in the UK Biobank to explore the causal effects of lifestyle-related risk factors, including smoking, alcohol consumption, physical activity, and sleep related traits, on benign nodular thyroid disease (BNTD) and differentiated thyroid carcinoma (DTC).

6.2 - Materials and Methods

6.2.1 – Study population

Unrelated participants in the UK Biobank, of European ancestry, formed the study population. The UK Biobank is a longitudinal study of 500,000 people in the UK aged 40-69 years. Participants were recruited between 2006 and 2010 and have been followed up since then (Sudlow *et al.* 2015). At enrolment, participants gave blood samples for genotyping and completed extensive health related questionnaires. They also consented to record linkage with national cancer and death registries and the Hospital Episode Statistics database (HES).

Genotyping was performed using DNA extracted from whole blood samples, using two single nucleotide polymorphism (SNP) arrays with over 95% content overlap. Around 450,000 participants underwent genotyping on the UK Biobank Axiom array® and the remaining 50,000 samples were analysed using the UK BiLEVE Axiom array® (Welsh *et al.* 2017). Quality control including exclusion of sex mismatches, duplicates and outliers of heterozygosity was performed centrally by the UK Biobank (Bycroft *et al.* 2018), and after exclusion of related individuals up to three degrees of separation (using Kinship coefficients provided by UK Biobank) a total of 379,708 individuals with available data remained for inclusion.

6.2.2 – Exposures and outcomes

Cancer registry and HES record linkage data was used to identify participants with a diagnosis of DTC. Initially the ICD-10 code “C73” was used to identify all

thyroid cancer cases, then this was refined using histology codes to exclude anaplastic thyroid carcinoma, medullary thyroid carcinoma and thyroid lymphoma, to leave 425 patients with a diagnosis of DTC. Similarly, a group of 1,812 participants with a diagnosis of BNTD was identified using the ICD-10 code "D34". Control groups of patients with no history of BNTD (n=377,896) and no history of any cancer diagnosis (n=310,176) were also identified amongst the 379,708 unrelated UK Biobank participants.

The exposures of interest in this study were cigarette smoking, alcohol consumption, sleep duration, physical activity and caffeinated coffee intake. Cigarette smoking and alcohol consumption were recorded in the baseline health questionnaire, as categorical and continuous variables (pack years and units per week). Sleep duration was also recorded on the health questionnaire, as hours per night. Self-reported intake of caffeinated coffee was measured in number of cups per day, and chronotype was recorded according to participants' response to the question "do you consider yourself to be 1. Definitely a morning person, 2. More a morning than evening person, 3. More an evening person than a morning person, 4. Definitely an evening person, 5. Do not know, or 6. Prefer not to answer.". In addition, accelerometer data was available for 95,776 participants from a wrist-worn accelerometer in order to objectively measure physical activity, measured as the average vector magnitude for each 30-minute period over 7 days (Doherty *et al.* 2017).

Genetic variants associated with the exposures of interest were identified from large genome wide association studies (GWAS) for the corresponding risk factor traits (Clarke *et al.* 2017; Erzurumluoglu *et al.* 2020; Cornelis *et al.* 2015; Jones

et al. 2019). For smoking, alcohol consumption, caffeine intake and chronotype, genetic risk scores (GRS) were produced using variant effect size and number of trait-raising alleles as reported in the respective GWAS. To obtain summary statistics for physical activity MR, a GWAS of the inverse-normalized overall physical activity measure was carried out using BOLT-LMM (Loh *et al.* 2015), adjusting for baseline age, sex, study center, and genotyping array. Variants with imputation quality (INFO) <0.3 or minor allele frequency (MAF) <0.1% were excluded.

6.2.3 – Statistical analysis

Observational analysis was performed using logistic regression, first to describe the demographic differences between cases and controls, then to calculate the associations between the risk factors of interest (smoking status, smoking pack years, units of alcohol per week, self-reported physical activity, accelerometer-measured physical activity, and self-reported sleep duration, chronotype and caffeinated coffee intake), and the outcomes of BNTD and DTC. These regression models were adjusted for body mass index, age at enrolment, sex and Townsend Deprivation Index.

One-sample MR was used for smoking, alcohol and caffeine exposure to avoid bias introduced by including non-smokers, non-drinkers and non-coffee drinkers in the genetic analysis. A two stage least squares regression model was utilised, whereby the GRS was regressed against the corresponding risk factor trait, and

the predicted values and residuals were used as the independent variables in a logistic regression model with robust standard errors, with the outcome trait of interest as the dependent variable. Results were adjusted for assessment centre, genotyping platform and ancestral principal components.

To investigate the causal association between chronotype and caffeine intake and BNTD and DTC, two-sample MR was utilised in addition to one-sample MR. This allowed inclusion of related participants and therefore a larger sample size of 451,025. First, the SNP-risk factor associations were taken from previously published GWAS (Jones *et al.* 2019; Cornelis *et al.* 2015). We then performed GWAS of the outcome traits of BNTD and DTC in the UK Biobank, with adjustment for interrelatedness using BOLT-LMM software (Loh *et al.* 2015). The SNP-outcome associations were then calculated using inverse variance weighted (IVW) MR, as well as Egger MR and weighted median MR techniques, which are more robust to pleiotropy (Bowden *et al.* 2016 *a;* Bowden *et al.* 2016 b).

To investigate the role of physical activity, we performed two-sample MR analyses using MR RAPS (Zhao *et al.* 2020), which uses a maximum profile likelihood estimator and is thus robust to deviations from the typical instrumental variable (IV) assumption (that the instrument affects the outcome only through the exposure). MR RAPS thus enables the use of “weak instruments” (variants that are typically below genome wide significance, with low explanatory power) that can bias the causal estimate towards the null. This allows causal analysis of exposures where too few genetic variants reach the genome-wide significance threshold ($P<5x10^{-8}$) required for standard MR methods.

Genetic variants for physical activity were selected by clumping the physical activity GWAS summary statistics with plink (v1.90 X) (Chang *et al.* 2015), with a P-value threshold of $1x10^{-4}$, a distance of 10Mb and R^2 threshold of 0.001. This resulted in 474 independent genetic variants (Supplementary data). As there was an overlap between our exposure and outcome, we applied a winner's curse correction to the summary statistics prior to MR analysis (Bowden & Dudbridge 2009).

We then used a variant-selection filter equivalent to Radial MR (Bowden *et al.* 2018) and based on the diagnostic procedures within the MR RAPS package. This method identifies variants whose effects on the outcome deviate from the estimated causal effect based on their effect on the exposure, thus selecting a subset of variants where pleiotropy is likely to be lowest. The variants that remained after this filtering step, are listed in the supplementary data.

6.3 – RESULTS

General demographic characteristics of the participants are shown in table 1. Participants with BNTD and those with DTC were more likely than controls to be women, to be older at the time of enrolment and to be obese. In addition, those with BNTD had a marginally higher Townsend deprivation index (TDI) score than the corresponding control group. The TDI uses a combination of unemployment, car ownership, house ownership and house overcrowding to give a single score,

with more positive values corresponding with higher deprivation, and more negative values corresponding to higher affluence (Townsend *et al.* 1988).

Characteristic	Benign nodular thyroid disease	Controls	p-value	Thyroid cancer	Cancer-free Controls	p-value
N	1,812	377,896		425	310,176	
Female sex	1,492 (82.34)	203,244 (53.78)	$<1x10^{-15}$	316 (74.35)	165,537 (53.37)	$<1x10^{-15}$
Mean age at recruitment (SD), years	58.92 (7.39)	57.23 (8.01)	$<1x10^{-15}$	57.72 (7.29)	56.54 (8.04)	$2.2x10^{-3}$
Mean Townsend Deprivation index (SD)	-1.16 (3.09)	-1.48 (2.99)	$4.8x10^{-8}$	-1.63 (2.94)	-1.45 (2.99)	0.42
Mean BMI (SD), kg/m^2	27.96 (5.23)	27.38 (4.78)	$4.1x10^{-12}$	27.55 (5.34)	27.38 (4.77)	0.16
Obese: BMI 30-40 kg/m^2	471 (25.99)	83,503 (22.10)	$1.6x10^{-9}$	111 (26.12)	68,563 (22.10)	$6.8x10^{-2}$

Chapter 6 Table 1: Summary of demographic characteristics of included participants
Values stated are numbers (percentages), unless otherwise stated. P-values calculated using logistic regression adjusted for age and sex.

6.3.1 - OBSERVATIONAL ANALYSIS

Observational analysis, as displayed in table 2, revealed an inverse association between alcohol intake and BNTD, with an odds ratio (OR) of 0.88 and 95% confidence interval (CI) of 0.84 – 0.94 per 1 standard deviation increase in mean units consumed per week. In addition, those reporting excess alcohol intake according to UK government guidance (to consume less than 14 units per week of alcohol) were less likely to have a diagnosis of BNTD (OR 0.71 95% CI 0.63 – 0.84). There was no association between self-reported alcohol intake and DTC. For cigarette smoking, current smokers were neither more or less likely than

former smokers or never smokers to have BNTD or DTC, however amongst current smokers, those with a higher pack year smoking history had a higher odds of BNTD (OR 1.71 95% CI 1.34 – 2.20) but not DTC. There were no observational associations between self-reported sleep duration, chronotype, caffeine intake or accelerometer measured physical activity and either BNTD or DTC, however those with self-reported exercise levels below those recommended by the UK government (150 minutes of moderate physical activity, or 75 minutes of vigorous physical activity per week) were more likely to have a diagnosis of DTC (OR 1.23 95% CI 1.00 – 1.50) but not BNTD.

Trait	Odds ratio benign nodular disease	95% Confidence interval	p-value	Odds ratio thyroid cancer	95% Confidence interval	p-value
Excess alcohol (binary) [a]	0.74	0.60 – 0.84	$3.7x10^{-5}$	0.92	0.68 – 1.23	0.58
Alcohol units per week	0.88	0.84 – 0.94	$1.4x10^{-5}$	0.96	0.86 – 1.08	0.52
Never smoker	1.04	0.97 – 1.12	0.28	1.01	0.86 – 1.17	0.95
Former smoker	1.04	0.94 – 1.15	0.46	1.08	0.88 – 1.33	0.48
Current smoker	1.09	0.92 – 1.29	0.34	0.91	0.62 – 1.33	0.63
Cigarette pack years	1.71	1.34 – 2.20	$2.2x10^{-5}$	1.23	0.75 – 2.01	0.42
Insufficient exercise (binary) [b]	1.04	0.94 – 1.15	0.44	1.23	1.00 – 1.50	$5.0x10^{-2}$
Average physical activity [c]	0.99	0.97 – 1.00	0.1	0.99	0.96 – 1.02	0.45
Sleep duration < 6 hours [d]	0.92	0.82 – 1.02	0.1	1.04	0.83 – 1.31	0.73
Chronotype [e]	0.98	0.95 – 1.02	0.27	0.94	0.87 – 1.01	$8.1x10^{-2}$

Chapter 6 Table 2: Observational associations between measured traits and benign nodular thyroid disease and thyroid cancer.

Continuous variables are reported as odds ratio per 1 standard deviation higher predictor. P-values calculated using logistic regression with age, sex, body mass index, and Townsend deprivation index as covariates.

[a] More than 14 units per week

[b] Less than 150 minutes moderate activity or 75 minutes vigorous activity per week.

[c] No wear time adjusted acceleration average based on wrist-worn accelerometer, measured in milligravity

[d] Self-reported sleep duration

[e] Based on responses to the question "do you consider yourself to be 1. Definitely a morning person, 2. More a morning than evening person, 3. More an evening person than a morning person, 4. Definitely an evening person, 5. Do not know, or 6. Prefer not to answer."

6.3.2 – MENDELIAN RANDOMISATION ANALYSIS

One-sample MR analysis did not support an association between genetic predisposition to higher alcohol intake and BNTD, but did not refute a possible association with DTC, albeit with a very wide 95% CI (OR 116.61 95% CI 1.88 – 7228.8). There was no evidence from one-sample MR to support causal associations between genetically instrumented tobacco or caffeine consumption, or genetically instrumented chronotype and either BNTD or DTC (Table 3).

Using two-sample MR we found no convincing evidence for an association between chronotype and either BNTD or DTC (Table 4).

Trait	Odds ratio benign nodular disease	95% Confidence interval	p-value	Odds ratio thyroid cancer	95% Confidence interval	p-value
Alcohol consumption [a]	1.94	0.17 – 22.7	0.6	116.61	1.88 – 7228.8	0.024
Cigarette smoking [b]	0.71	0.16 – 3.14	0.65	1.63	0.05 – 53.2	0.78
Caffeine consumption [c]	1.06	0.38 – 2.94	0.91	1.25	0.16 – 9.70	0.83
Chronotype[d]	2.12	0.75 – 5.96	0.16	2.38	0.28 – 20.10	0.43

Chapter 6 Table 3: Results of one-sample MR analysis for benign nodular thyroid disease (BNTD) and differentiated thyroid cancer (DTC)

[a] Excluding non-drinkers; F statistic 9.15

[b] Excluding never smokers; F statistic 317.49

[c] Excluding participants who report no caffeinated coffee consumption; F statistic 1077.55

[d] F statistic 751.55

		IVW		Egger			WM	
Trait	Outcome	OR (95% CI)	p-value	OR (95% CI)	p-value	Int P	OR (95% CI)	p-value
Chronotype	BNTD	0.97 (0.78-1.19)	0.74	0.66 (0.35-1.22)	0.19	0.19	0.90 (0.67-1.23)	0.51
Chronotype	DTC	0.77 (0.52-1.13)	0.18	0.83 (0.26-2.60)	0.75	0.89	0.98 (0.55-1.74)	0.93

Chapter 6 Table 4: Results of two-sample MR analysis for benign nodular thyroid disease (BNTD) and differentiated thyroid cancer (DTC)

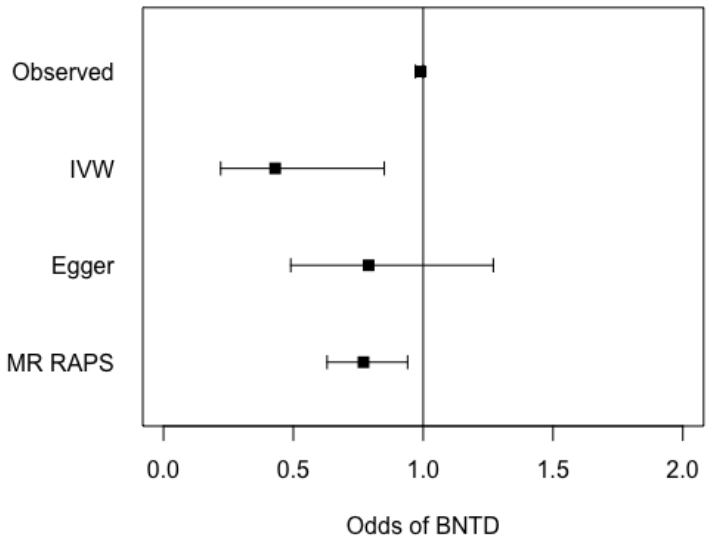
Odds ratios, 95% confidence intervals and p-values for the inverse variant weighted (IVW), Egger, and weighted median (WM) analyses are displayed. Int P represents the p-intercept of Egger analysis, which is a measure of horizontal pleiotropy. Mean F statistic for genetic instrument = 33.46.

Using two sample MR to investigate the role of physical activity on BNTD and DTC, we found an inverse association between genetically instrumented physical activity levels and BNTD (OR 0.43 95% CI 0.23-0.82) with directionally consistent results using Egger analysis to account for horizontal pleiotropy, and MR RAPS analysis to account for weak instrument bias (Table 5, Figure 1). There was however no evidence for a causal association between physical activity and DTC.

		IVW		Egger		MR RAPS	
Trait	Outcome	OR (95% CI)	p-value	OR (95% CI)	p-value	OR (95% CI)	p-value
Physical activity	BNTD	0.43 (0.22-0.85)	$1.46x10^{-2}$	0.79 (0.49-1.27)	0.33	0.77 (0.63-0.94)	$1.0x10^{-2}$
Physical activity	DTC	1.41 (0.18-11.42)	0.75	0.72 (0.30-1.72)	0.45	1.13 (0.80-1.58)	0.50

Chapter 6 Table 5: Weak-instrument MR analysis for benign nodular thyroid disease (BNTD) and differentiated thyroid cancer (DTC) and physical activity

Odds ratios, 95% confidence intervals and p-values for the inverse variant weighted (IVW), Egger, and MR RAPS analyses are displayed. Median F statistics for physical activity genetic instrument 16.7 and 17.6 for BNTD and DTC respectively.



Chapter 6 Figure 1: Forest plot showing the odds of benign nodular thyroid disease per 1 standard deviation increase in physical activity on observational analysis, and using weak instrument MR with inverse variance weighted (IVW), Egger and MR RAPS analyses

6.4 – DISCUSSION

We have used mendelian randomisation to explore the effect of smoking, alcohol, physical activity, caffeine intake and sleep on the risk of BNTD and DTC amongst UK Biobank participants. Our observational results suggest an association between insufficient physical activity levels and DTC, but no association with smoking, alcohol consumption, chronotype or sleep duration. In the case of BNTD, we observed an inverse association with alcohol intake, and whilst smokers had the same risk as non-smokers, the risk of BNTD increased with increasing lifetime cigarette consumption amongst smokers. Mendelian randomisation analysis did not show evidence for a causal role for cigarette smoking, alcohol consumption, chronotype, sleep duration or caffeine intake in either increasing or reducing the risk of BNTD or DTC. Weak instrument MR

suggested an inverse association between genetically instrumented physical activity levels and BNTD.

Many authors have reported an inverse association between smoking and thyroid cancer (Kitahara *et al.* 2012; Cho *et al.* 2018), however the biological explanation for this observation is not clear. One possible explanation is the confounding effect of body mass index (BMI) which tends to be lower in smokers, and has been shown to be positively associated with thyroid cancer (Kitahara *et al.* 2011). Another proposed confounder is the effect of cigarette smoking on thyroid function, and serum TSH levels in particular (Wiersinga 2013). However, although evidence from observational studies have demonstrated a positive association between serum TSH and thyroid cancer (McCleod *et al.* 2012) more recent Mendelian randomisation studies have reported an inverse association (Yuan *et al.* 2020a; Zhou *et al.* 2020; Fussey *et al.* 2020). There is significant variation in adjustment for confounders in the observational literature on cigarette smoking and thyroid cancer (Kitahara *et al.* 2012), however one study which adjusted for BMI and TSH still reported an inverse association (Cho *et al.* 2018). Another possible explanation for the observational reports of an inverse association is the anti-oestrogenic effect of cigarette smoking (Kreiger & Parkes 2000). Oestrogen is known to promote thyrocyte growth and proliferation, and women of child-bearing age have a much higher incidence of both BNTD and DTC than age-matched men (Derwahl & Nicula 2014). Whilst our observational data is likely to be limited by relatively small numbers of DTC cases, our MR findings do not suggest a causal association between smoking and DTC. The only other MR study on the effect of cigarette smoking on thyroid cancer risk also failed to demonstrate a causal role (Huang *et al.* 2022).

As with cigarette smoking, alcohol consumption has been found to be inversely associated with thyroid cancer (Hong *et al.* 2017). Once again, several mechanisms have been proposed for this relationship, including an inhibitory effect on TSH secretion caused by alcohol consumption (Zoller *et al.* 1996), and thyrocyte toxicity resulting in reduced gland volume (Knudsen *et al.* 2001). However, once again there are many possible confounding variables for which it is hard to fully account in observational study design. We did observe an inverse association between reported alcohol intake and BNTD but not DTC, and MR analysis revealed no evidence for a causal association, with wide confidence intervals suggesting an inconclusive result. The MR work by Huang *et al.* also investigated the effect of alcohol use on thyroid cancer risk, and reported no conclusive association in either direction, with confidence intervals crossing the null on 2-sample analysis which was, as the authors concede, somewhat underpowered (Huang *et al.* 2022).

