Causal Effect of Body Mass Index on Thyroid Cancer Risk: A Mendelian Randomization Analysis

A Thesis Submitted to
The Department of Cancer Control and Population Health
in Partial Fulfillment of the Requirements
for the Master's Degree of Public Health

Quy Ngoc Nguyen

July 2018

GRADUATE SCHOOL OF CANCER SCIENCE AND POLICY

Professor Jeongseon Kim

We hereby approve the M.P.H. thesis of
Quy Ngoc Nguyen.

July 2018

EUN KYUNG LEE sign
Chairman of Thesis Committee

YuL Hwangbo sign
Thesis Committee Member

Jeongseon Kim sign
Thesis Committee Member

NATIONAL CANCER CENTER
GRADUATE SCHOOL OF CANCER SCIENCE AND POLICY

ABSTRACT

Causal Effect of Body Mass Index on Thyroid Cancer Risk: A Mendelian Randomization Analysis

Background: It is indicated that high body mass index (BMI) was associated with an increased risk of thyroid cancer (TC) in both men and women in observational epidemiological studies. However, establishing a causal relationship has been difficult because of the possibility of unmeasured confounding effects or reverse causality. We conducted a Mendelian randomization (MR) with the use of genetic instrument variables (IVs) to estimate the causal effect between BMI and TC risk.

Methods: We used data from 744 cases from the National Cancer Center (NCC), South Korea and 6,216 healthy controls including the Korean Genome Epidemiology Study and NCC. BMI genetic risk score, which comprises 55 BMI-associated genetic variants, was included in the MR analysis as the instrumental variable. The Wald/ratio was used to find the causal odds ratio (OR) for the effect of BMI on TC.

Results: The F-statistic from the regression of BMI on genetic risk score for BMI (IV) was 78.5, and the IV explained 1.25% of the phenotypic variance for BMI. The causal OR for a 1 kg/m^2 genetically instrumented BMI was 1.08 [95% confidence interval (CI), 0.87 to 1.35]. Our result of IV estimation suggests that BMI is not a causal risk factor of TC. When adjusting for age, genetically influenced BMI was not associated with TC (OR, 1.02; 95% CI, 0.82-1.25) for neither women nor (OR, 1.44; 95% CI, 0.75-2.75) men.

Conclusions: Genetically predicted BMI is not associated with an increased risk of TC. This result suggests that high BMI is not a causal risk factor for TC and the corresponding observational association is likely explained by reverse causation or confounding effects.

Contents

List of Tables

List of Figures

1. Introduction

1.1 Background

Over the past few decades, the incidence of thyroid cancer worldwide has increased at a higher rate than any other cancers [1]. In South Korea, there has been a rapid acceleration in the number of thyroid diagnoses, and currently, the thyroid cancer incidence rate in South Korea is the greatest in the world [2]. According to the cancer statistics in Korea, thyroid cancer was shown to be the most common cancer among women, with the crude incidence rate of 97.0 per 100,000; followed by breast, colorectal, stomach, and lung cancer. These five cancers accounted for 67.9% of cases in women [3].

According to cancer statistics from 1999 to 2015, the age-standardized rate (ASR) for thyroid cancer incidence has notably increased in both men from 2.1 per 100,000 in 1999 to 16.9 in 2015, and in women from 10.4 per 100,000 in 1999 to 55.6 in 2015. Especially, from 1999 to 2011 ASR for thyroid cancer has increased rapidly to 22.8%, but then dropped to 13.3% annually starting in 2011 (**Figure 1**) [4]. South Korea's thyroid cancer prevalence has increased approximately 15 times over the past 20 years, and ten times the average prevalence rate worldwide [5]. However, there has been no feasible explanation for the increasing prevalence of thyroid cancer. Some previous studies have suggested that the increase in thyroid cancer incidence is mainly due to the clinically advanced detection of thyroid cancer using thyroid ultrasonography [6].

While the substantial increase of incidence of thyroid cancer may solely be the result of improved detection, some risk factors (including tobacco use, alcohol drinking, and obesity) have been suggested to be associated with the risk of developing thyroid cancer from the results of some observational studies [7-9].