The effect of physical activity and sleep on thyroid cancer has been less extensively studied. There is some evidence that increased physical activity levels reduce the risk of DTC in women (Rossing *et al.* 2001), and that walking for more than 60 minutes a day reduces the risk of overall thyroid cancer (Fiore *et al.* 2019). A pooled analysis of two case control studies also supports these findings, with recreational physical activity being inversely associated with thyroid cancer (Xhaard *et al.* 2016). These findings may be due to the fact that people who exercise are more likely to have a lower BMI, and BMI is associated with DTC risk, or alternatively they may be a result of the effects of exercise in reducing systemic inflammation which is itself is carcinogenic (Booth *et al.* 2015).

Our observational findings do support an association between insufficient physical activity and DTC with adjustment for BMI, but not BNTD. Conversely, our MR analyses including correction for winner's curse and weak instrument bias identified tentative evidence that physical activity is protective against BNTD but not DTC. This may be a reflection of the inherent confounding in our observational analyses, as one would expect an observational association with BNTD if there is indeed a casual association. The finding of an inverse association between genetically instrumented physical activity levels and BNTD may simply represent the fact that participants predisposed to higher physical activity levels throughout their lifetime are less likely to be obese and therefore more likely to notice small thyroid nodules. This can be further investigated using multivariable MR approaches.

Sleep deprivation has been shown to be associated with a higher risk of thyroid cancer in post-menopausal women (Luo *et al.* 2013), and there is some evidence that sleep cycle disruption affects thyroid function (Ikegami *et al.* 2019). Sleep cycle disturbance is increasingly being recognised as a contributory factor to many health conditions including type 2 diabetes, obesity, cardiovascular disease and cancers (Kettner *et al.* 2014). There are therefore several possible explanations for a plausible link between sleep deprivation and DTC, however we were not able to demonstrate a causal relationship using MR.

Our one-sample results failed to demonstrate any association between genetic variants associated with higher lifetime caffeine intake amongst coffee drinkers, and BNTD and DTC. We used one-sample rather than two-sample techniques for this analysis, as two-sample MR did not allow for exclusion of non-coffee

drinkers. Previous studies have failed to show an association between coffee consumption and thyroid cancer risk; however, caffeine has some biological effects which would be expected to protect against carcinogenesis such as increasing intracellular cyclic adenosine monophosphate (Han & Kim. 2017).

This study is the first to use Mendelian randomisation to investigate the causal role of lifestyle-related risk factors on the risk of DTC. We acknowledge that there are some limitations to the current study, including the relatively small number of patients with DTC and BNTD, and the previously reported healthy volunteer bias in the UK Biobank participants (Fry *et al.* 2017). In addition, whilst cancers are well documented in cancer registry data, benign nodules are only confirmed as such on surgical resection, and the prevalence of BNTD in the UK Biobank cohort (0.5%) is lower than the generally accepted prevalence of clinically palpable thyroid nodules (2-6%) (Dean & Gharib 2008). This may result in detection bias, whereby control groups are contaminated by undiagnosed cases. Finally, some of the 95% confidence intervals for our MR analyses were wide, meaning that the exact effect of the genetic instruments on the outcomes of interest are somewhat unclear. Equally, the wide confidence intervals mean that p-values of greater than 0.05 should not be taken of evidence for the lack of a causal role but rather an inconclusive result.

6.4.1 – Conclusion

Using MR analysis in a large cohort of European participants we have not been able to identify any causative or protective role for smoking, alcohol and caffeine consumption on BNTD or DTC. MR did not support our observational finding that

reduced physical activity was associated with a higher risk of DTC, but neither did it provide enough evidence to refute it. These findings suggest that more work is required to unpick the causal associations between lifestyle-related risk factors and DTC in order to inform public health policy, health education and risk stratification of patients with indeterminate thyroid nodules.

CHAPTER 7:

DISCUSSION AND CONCLUSIONS

7.1 – SUMMARY OF KEY FINDINGS

The present thesis investigates the use of genetics in the diagnosis and management of thyroid cancer. This was a broad remit and clearly, an exhaustive analysis of the entirety of thyroid cancer genetics was beyond the scope of the thesis, so the research focussed on two main themes: the clinical application of *RET* analysis in both familial and sporadic MTC, and the use of genetic epidemiology to advance our understanding of modifiable risk factors for DTC.

On the topic of MTC, the work reported in Chapter 2 represents the largest UK series of germline *RET* analysis, and provides an insight into the mutation rate in a variety of clinical presentations in a real-life UK clinical setting. In addition to supporting the work of other authors, for example in identifying a pathogenic *RET* mutation in 8.5% of patients with presumed sporadic MTC, the study gave some insight into changing practices in germline *RET* analysis over the last two decades. Amongst the findings of this study was the fact that sporadic MTC makes up the majority of clinical presentations, with the single largest group being those with isolated MTC and no personal or family history of associated endocrine malignancy or endocrinopathy. Although current guidelines in the UK do not recommend routine somatic *RET* analysis in sporadic MTC (Perros *et al.* 2014), it is plausible that with the increasing availability of targeted therapy for MTC,

somatic mutation status may help to guide adjuvant treatment. With this in mind, I embarked upon a systematic review of the literature concerning the role of somatic genetic and epigenetic analysis in the management of sporadic MTC. This work forms Chapter 3 of this thesis, which describes the comprehensive synthesis of 23 articles focusing on a variety of genetic and epigenetic markers, with the key findings including the substantial evidence for a role for somatic *RET* mutations – especially the M918T mutation – in predicting a poorer prognosis, and the emerging evidence for the role of epigenetic markers including miR-21, miR-183 and miR-375 as predictors of lymph node metastasis. This work added to the literature on the topic of risk stratification and prognostication in sporadic MTC by highlighting the possible usefulness of these somatic genetic and epigenetic markers in risk stratification and management planning. The chapter also explored the potential use of these markers in providing personalised adjuvant treatment to patients in the form of TKIs.

Continuing with the general theme of genetics in thyroid cancer, but moving on to DTC, I embarked upon a series of projects using genetic epidemiological techniques to investigate modifiable risk factors for DTC and benign nodular thyroid disease (BNTD). Mendelian randomisation (MR) encompasses a range of techniques using genetic information to investigate the associations between exposures and disease outcomes, and at the time of my first publication on MR, on the effect of obesity and type 2 diabetes mellitus (T2DM), it had not previously been used to investigate risk factors for thyroid cancer. Chapter 4 describes the techniques and findings of this study, which found no evidence for a causal role for genetically instrumented obesity on either DTC or BNTD, but a possible role for T2DM, with participants with the highest genetically instrumented risk of T2DM

being more likely to develop DTC. This study raised questions about the true relationship between obesity and thyroid cancer, and although not significant enough to change clinical practice, the findings regarding T2DM and DTC present an interesting area for further research which may in future help to support public health initiatives and risk stratify indeterminate nodules.

The work described in Chapter 4 resulted in further questions relating to the possible explanations for the widely reported observational associations between obesity and DTC, the possible role of confounding factors. Thyroid stimulating hormone (TSH) and its association with thyroid cancer has been investigated by a number of authors, and the close relationships between TSH, obesity and T2DM as well as benign nodular thyroid disease made it an attractive subject for an MR study using similar techniques to those described in Chapter 4. This work resulted in unexpected findings of an inverse association between genetically instrumented TSH levels and both BNTD and DTC. This was published shortly after work by two other teams also using MR to investigate the effect of TSH on thyroid cancer, with similar results (Zhou *et al.* 2020; Yuan *et al.* 2020a). The findings of three separate studies using MR all supporting an inverse association between TSH and thyroid cancer lent weight to the argument that there may be a legitimate explanation for the finding which was at odds with the majority of the observational research on the subject. This raises questions about the possible different effects of TSH at different times in life versus lifetime TSH exposure on thyroid cancer risk. We postulated that as MR techniques use genetic predisposition to raised TSH they provide an estimate of the association between lifetime exposure to TSH and thyroid cancer, whereas observational studies measure TSH at the time of diagnosis. This question requires further research to

resolve and is potentially clinically relevant as it raises the possibility of using genetic instruments for TSH to help risk stratify indeterminate nodules.

The final project in the series of work using genetic epidemiology to investigate risk factors for thyroid cancer focussed on lifestyle risk factors. This was a logical progression from the previous work in this thesis on risk factors for structural thyroid disease and thyroid cancer, aiming to clarify the roles of various lifestyle risk factors for thyroid cancer. In addition to an observational analysis of the association between smoking, alcohol, physical exercise and sleep duration with both BNTD and DTC, I used mendelian randomisation techniques to explore the underlying association with genetically instrumented exposure variables. This required some modification from the techniques I had used in previous work. For example, when examining the role of smoking and alcohol intake, it was not appropriate to use 2-sample MR, as the genetic instruments for cigarette and alcohol consumption can only be valid in participants who smoke and drink alcohol. This required a 1-sample MR approach, thus allowing exclusion of the participants in the UK Biobank cohort who reported no alcohol consumption or cigarette smoking. A similar approach was taken with caffeine consumption. When examining physical activity and its effect on BNTD and DTC I worked with co-authors to use a novel weak instrument MR approach to explore the effect of a genetic instrument consisting of variants with marginal associations with measured physical activity on both BNTD and DTC.

The findings of this work were largely null, with no evidence for a causal or protective role of either smoking, alcohol consumption or sleep duration on DTC. This was largely expected, and the various possible mechanisms for the

observational findings reported elsewhere in the literature are discussed in detail. The weak instrument MR for physical activity found no association with DTC, but an inverse association with BNTD. This may represent a true inverse association, due to possible protective effects of physical exercise on thyroid nodule formation and growth, although it could represent a reflection of the fact that those genetically predisposed to be physically active are less likely to be obese and therefore more likely to notice small palpable thyroid nodules. Although our analyses did not indicate an association between physical activity and DTC, they were limited by the relatively small numbers of participants with DTC. The fact that BNTD is a well-recognised factor for DTC, along with the fact that our observational analysis suggested an inverse association between self-reported physical exercise levels and DTC, makes this an interesting area for future research, which given the multitude of inter-related potential exposures predisposing to DTC should include multivariate MR techniques (Burgess & Thompson 2015).

7.2 – KEY FINDINGS IN THE CONTEXT OF EXISTING KNOWLEDGE

7.2.1 - MTC

The initial study in this thesis (chapter 2) reported a very similar rate of germline *RET* mutation in patients with supposed sporadic MTC to that reported by authors elsewhere, confirming the applicability to the UK population of international guidelines for universal germline *RET* testing in patients with MTC. In addition, our results showed a lower-than-expected rate of detection of germline *RET* mutation in patients with MTC and a family history of MTC or other MEN2-associated conditions. This may be a reflection of the fact that family history for

referred patients was not always accurate, but also raises the possibility of other as yet undiscovered variants either in *RET* or other genes which may be responsible for familial MTC. An example of this is the *ESR2* gene which was found to harbour a frameshift deletion in a family of patients with MTC and CCH with no identified *RET* mutation (Smith *et al.* 2016).

The investigation of the somatic genetic landscape of sporadic MTC is an area of current interest in the new era of personalised medicine. Although the American Thyroid Association guidelines on MTC do not recommend the routine analysis of tumour tissue for somatic mutations (Wells *et al.* 2015), the European Thyroid Association does, noting that the results should be recorded for future therapeutic purposes (Elisei *et al.* 2012). There is a precedent for using targeted therapy based on somatic mutation profile in other cancers, such as malignant melanoma (Robert *et al.* 2019), and the apparent variation in clinical behaviour according to somatic mutation status in sporadic MTC suggests that it too would be amenable to targeted therapy. Indeed, two multi-kinase inhibitors have been studied in progressive metastatic MTC (Wells *et al.* 2012; Elisei *et al.* 2013), and in the case of vandetanib, the authors found an improved response rate in patients with somatic *RET* M918T mutations (Wells *et al.* 2012). Carbozantinib has also been found to be more effective in those with somatic *RET* M918T mutations or *RAS* mutations (Sherman *et al.* 2016). The imprecise mechanism of these multi-kinase inhibitors has led other researchers to investigate the use of targeted highly selective *RET* inhibitors in metastatic MTC, with promising results reported with two agents, selpercatinib (LOXO-292) and pralsetinib BLU-667 (Wirth *et al.* 2020; Subbiah *et al.* 2021), which have both recently been granted United States Food and Drug Administration (FDA) approval.

Despite recent advances in the field of *RET*-targeted therapy in MTC, it remains a fact that almost half of all patients with sporadic MTC have no *RET* mutation. The findings of my systematic review in chapter 3 highlight the emerging role for epigenetic markers in the prognostication of sporadic MTC, with changes in expression levels of certain microRNAs being associated with more aggressive disease as a result of their effect on tumour suppressor genes. These have been identified as possible targets for cancer therapy, and there are a number of early phase clinical trials using these microRNAs as targets for novel therapies and other cancer types (Forterre *et al.* 2020). There is also evidence from *in vitro* research that overexpression of miR-375 results in improved efficacy of vandetanib on MTC cancer cells (Lassalle *et al.* 2016). Future research is needed in the field of MTC in order to elucidate the exact relationship between microRNA profiles and tumour behaviour in order to inform trials using microRNA-modifying agents in the management of these tumours.

7.2.2 - DTC

Chapter 4 of this thesis describes my use of MR to investigate the causative effects of obesity and T2DM on DTC. This is an area which has been extensively studied using observational techniques, and there is a precedent for using MR to investigate the effect of obesity on cancer risk in other types of cancer, but it has never been used in thyroid cancer. Recent MR studies have helped to confirm the positive association between obesity and oesophageal, colorectal, ovarian, endometrial, kidney and pancreatic cancers, and the inverse association between early life obesity and breast cancer (Fang *et al.* 2021). Whilst traditionally, genetically instrumented BMI has been seen as a surrogate marker for lifetime

obesity, there is some evidence that certain variants affect BMI differently at different ages (Winkler *et al.* 2015). This raises the possibility of investigating the effect of BMI at different ages on cancer, as has been shown in both breast (associated with early life obesity) and oesophageal (associated with adult obesity) cancers (Gao *et al.* 2016; Thrift *et al.* 2014). The effect of obesity at different ages with thyroid cancer was not explored in the current research, however given the finding that our BMI instrument is associated not only with BMI in adulthood but also with self-reported obesity in childhood (Yaghootkar *et al.* 2016), and the reports of observational studies suggesting that body shape in middle age is an important risk factor for thyroid cancer (Clavel-Chapelon *et al.* 2010), this would be an interesting area for future research.

The clinical relevance of any link between obesity or adiposity and thyroid cancer is clear. If such a link were proven to be causal it would have significant implications for public health, patient education and health policy. Many authors have proposed physiological mechanisms for a link between obesity and thyroid cancer, including the effect of various adiposity-related tissue proteins such as leptin and adiponectin (Hebbard & Ranscht 2014; Uddin *et al.* 2010), and this is an area of ongoing study. The difficulty of testing an association between a serum protein measured at a single point in time and an outcome such as thyroid cancer is the high risk of confounding and reverse causality, both of which introduce a significant element of bias. For this reason, MR would be a suitable approach to test these associations, if suitable genetic instruments could be validated for this purpose.

Whilst our MR findings did not show an association between genetically instrumented T2DM and DTC *per se*, we did demonstrate an increasing incidence of DTC with increasing genetic risk of T2DM when the study population was divided into quartiles according to genetic risk score for T2DM. Several authors have identified an observational association between T2DM and thyroid cancer (Li & Qian 2017; Yeo *et al.* 2014), and it has been hypothesised that insulin resistance, which is increasing in global incidence in parallel with thyroid cancer, may be the mechanism (Gursoy 2010). This theory is supported by the finding that insulin resistance is associated with thyroid nodules (Tang *et al.* 2017). The insulin resistance seen in T2DM results in increased production of insulin like growth factor 1 (IGF-1), and IGF-1 receptors have been shown to be overexpressed in thyroid cancer cells (Vella & Malaguarnera 2018). Certainly, further research is required to explore the causative role for T2DM on thyroid cancer, as given the increasing evidence for a link between T2DM and other cancers, a positive association could help to inform cancer screening and surveillance in T2DM patients (Yuan *et al.* 2020b).

The findings of our paper on TSH and thyroid cancer, reported in chapter 5, are at odds with much of the published observational evidence on the subject. For example, two meta-analyses of observational studies have reported a positive association between TSH levels and DTC risk (McLeod *et al.* 2012; Zheng *et al.* 2016). There are however limitations to this data, for example the fact that all but two of the studies included by McLeod *et al.* were cross-sectional and retrospective in design, and that only one study excluded patients with functioning thyroid nodules (known to be associated with both suppressed TSH levels and reduced risk of thyroid cancer). Similarly, all but two studies included in the meta-

analysis of Zheng *et al.* were retrospective in design. Another potential weakness of the available observational data on the relationship between TSH and thyroid cancer, is that the vast majority of studies explore the link between TSH levels at the time of diagnosis of a thyroid nodule, and the risk of thyroid cancer. It is therefore not reasonable to draw any conclusions regarding the role of TSH in thyroid cancer pathogenesis from this data.

There is however some observational data supporting an inverse association between serum TSH and thyroid cancer risk. For example, Huang *et al.,* in their matched case-control study identified an inverse association between TSH within the normal range and thyroid cancer, and whilst this trend continued in the below normal range for women, they saw a paradoxical increase in the odds of thyroid cancer as TSH rose above the normal range in men (Huang *et al.* 2017). They relied upon blood drawn before diagnosis for their analysis, with a median interval between blood draw and diagnosis of 1,454 days, in contrast to many other observational studies. The EPIC case control study also identified an inverse association between TSH and thyroid cancer risk, with a mean interval between blood draw and outcome analysis of 6.4 years (Rinaldi *et al.* 2014).

The earliest use of genetic analysis to investigate the link between TSH and thyroid cancer risk was in 2012, when Gudmundsson *et al.* identified variants identified with both low TSH and thyroid cancer (Gudmundsson *et al.* 2012). The authors concluded that for carriers of those variants, the pathophysiology of non-medullary thyroid cancer is endocrinological, i.e., related to low TSH levels. They hypothesised that this may be due to reduced differentiation in thyroid parenchyma in the setting of low TSH, predisposing to malignant transformation.

A subsequent GWAS of TSH-associated variants (Zhou *et al.* 2020) used data from UK, US and Icelandic populations and reported that in 63 of 94 TSH-associated variants, the TSH-increasing allele was associated with a reduced risk of thyroid cancer. Furthermore, their 2-sample MR analysis suggested an inverse association between genetically instrumented TSH and thyroid cancer risk, with a one standard deviation increase in TSH resulting in a 45% reduction in thyroid cancer risk. The authors once again referred to the theory that lower TSH may result in less differentiation of thyroid cells and thus a higher risk of malignant transformation, but also raised the possibility that their instrument for TSH was responsible for both TSH and thyroid growth (and thus tumorigenesis) rather than solely TSH with a resultant downstream effect on tumorigenesis. Finally, a separate group published their findings of MR analysis on the effect of TSH on both breast and thyroid cancer shortly after publication of our own findings and those of Zhou *et al.* Yuan *et al.* also used data from the UK Biobank and reported very similar findings to our own and those of Zhou *et al.* with a 53% reduction in risk of thyroid cancer per one standard deviation higher genetically instrumented TSH (Yuan *et al.* 2020a). They postulated that a possible explanation for the discrepancy between their results and some of the observational evidence suggesting a positive association between TSH and thyroid cancer risk is the possibility that TSH levels have different effects on normal thyroid tissue and thyroid tumour cells. Thus, in patients with a thyroid nodule or a thyroid tumour, TSH may promote tumorigenesis and tumour growth, whereas higher TSH levels in patients with normal thyroid parenchyma may prevent tumorigenesis. This is an area of interest for future research with UK Biobank data, which provides the ability to follow-up patients with benign thyroid nodules and stratify them

according to both genetically instrumented TSH, and measured TSH as that data becomes available.