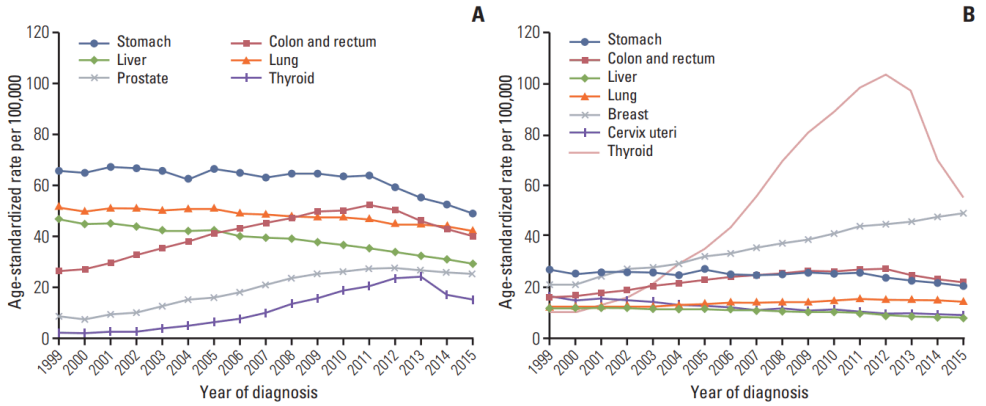

Fig 1. Trends in age-standardized incidences of selected cancers by gender in Korea [4]
(A) Men (B) Women, 1999-2015
Available from: https://www.e-crt.org/journal/view.php?number=2850

Thyroid cancer consists of two major types of neoplasia which depend on the malignant transformation progression. Neoplasia which arises from follicular cells or thyrocytes is the most common type of thyroid cancer (TC). Specifically, 85% and 10% of subtypes are papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC), respectively. The second group is medullary thyroid carcinoma (MTC), which accounts for 5% of thyroid cancer cases, affects the parafollicular or C-cells [10].

1.2 The risk factors for thyroid cancer

1.2.1 The environmental and lifestyle risk factors

Even with early detection and treatment, the incidence of thyroid cancer increased, but it is still uncertain if increased detection is indeed the sole cause. It is suggested that other factors are leading to the growth of thyroid cancer incidence. Therefore, environmental factors (e.g., radiation, iodine intake, and nitrates), as well as lifestyles (e.g., alcohol, smoking, diabetes, and obesity), and perhaps interactions between multiple factors are considered possible causes for a real increase in thyroid cancer incidence.

Radiation exposure. An increased occurrence of thyroid cancer is well-recognized to be caused by radiation exposure. The thyroid cancer progression is mainly due to two risk factors, namely, radiative dose to the thyroid gland and age of exposure. In particular, the risk is significantly increased with over 0.05-0.1 Gy of mean dose [11]. Younger age is most sensitive to radiation and diminishes with a higher age of exposure, and is lower among exposed adults [12]. The majority of these cases are papillary carcinoma (PTC), which young children perform the higher prevalence of solid subtype in the absence of any exposure with radiation. On the molecular level, the occurrence of chromosomal rearrangements is frequently found with RET-PTC, while BRAF and other genes rarely occur [13].

Iodine Intake. Both inadequate and excessive iodine intake lead to thyroid disease. In contrast, the role of iodine intake in thyroid cancer remains unclear, in spite of decades of study and debate. Epidemiological studies have also had

conflicting results, including several studies conducted in the U.S, China, and Argentina. Over the past four decades in the US, a greater dietary iodine intake was significantly associated with a reduced risk of thyroid cancer [14]. In a five-year longitudinal study of thyroid cognitive in three different areas of Han population, the region with redundant iodine consumption had more than 10 new cases of thyroid cancer, while the remaining two regions found no case [15]. When iodized salt was introduced in 1963, thyroid cancer incidence exhibited a more than two times increase in the study from northern Argentina during the period from 1960 to 2007 [16]. In Korea, PTC is the most popular endocrine malignancy, accounting for over 97% of overall thyroid cancers. Genetic variations including the mitogen-activated protein kinase (MAPK) pathway are commonly described in PTC, such as the mutations of RET/PTC, RAS, and B-type Raf kinase (*BRAF*) [17]. More than of 80% PTC cases harbor *BRAF* mutations in Korea. There is plenty of research considering the relationship between iodine consumption and the prevalence of *BRAF* mutations in PTC. A study on a large sample of PTC patients from distinct regions in China pointed out a significant correlation of the *BRAF* mutation in PTC with high iodine consumption [18]. Recently, a study in Korea showed that relatively low and more than excessive iodine intakes played as risk factors for *BRAF* mutations in PTC patients [19].

Smoking. Many studies have demonstrated that cigarette smoking is associated with thyroid cancer risk. In a meta-analysis of 31 non-interventional studies, the risk of thyroid cancer was reduced in persons who had a history of smoking (Relative Risk, RR 0.79; 95% CI 0.70-0.88) in comparison with never-

smokers [20]. Additionally, a study on Korean adults has shown a dose-response association in the ever-smokers during the years of cigarette consumption and a declined thyroid cancer progression, as compared to the non-smoker [21]. It can be explained that the potential factor influencing thyroid cancer risk is thyroid-stimulating hormone (TSH) [22]. As compared with non-smokers or past smokers, TSH, thyroid hormones Thyroxine (T_4), and Triiodothyronine (T_3) levels were much lower in present smokers [22]. Another possible reason could be an anti-estrogenic effect of smoking, which can weaken thyroid cancer development [23].