In contrast to obesity and TSH levels, the existing literature on the link between other lifestyle risk factors and thyroid cancer is limited. Observational studies investigating the link between smoking and DTC have however consistently reported an inverse association, with smoking appearing to reduce the risk of DTC (Lee *et al.* 2021; Yeo *et al.* 2020). Similarly, alcohol consumption has been found by several authors to be associated with a lower risk of DTC (Yeo *et al.* 2020; Hong *et al.* 2017). Whilst our observational analysis in the UK Biobank did show an odds ratio of DTC of 0.91 in current smokers, the 95% confidence interval crossed the null (0.62-1.33) and our genetic analyses failed to show a causal association between smoking and DTC. Our observational analyses similarly showed a nonsignificant inverse association between alcohol consumption and DTC, and our MR analyses were unable to support or refute this, with a positive effect estimate but very wide confidence intervals. This may be a reflection of our limited sample size, and the fact that the UK Biobank participants are healthier and less likely to be smokers than the general population (Munafò *et al.* 2018). In addition, our genetic instruments, especially for alcohol, lacked precision. Other authors have used MR to investigate the causal association between both smoking and alcohol consumption and various cancers, with some success (Larsson *et al.* 2020), however there is to date no existing precedent for the use of MR to investigate the effect of smoking and alcohol consumption on thyroid cancer risk.

The final aspect of our work on lifestyle risk factors and thyroid cancer included an investigation of the effect of sleep and physical activity on thyroid cancer. The rationale for studying these particular lifestyle risk factors was the observational evidence from other authors suggesting an inverse association between physical activity and DTC (Xhaard *et al.* 2016; Rossing *et al.* 2001; Fiore *et al.* 2019), and a positive association between sleep disturbance and DTC in post-menopausal women (Luo *et al.* 2015). Whilst our observational results suggested an association between self-reported physical inactivity and DTC, the data from wrist-worn accelerometers did not support this, and our MR analyses were unable to provide evidence for a causal association between either physical activity or sleep duration and DTC. There is recent MR evidence that genetic liability to shorter sleep duration is associated with increased risk of stomach and pancreas cancers, and genetic liability to long sleep duration is associated with reduced risk of pancreatic cancer (Titova *et al.* 2021). The authors highlight the many plausible physiological explanations for a link between sleep and cancer risk, including the anti-inflammatory and anticancer actions of melatonin, the effect of sleep deprivation on glucose metabolism and its association with obesity an insulin resistance. The use of MR to investigate the association between sleep and thyroid cancer has not previously been reported, however in light of our results it may be that larger samples are required to overcome the limitations posed by the relative rarity of thyroid cancer, and the small effect of genetic liability on variance in sleep behaviours. Similarly, there is no precedent for the use of MR to investigate the effect of physical activity on thyroid cancer risk, however there is evidence from a recent MR study for a causal relationship between reduced genetically instrumented physical activity and colorectal cancer (Zhang *et al.* 2021).

7.3 – LIMITATIONS

The broad remit of my research has meant that by necessity I have had to narrow down and focus my attention on two discrete themes related to genetics in thyroid cancer. My work started with a focus on MTC and in particular the clinical usefulness of germline and somatic genetic testing in sporadic MTC, with a view to continuing with prospective research on the clinical applicability of somatic genetic testing in these patients, however a parallel stream of study focussing on the genetic epidemiology of thyroid cancer grew to become the main focus of the thesis. Nevertheless, the work presented here in chapters 2 and 3 has added to the existing literature on MTC with respect to the incidence and types of germline *RET* mutation according to clinical presentation in a UK population, and my systematic review (chapter 3) has formed an important foundation on which to base subsequent work on risk stratification in sporadic MTC which is already underway (see section 7.4).

The work presented in chapters 4-6 uses broadly similar methodology with some subtle but important variations, and therefore the studies share some common limitations. Whilst the UK Biobank is a hugely valuable resource for researchers, it does introduce some bias to any research concerning health conditions in the wider population. It consists only of participants recruited between the ages of 40 and 69 and living in the UK, and the low response rate (5.5%) introduces some selection bias despite the large number of participants (Swanson 2012). There is also evidence of some healthy volunteer bias, with the rates of cancer incidence and obesity being lower than in the general UK population (Fry *et al.* 2017). The identification of participants with diseases of interest relies upon a combination of

self-reported diagnoses, ICD-10 codes and cancer registry data. This means that for cancer diagnoses the data is robust (as long as the diagnosis was made in the UK), but for benign conditions such as BNTD, the coding is likely to be less accurate, and doesn't account for the many people in the general population (and therefore also in the UK Biobank population) with undiagnosed BNTD, which has been reported by authors of autopsy studies to be up to 50% of the population (Mortensen *et al.* 1955). Perhaps the most important limitation of the UK Biobank in the context of the work reported in chapters 4-6 is the limited number of participants with DTC. This reduces the power of both observational and Mendelian randomisation analyses. The UK Biobank is a longitudinal study, and thus the number of participants with DTC is likely to increase with time, it is clear that collaboration with other similar resources around the world will be required in order to adequately increase statistical power.

Although we included observational analysis in our investigation of risk factors for DTC and BNTD in the UK Biobank, it must be acknowledged that for the reasons explored in chapter 1, observational analysis alone is not a reliable way to investigate such a complex and likely multifactorial causal relationship. Several of the proposed risk factors are closely related to one another, for example obesity, TSH levels and T2DM. Furthermore, the UK Biobank is an ongoing project, so cases of DTC and BNTD are a mixture of incident (being diagnosed since participant enrolment) and prevalent (being previously diagnosed). Therefore, the temporal association between a risk factor and outcome of interest is not always relevant in the context of establishing causation. It is primarily for these reasons that we decided to use Mendelian randomisation using genetic variants, however this methodology is also not without its limitations.

The most obvious such limitation is the fact that most of our genetic instruments are responsible for only a small proportion of variance in their corresponding phenotypic trait, especially for behavioural risk factors such as smoking and alcohol consumption. This problem is partially overcome by using very large sample sizes, meaning that even with weak instruments the distribution of confounding factors should be so even between groups that any difference in outcome is more likely to be a result of the genetic instrument of interest than confounders, i.e., a causal effect. In our work we were able to use data from almost 500,000 individuals to counter the limitation imposed by relatively weak genetic instruments. In the case of physical activity, for which weak instrument bias is a particular problem due to the low variance in measured physical activity attributable to genetic variation, we were able to use a novel technique known as weak instrument MR in order to minimise the effect of this bias as reported in Chapter 6. The comprehensive clinical data available for the UK Biobank participants also allowed us to improve the precision of our analyses by adjusting for covariates associated with our outcomes of interests in the observational literature, for example BMI, body fat and body shape when investigating the causal role of T2DM in DTC risk.

Pleiotropy, whereby a genetic variant has an effect on the outcome of interest by a route other than its effect on the risk factor trait of interest, is another possible limitation of any MR study. In the work reported here, we have tried to minimise the effect of pleiotropy by using multiple variants for each genetic instrument, and employing multiple techniques such as Egger and median weighted MR to reduce

the effect of horizontal pleiotropy (Bowden *et al.* 2016 *a;* Bowden *et al.* 2016 b), however it cannot be completely excluded.

Two-sample MR, which uses different study populations to ascertain the SNP-risk factor association and the SNP-outcome association has the advantage that it can further increase the number of study subjects and thus the power of the study. It also allows the investigation of genetic instruments for which there is no corresponding observed variable in the study group. This allowed us to study the effect of genetically instrumented TSH levels in the UK Biobank cohort despite the participants not having serum TSH levels available. The limitation of this approach is the reliance on the work of other researchers in the form of GWAS and meta-analysis of GWAS, and the fact that as the study populations are by definition different, the assumption that the same SNP-risk factor associations will exist in both is a potential source of bias. We tried to address this by using large meta-analyses of many thousands of participants of European descent to generate our instruments for 2-sample MR.

7.4 – Future directions and considerations

The two main streams of my research have each opened up further areas of interest worthy of continued study. In sporadic MTC, it is clear that there is a lack of high-quality evidence regarding the genetic and epigenetic factors responsible for variation in prognosis, and the clinical usefulness of somatic genetic analysis in planning management and follow up. The result of the work presented in Chapters 2 and 3 was the implementation of a prospective study to investigate the clinical, pathological, genetic and epigenetic determinants of disease behaviour and response to treatment in sMTC. This work will require national

collaboration in order to generate significant patient numbers, and will take a number of years to complete, however it will help to explore the underlying mechanisms for the wide variation in clinical behaviour of sMTC observed in practice. It will also help to develop a risk stratification system to help inform the aggressiveness of surgical treatment, the intensity of post-treatment follow-up and the use of adjuvant systemic therapy in these patients.

The use of Mendelian randomisation to investigate risk factors for thyroid cancer is currently a topic of significant interest, with several groups publishing work in the field in the time that I have been undertaking my research (Zhou *et al.* 2020; Yuan *et al.* 2020a). The main limitations of the use of MR in thyroid cancer is the relatively low incidence of thyroid cancer, and the fact that the genetic instruments for many of the proposed risk factors only account for a small proportion of variance in their corresponding risk factor trait. To overcome these limitations, larger datasets will be required, which will mean collaboration between groups. In addition to this, the UK Biobank is a dynamic resource, with frequent updates as data becomes available for the enrolled participants. As participants continue to be diagnosed with BNTD and DTC, the analyses described here can be repeated to test the findings as they currently stand. It will also be possible to focus on the larger group of UK Biobank participants with BNTD, to examine the effect of TSH on their risk of developing DTC. This may help to resolve the uncertainties regarding the differing effects of TSH on normal thyroid gland and thyroid nodules.

Genetic variants have been used in the study of thyroid cancer to produce polygenic risk scores (PRS) in an attempt to estimate the risk of thyroid cancer

(Liyanarachchi *et al.* 2020), and the genetic data in the UK Biobank is a valuable resource in contributing to this research, with the opportunity to extend the clinical applicability by focusing on patients with BNTD as well as thyroid cancer, in order to develop a risk stratification tool to help distinguish between benign and malignant thyroid nodules.

7.5 – Final Conclusions

The initial aims of this project were to explore the role of genetics in the diagnosis and management of MTC, and to add to current knowledge om the risk factors for DTC, using genetic variants and Mendelian randomisation techniques. The main outcomes of this thesis in working towards these aims are as follows: Firstly, I have helped to shed light on the clinical presentation of patients with *RET* mutations in the UK and highlighted the significant rate of germline *RET* mutation in those with presumed sporadic MTC, and explored the evidence for the use of somatic *RET* testing in patients lacking a germline mutation. This has attracted funding and approval for further study into the clinical significance of somatic *RET* mutations in sporadic MTC. Secondly, I have used novel techniques to explore causative roles for several proposed lifestyle-related risk factors for DTC. Whilst the results are far from conclusive, they do raise questions about the validity of previously accepted observational associations, and pave the way for future work to unpick the complex network of modifiable risk factors for thyroid cancer.

BIBLIOGRAPHY

Abraham D, Jackson NE, Gundara JS *et al*. MicroRNA profiling of sporadic and hereditary medullary thyroid cancer identifies predictors of nodal metastasis, prognosis and potential therapeutic targets. Clinical Cancer Research. 2011;17(14):4772-81.

Aschebrook-Kilfoy B, Schechter RB, Shih YC *et al*. The clinical and economic burden of a sustained increase in thyroid cancer incidence. Cancer Epidemiollogy, Biomarkers & Prevention. 2013;22:1252-9.

Aubert S, Berdelou A, Gnemmi V *et al*. Large sporadic thyroid medullary carcinomas: predictive factors for lymph node involvement. Virchows Archiv. 2018;472(3):461-8.

Barbieri RB, Bufalo NE, Secolin R *et al*. Polymorphisms of cell cycle control genes influence the development of sporadic medullary thyroid carcinoma. European Journal of Endocrinology. 2014;171(6):761-7.

Basen-Engquist K, Chang M. Obesity and cancer risk: recent review and evidence. Current Oncology Reports. 2011;13(1):71-6.

Beg MS, Brenner AJ, Sachdev J *et al*. Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. Investigational New dDrugs. 2017;35(2):180-8.

Belfiore A, La Rosa GL, La Porta GA *et al*. Cancer risk in patients with cold thyroid nodules: relevance of iodine intake, sex, age, and multinodularity. The American Journal of Medicine. 1992;93(4):363-9.

Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nature Genetics. 2008;40(6):695-701.

Boelaert K. The association between serum TSH concentration and thyroid cancer. Endocrine-Related Cancer. 2009;16(4):1065-72.

Boelaert K, Horacek J, Holder RL *et al*. Serum thyrotropin concentration as a novel predictor of malignancy in thyroid nodules investigated by fine-needle aspiration. The Journal of Clinical Endocrinology & Metabolism. 2006;91(11):4295-301.

Bongiovanni M, Spitale A, Faquin WC *et al*. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. Acta Cytologica. 2012;56(4):333-9.

Booth A, Magnuson A, Fouts J *et al*. Adipose tissue, obesity and adipokines: role in cancer promotion. Hormone Molecular Biology and Clinical Investigation. 2015;21(1):57-74.

Bowden J, Davey Smith G, Haycock PC *et al*. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic Epidemiology. 2016;40(4):304–314.

Bowden J, Del Greco MF, Minelli C *et al*. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. International Journal of Epidemiology. 2016;45(6):1961–1974.

Bowden J, Dudbridge F. Unbiased estimation of odds ratios: combining genome wide association scans with replication studies. Genetic Epidemiology. 2009;33:406-18.

Bowden J, Spiller W, Del Greco MF *et al*. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. International Journal of Epidemiology. 2018;47(4):1264-78.

Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. International Journal of Epidemiology. 2013;42(5):1497-501.

Brito JP, Ito Y, Miyauchi A *et al*. A clinical framework to facilitate risk stratification when considering an active surveillance alternative to immediate biopsy and surgery in papillary microcarcinoma. Thyroid. 2016;26(1):144-9.

Bruno DS, Berger NA. Impact of bariatric surgery on cancer risk reduction. Annals of Translational Medicine. 2020;8(Suppl 1):S13.

Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. American Journal of Epidemiology. 2015;181(4):251-60.

Buxton OM, Cain SW, O'Connor SP *et al*. Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. Science Translational Medicine. 2012;4(129):129ra43-.

Bycroft C, Freeman C, Petkova D *et al*. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562(7726):203-209.

Cancer Research UK, https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/thyroid-cancer/incidence#heading-Two, Accessed October 2020.

Cao Y, Wang Z, Gu J *et al*. Reproductive factors but not hormonal factors associated with thyroid cancer risk: a systematic review and meta-analysis. BioMed research international. 2015;103515.

Castellone MD, Verrienti A, Magendra Rao D *et al*. A novel de novo germ-line V292M mutation in the extracellular region of RET in a patient with phaeochromocytoma and medullary thyroid carcinoma: functional characterization. Clinical Endocrinology. 2010;73(4):529-34.

Cavedon E, Barollo S, Bertazza L *et al*. Prognostic impact of miR-224 and RAS mutations in medullary thyroid carcinoma. International Journal of Endocrinology. 2017;2017:4915736

Ceolin L, Goularte AP, Ferreira CV *et al*. Global DNA methylation profile in medullary thyroid cancer patients. Experimental and Molecular Pathology. 2018;105(1):110-114.

Cerrato A, De Falco V, Santoro M. Molecular genetics of medullary thyroid carcinoma: the quest for novel therapeutic targets. Journal of Molecular Endocrinology. 2009;43(4):143-55.

Chaker L, Ligthart S, Korevaar TI *et al*. Thyroid function and risk of type 2 diabetes: a population-based prospective cohort study. BMC medicine. 2016;14(1):150.

Chang CC, Chow CC, Tellier LC *et al*. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:s13742-015.

Chen Y, Chen Y, Wang N *et al*. Are thyroid nodules associated with sex-related hormones? A cross-sectional SPECT-China study. BMJ Open. 2017;7(8):e015812.

Chen GG, Vlantis AC, Zeng Q *et al*. Regulation of cell growth by estrogen signaling and potential targets in thyroid cancer. Current Cancer Drug Targets. 2008;8(5):367-77.

Cheng Y, He C, Wang M *et al*. Targeting epigenetic regulators for cancer therapy: Mechanisms and advances in clinical trials. Signal Transduction and Targeted Therapy. 2019;4(1):1-39.

Chiu HK, Sanda S, Fechner PY *et al*. Correlation of TSH with the risk of paediatric thyroid carcinoma. Clinical Endocrinology. 2012;77(2):316-22.

Cho A, Chang Y, Ahn J *et al*. Cigarette smoking and thyroid cancer risk: a cohort study. British Journal of Cancer. 2018;119(5):

Cho SW, Yi KH, Han SK *et al*. Therapeutic potential of metformin in papillary thyroid cancer in vitro and in vivo. Molecular and Cellular Endocrinology. 2014;393(1-2):24-9.

Cho YA, Kim J. Thyroid cancer risk and smoking status: a meta-analysis. Cancer Causes & Control. 2014;25(9):1187-95.

Choi JS, Kim EK, Moon HJ *et al*. Higher body mass index may be a predictor of extrathyroidal extension in patients with papillary thyroid microcarcinoma. Endocrine. 2015;48(1):264-71.

Choi WJ, Kim J. Dietary factors and the risk of thyroid cancer: a review. Clinical Nutrition Research. 2014;3(2):75.

Ciampi R, Mian C, Fugazzola L *et al*. Evidence of a low prevalence of RAS mutations in a large medullary thyroid cancer series. Thyroid. 2013;23(1):50-7.

Ciampi R, Romei C, Ramone T *et al*. Genetic landscape of somatic mutations in a large cohort of sporadic medullary thyroid carcinomas studied by next-generation targeted sequencing. iScience. 2019;20:324-36.

Cibas ES, Ali SZ. The Bethesda system for reporting thyroid cytopathology. Thyroid. 2009;19(11):1159-65.

Clarke TK, Adams MJ, Davies G *et al*. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N= 112 117). Molecular Psychiatry. 2017;22:1376-84.

Clavel-Chapelon F, Guillas G, Tondeur L *et al*. Risk of differentiated thyroid cancer in relation to adult weight, height and body shape over life: the

French E3N cohort. International Journal of Cancer. 2010;126(12):2984-2990.

Collins R. What makes UK Biobank special? Lancet 2012;379(9822):1173-1174.

Cooper DS, Doherty GM, Haugen BR *et al*. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association (ATA) guidelines taskforce on thyroid nodules and differentiated thyroid cancer. Thyroid. 2009;19(11):1167-214.

Cornelis MC, Byrne EM, Esko T *et al.* Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. Molecular Psychiatry. 2015;20:647-56.

Cote GJ, Evers C, Hu MI *et al*. Prognostic significance of circulating RET M918T mutated tumor DNA in patients with advanced medullary thyroid carcinoma. The Journal of Clinical Endocrinology & Metabolism. 2017;102(9):3591-9.

Cross P, Chandra A, Giles T *et al*. Guidance on the reporting of thyroid cytology specimens. 2nd edition. London, United Kingdom: Royal College of Pathologists, 2016.

Cust AE, Kaaks R, Friedenreich C *et al*. Plasma adiponectin levels and endometrial cancer risk in pre-and postmenopausal women. The Journal of Clinical Endocrinology & Metabolism. 2007 Jan 1;92(1):255-63.

Dalamaga M, Diakopoulos KN, Mantzoros CS. The role of adiponectin in cancer: a review of current evidence. Endocrine Reviews. 2012;33(4):547-94.

Dal Maso L, Bosetti C, La Vecchia C *et al*. Risk factors for thyroid cancer: an epidemiological review focused on nutritional factors. Cancer Causes & Control. 2009;20(1):75-86.

Dal Maso L, La Vecchia C, Franceschi S *et al*. A pooled analysis of thyroid cancer studies. V. Anthropometric factors. Cancer Causes & Control. 2000;11(2):137-44.

Dabir S, Babakoohi S, Kluge A *et al*. RET mutation and expression in small-cell lung cancer. Journal of Thoracic Oncology. 2014;9(9):1316-23.

Davey RA, Findlay DM. Calcitonin: physiology or fantasy? Journal of Bone and Mineral Research. 2013;28(5):973-9.

Davey Smith G, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease?. International Journal of Epidemiology. 2003;32(1):1-22.

Davey Smith G, Phillips AN. Confounding in epidemiological studies:why "independent" effects may not be all they seem. BMJ. 1992;305:757–9.

Davies L, Ouellette M, Hunter M *et al*. The increasing incidence of small thyroid cancers: where are the cases coming from? Laryngoscope. 2010;120:2446–2451.

Davies L, Welch HG. Current thyroid cancer trends in the United States. JAMA Otolaryngology Head and Neck Surgery. 2014;140:317–322.

Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1):12-27.

Dean DS, Gharib H. Epidemiology of thyroid nodules. Best practice & research Clinincal Endocrinology & Metabolism 2008;22(6):901-11.

Derwahl M, Nicula D. Estrogen and its role in thyroid cancer*. Endocrine-Related Cancer*. 2014;21:T273-83.

Doherty A, Jackson D, Hammerla N *et al*. Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. PloS one. 2017;12:e0169649.

Donis-Keller H, Dou S, Chi D *et al*. Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. Human Molecular Genetics. 1993;2(7):851-6.

Dossus L, Franceschi S, Biessy C *et al*. Adipokines and inflammation markers and risk of differentiated thyroid carcinoma: The EPIC study. International Journal of Cancer. 2018;142(7):1332-42.

Duggal P, Ladd-Acosta C, Ray D *et al*. The Evolving Field of Genetic Epidemiology: From Familial Aggregation to Genomic Sequencing. American Journal of Epidemiology. 2019;188(12):2069-77.

Dvorakova S, Vaclavikova E, Sykorova V *et al*. Somatic mutations in the RET proto-oncogene in sporadic medullary thyroid carcinomas. Molecular and Cellular Endocrinology. 2008;284(1-2):21-7.

El Naofal M, Kim A, Yon HY *et al*. Role of CDKN2C Fluorescence In Situ Hybridization in the Management of Medullary Thyroid Carcinoma. Annals of Clinical & Laboratory Science. 2017;47(5):523-8.

Elisei R, Alevizaki M, Conte-Devolx B *et al*. European thyroid association guidelines for genetic testing and its clinical consequences in medullary thyroid cancer. European Thyroid Journal. 2012;1(4):216-31.

Elisei R, Cosci B, Romei C *et al*. Prognostic significance of somatic RET oncogene mutations in sporadic medullary thyroid cancer: a 10- year follow-up study. Journal of Clinical Endocrinology & Metabolism. 2008;93(3):682-687.

Elisei R, Schlumberger MJ, Müller SP *et al*. Cabozantinib in progressive medullary thyroid cancer. Journal of Clinical Oncology. 2013;31(29):3639.

Elisei R, Tacito A, Ramone T *et al*. Twenty-Five Years' Experience on RET Genetic Screening on Hereditary MTC: An Update on The Prevalence of Germline RET Mutations. Genes. 2019;10(9):698.

Esfandiari NH, Hughes DT, Yin H *et al*. The effect of extent of surgery and number of lymph node metastases on overall survival in patients with medullary thyroid cancer. The Journal of Clinical Endocrinology & Metabolism. 2014;99(2):448-54.

Eng C, Clayton D, Schuffenecker I *et al*. The relationship between specific RET proto-oncogene mutations and disease phenotype in multiple endocrine neoplasia type 2: International RET Mutation Consortium analysis. Journal of the American Medical Association. 1996;276(19):1575-9.

Erzurumluoglu AM, Liu M, Jackson VE *et al*. Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci. Molecular Psychiatry. 2020;25:2392-409.

Fallah M, Pukkala E, Tryggvadottir L *et al*. Risk of thyroid cancer in first-degree relatives of patients with non-medullary thyroid cancer by histology type and age at diagnosis: a joint study from five Nordic countries. Journal of Medical Genetics. 2013;50(6)373-82.

Fan W, Xiao C, Wu F. Analysis of risk factors for cervical lymph node metastases in patients with sporadic medullary thyroid carcinoma. Journal of International Medical Research. 2018;46:1982–1989.

Fang Z, Song M, Lee D *et al*. The Role of Mendelian Randomization Studies in Deciphering the Effect of Obesity on Cancer. Journal of the National Cancer Institute. 2021;114(3):361-371.

Fazioli F, Piccinini G, Appolloni G *et al*. A new germline point mutation in Ret exon 8 (cys515ser) in a family with medullary thyroid carcinoma. Thyroid. 2008;18(7):775-82.

Ferrara R, Auger N, Auclin E *et al*. Clinical and translational implications of RET rearrangements in non–small cell lung cancer. Journal of Thoracic Oncology. 2018;13(1):27-45.

Fiore M, Cristaldi A, Okatyeva V *et al*. Dietary habits and thyroid cancer risk: A hospital-based case–control study in Sicily (South Italy). Food and Chemical Toxicology. 2020;146:111778.

Fiore E, Rago T, Latrofa F *et al.* Hashimoto's thyroiditis is associated with papillary thyroid carcinoma: role of TSH and of treatment with L-thyroxine. Endocrine Related Cancer. 2011;18(4):429.

Fiore E, Rago T, Provenzale MA *et al*. Lower levels of TSH are associated with a lower risk of papillary thyroid cancer in patients with thyroid nodular disease: thyroid autonomy may play a protective role. Endocrine-Related Cancer. 2009;16(4):1251-60.

Fighera TM, Perez CL, Faris N *et al*. TSH levels are associated with increased risk of thyroid carcinoma in patients with nodular disease. Endokrynologia Polska. 2015;66(6):480-5.

Forbes SA, Bindal N, Bamford S *et al*. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Research. 2010;39(suppl_1):D945-50.

Forterre A, Komuro H, Aminova S *et al*. A comprehensive review of cancer MicroRNA therapeutic delivery strategies. Cancers. 2020;12(7):1852.

Franceschi S, Preston-Martin S, Dal Maso L *et al*. A pooled analysis of case–control studies of thyroid cancer. IV. Benign thyroid diseases. Cancer Causes Control 1999;10(6):583-595.

Franco AT, Malaguarnera R, Refetoff S *et al*. Thyrotrophin receptor signaling dependence of Braf-induced thyroid tumor initiation in mice. Proceedings of the National Academy of Sciences. 2011;108(4):1615-20.

Frank-Raue K, Höppner W, Frilling A *et al*. Mutations of the ret protooncogene in German multiple endocrine neoplasia families: relation between genotype and phenotype. German Medullary Thyroid Carcinoma Study Group. The Journal of Clinical Endocrinology & Metabolism. 1996;81(5):1780-3.

Fry A, Littlejohns TJ, Sudlow C *et al*. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. American Journal of Epidemiology. 2017;186(9):1026-1034.

Furukawa K, Preston D, Funamoto S *et al*. Long-term trend of thyroid cancer risk among Japanese atomic-bomb survivors: 60 years after exposure. International Journal of Cancer. 2013;132(5):1222-6.

Fussey JM, Beaumont RN, Wood AR *et al*. Does obesity cause thyroid cancer? A Mendelian randomization study. Journal of Clinical Endocrinology & Metabolism. 2020;105(7):e2398-407.

Fussey JM, Bradley PJ, Smith JA. Controversies in the surgical management of sporadic medullary thyroid carcinoma. Current Opinion in Otolaryngology & Head and Neck Surgery. 2020;28(2):68-73.

Fussey JM, Beaumont RN, Wood AR *et al*. Mendelian randomization to investigate the link between TSH and thyroid cancer. Endocrine-Related Cancer. 2021;28:L11-4.

Fussey JM, Smith JA, Cleaver R *et al*. Diagnostic RET genetic testing in 1,058 index patients: a UK centre perspective. Clinical Endocrinology. 2021;95(2):295-302.

Fussey JM, Vaidya B, Kim D *et al*. The role of molecular genetics in the clinical management of sporadic medullary thyroid carcinoma: A systematic review. Clinical Endocrinology. 2019;91(6):697-707.

Gabriel SB, Schaffner SF, Nguyen H *et al*. The structure of haplotype blocks in the human genome. Science. 2002;296(5576):2225-9.

Gao R, Jia X, Liang Y *et al*. Papillary Thyroid Micro Carcinoma: The Incidence of High-Risk Features and Its Prognostic Implications. Frontiers in Endocrinology. 2019;10:74.

Gao C, Patel CJ, Michailidou K *et al*. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. International Journal of Epidemiology. 2016;45(3):896-908.

Galuppini F, Bertazza L, Barollo S *et al*. MiR-375 and YAP1 expression profiling in medullary thyroid carcinoma and their correlation with clinical–pathological features and outcome. Virchows Archiv. 2017;471(5):651-8.

Gerlinger M, Rowan AJ, Horswell S *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. New England Journal of Medicine. 2012;366:883-92.

Gilliland FD, Hunt WC, Morris DM *et al*. Prognostic factors for thyroid carcinoma: A population-based study of 15,698 cases from the Surveillance, Epidemiology and End Results (SEER) Program 1973–1991. Cancer: Interdisciplinary International Journal of the American Cancer Society. 1997;79(3):564-73.

Giovannucci E, Harlan DM, Archer MC *et al*. Diabetes and cancer: a consensus report. Diabetes Care. 2010;33(7):1674-85.

Golbert L, de Cristo AP, Faccin CS *et al*. Serum TSH levels as a predictor of malignancy in thyroid nodules: A prospective study. PloS one. 2017;12(11):e0188123.

Goto M, Miller RW, Ishikawa Y *et al*. Excess of rare cancers in Werner syndrome (adult progeria). Cancer Epidemiology, Biomarkers & Prevention. 1996;5:239–246.

Griseri P, Garrone O, Sardo AL *et al*. Genetic and epigenetic factors affect RET gene expression in breast cancer cell lines and influence survival in patients. Oncotarget. 2016;7(18):26465.

Grubbs EG, Williams MD, Scheet P *et al*. Role of CDKN2C copy number in sporadic medullary thyroid carcinoma. Thyroid. 2016;26(11):1553-62.

Gudmundsson J, Sulem P, Gudbjartsson DF *et al*. Discovery of common variants associated with low TSH levels and thyroid cancer risk. Nature Genetics. 2012;44(3):319.

Gudmundsson J, Thorleifsson G, Sigurdsson JK *et al*. A genome-wide association study yields five novel thyroid cancer risk loci. Nature Communications. 2017;8(1):1-8.

Gul K, Ozdemir D, Dirikoc A *et al*. Are endogenously lower serum thyroid hormones new predictors for thyroid malignancy in addition to higher serum thyrotropin?. Endocrine. 2010;37(2):253-60.

Guignard R, Truong T, Rougier Y *et al*. Alcohol drinking, tobacco smoking, and anthropometric characteristics as risk factors for thyroid cancer: a countrywide case-control study in New Caledonia. American Journal of Epidemiology. 2007;166(10):1140-9.

Guo Y, Warren Andersen S, Shu XO *et al*. Genetically predicted body mass index and breast cancer risk: Mendelian randomization analyses of data from 145,000 women of European descent. PLoS Medicine. 2016;13(8):e1002105.

Gursoy A. Rising thyroid cancer incidence in the world might be related to insulin resistance. Medical Hypotheses. 2010;74(1):35-6.

Guth S, Theune U, Aberle J *et al*. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. European Journal of Clinical Investigation. 2009;39(8):699-706.

Han MA, Kim JH. Coffee consumption and the risk of thyroid cancer: a systematic review and meta-analysis. International Journal of Environmental Research and Public Health. 2017;14(2):129.

Han TS, Lean ME. A clinical perspective of obesity, metabolic syndrome and cardiovascular disease. JRSM Cardiovascular Disease. 2016;5:2048004016633371.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.

Hannibal CG, Jensen A, Sharif H *et al*. Risk of thyroid cancer after exposure to fertility drugs: results from a large Danish cohort study. Human Reproduction. 2007;23(2):451-6.

Harris H, Miller OJ, Klein G *et al*. Suppression of malignancy by cell fusion. Nature. 1969;223(5204):363-8.

Haugen BR. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: what is new and what has changed?. Cancer. 2017;123(3):372-81.

Haymart MR, Glinberg SL, Liu J *et al*. Higher serum TSH in thyroid cancer patients occurs independent of age and correlates with extrathyroidal extension. Clinical Endocrinology. 2009:71(3):434-9.

Haymart MR, Repplinger DJ, Leverson GE *et al*. Higher serum thyroid stimulating hormone level in thyroid nodule patients is associated with greater risks of differentiated thyroid cancer and advanced tumor stage. The Journal of Clinical Endocrinology & Metabolism. 2008;93(3):809-14.

Hebbard L, Ranscht B. Multifaceted roles of adiponectin in cancer. Best Practice & Research Clinical Endocrinology & Metabolism. 2014;28(1):59-69.

Hedayati M, Zarif Yeganeh M, Sheikholeslami S *et al*. Diversity of mutations in the RET proto-oncogene and its oncogenic mechanism in medullary thyroid cancer. Critical Reviews in Clinical Laboratory Sciences. 2016;53(4):217-27.

Hegedüs L. The thyroid nodule. New England Journal of Medicine. 2004;351:1764-1771.

Hong SH, Myung SK, Kim HS, Korean Meta-Analysis (KORMA) Study Group. Alcohol intake and risk of thyroid cancer: a meta-analysis of observational studies. Cancer Research & Treatment. 2017;49:534.

Horn-Ross PL, Canchola AJ, Ma H *et al*. Hormonal factors and the risk of papillary thyroid cancer in the California Teachers Study cohort. Cancer Epidemiology and Prevention Biomarkers. 2011;20(8):1751-9.

Howell GM, Hodak SP, Yip L. RAS mutations in thyroid cancer. The Oncologist. 2013;18(8):926.

Howlader N, Noone AM, Krapcho M *et al*. (eds). SEER Cancer Statistics Review, 1975-2017, National Cancer Institute, Bethesda, MD, https://seer.cancer. gov/csr/1975_2017/, based on November 2019 SEER

data submission, posted to the SEER website, April 2020 (Last accessed 21/02/2021)

Huang H, Rusiecki J, Zhao N *et al*. Thyroid-stimulating hormone, thyroid hormones and risk of papillary thyroid cancer: a nested case-control study. Cancer Epidemiology, Biomarkers & Prevention. 2017;26(8):1209-18.

Huang L, Feng X, Yang W *et al.* Appraising the Effect of Potential Risk Factors on Thyroid Cancer: A Mendelian Randomization Study. Journal of Clinical Endocrinology and Metabolism. 2022;107(7):e2783-e2791.

Huang Y, Qu S, Zhu G *et al*. BRAF V600E mutation-assisted risk stratification of solitary intrathyroidal papillary thyroid cancer for precision treatment. JNCI: Journal of the National Cancer Institute. 2018;110(4):362-70.

Husson O, Haak HR, van Steenbergen LN *et al*. Rising incidence, no change in survival and decreasing mortality from thyroid cancer in The Netherlands since 1989. Endocrine-Related Cancer. 2013; 20:263-71.

Ichikawa Y, Saito E, Abe Y *et al*. Presence of TSH receptor in thyroid neoplasms. The Journal of Clinical Endocrinology & Metabolism. 1976;42(2):395-8.

Ikegami K, Refetoff S, Van Cauter E *et al*. Interconnection between circadian clocks and thyroid function. Nature Reviews Endocrinology. 2019;15(10):590-600.

Ito Y, Miyauchi A, Kihara M *et al*. Static prognostic factors and appropriate surgical designs for patients with medullary thyroid carcinoma: the second report from a single-institution study in Japan. World Journal of Surgery. 2018;42(12):3954-66.

Ito Y, Miyauchi A, Kihara M *et al*. Patient age is significantly related to the progression of papillary microcarcinoma of the thyroid under observation. Thyroid. 2014;24(1):27-34.

Ito Y, Miyauchi A, Kihara M *et al*. Overall survival of papillary thyroid carcinoma patients: a single-institution long-term follow-up of 5897 patients. World journal of surgery. 2018;42(3):615-22.

Jeong JS, Kim HK, Lee CR, Park S, Park JH, Kang SW, Jeong JJ, Nam KH, Chung WY, Park CS. Coexistence of chronic lymphocytic thyroiditis with papillary thyroid carcinoma: clinical manifestation and prognostic outcome. Journal of Korean medical science. 2012;27(8):883-9.

Jones SE, Lane JM, Wood AR *et al*. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. Nature Communications. 2019;10:1-1.

Jonklaas J, Nsouli-Maktabi H, Soldin SJ. Endogenous thyrotropin and triiodothyronine concentrations In individuals with thyroid cancer. Thyroid. 2008;18(9):943-52.

Juiz-Valiña P, Outeiriño-Blanco E, Pértega S *et al*. Effect of Weight Loss after Bariatric Surgery on Thyroid-Stimulating Hormone Levels in Euthyroid Patients with Morbid Obesity. Nutrients. 2019;11(5):1121.

Kanno J, Onodera H, Furuta K *et al*. Tumor-promoting effects of both iodine deficiency and iodine excess in the rat thyroid. Toxicologic Pathology. 1992;20(2):226-35.

Kebebew E, Greenspan FS, Clark OH *et al*. Extent of disease and practice patterns for medullary thyroid cancer. Journal of the American College of Surgeons. 2005;200(6):890-6.

Kebebew E, Weng J, Bauer J *et al*. The prevalence and prognostic value of BRAF mutation in thyroid cancer. Annals of surgery. 2007;246(3):466.

Kettner NM, Katchy CA, Fu L. Circadian gene variants in cancer. Annals of medicine. 2014;46(4):208-20.

Kihara M, Miyauchi A, Yoshioka K *et al*. Germline RET mutation carriers in Japanese patients with apparently sporadic medullary thyroid carcinoma: a single institution experience. Auris Nasus Larynx. 2016;43(5):551-5.

Kilfoy BA, Devesa SS, Ward MH *et al*. Gender is an age-specific effect modifier for papillary cancers of the thyroid gland. Cancer Epidemiology and Prevention Biomarkers. 2009;18(4):1092-100.

Kim HJ, Kim NK, Choi JH *et al*. Associations between body mass index and clinico-pathological characteristics of papillary thyroid cancer. Clinical Endocrinology. 2013;78(1):134-40.

Kinzler KW, Vogelstein B. Gatekeepers and caretakers. Nature. 1997;386(6627):761-3.

Kitahara CM, Körmendiné Farkas D, Jørgensen JO *et al*. Benign Thyroid Diseases and Risk of Thyroid Cancer: A Nationwide Cohort Study. Journal of Clinical Endocrinology & Metabolism. 2018;103(6):2216-2224.

Kitahara CM, Linet MS, Freeman LE *et al*. Cigarette smoking, alcohol intake, and thyroid cancer risk: a pooled analysis of five prospective studies in the United States. Cancer Causes & Control. 2012;23(10):1615-24.

Kitahara CM, Platz EA, Freeman LB *et al*. Obesity and thyroid cancer risk among US men and women: a pooled analysis of 5 prospective studies. Cancer Epidemiology and Prevention Biomarkers. 2011:cebp-1220.

Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. Nature Reviews Endocrinology. 2016;12(11):646-53.

Klubo-Gwiezdzinska J, Costello Jr J, Patel A *et al*. Treatment with metformin is associated with higher remission rate in diabetic patients with thyroid cancer. The Journal of Clinical Endocrinology & Metabolism. 2013;98(8):3269-79.

Knudsen N, Bülow I, Laurberg P *et al*. Alcohol consumption is associated with reduced prevalence of goitre and solitary thyroid nodules. Clinical Endocrinology. 2001;55:41-6.

Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proceedings of the National Academy of Sciences. 1971;68(4):820-3.

Köhler A, Chen B, Gemignani F *et al*. Genome-wide association study on differentiated thyroid cancer. The Journal of Clinical Endocrinology & Metabolism. 2013;98(10):E1674-81.

Krátký J, Ježková J, Kosák M *et al*. Positive antithyroid antibodies and nonsuppressed TSH are Associated with thyroid cancer: a retrospective cross-sectional study. International Journal of Endocrinology. 2018;9793850

Kreiger N, Parkes R. Cigarette smoking and the risk of thyroid cancer. European Journal of Cancer. 2000;36:1969-73.

Kung AW, Chau MT, Lao TT *et al*. The effect of pregnancy on thyroid nodule formation. The Journal of Clinical Endocrinology & Metabolism. 2002;87(3):1010-4.

Lang M, Pelkonen O. Metabolism of xenobiotics and chemical carcinogenesis. IARC scientific publications. 1999;(148):13-22.

Lange K. An approximate model of polygenic inheritance. Genetics. 1997;147(3):1423-30.

Larsen LV, Mirebeau-Prunier D, Imai T *et al*. Primary hyperparathyroidism as first manifestation in MEN 2A: an international multicenter study. Endocrine Connections. 2020;9(6):489-497.

Larsson SC, Carter P, Kar S, *et al*. Smoking, alcohol consumption, and cancer: a mendelian randomisation study in UK Biobank and international genetic consortia participants. PLoS medicine. 2020;17(7):e1003178.

Larsson SC, Traylor M, Malik R *et al*. Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. British Medical Journal. 2017;359:j5375

Lassalle S, Zangari J, Popa A *et al*. MicroRNA-375/SEC23A as biomarkers of the in vitro efficacy of vandetanib. Oncotarget. 2016;7(21):30461.

Latteyer S, Klein-Hitpass L, Khandanpour C *et al*. A 6-base pair in frame germline deletion in exon 7 of RET leads to increased RET phosphorylation, ERK activation, and MEN2A. The Journal of Clinical Endocrinology & Metabolism. 2016;101(3):1016-22.

Lawlor DA, Harbord RM, Sterne JA *et al*. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in medicine. 2008;27(8):1133-63.

Lebeault M, Pinson S, Guillaud-Bataille M *et al*. Nationwide French study of RET variants detected from 2003 to 2013 suggests a possible influence of polymorphisms as modifiers. Thyroid. 2017;27(12):1511-22.

Lehrer S, Diamond EJ, Stone NN *et al*. Serum triiodothyronine is increased in men with prostate cancer and benign prostatic hyperplasia. The Journal of Urology. 2002;168(6):2431-3.

Li H, Qian J. Association of diabetes mellitus with thyroid cancer risk: A meta-analysis of cohort studies. Medicine. 2017 Nov;96(47).

Li J, Fu H, Xu C *et al*. miR-183 inhibits TGF-β1-induced apoptosis by downregulation of PDCD4 expression in human hepatocellular carcinoma cells. BMC cancer. 2010;10(1):354.

Lim H, Devesa SS, Sosa JA *et al*. Trends in thyroid cancer incidence and mortality in the United States, 1974-2013. *JAMA*. 2017;317:1338-48.

Lin WM, Baker AC, Beroukhim R *et al*. Modeling genomic diversity and tumor dependency in malignant melanoma. Cancer Research. 2008;68(3):664-73.

Liu T, Wang N, Cao J *et al*. The age-and shorter telomere-dependent TERT promoter mutation in follicular thyroid cell-derived carcinomas. Oncogene. 2014;33(42):4978.

Liyanarachchi S, Gudmundsson J, Ferkingstad E *et al*. Assessing thyroid cancer risk using polygenic risk scores. Proceedings of the National Academy of Sciences. 2020;117(11):5997-6002.

Loh PR, Tucker G, Bulik-Sullivan BK *et al*. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nature Genetics. 2015;47(3):284–290.

Lun Y, Wu X, Xia Q, Han Y, Zhang X, Liu Z, Wang F, Duan Z, Xin S, Zhang J. Hashimoto's thyroiditis as a risk factor of papillary thyroid cancer may improve cancer prognosis. Otolaryngology--Head and Neck Surgery. 2013 Mar;148(3):396-402.

Lundgren CI, Hall P, Dickman PW *et al*. Clinically significant prognostic factors for differentiated thyroid carcinoma: a population-based, nested case–control study. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2006;106(3):524-31.

Luo J, Sands M, Wactawski-Wende J *et al*. Sleep disturbance and incidence of thyroid cancer in postmenopausal women the Women's Health Initiative. Americal Journal of Epidemiology. 2013;177:42-9.

Livhits MJ, Zhu CY, Kuo EJ *et al*. Effectiveness of molecular testing techniques for diagnosis of indeterminate thyroid nodules: a randomized clinical trial. JAMA oncology. 2021;7(1):70-7.

Locke AE, Kahali B, Berndt SI *et al*. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206.

Lu C, Zhao L, Ying H *et al*. Growth activation alone is not sufficient to cause metastatic thyroid cancer in a mouse model of follicular thyroid carcinoma. Endocrinology. 2010;151(4):1929-39.

Lyra J, Vinagre J, Batista R *et al*. mTOR activation in medullary thyroid carcinoma with RAS mutation. European Journal of Endocrinology. 2014;171(5):633-40.

Ma J, Huang M, Wang L *et al*. Obesity and risk of thyroid cancer: evidence from a meta-analysis of 21 observational studies. Medical science monitor: International Medical Journal of Experimental and Clinical Research. 2015;21:283.

Machens A, Dralle H. Biomarker-based risk stratification for previously untreated medullary thyroid cancer. Journal of Clinical Endocrinology & Metabolism. 2010;95:2655–2663.

Machens A, Gimm O, Hinze R *et al*. Genotype-phenotype correlations in hereditary medullary thyroid carcinoma: oncological features and biochemical properties. The Journal of Clinical Endocrinology & Metabolism. 2001;86(3):1104-9.

Machens A, Niccoli-Sire P, Hoegel J *et al*. Early malignant progression of hereditary medullary thyroid cancer. New England Journal of Medicine. 2003;349(16):1517-25.

Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31(21):3555-7.

Maciel RM, Camacho CP, Assumpção LV *et al*. Genotype and phenotype landscape of MEN2 in 554 medullary thyroid cancer patients: the BrasMEN study. Endocrine Connections. 2019;8(3):289-98.

Mahajan A, Taliun D, Thurner M *et al*. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nature Genetics. 2018;50(11):1505-13.

Mannic T, Meyer P, Triponez F *et al*. Circadian clock characteristics are altered in human thyroid malignant nodules. The Journal of Clinical Endocrinology & Metabolism. 2013;98(11):4446-56.

Marsh DJ, Learoyd DL, Andrew SD *et al*. Somatic mutations in the RET proto-oncogene in sporadic medullary thyroid carcinoma. Clinical Endocrinology. 1996;44(3):249-57.

Mathiesen JS, Kroustrup JP, Vestergaard P *et al*. Distribution of RET mutations in multiple endocrine neoplasia 2 in Denmark 1994–2014: a nationwide study. Thyroid. 2017;27(2):215-23.

Mazzaferri EL. Management of a solitary thyroid nodule. New England Journal of Medicine. 1993;328(8):553-9.

McGlashan JA (2018). ‘Developmental anatomy of the thyroid and parathyroid glands’ in Watkinson JC & Clarke RW (ed.) Scott-Brown’s Otolaryngology Head & Neck Surgery. Boca Raton: CRC Press, pp. 603-610

McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. Cell. 2017;168(4):613-28.

McLeod DS, Cooper DS, Ladenson PW *et al*. Prognosis of differentiated thyroid cancer in relation to serum thyrotropin and thyroglobulin antibody status at time of diagnosis. Thyroid. 2014;24(1):35-42.

McLeod DS, Watters KF, Carpenter AD *et al*. Thyrotropin and thyroid cancer diagnosis: a systematic review and dose-response meta-analysis. Journal of Clinical Endocrinology & Metabolism. 2012;97(8):2682–92

Mehanna H, Al-Maqbili T, Carter B *et al*. Differences in the recurrence and mortality outcomes rates of incidental and nonincidental papillary thyroid microcarcinoma: a systematic review and meta-analysis of 21 329 person-years of follow-up. The Journal of Clinical Endocrinology & Metabolism. 2014;99(8):2834-43.

Mian C, Pennelli G, Barollo S *et al*. Combined RET and Ki-67 assessment in sporadic medullary thyroid carcinoma: a useful tool for patient risk stratification. European Journal of Endocrinology. 2011;164(6):971-6.

Mian C, Pennelli G, Fassan M *et al*. MicroRNA profiles in familial and sporadic medullary thyroid carcinoma: preliminary relationships with RET status and outcome. Thyroid. 2012;22(9):890-6.

Miranda-Filho A, Lortet-Tieulent J, Bray F *et al.* Thyroid cancer incidence trends by histology in 25 countries: a population-based study. Lancet Diabetes Endocrinology. 2021;9:225-234.

Mitsiades CS, Poulaki V, McMullan C *et al*. Novel histone deacetylase inhibitors in the treatment of thyroid cancer. Clinical Cancer Research. 2005;11(10):3958-65.

Moeller LC, Führer D. Thyroid hormone, thyroid hormone receptors, and cancer: a clinical perspective. Endocrine-Related Cancer. 2013;20(2):R19-29.

Mohebati A, Shaha AR. Anatomy of thyroid and parathyroid glands and neurovascular relations. Clinical Anatomy. 2012;25(1):19-31.

Moon JH, Hyun MK, Lee JY *et al*. Prevalence of thyroid nodules and their associated clinical parameters: a large-scale, multicenter-based health checkup study. The Korean journal of internal medicine. 2018;33(4):753.

Moriggi G, Falzacappa CV, Mangialardo C *et al*. Thyroid hormones (T3 and T4): dual effect on human cancer cell proliferation. Anticancer Research. 2011;31(1):89-96.

Mortensen JD, Woolner LB, Bennett WA. Gross and microscopic findings in clinically normal thyroid glands. The Journal of Clinical Endocrinology & Metabolism. 1955;15(10):1270-80.

Motoyama K, Inoue H, Takatsuno Y *et al*. Over-and under-expressed microRNAs in human colorectal cancer. International journal of Oncology. 2009;34(4):1069-75.

Moura MM, Cavaco BM, Leite V. RAS proto-oncogene in medullary thyroid carcinoma Endocrine-Related Cancer. 2015;22(5):R235-52.

Moura MM, Cavaco BM, Pinto AE *et al*. Correlation of RET somatic mutations with clinicopathogenic features in sporadic medullary thyroid carcinomas. British Journal of Cancer. 2009;100(11):1777.

Moura MM, Cavaco BM, Pinto AE *et al*. High prevalence of RAS mutations in RET-negative sporadic medullary thyroid carcinomas. The Journal of Clinical Endocrinology & Metabolism. 2011;96(5):E863-8.

Mulligan LM. RET revisited: expanding the oncogenic portfolio. Nature Reviews Cancer. 2014;14(3):173-86.

Munafò MR, Tilling K, Taylor AE *et al*. Collider scope: when selection bias can substantially influence observed associations. International Journal of Epidemiology. 2018;47(1):226-35.

Nar A, Gedik O. The effect of metformin on leptin in obese patients with type 2 diabetes mellitus and nonalcoholic fatty liver disease. Acta Diabetologica. 2009;46(2):113-8.

Neel J, Schull W. Human Heredity. Chicago, IL: University of Chicago Press; 1954

Negri E, Dal Maso L, Ron E *et al*. A pooled analysis of case-control studies of thyroid cancer: II. Menstrual and reproductive factors. Cancer Causes & Control. 1999;10(2):143-55.

Nelen MR, van Staveren CG, Peeters EAJ *et al*. Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease. Human Molecular Genetics. 1997;6:1383–1387

Neumann HP, Bausch B, McWhinney SR *et al*. A. Germ-line mutations in nonsyndromic pheochromocytoma. New England Journal of Medicine. 2002;346(19):1459-66.

Nikiforov YE. Radiation-induced thyroid cancer: what we have learned from Chernobyl. Endocrine Pathology. 2006;17(4):307-18.

Nikiforov YE, Baloch ZW. Clinical validation of the ThyroSeq v3 genomic classifier in thyroid nodules with indeterminate FNA cytology. Cancer Cytopathology. 2019;127(4):225-30

Oberaigner W, Ebenbichler C, Oberaigner K *et al*. Increased cancer incidence risk in type 2 diabetes mellitus: results from a cohort study in Tyrol/Austria. BMC Public Health. 2014;14(1):1058.

Oakley GM, Curtin K, Pimentel R *et al*. Establishing a familial basis for papillary thyroid carcinoma using the Utah Population Database. JAMA Otolaryngology–Head & Neck Surgery. 2013;139(11):1171-4.

Okbay A, Beauchamp JP, Fontana MA *et al*. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539-42.

Owens M, Kivuva E, Quinn A *et al*. SOS 1 frameshift mutations cause pure mucosal neuroma syndrome, a clinical phenotype distinct from multiple endocrine neoplasia type 2B. Clinical Endocrinology. 2016;84(5):715-9.

Page C, Biet A, Boute P *et al*. 'Aggressive papillary'thyroid microcarcinoma. European Archives of Oto-Rhino-Laryngology. 2009;266(12):1959.

Park JW, Han CR, Zhao L *et al*. Inhibition of STAT3 activity delays obesity-induced thyroid carcinogenesis in a mouse model. Endocrine-Related Cancer. 2016;23(1):53-63.

Park S, Willingham MC, Qi J *et al*. Metformin and JQ1 synergistically inhibit obesity-activated thyroid cancer. Endocrine-Related Cancer. 2018;25(10):865-77.

Paszko Z, Sromek M, Czetwertynska M *et al*. The occurrence and the type of germline mutations in the RET gene in patients with medullary thyroid carcinoma and their unaffected kindreds from Central Poland. Cancer Investigation. 2007;25(8):742-9.

Pelizzo MR, Boschin IM, Bernante P *et al*. Natural history, diagnosis, treatment and outcome of medullary thyroid cancer: 37 years' experience on 157 patients. European Journal of Surgical Oncology (EJSO). 2007;33(4):493-7.

Pellegriti G, Mannarino C, Russo M, Terranova R, Marturano I, Vigneri R, Belfiore A. Increased mortality in patients with differentiated thyroid cancer

associated with Graves' disease. The Journal of Clinical Endocrinology & Metabolism. 2013 Mar 1;98(3):1014-21.

Pennelli G, Galuppini F, Barollo S *et al*. The PDCD4/miR-21 pathway in medullary thyroid carcinoma. Human Pathology. 2015;46(1):50-7.

Perros P, Boelaert K, Colley S *et al*. Guidelines for the management of thyroid cancer. Clinical Endocrinology. 2014; 81:1-22.

Peschansky VJ, Wahlestedt C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. Epigenetics. 2014;9(1):3-12.

Piirtola M, Jelenkovic A, Latvala A *et al*. Association of current and former smoking with body mass index: A study of smoking discordant twin pairs from 21 twin cohorts. PLoS One. 2018;13(7):e0200140.

Plail RO, Bussey HJR, Glazer G *et al*. Adenomatous polyposis: an association with carcinoma of the thyroid. British Journal of Surgery 1987;74:377-380

Policeni BA, Smoker WR, Reede DL. Anatomy and embryology of the thyroid and parathyroid glands. Seminars in Ultrasound, CT and MRI. 2012;33(2):104-114.

Popoveniuc G, Jonklaas J. Thyroid nodules. Medical Clinics. 2012;96(2):329-49.

Porcu E, Medici M, Pistis G *et al*. A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. PLoS Genetics. 2013;9(2):e1003266.

Prescott JD, Zeiger MA. The RET oncogene in papillary thyroid carcinoma. Cancer. 2015;121(13):2137-46.

Pulit SL, Stoneman C, Morris AP *et al*. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Human Molecular Genetics. 2018;28(1):166-174.

Radwanska P, Kosior-Korzecka U. Effect of leptin on thyroid-stimulating hormone secretion and nitric oxide release from pituitary cells of ewe lambs in vitro. Journal of Physiology & Pharmacology. 2014;65(1):145-51.

Rahbari R, Zhang L, Kebebew E. Thyroid cancer gender disparity. Future Oncology. 2010;6(11):1771-9.

Raue F, Frank-Raue K. Genotype-phenotype correlation in multiple endocrine neoplasia type 2. Clinics. 2012;67:69-75.

Richards S, Aziz N, Bale S *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine. 2015;17:405–423.

Richardson TG, Sanderson E, Palmer TM *et al*. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. PLoS medicine. 2020;17(3):e1003062.

Rinaldi S, Plummer M, Biessy C *et al*. Thyroid-stimulating hormone, thyroglobulin, and thyroid hormones and risk of differentiated thyroid carcinoma: the EPIC study. JNCI: Journal of the National Cancer Institute. 2014;106(6):dju097.

Riza E, Linos A, Petralias A *et al*. The effect of Greek herbal tea consumption on thyroid cancer: a case-control study. European Journal of Public Health. 2015;25:1001-5.

Robert C, Grob JJ, Stroyakovskiy D *et al*. Five-year outcomes with dabrafenib plus trametinib in metastatic melanoma. New England Journal of Medicine. 2019;381(7):626-36.

Roman S, Lin R, Sosa JA. Prognosis of medullary thyroid carcinoma: demographic, clinical, and pathologic predictors of survival in 1252 cases. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2006;107(9):2134-42.

Romei C, Casella F, Tacito A *et al*. New insights in the molecular signature of advanced medullary thyroid cancer: evidence of a bad outcome of cases with double RET mutations. Journal of Medical Genetics. 2016;53(11):729-34.

Romei C, Cosci B, Renzini G *et al*. RET genetic screening of sporadic medullary thyroid cancer (MTC) allows the preclinical diagnosis of unsuspected gene carriers and the identification of a relevant percentage of hidden familial MTC (FMTC). Clinical Endocrinology. 2011;74(2):241-7.

Romei C, Elisei R, Pinchera A *et al*. Somatic mutations of the RET protooncogene in sporadic medullary thyroid carcinoma are not restricted to exon 16 and are associated with tumor recurrence. The Journal of Clinical Endocrinology & Metabolism. 1996;81(4):1619-22.

Romei C, Ugolini C, Cosci B *et al*. Low prevalence of the somatic M918T RET mutation in micro-medullary thyroid cancer. Thyroid. 2012;22(5):476-81.

Romeo P, Colombo C, Granata R *et al*. Circulating miR-375 as a novel prognostic marker for metastatic medullary thyroid cancer patients. Endocrine-Related Cancer. 2018;25(3):217-31.

Ron E, Lubin JH, Shore RE *et al*. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. Radiation Research. 1995;141(3):259-77.

Rossing MA, Remler R, Voigt LF *et al*. Recreational physical activity and risk of papillary thyroid cancer (United States). Cancer Causes & Control. 2001;12:881-5.

Roti E, degli Uberti EC, Bondanelli M *et al*. Thyroid papillary microcarcinoma: a descriptive and meta-analysis study. European Journal of Endocrinology. 2008;159(6):659.

Rotondi M, Castagna MG, Cappelli C *et al*. Obesity does not modify the risk of differentiated thyroid cancer in a cytological series of thyroid nodules. European Thyroid Journal. 2016;5(2):125-31.

Ruiz-Ferrer M, Fernández RM, Navarro E, Antiñolo G *et al*. ESR2 Gene and Medullary Thyroid Carcinoma. Thyroid. 2017;27(11):1456-7.

Sarika HL, Papathoma A, Garofalaki M *et al*. Genetic screening of patients with medullary thyroid cancer in a referral center in Greece during the past two decades. European Journal of Endocrinology. 2015;172(4):501-9.

Satta MA, De Rosa G, Testa A *et al*. Thyroid cancer in suppressed contralateral lobe of patients with hot thyroid nodule. European Journal of Cancer. 1993;29(8):1190-2.

Sattar N, Preiss D. Reverse causality in cardiovascular epidemiological research: more common than imagined? Circulation 2017;135:2369–72.

Schilling T, Bürck J, Sinn HP *et al*. Prognostic value of codon 918 (ATG→ ACG) RET proto-oncogene mutations in sporadic medullary thyroid carcinoma. International Journal of Cancer. 2001;95(1):62-66.

Schmid D, Behrens G, Jochem C *et al*. Physical activity, diabetes, and risk of thyroid cancer: a systematic review and meta-analysis. European Journal of Epidemiology. 2013;28(12):945-58.

Schmid D, Ricci C, Behrens G *et al*. Adiposity and risk of thyroid cancer: a systematic review and meta-analysis. Obesity Reviews. 2015;16(12):1042-54.

Schneider TC, De Wit D, Links TP *et al*. Beneficial effects of the mTOR inhibitor everolimus in patients with advanced medullary thyroid carcinoma: subgroup results of a phase II trial. International Journal of Endocrinology. 2015;2015:348124.

Shabani N, Razaviyan J, Paryan M *et al*. vEvaluation of miRNAs expression in medullary thyroid carcinoma tissue samples: miR-34a and miR-144 as promising overexpressed markers in MTC. Human Pathology. 2018; 79:212-221.

Shah JP. Thyroid carcinoma: epidemiology, histology, and diagnosis. Clinical Advances in Hematology & Oncology: H&O. 2015;13(4 Suppl 4):3.

Shen CT, Zhang XY, Qiu ZL *et al.* Thyroid autoimmune antibodies in patients with papillary thyroid carcinoma: a double-edged sword? Endocrine. 2017;58(1):176-83.

Sherman SI, Clary DO, Elisei R *et al*. Correlative analyses of RET and RAS mutations in a phase 3 trial of cabozantinib in patients with progressive, metastatic medullary thyroid cancer. Cancer. 2016;122(24):3856-64.

Simbolo M, Mian C, Barollo S *et al*. High-throughput mutation profiling improves diagnostic stratification of sporadic medullary thyroid carcinomas. Virchows archiv. 2014;465(1):73-8.

Singh B, Shaha AR, Trivedi H *et al*. Coexistent Hashimoto's thyroiditis with papillary thyroid carcinoma: impact on presentation, management, and outcome. Surgery. 1999;126(6):1070-7.

Smallridge RC, Copland JA. Anaplastic thyroid carcinoma: pathogenesis and emerging therapies. Clinical Oncology. 2010;22(6):486-97.

Smith J, Read ML, Hoffman J, et al. Germline ESR2 mutation predisposes to medullary thyroid carcinoma and causes up-regulation of RET expression. Human Molecular Genetics. 2016;25(9):1836-1845.

Smittenaar CR, Petersen KA, Stewart K *et al*. Cancer incidence and mortality projections in the UK until 2035. British Journal of Cancer 2016;115(9):1147-1155.

Sponziello M, Durante C, Boichard A *et al*. Epigenetic-related gene expression profile in medullary thyroid cancer revealed the overexpression of the histone methyltransferases EZH2 and SMYD3 in aggressive tumours. Molecular and Cellular Endocrinology. 2014;392(1-2):8-13.

Spyer G, Ellard S, Turnpenny PD *et al*. Phenotypic multiple endocrine neoplasia type 2b, without endocrinopathy or RET gene mutation: implications for management. Thyroid. 2006;16(6):605-8.

Staniforth JU, Erdirimanne S, Eslick GD. Thyroid carcinoma in Graves' disease: a meta-analysis. International Journal of Surgery. 2016;27:118-25.

Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458(7239):719-24.

Subbiah V, Velcheti V, Tuch BB *et al*. Selective RET kinase inhibition for patients with RET-altered cancers. Annals of Oncology. 2018a;29(8):1869-76.

Subbiah V, Gainor JF, Rahal R *et al*. Precision targeted therapy with BLU-667 for RET-driven cancers. Cancer Discovery. 2018b;8(7):836-49.

Subbiah V, Hu MI, Wirth LJ *et al*. Pralsetinib for patients with advanced or metastatic RET-altered thyroid cancer (ARROW): a multi-cohort, open-label,

registrational, phase 1/2 study. The Lancet Diabetes & Endocrinology. 2021;9(8):491-501.

Sudlow C, Gallacher J, Allen N *et al*. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine 2015;12(3):e1001779.

Swanson JM. The UK Biobank and selection bias. The Lancet. 2012;380(9837):110.

Takahashi M, Ritz J, Cooper GM. Activation of a novel human transforming gene, ret, by DNA rearrangement. Cell. 1985;42(2):581-8.

Takezaki T, Hirose K, Inoue M *et al*. Risk factors of thyroid cancer among women in Tokai, Japan. Journal of Epidemiology. 1996;6:140-7.

Tamburrino A, Molinolo AA, Salerno P *et al*. Activation of the mTOR pathway in primary medullary thyroid carcinoma and lymph node metastases. Clinical Cancer Research. 2012;18(13):3532-40.

Tamez-Perez HE, Martínez E, Quintanilla-Flores DL *et al*. The rate of primary hypothyroidism in diabetic patients is greater than in the non-diabetic population. An observational study. Medicina Clinica. 2012;138(11):475-7.

Tang Y, Yan T, Wang G *et al*. Correlation between Insulin Resistance and Thyroid Nodule in Type 2 Diabetes Mellitus. International Journal of Endocrinology. 2017;2017:1617458.

Tee MC, Cao Y, Warnock GL *et al*. Effect of bariatric surgery on oncologic outcomes: a systematic review and meta-analysis. Surgical Endoscopy. 2013;27(12):4449-56.

Thakur S, Tobey A, Daley B *et al*. Limited utility of circulating cell-free DNA integrity as a diagnostic tool for differentiating between malignant and

benign thyroid nodules with indeterminate cytology (Bethesda Category III). Frontiers in oncology. 2019;9:905.

Thrift AP, Shaheen NJ, Gammon MD *et al*. Obesity and risk of esophageal adenocarcinoma and Barrett's esophagus: a Mendelian randomization study. JNCI: Journal of the National Cancer Institute. 2014;106(11).

Titova OE, Michaëlsson K, Vithayathil M *et al*. Sleep duration and risk of overall and 22 site-specific cancers: A Mendelian randomization study. International Journal of Cancer. 2021;148(4):914-20.

Townsend P, Phillimore P, Beattie A. Health and deprivation: inequality and the North. Routledge; London, 1988.

Trésallet C, Seman M, Tissier F *et al*. The incidence of papillary thyroid carcinoma and outcomes in operative patients according to their body mass indices. Surgery. 2014;156(5):1145-52.

Trimboli P, Castellana M, Virili C *et al*. Efficacy of Vandetanib in Treating Locally Advanced or Metastatic Medullary Thyroid Carcinoma According to RECIST Criteria: A Systematic Review and Meta-Analysis. Frontiers in Endocrinology. 2018;9:224.

Trimboli P, Cremonini N, Ceriani L *et al*. Calcitonin measurement in aspiration needle washout fluids has higher sensitivity than cytology in detecting medullary thyroid cancer: a retrospective multicentre study. Clinical Endocrinology. 2014;80(1):135-40.

Trimboli P, Scappaticcio L, Treglia G *et al*. Testing for BRAF (V600E) Mutation in Thyroid Nodules with Fine-Needle Aspiration (FNA) Read as Suspicious for Malignancy (Bethesda V, Thy4, TIR4): a Systematic Review and Meta-analysis. Endocrine Pathology. 2020;31(1):57-66.

Trimboli P, Treglia G, Guidobaldi L *et al*. Detection rate of FNA cytology in medullary thyroid carcinoma: a meta-analysis. Clinical Endocrinology. 2015;82(2):280-5.

Tseng CH. Metformin reduces thyroid cancer risk in Taiwanese patients with type 2 diabetes. PloS one. 2014;9(10):e109852.

Tworoger SS, Eliassen AH, Kelesidis T *et al*. Plasma adiponectin concentrations and risk of incident breast cancer. The Journal of Clinical Endocrinology & Metabolism. 2007;92(4):1510-6.

Tyrrell J, Mulugeta A, Wood AR *et al*. Using genetics to understand the causal influence of higher BMI on depression. International Journal of Epidemiology. 2019;48(3):834-48.

Tyrrell JS, Yaghootkar H, Freathy RM *et al*. Parental diabetes and birthweight in 236 030 individuals in the UK Biobank Study. International Journal of Epidemiology. 2013;42(6):1714-1723.

Uddin S, Bavi P, Siraj AK *et al*. Leptin-R and its association with PI3K/AKT signaling pathway in papillary thyroid carcinoma. Endocrine-Related Cancer. 2010;17(1):191-202.

Valenciaga A, Saji M, Yu L *et al*. Transcriptional targeting of oncogene addiction in medullary thyroid cancer. JCI insight. 2018;3(16).

Vardarli I, Weber M, Weidemann F *et al*. Diagnostic accuracy of routine calcitonin measurement for detection of medullary thyroid carcinoma in the management of patients with nodular thyroid disease: a meta-analysis. Endocrine Connections. 2021;1(aop).

Vella V, Malaguarnera R. The emerging role of insulin receptor isoforms in thyroid cancer: clinical implications and new perspectives. International Journal of Molecular Sciences. 2018;19(12):3814.

Verbeek HH, de Groot JW, Sluiter WJ *et al*. Calcitonin testing for detection of medullary thyroid cancer in people with thyroid nodules. Cochrane Database of Systematic Reviews. 2020;3(3):CD010159.

Vigneri P, Frasca F, Sciacca L *et al*. Diabetes and cancer. Endocrine-Related Cancer. 2009;16(4):1103-23.

Vigneri R, Malandrino P, Russo M. Is thyroid cancer increasing in incidence and aggressiveness? The Journal of Clinical Endocrinology & Metabolism. 2020;105(7):dgaa223.

Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nature Medicine. 2004;10(8):789-99.

Vriens MR, Suh I, Moses W *et al*. Clinical features and genetic predisposition to hereditary nonmedullary thyroid cancer. Thyroid. 2009;19(12):1343-9.

Vuong HG, Odate T, Ngo HT *et al*. Clinical significance of RET and RAS mutations in sporadic medullary thyroid cancer-a meta-analysis. Endocrine-Related Cancer. 2018;25(6):633-41.

Wang N, Kjellin H, Sofiadis A *et al*. Genetic and epigenetic background and protein expression profiles in relation to telomerase activation in medullary thyroid carcinoma. Oncotarget. 2016;7(16):21332.

Watson JD, Crick FH. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature. 1953;171(4356):737-8.

Weickert MO (2018). 'Developmental anatomy of the thyroid and parathyroid glands' in Watkinson JC & Clarke RW (ed.) Scott-Brown's Otolaryngology Head & Neck Surgery. Boca Raton: CRC Press, pp. 619-627

Weller A, Sharif B, Qarib MH *et al*. British Thyroid Association 2014 classification ultrasound scoring of thyroid nodules in predicting malignancy:

Diagnostic performance and inter-observer agreement. Ultrasound. 2020;28(1):4-13.

Wells Jr SA, Asa SL, Dralle H *et al*. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma: the American Thyroid Association Guidelines Task Force on medullary thyroid carcinoma. Thyroid. 2015;25(6):567-610.

Wells SA, Robinson BG, Gagel RF, et al. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. Journal of Clinical Oncology. 2012;30(2):134.

Welsh S, Peakman T, Sheard S *et al*. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. BMC genomics 2017;18(1):26.

Wémeau JL, Caron P, Schvartz C *et al*. and French Thyroid Research Group, 2002. Effects of thyroid-stimulating hormone suppression with levothyroxine in reducing the volume of solitary thyroid nodules and improving extranodular nonpalpable changes: a randomized, double-blind, placebo-controlled trial by the French Thyroid Research Group. The Journal of Clinical Endocrinology & Metabolism. 2002;87(11):4928-4934.

Wiersinga WM. Smoking and thyroid. Clinical Endocrinology. 2013;79:145-51.

Wiltshire JJ, Drake TM, Uttley L *et al*. Systematic review of trends in the incidence rates of thyroid cancer. Thyroid. 2016;26(11):1541-52.

Winkler TW, Justice AE, Graff M *et al*. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. PLoS Genetics. 2015;11(10):e1005378.

Wirth LJ, Sherman E, Robinson B *et al*. Efficacy of selpercatinib in RET-altered thyroid cancers. New England Journal of Medicine. 2020;383(9):825-35.

Witczak J, Taylor P, Chai J *et al*. Predicting malignancy in thyroid nodules: feasibility of a predictive model integrating clinical, biochemical, and ultrasound characteristics. Thyroid Research. 2016;9(1):4.

Wolin KY, Carson K, Colditz GA. Obesity and cancer. Oncologist 2010;15:556–65.

Wu L, Zhu J. Linear reduction in thyroid cancer risk by oral contraceptive use: a dose–response meta-analysis of prospective cohort studies. Human Reproduction. 2015;30(9):2234-40.

Wu X, Lun Y, Jiang H *et al.* Coexistence of thyroglobulin antibodies and thyroid peroxidase antibodies correlates with elevated thyroid-stimulating hormone level and advanced tumor stage of papillary thyroid cancer. Endocrine. 2014;46(3):554-60.

Xhaard C, Lence-Anta JJ, Ren Y *et al*. Recreational physical activity and differentiated thyroid cancer risk: a pooled analysis of two case-control studies. European Thyroid Journal. 2016;5(2):132-8.

Yaghootkar H, Lotta LA, Tyrrell J *et al*. Genetic evidence for a link between favorable adiposity and lower risk of type 2 diabetes, hypertension, and heart disease. Diabetes 2016;65(8):2448-2460.

Yan KL, Li S, Tseng CH *et al*. Rising Incidence and Incidence-Based Mortality of Thyroid Cancer in California, 2000-2017. The Journal of Clinical Endocrinology & Metabolism. 2020;105(6):dgaa121.

Yang SP, Ngeow J. Familial non-medullary thyroid cancer: unraveling the genetic maze. Endocrine-Related Cancer. 2016;23:R577-95.

Yarmolinsky J, Wade KH, Richmond RC *et al*. Causal inference in cancer epidemiology: what is the role of Mendelian randomization? Cancer Epidemiology and Prevention Biomarkers. 2018;27(9):995-1010.

Ye S, Zhao XY, Hu XG *et al*. TP53 and RET may serve as biomarkers of prognostic evaluation and targeted therapy in hepatocellular carcinoma. Oncology Reports. 2017;37(4):2215-26.

Ye ZQ, Gu DN, Hu HY *et al*. Hashimoto's thyroiditis, microcalcification and raised thyrotropin levels within normal range are associated with thyroid cancer. World Journal of Surgical Oncology. 2013;11(1):56.

Yeo Y, Ma SH, Hwang Y *et al*. Diabetes mellitus and risk of thyroid cancer: a meta-analysis. PloS one. 2014;9(6):e98135.

Yeo Y, Shin DW, Han KD *et al*. Smoking, alcohol consumption, and the risk of thyroid cancer: a population-based Korean cohort study of 10 million people. Thyroid. 2020(ja).

Yengo L, Sidorenko J, Kemper KE *et al*. GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in~ 700000 individuals of European ancestry. Human Molecular Genetics. 2018;27(20):3641-3649.

Yi X, Zhu J, Zhu X *et al*. Breastfeeding and thyroid cancer risk in women: A dose-response meta-analysis of epidemiological studies. Clinical Nutrition. 2016;35(5):1039-46.

You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin?. Cancer Cell. 2012;22(1):9-20.

Yu N, Donnan PT, Murphy MJ *et al*. Epidemiology of primary hyperparathyroidism in Tayside, Scotland, UK. Clinical Endocrinology. 2009;71(4):485-93.

Yu Q, Lv X, Liu K *et al*. Fertility Drugs Associated with Thyroid Cancer Risk: A Systematic Review and Meta-Analysis. BioMed Research International. 2018;2018:7191704.

Yuan S, Kar S, Vithayathil M *et al*. Causal associations of thyroid function and dysfunction with overall, breast and thyroid cancer: A two-sample Mendelian randomization study. International Journal of Cancer. 2020a;147(7):1895-903.

Yuan S, Kar S, Carter P *et al*. Is type 2 diabetes causally associated with cancer risk? Evidence from a two-sample Mendelian randomization study. Diabetes. 2020b;69(7):1588-96.

Zafón C, Obiols G, Mesa J. Preoperative TSH level and risk of thyroid cancer in patients with nodular thyroid disease: nodule size contribution. Endocrinología Y Nutrición. 2015;62(1):24-8.

Zalzali M, Debreuve A, Richard C *et al*. Micropapillary carcinoma: Description and rise in incidence in the French Marne-Ardennes thyroid cancer registry. Annales d'endocrinologie. 2019;80:(4)229-233

Zayed AA, Amarin JZ, Al-Ani AT *et al*. Association of parental consanguinity with papillary thyroid carcinoma: a case–control study. The Journal of Clinical Endocrinology & Metabolism. 2021;106(7):e2769-e2774.

Zhang X, Theodoratou E, Li X *et al*. Genetically predicted physical activity levels are associated with lower colorectal cancer risk: a Mendelian randomisation study. British Journal of Cancer. 2021;124(7):1330-8.

Zhao Q, Wang J, Hemani G *et al*. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. Annals of Statistics. 2020:48:1742–1769.

Zheng J, Li C, Lu W *et al*. Quantitative asasessment of preoperative serum thyrotropin level and thyroid cancer. Oncotarget. 2016;7(23):34918.

Zhou W, Brumpton B, Kabil O *et al*. GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. Nature communications. 2020;11(1):1-3.

Zou W, Restifo NP. TH17 cells in tumour immunity and immunotherapy. Nature Reviews Immunology. 2010;10(4):248-56.

Zhu J, Zhu X, Tu C *et al*. Parity and thyroid cancer risk: a meta-analysis of epidemiological studies. Cancer Medicine. 2016;5(4):739-52.

Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. The Lancet Diabetes & Endocrinology. 2015;3(4):286-95.

Zimmermann MB, Galetti V. Iodine intake as a risk factor for thyroid cancer: a comprehensive review of animal and human studies. Thyroid Research. 2015;8(1):1-21.

Zoeller RT, Fletcher DL, Simonyi A *et al*. Chronic ethanol treatment reduces the responsiveness of the hypothalamic-pituitary-thyroid axis to central stimulation. Alcoholism: Clinical and Experimental Research. 1996;20(5):954-60.

APPENDICES AND SUPPLEMENTARY DATA

Chapter 3 Appendix 1: Search Log

#	Searches	Results	Type
1	Medullary carcinoma.sh	2473	Advanced
2	Medullary carcinoma.af	13184	Advanced
3	Sporadic medullary thyroid cancer.ti	51	Advanced
4	Sporadic medullary thyroid carcinoma.ti	224	Advanced
5	1 or 2 or 3 or 4	13315	Advanced
6	Epigenetics.sh	63624	Advanced
7	Molecular genetics.sh	147654	Advanced
8	RET.ti	5511	Advanced
9	RAS.ti	35313	Advanced
10	microRNAs.sh	64738	Advanced
11	miR.ti	54590	Advanced
12	6 or 7 or 8 or 9 or 10 or 11	352177	Advanced
13	5 and 12	1036	Advanced
14	Limit 13 to English language	966	Advanced
15	Limit 14 to human	879	Advanced
16	Limit 15 to yr= '1994-current'	868	Advanced
17	Limit 16 to humans	868	Advanced
18	Remove duplicates from 17	788	Advanced
19	From 17 keep 3, 8, 11, 13, 19-20, 63-64…	34	Advanced
20	From 18 keep 3, 8, 10-11, 13, 19-20, 30…	40	Advanced

Chapter 3 Appendix 2: Modified Newcastle Ottawa scale for assessing risk of bias

1. Patient selection	**Score**

Was case definition adequate?	
Was selection consecutive?	
Were community controls used?	
Were controls defined?	
Were controls comparable to cases?	
2. Data collection	
Was outcome assessed by record linkage?	
Was follow-up long enough for events to occur?	
Were all subjects accounted for in follow-up?	
Total	

Chapter 6 Supplementary data: physical activity variants remaining following variant selection filter for pleiotropy.
SNP=single nucleotide polymorphism; CHR=chromosome; BP=base position; SE=standard error; BNTD=benign nodular thyroid disease; DTC=differentiated thyroid carcinoma

SNP	CHR	BP	ALLELE1	ALLELE0	A1FREQ	BETA	SE	P_BOLT_LMM	Passing variant selection filter for BNTD	Passing variant selection filter for DTC
rs10915663	1	4966393	G	C	0.977745	0.0638888	0.0151049	0.000022	1	1
rs4908685	1	7779957	C	A	0.958445	0.0460346	0.0111019	0.000028	1	1
rs301816	1	8505058	G	A	0.417243	0.019773	0.00448906	0.000011	1	1
rs11810507	1	9063238	G	A	0.847274	0.0300417	0.00619556	0.0000013	1	1
rs820623	1	14661294	C	T	0.715015	0.0205663	0.00488265	0.000025	1	1
rs12042960	1	19125784	G	A	0.501949	-0.0186967	0.00441268	0.000023	1	1
rs140569782	1	24095576	T	C	0.989081	0.0916668	0.0216765	0.000022	1	1
rs72650433	1	24524240	G	A	0.989283	-0.0879122	0.0221457	0.000069	1	1
rs141010981	1	42877403	G	GGCACAATCTCC	0.833431	0.0294863	0.00595982	0.00000066	1	1
rs147647928	1	48137363	T	A	0.986983	-0.0870449	0.0206503	0.000025	1	1
rs11579616	1	65304470	A	C	0.93416	0.0410498	0.00890157	0.0000045	1	1
rs111205056	1	84358066	T	C	0.962229	0.0511458	0.0129096	0.000078	1	1
1:91153766_TA_T	1	91153766	TA	T	0.617395	0.0193457	0.00455326	0.00002	1	1
rs35102250	1	93889568	G	C	0.436151	-0.0178542	0.00445745	0.00006	1	1
1:111099924_CTT_C	1	111099924	CTT	C	0.79229	0.0270538	0.00548909	0.0000007	1	1
rs12033257	1	112318484	A	G	0.620146	-0.0181004	0.00457823	0.000075	1	1
rs112998388	1	114126365	T	TAGAAAGTTTGC	0.238868	0.0225182	0.00520915	0.000015	1	1
rs546142907	1	115609300	A	AT	0.701034	-0.0195688	0.00484492	0.000064	1	1
rs4480415	1	174800121	G	A	0.725427	0.0249095	0.00511315	0.0000011	1	1

rs757711024	1	187975472	TA	T	0.412507	0.0204107	0.00450515	0.0000077	1	1
rs115009143	1	188827543	G	A	0.958515	-0.0480232	0.0120639	0.000063	1	1
rs181247232	1	190209301	T	C	0.98989	-0.0962173	0.0228018	0.000026	1	1
rs181338442	1	217130734	G	C	0.989941	-0.0903116	0.022525	0.000061	1	1
rs146566355	1	230531657	C	CCT	0.710473	0.0220757	0.00503767	0.000012	1	1
rs35661503	1	232512152	A	G	0.770587	-0.0217411	0.00522685	0.000031	1	1
rs1345496	1	237457970	C	T	0.500057	-0.0172589	0.00439095	0.000083	1	1
rs377203748	2	370872	AGT	A	0.134494	0.0306101	0.00699476	0.000012	1	1
rs28969622	2	12856781	G	C	0.850096	0.0263011	0.00625177	0.000024	1	1
rs116747784	2	24511614	T	C	0.98257	0.0727825	0.0170144	0.000019	1	1
rs565724031	2	32762422	G	T	0.984679	-0.0721986	0.0180033	0.000058	1	1
rs3732075	2	36593060	G	A	0.752602	0.0221847	0.00511501	0.000016	1	1
rs4646430	2	38306415	C	G	0.72367	-0.0196371	0.00491979	0.00006	1	1
rs116804952	2	39274832	C	A	0.966251	-0.0496303	0.0121978	0.000045	1	1
2:40346474_TA_T	2	40346474	TA	T	0.0546078	-0.0445323	0.0108057	0.000037	1	1
rs4952440	2	41385771	A	C	0.632987	-0.0193744	0.00457931	0.000026	1	1
rs138359441	2	45022763	G	A	0.958754	-0.0461258	0.0112408	0.00004	1	1
rs35006042	2	54941327	T	C	0.918932	-0.033287	0.00806645	0.000039	1	1
rs138065104	2	57837017	C	G	0.913024	-0.0332077	0.00780122	0.000021	1	1
rs7609303	2	58101385	A	T	0.782118	0.0246182	0.0053391	0.0000041	1	1
rs12466227	2	60148940	A	G	0.450789	0.0182405	0.00442519	0.000037	1	1
rs374429231	2	66285281	T	TTG	0.286585	-0.0212558	0.00498244	0.00002	1	1
rs112332086	2	67453726	T	A	0.948241	0.0396985	0.00993953	0.000062	1	1
rs55839368	2	117299962	G	A	0.98253	-0.0823152	0.0173424	0.0000019	1	1

rs116161166	2	119019012	C	G	0.987964	-0.0899879	0.0215614	0.000029	1	1
rs2565762	2	125143225	G	T	0.584235	0.0194292	0.00452037	0.000018	1	1
rs72831593	2	127399523	A	G	0.98726	-0.0838931	0.0199447	0.000028	1	1
rs145906554	2	145258645	G	C	0.988226	-0.0889781	0.0221176	0.000052	1	1
rs35845966	2	152205728	G	A	0.665126	0.0186781	0.00466659	0.000066	1	1
rs1220114	2	158543700	T	A	0.743491	0.0244747	0.00515174	0.0000021	1	1
rs7557793	2	171665325	T	C	0.779149	-0.0220616	0.00532949	0.000035	1	1
rs259849	2	180732693	C	T	0.606561	-0.0203145	0.00454524	0.0000071	1	1
rs34840426	2	201084881	C	CT	0.677812	-0.018926	0.00474343	0.000062	1	1
2:212352422_TA_T	2	212352422	TA	T	0.932325	0.0386804	0.00902619	0.000016	1	1
rs6723971	2	217390389	G	A	0.483173	0.018844	0.0044107	0.000022	1	1
rs1454556	2	221156441	T	C	0.625315	-0.0221805	0.00453921	0.00000089	1	1
rs143304851	2	225272556	C	CG	0.856451	-0.0265154	0.00638909	0.000029	1	1
rs56122980	2	228077989	G	A	0.611203	-0.0186599	0.00452485	0.000036	1	1
rs114526745	2	228635917	A	T	0.915553	0.0388588	0.00888734	0.000013	1	1
2:230870821_GT_G	2	230870821	GT	G	0.731422	0.0227132	0.00526908	0.000017	1	1
rs116009547	3	1754060	A	G	0.979952	-0.0672419	0.0159788	0.000026	1	1
rs42445	3	10455784	T	G	0.775204	-0.0221638	0.00531449	0.000033	1	1
rs9840917	3	12934141	A	G	0.680696	0.0216345	0.00480026	0.0000061	1	1
rs6775319	3	18758501	A	T	0.269684	0.023702	0.00495687	0.0000018	1	1
rs4283589	3	23206359	G	C	0.440213	0.0190501	0.00443732	0.000016	1	1
rs376297449	3	37020686	C	T	0.517956	0.0213702	0.00501149	0.000023	1	1
rs76987077	3	78547955	T	C	0.909166	-0.0335044	0.00777688	0.000015	1	1
rs115803446	3	82062169	A	G	0.949904	0.048143	0.0101963	0.0000023	1	1

rs683403	3	101584469	C	A	0.714813	-0.0196781	0.00489928	0.000056	1	1
rs7620022	3	105206823	C	A	0.507197	0.0181953	0.00440608	0.000032	1	1
rs77743004	3	107039602	C	T	0.943658	0.0390236	0.00974712	0.000071	1	1
rs11719262	3	119889300	G	A	0.618512	-0.0194544	0.00453985	0.000017	1	1
rs17702257	3	126869370	C	T	0.970747	-0.0527306	0.0130324	0.000054	1	1
rs28575712	3	155221319	C	T	0.757548	-0.0207577	0.00513306	0.000048	1	1
3:162527549_C_A	3	162527549	C	A	0.973288	0.0821334	0.0206028	0.000065	1	1
rs141987884	3	164910360	A	G	0.980778	-0.0699752	0.0163694	0.00002	1	1
rs182379099	3	167218768	T	C	0.987616	0.0820403	0.0204441	0.00006	1	1
3:171861250_CAA_C	3	171861250	CAA	C	0.987441	-0.0972466	0.022735	0.000018	1	1
rs9858817	3	175025004	A	C	0.83229	-0.0235696	0.00592398	0.000059	1	1
rs7626095	3	176320312	C	T	0.563545	-0.0216255	0.00450597	0.0000018	1	1
rs113796964	3	188036736	C	T	0.962323	-0.0485194	0.011744	0.000032	1	1
rs7658462	4	3283422	C	T	0.828242	0.026782	0.00585177	0.000005	1	1
rs61147010	4	4910994	C	T	0.927844	0.0344026	0.00850982	0.00006	1	1
rs869124306	4	7831842	AAC	A	0.93847	0.0492209	0.0117168	0.00003	1	1
rs115835824	4	10502576	A	G	0.968464	-0.050026	0.0126599	0.000079	1	1
rs140715375	4	18538579	T	TA	0.986742	-0.0780153	0.0193148	0.000051	1	1
4:19566525_AAG_A	4	19566525	AAG	A	0.894175	-0.0347321	0.00774514	0.0000079	1	1
rs35662554	4	28904877	G	A	0.761954	-0.0209945	0.00516019	0.000049	1	1
rs562009500	4	37510625	T	C	0.984311	0.0875154	0.0214324	0.000043	1	1
rs1139089	4	38946486	A	G	0.893871	0.0303541	0.00719812	0.000023	1	1
rs13128197	4	40219917	A	G	0.891702	0.0298226	0.00716831	0.000026	1	1
rs201118475	4	49107061	A	C	0.248267	-0.0246759	0.00527203	0.0000033	1	1

rs72642890	4	63482431	G	A	0.805026	0.0261779	0.00565174	0.0000034	1	1
rs7672835	4	79520315	T	C	0.354608	0.0200641	0.00462167	0.000016	1	1
4:83720306_CA_C	4	83720306	CA	C	0.248659	-0.0239324	0.00525005	0.000005	1	1
4:102670257_AAATAT_A	4	102670257	AAATAT	A	0.970072	0.0534817	0.0133998	0.000059	1	1
rs17605555	4	113479971	G	A	0.908195	-0.0308824	0.00759128	0.000048	1	1
rs138061731	4	121713536	A	G	0.983423	-0.0770991	0.0174449	0.0000097	1	1
rs13113505	4	132163461	C	A	0.412045	0.0184772	0.00447229	0.000039	1	1
rs6824262	4	139876829	G	A	0.853999	0.0248708	0.00624805	0.000065	1	1
rs3733385	4	140951417	T	C	0.641575	-0.0188618	0.00461425	0.000041	1	1
rs28445273	4	141924100	T	C	0.270895	0.0202657	0.00496864	0.000043	1	1
rs33962189	4	157266122	A	C	0.631815	0.0181811	0.00457397	0.000078	1	1
rs115354001	4	159474319	C	T	0.982208	-0.0731267	0.018231	0.000056	1	1
rs10857376	4	166290273	A	T	0.799418	-0.0261001	0.00587173	0.0000077	1	1
rs12649204	4	182064241	G	T	0.947985	-0.0426484	0.00994799	0.000017	1	1
rs7667358	4	189187088	A	G	0.365169	0.0186041	0.00462318	0.000062	1	1
rs11132571	4	189622488	A	T	0.0333579	0.0496964	0.0123052	0.00005	1	1
rs7701730	5	2909826	A	G	0.594865	0.019195	0.00451931	0.00002	1	1
rs115946887	5	6630674	C	T	0.986966	0.0788393	0.0198054	0.000067	1	1
rs3094309	5	8800261	C	T	0.592481	0.0193759	0.00448416	0.000015	1	1
rs10472492	5	23057391	C	A	0.776183	0.0218263	0.0052918	0.00004	1	1
rs6888342	5	53704572	C	T	0.712721	-0.0198063	0.00490882	0.000049	1	1
rs62356593	5	56287108	G	A	0.97604	-0.0581115	0.0145146	0.000075	1	1
5:94560760_CCA_C	5	94560760	CCA	C	0.972789	-0.0557827	0.0138359	0.000059	1	1
rs6895462	5	101258083	C	T	0.385645	-0.0197671	0.00454444	0.000014	1	1

rs677797	5	106896484	G	A	0.624533	0.0185739	0.00458901	0.000051	1	1
rs17135735	5	112917803	C	T	0.929079	0.0426489	0.00859028	0.00000062	1	1
rs144728538	5	122017366	C	T	0.972215	-0.0541707	0.01384	0.000091	1	1
rs34989027	5	133839271	C	T	0.821561	-0.025195	0.005795	0.000012	1	1
rs13171519	5	137859914	C	T	0.722266	0.0197773	0.00493585	0.000072	1	1
rs116345623	5	143439060	T	C	0.986128	0.0809125	0.0191361	0.000026	1	1
rs776788064	5	148093108	TATTTA	T	0.741208	-0.0204241	0.00505525	0.000058	1	1
rs545614337	5	155656275	T	C	0.986997	0.0956045	0.0221993	0.000017	1	1
rs75312759	5	155660323	G	A	0.985866	-0.0743706	0.0187267	0.000072	1	1
rs76083734	5	163046955	G	A	0.689165	-0.0210087	0.00532216	0.000077	1	1
rs17065867	5	165205386	T	C	0.961645	0.0512671	0.0114928	0.0000081	1	1
rs560101167	5	167029381	G	GTTCTGA	0.988942	-0.0984978	0.0227186	0.000014	1	1
rs62376904	5	173279464	G	A	0.970693	0.0573827	0.01308	0.00001	1	1
6:14698151_TTTTTA_T	6	14698151	TTTTTA	T	0.131011	0.0265544	0.00655426	0.000052	1	1
6:22215714_TAAAC_T	6	22215714	TAAAC	T	0.951426	0.0421793	0.0105946	0.000065	1	1
rs199004	6	23466670	A	G	0.865681	0.0266897	0.00649111	0.00004	1	1
rs172326	6	28849790	G	A	0.608827	-0.0194322	0.00450824	0.000016	1	1
rs71558740	6	41365626	G	A	0.846952	-0.0249124	0.00615498	0.00005	1	1
rs4260753	6	51399953	C	A	0.180543	-0.024385	0.00573967	0.000023	1	1
rs546264369	6	52513418	G	A	0.93321	0.0431343	0.00936887	0.0000037	1	1
rs4715623	6	56279091	A	T	0.571472	0.0194194	0.00445737	0.000013	1	1
rs79557053	6	67500388	C	T	0.765104	0.0232069	0.00519879	0.000008	1	1
rs13196061	6	94020239	C	T	0.859291	-0.0273918	0.0064525	0.000024	1	1
rs12205221	6	98790100	C	G	0.622304	-0.0186772	0.00454051	0.000039	1	1
rs190225848	6	99700038	A	G	0.897003	0.0299753	0.00730235	0.000043	1	1

rs144277946	6	120120386	C	T	0.976964	0.0683096	0.0149284	0.0000057	1	1
rs180947147	6	123176726	G	T	0.988391	0.100958	0.021037	0.0000015	1	1
rs17052852	6	125523197	G	C	0.967514	0.0579245	0.0133551	0.000016	1	1
rs77732289	6	150710681	T	G	0.987592	-0.114603	0.0290231	0.000067	1	1
rs369914136	6	152459647	AAAAT	A	0.789771	0.0230366	0.00550404	0.000028	1	1
rs76107962	6	153545700	T	G	0.980321	0.0686728	0.0164978	0.000031	1	1
rs79575917	6	162977857	C	T	0.895658	-0.0281494	0.00705312	0.000071	1	1
rs75713515	7	3247039	G	A	0.958047	-0.0450337	0.0111751	0.000052	1	1
rs10499509	7	18022771	C	A	0.320611	-0.0188088	0.00473299	0.000066	1	1
rs369173787	7	24127656	T	C	0.727861	-0.0230196	0.0058219	0.00008	1	1
rs112710860	7	35382860	C	T	0.939031	0.0374209	0.00927123	0.000057	1	1
rs10228416	7	39303261	C	T	0.456421	0.0194243	0.004443	0.000011	1	1
rs17860070	7	44839989	A	G	0.924414	-0.037972	0.00833624	0.000005	1	1
rs36045044	7	64014119	A	ACGCC	0.922458	-0.0364172	0.00824455	0.000011	1	1
rs7792468	7	78532586	C	T	0.523284	-0.0185686	0.00448472	0.00003	1	1
rs4728495	7	81700177	T	A	0.188813	0.0231115	0.00564135	0.000041	1	1
7:99349757_GATTAGT_G	7	99349757	GATTAGT	G	0.896537	-0.0307898	0.00725709	0.000021	1	1
rs35333197	7	114763404	C	T	0.985764	0.0760398	0.0186871	0.000043	1	1
7:129740849_AAC_A	7	129740849	AAC	A	0.899538	0.0361514	0.00896329	0.000062	1	1
rs118121567	8	4579010	G	T	0.976269	0.0608976	0.0145307	0.000031	1	1
rs767243853	8	4856709	CAA	C	0.431862	-0.022138	0.00463286	0.0000019	1	1
rs139194318	8	18653317	T	G	0.986456	-0.0797295	0.0195538	0.00004	1	1
rs2442982	8	20590386	G	A	0.731999	-0.0202797	0.0050573	0.000067	1	1
rs72641016	8	47336387	G	A	0.978413	0.0645824	0.0151408	0.000022	1	1
rs75097833	8	59287948	T	C	0.844431	0.0267366	0.00659864	0.000052	1	1

rs527707170	8	78255712	C	T	0.716798	0.0220238	0.00530988	0.000039	1	1
rs77824520	8	84572617	G	C	0.971887	-0.0560656	0.0133373	0.000024	1	1
rs72667900	8	92781415	A	T	0.945324	0.0408282	0.00978653	0.000026	1	1
rs142561440	8	96574531	C	T	0.98709	0.0961899	0.0197201	0.00000097	1	1
rs117893126	8	116447211	A	C	0.963152	-0.0488481	0.0121221	0.000055	1	1
rs59590279	8	124071909	C	CT	0.802263	0.0242018	0.00564021	0.000019	1	1
rs62522796	8	126818335	G	A	0.94363	-0.0389078	0.00968663	0.000071	1	1
rs77175025	8	130192653	A	G	0.989879	-0.101826	0.0232934	0.00001	1	1
8:132396389_ATT_A	8	132396389	ATT	A	0.987802	-0.0864849	0.0209983	0.000036	1	1
rs34093077	8	133292100	C	CGT	0.552601	0.0206871	0.00474889	0.000014	1	1
rs73715570	8	145025298	C	T	0.771096	-0.0222545	0.00528865	0.000022	1	1
rs10511609	9	15335996	A	G	0.95934	-0.045697	0.0114413	0.000066	1	1
rs77707476	9	22582974	G	A	0.929249	-0.0359892	0.00858308	0.000027	1	1
rs10970198	9	31193666	C	A	0.266192	-0.0220998	0.00497519	0.0000089	1	1
9:31913462_TA_T	9	31913462	TA	T	0.807667	-0.0236845	0.00560564	0.000024	1	1
rs56059667	9	38708301	C	T	0.919034	-0.0331022	0.00806757	0.000039	1	1
rs188092309	9	73511705	C	A	0.976568	-0.0589867	0.0146746	0.000056	1	1
rs144813308	9	77372112	C	CA	0.914528	0.0344596	0.00788558	0.000011	1	1
rs71361141	9	85353560	A	AG	0.525305	0.0190014	0.00443975	0.000018	1	1
rs773281	9	93868693	G	A	0.476745	0.0198536	0.00441995	0.0000074	1	1
rs13284114	9	96129132	A	G	0.441519	0.0178495	0.00443532	0.000061	1	1
rs142198996	9	105080221	C	A	0.954902	0.0447179	0.0108337	0.000038	1	1
rs34064803	9	130165973	T	C	0.898046	-0.029886	0.00739983	0.000051	1	1
rs9330466	9	136414831	A	G	0.077139	0.0327465	0.0082515	0.000068	1	1

rs4528211	10	6839290	T	C	0.0364733	-0.0528366	0.0122019	0.000016	1	1
rs11258233	10	13222207	C	T	0.770009	-0.0231395	0.00524568	0.0000094	1	1
rs78496751	10	18868982	G	T	0.983978	-0.0721523	0.0175664	0.000039	1	1
rs35654349	10	20140666	C	A	0.748454	0.0223662	0.00510093	0.000015	1	1
rs181085835	10	46191657	C	A	0.986941	0.102052	0.0221691	0.000004	1	1
rs150736114	10	56652107	G	A	0.968556	0.0528866	0.0131915	0.000055	1	1
rs118015950	10	74628844	A	G	0.956271	-0.0432235	0.0107601	0.000054	1	1
rs61864473	10	75561972	C	T	0.976338	-0.0650075	0.0144585	0.0000061	1	1
rs74142329	10	80949988	C	T	0.968522	0.0546345	0.0126857	0.000016	1	1
rs1274179	10	91772949	T	C	0.428459	-0.0186061	0.00470946	0.000083	1	1
rs147281191	10	95778542	G	A	0.989395	0.096607	0.0224699	0.000018	1	1
rs10882874	10	98968783	T	A	0.421162	-0.0183571	0.00447192	0.000037	1	1
10:99794208_CTT_C	10	99794208	CTT	C	0.806866	0.0243799	0.00557774	0.000012	1	1
rs7898420	10	105054192	T	C	0.610904	-0.0187864	0.00459399	0.00004	1	1
rs10884028	10	106396917	C	T	0.293562	0.0191375	0.00483151	0.000066	1	1
rs58315919	10	115363664	C	T	0.766631	0.0207587	0.00521655	0.000064	1	1
rs578061988	10	123153793	A	G	0.973545	0.0603248	0.0143783	0.000027	1	1
rs17618337	10	125679970	T	A	0.632898	-0.0200654	0.00467596	0.000016	1	1
rs2289432	10	126687145	T	G	0.716922	0.0203328	0.00492618	0.00003	1	1
rs2890087	10	128845623	C	T	0.509033	0.0204515	0.00442026	0.0000039	1	1
rs3793670	10	134539711	C	A	0.747348	-0.0203706	0.00506086	0.000055	1	1
rs6578412	11	1482582	T	C	0.0570752	-0.040171	0.00956976	0.00003	1	1
rs77167127	11	12741030	C	T	0.987673	-0.078673	0.019888	0.000085	1	1
rs199983929	11	18273583	T	C	0.983336	0.0760248	0.019044	0.000063	1	1

rs2055014	11	29195732	G	A	0.705071	-0.0205954	0.004837	0.000022	1	1
rs72892570	11	31626807	A	G	0.947975	0.0493923	0.00994865	0.00000062	1	1
rs17392289	11	40464309	C	T	0.9097	-0.0308293	0.00769228	0.000061	1	1
rs185332274	11	41670838	T	C	0.985584	-0.0849042	0.0193487	0.000013	1	1
11:47646988_AT_A	11	47646988	AT	A	0.308571	-0.0196133	0.00484361	0.000051	1	1
rs777529284	11	57486502	TATA	T	0.684294	0.0223789	0.00489347	0.0000042	1	1
rs11607366	11	64575769	C	T	0.974941	-0.0575771	0.0144517	0.000067	1	1
rs74800845	11	72380207	G	A	0.942836	-0.0443224	0.00957097	0.0000039	1	1
rs80289431	11	85581567	T	G	0.968484	-0.0548666	0.0126286	0.000015	1	1
rs142641476	11	88674436	G	A	0.979544	0.0734703	0.0178749	0.000043	1	1
rs1046654	11	94231847	G	A	0.866629	0.0278663	0.00648366	0.000017	1	1
rs10831321	11	94964343	A	G	0.697681	0.0225388	0.00485132	0.0000035	1	1
rs75545062	11	99495442	A	G	0.935745	0.0358127	0.00895651	0.000061	1	1
rs17662199	11	107090726	C	A	0.852992	-0.0258991	0.00623343	0.000031	1	1
rs144680902	11	110229873	T	TTG	0.89663	0.0314429	0.00748382	0.000029	1	1
rs1061	11	112989298	T	C	0.429635	0.0178801	0.00445952	0.000071	1	1
rs17119029	11	115857374	T	C	0.986727	0.0804338	0.0191866	0.000027	1	1
rs7950728	11	126644477	T	C	0.700486	0.0209755	0.00482582	0.000013	1	1
rs142055384	12	24304590	A	G	0.986916	0.087667	0.0196711	0.0000079	1	1
rs794154	12	30737459	A	G	0.297517	0.020011	0.00482899	0.000037	1	1
rs574171779	12	34301192	A	AT	0.447112	-0.0216986	0.00460175	0.0000024	1	1
rs117677831	12	47215359	A	G	0.906863	-0.03367	0.00769939	0.000015	1	1
rs7309997	12	49756665	C	T	0.899105	-0.0319133	0.00736006	0.000015	1	1
rs370790370	12	58531891	GTT	G	0.476893	0.0186972	0.00463424	0.000052	1	1
rs78982639	12	70604120	G	A	0.988523	0.0988925	0.0208121	0.000002	1	1

rs78672908	12	92605897	G	C	0.953621	-0.0437929	0.0106719	0.000043	1	1
rs10507037	12	94769945	A	G	0.839665	0.0243601	0.00599708	0.000047	1	1
rs35058145	12	98134238	G	A	0.963389	-0.0582057	0.0117547	0.00000082	1	1
rs147175344	12	103484508	C	T	0.972968	-0.0591341	0.0139729	0.000026	1	1
rs34570416	12	117354034	G	A	0.974964	-0.05747	0.0140809	0.000045	1	1
rs150740421	12	121700776	G	GTT	0.950448	-0.0424033	0.0102652	0.000034	1	1
rs9576057	13	31985220	G	C	0.973045	-0.0563526	0.0141161	0.000068	1	1
rs9544535	13	36095039	A	G	0.425196	-0.0180718	0.00450088	0.000062	1	1
rs7324697	13	58259492	C	A	0.675626	-0.0190176	0.00470626	0.000053	1	1
rs184099776	13	64989530	G	A	0.989085	0.0931871	0.0215656	0.000015	1	1
rs9529099	13	67151771	C	G	0.705733	-0.0219888	0.00484085	0.0000042	1	1
rs12869487	13	73628682	A	G	0.850168	-0.0251653	0.00623767	0.000053	1	1
rs9561469	13	94565017	A	G	0.383927	0.0193627	0.00457086	0.000021	1	1
rs61968328	13	101607321	T	C	0.909397	0.0318251	0.00793572	0.00006	1	1
rs61969123	13	111420599	A	C	0.986772	-0.0842291	0.0209162	0.000059	1	1
rs9590511	13	115066004	T	C	0.727121	-0.0201232	0.00501532	0.00005	1	1
rs11621360	14	23911407	T	C	0.336272	0.0184421	0.00466192	0.000078	1	1
rs17490849	14	26668275	T	A	0.901873	-0.0304855	0.00759696	0.000058	1	1
14:36992180_ATT_A	14	36992180	ATT	A	0.609567	-0.0208435	0.00461058	0.0000064	1	1
14:38332163_CT_C	14	38332163	CT	C	0.906647	0.0347738	0.00850283	0.000044	1	1
rs8005455	14	51779969	C	G	0.54883	-0.017724	0.00448346	0.000087	1	1
14:55654345_CT_C	14	55654345	CT	C	0.0329592	-0.0557239	0.0139183	0.000061	1	1
rs80336833	14	70366990	G	T	0.954501	-0.0491394	0.0109768	0.0000084	1	1
rs390316	14	78596887	A	C	0.380245	-0.0195693	0.00454218	0.000017	1	1

14:93693138_CAAAAAAAAAAAAA_C	14	93693138	CAAAAAAAAAAAAA	C	0.189458	-0.0241977	0.00610829	0.000072	1	1
rs76587776	14	95946901	C	T	0.961563	-0.0503293	0.0123305	0.000049	1	1
14:98658596_AT_A	14	98658596	AT	A	0.973107	0.0655365	0.0137028	0.0000017	1	1
rs117244059	14	100157934	G	A	0.924222	-0.0345835	0.00833511	0.000033	1	1
rs12878003	14	101124721	G	A	0.244681	-0.0203691	0.00513054	0.000065	1	1
rs8013188	14	103122736	G	A	0.365438	0.0186722	0.00457136	0.000046	1	1
rs78694677	15	24320593	A	C	0.987707	0.0925923	0.0217973	0.000021	1	1
15:33966992_GA_G	15	33966992	GA	G	0.304951	-0.0220764	0.00511177	0.000018	1	1
rs12708539	15	36652459	A	C	0.60546	0.0186014	0.0045482	0.000038	1	1
rs139552414	15	55435501	C	T	0.976691	-0.0629741	0.0146101	0.000018	1	1
rs150460345	15	68239459	G	GT	0.924939	0.0364737	0.00847722	0.000017	1	1
rs743580	15	74328116	A	G	0.510812	0.0197744	0.00440159	0.0000083	1	1
rs144914807	15	78016560	C	T	0.981716	0.0687631	0.0169253	0.000046	1	1
rs1848707	15	81421792	A	G	0.373129	-0.0191475	0.00468679	0.000042	1	1
rs560187634	15	83465785	A	AC	0.288651	-0.0226212	0.0049136	0.000004	1	1
15:90614493_TA_T	15	90614493	TA	T	0.593659	0.0211419	0.00471788	0.0000081	1	1
rs75467505	15	98653511	G	T	0.985311	-0.0769617	0.019268	0.000069	1	1
16:8513001_GT_G	16	8513001	GT	G	0.970548	-0.0591844	0.0137676	0.000015	1	1
rs8050915	16	17438906	A	G	0.46235	0.0181827	0.00444868	0.000048	1	1
rs117594380	16	25439272	T	C	0.943411	0.0391778	0.00976212	0.000058	1	1
rs199851257	16	30930263	T	C	0.837605	-0.0293845	0.00677297	0.000017	1	1
rs145803819	16	32598117	G	A	0.985879	-0.118023	0.0277745	0.000024	1	1
rs28529261	16	47197865	T	C	0.986063	-0.0895599	0.0213558	0.000027	1	1
rs9938281	16	49625336	A	G	0.475023	-0.0203056	0.00444508	0.0000049	1	1

rs111555721	16	56959558	C	CTA	0.913759	0.0355994	0.00835541	0.000021	1	1
rs1366537	16	62873194	A	G	0.687203	0.0201186	0.00476963	0.000025	1	1
rs7202592	16	64550767	C	T	0.562389	0.0176115	0.00444092	0.000068	1	1
rs4473187	16	65190902	G	A	0.949849	-0.0474745	0.0100822	0.0000024	1	1
rs62049375	16	70236911	G	A	0.679997	0.0220603	0.00502358	0.0000099	1	1
16:71673349_CT_C	16	71673349	CT	C	0.5413	-0.0217149	0.00451977	0.0000014	1	1
rs8051200	16	73468175	C	T	0.799888	-0.0253371	0.00552387	0.0000041	1	1
rs825682	16	73608290	T	C	0.608101	-0.0197657	0.00452163	0.000013	1	1
rs72792397	16	78372637	G	A	0.972378	0.0643663	0.0142097	0.0000058	1	1
16:83285681_AGTATATAT_A	16	83285681	AGTATATAT	A	0.678991	0.0215775	0.00475587	0.0000053	1	1
rs67659612	16	85682471	C	T	0.553098	-0.0185378	0.00450869	0.00004	1	1
rs72816573	16	86507923	A	G	0.881719	-0.0298914	0.00685678	0.000014	1	1
rs4640189	16	86952035	C	T	0.934206	-0.0364548	0.00892406	0.000046	1	1
rs142689545	16	87144067	T	C	0.983903	0.0703639	0.0176667	0.000081	1	1
rs12149336	16	89423041	T	G	0.487571	0.0201207	0.00444713	0.0000058	1	1
rs34176755	17	1877754	C	G	0.864558	-0.0267335	0.00653813	0.00005	1	1
rs8076040	17	30590635	G	A	0.831685	-0.0253174	0.00589203	0.000017	1	1
rs77336549	17	35191248	TAA	T	0.207264	0.0236015	0.00575864	0.000038	1	1
rs2271308	17	37817482	T	C	0.268516	-0.021523	0.00496592	0.000016	1	1
rs141088836	17	39258221	G	A	0.978176	0.0667045	0.0154186	0.000019	1	1
rs113126189	17	56761824	A	C	0.809809	-0.0229861	0.00570038	0.000052	1	1
rs2683156	17	71747610	T	C	0.560313	0.0186348	0.00444362	0.000024	1	1
rs58663604	17	74893027	G	A	0.837265	0.0249596	0.00599713	0.000033	1	1
rs117969406	17	77688797	A	G	0.943357	-0.0431416	0.00982545	0.00001	1	1

rs186056728	18	221643	C	G	0.967374	-0.0594248	0.0129121	0.0000043	1	1
18:21096690_TC_T	18	21096690	TC	T	0.574757	-0.0192278	0.0047891	0.000063	1	1
rs1668835	18	22478952	T	A	0.686952	-0.0223017	0.00476266	0.000003	1	1
rs76347875	18	30146899	C	G	0.981144	0.0683218	0.0166845	0.000039	1	1
rs35875326	18	41658694	C	T	0.373427	0.0183676	0.00456398	0.000051	1	1
rs12965648	18	44819660	C	A	0.574242	-0.0178449	0.00449483	0.000073	1	1
rs916996	19	1340499	A	G	0.792011	0.0247967	0.00623934	0.000063	1	1
rs10401917	19	3315384	G	A	0.717757	-0.0212448	0.00488563	0.000012	1	1
rs4806995	19	4758436	C	A	0.644107	0.0216639	0.00462512	0.0000027	1	1
rs76589773	19	7559407	G	A	0.968181	-0.0508782	0.0125274	0.000053	1	1
19:7942830_CATATATACAT_C	19	7942830	CATATATACAT	C	0.781884	0.0230993	0.00558736	0.00004	1	1
rs1060463	19	16025176	C	T	0.382019	-0.0189086	0.00452822	0.00003	1	1
rs10330	19	19017862	G	A	0.871584	0.0280814	0.0065781	0.000018	1	1
rs73529173	19	19060490	C	T	0.842955	0.0283165	0.00704218	0.000057	1	1
rs7250389	19	30325332	G	T	0.676997	0.0194828	0.00473621	0.000041	1	1
rs10422015	19	33649995	G	T	0.85492	0.0250222	0.00627884	0.000066	1	1
rs7254947	19	38726617	G	A	0.825002	-0.024178	0.0059365	0.000046	1	1
rs157595	19	45425460	A	G	0.383928	-0.0189079	0.00462591	0.000041	1	1
rs10415062	19	46100752	C	T	0.659342	-0.0189032	0.00471937	0.000055	1	1
19:46821467_AT_A	19	46821467	AT	A	0.105549	0.0328669	0.00752483	0.000012	1	1
rs3760731	19	51375716	C	T	0.963821	-0.0549576	0.0123604	0.0000075	1	1
rs8101411	19	53314033	C	A	0.970497	0.0660738	0.0151429	0.000012	1	1
rs7272683	20	17620092	T	C	0.634038	0.0187007	0.00466679	0.000055	1	1
rs62208117	20	35538539	T	C	0.834359	0.0237644	0.0059268	0.000076	1	1
rs4810326	20	40220214	A	C	0.0938173	0.0313414	0.00774482	0.000046	1	1

rs117999688	20	45963666	T	C	0.96816	0.0585604	0.0127721	0.0000044	1	1
rs4811486	20	52561702	C	T	0.823136	0.0253411	0.00576914	0.000012	1	1
rs2829752	21	26830041	A	G	0.989604	0.103956	0.0219138	0.0000017	1	1
rs2836318	21	39716466	G	A	0.885521	0.0291059	0.00692451	0.000026	1	1
rs58651699	22	23916271	C	T	0.514996	-0.0207082	0.00478092	0.000014	1	1
rs78338316	22	29079238	C	A	0.985221	-0.113173	0.0268242	0.000025	1	1
rs9606733	22	30874324	T	C	0.978901	0.0894361	0.0223334	0.000059	1	1
rs2014137	22	32728966	A	G	0.401023	-0.0183837	0.00456374	0.000056	1	1
rs133309	22	42398606	G	A	0.0534254	0.0465166	0.00985755	0.0000021	1	1
rs2413721	22	43198836	T	G	0.541709	0.0188403	0.00441478	0.00002	1	1