Alcohol. Similar to smoking status, past observational studies have described a correlation between alcohol drinking and risk of thyroid cancer. A retrospective study indicated that in comparison with non-drinkers, current drinkers have 54% decrease in the risk of thyroid cancer; analogously, people who drink 1 to 2 drinks per day have 42% decreased risk compared to those who drink < 1 drink per day [24]. A combined analysis of five prospective cohort studies conducted in the U.S showed that more than seven drinks of alcohol per week were inversely associated with the risk of thyroid cancer after adjustment for smoking [25]. The reason can be the reduction of nodules and goiter or the constraint of dysfunction in thyroid gland which is responsible by frequent drinking in many human and animal studies [26].

Diabetes. Diabetes is associated with increased risks of various types of cancer, such as breast, colon, liver, pancreas, and endometrium [27-31]. Hyperglycaemia and hyperinsulinaemia are the most conclusive causes of the link between diabetes and cancer, with the latter being more definitive [32, 33]. There is, however, no established association of diabetes with thyroid cancer risk. Studies in

the past raised controversy about the risk of diabetes on thyroid cancer. A meta-analysis of 14 cohorts and 3 case-control studies showed that women with preexisting diabetes had an increased risk of thyroid cancer, compared with their nondiabetic counterparts [34]. However, a diagnosis of diabetes may be associated with increased screening that led to increased detection of thyroid cancer, rather than contributing to a true increase in cancer incidence. Furthermore, a large prospective research study in the U.S has shown no significant correlation between thyroid cancer and diabetes [35]. In addition, a recent pooled study of 5 prospective studies from the U.S, including NIH-AARP study, indicated there was not enough evidence to support the association between diabetes history and thyroid cancer [36]. In addition, a previous literature review reported that the findings are uncertain - any association between diabetes and thyroid cancer, if exists, was not robust [37, 38].

Obesity. In epidemiological studies, obesity is usually determined by body mass index (BMI) as a singular measure that can be examined in crossed studies and community. According to the World Health Organization (WHO), a BMI higher than or equal to 25 kg/m^2 is overweight, while obesity is described as having a BMI equal to or greater than 30 kg/m^2. The epidemiological studies have noted that obesity is associated with higher risk of numerous cancer types, including kidney, endometrium, colon, esophagus, postmenopausal breast, gastric, and liver cancer [39]. The association between obesity and risk of thyroid cancer has been considered for a long time. Various studies have estimated the relationship between obesity and thyroid cancer. Some of them found that obesity is a risk factor for thyroid cancer, while others showed no association between obesity and risk of thyroid cancer. In

particular, the meta-analysis of 21 observational studies observed obesity was associated with a significantly developed risk of thyroid cancer (RR, 1.33) and decreased the risk of medullary thyroid cancer [40]. A combined analysis of five prospective studies found the hazard ratio for overweight and obesity compared with normal-weight were 1.20 and 1.53, respectively [41]. Seven cohort studies found the combined RR of thyroid cancer was 1.18 for overweight and obesity combined [42]. However, an extended cohort study on 200,000 people for the duration of 20 years found no significant correlation between overweight or obesity and thyroid cancer [43]. Recent retrospective research analyzing Midwest cases with thyroid cancer also showed no remarkable difference between current BMI, median BMI in 20-year-old people, or lifetime maximum BMI among disease and non-disease groups [24].

The biological mechanisms of the association between obesity, diabetes, and thyroid cancer is complex and not well explained. Overweight subjects are at a 10-time enhanced risk of diabetes [41] and they are at increased risk to develop thyroid cancer. Although changes in BMI slightly decreases thyroid cancer possibility related to diabetes, BMI alone was difficult to reveal the relationship between diabetes and thyroid cancer [37]. There were several potential links between obesity and thyroid cancer. First, some studies indicated BMI and TSH are proportional to each other [41]. Similarly, Fiore et al. have found remarkably higher TSH concentrations among individuals who were later confirmed with thyroid cancer in comparison with people having the benign disease [44]. Additionally, autonomous thyroid function (TSH<0.4 μU/ml) growth was observed to relate to a decrease in the possibility of PTC [45]. Second, insulin disorder, a general metabolic disorder in obesity, may

affect the proliferation and differentiation of follicular cells [46]. Insulin resistance may stimulate insulin and the Insulin-like growth factor (IGF) signal way, which are significant to cell growth and death. The chronically elevated circulating insulin levels may affect thyroid cancer risk through by insulin receptors overexpressed by carcinoma cells [47]. Third, adipokines such as adiponectin, leptin, and hepatocyte growth factor can control carcinoma cell reproduction and are possibly linked to improvement of the tumor [48]. Enhanced expression of leptin and its receptor were described in thyroid carcinoma [49].

Physical activity. There is a huge body of evidence showing that physical activity has beneficial effects on varied regards of health. Physical activity has been identified as a method to reduce the incidence of coronary heart disease, diabetes mellitus, stroke, obesity, and to reduce the impact of chronic diseases [50]. One of the identified risk factors for thyroid cancer was physical activity. In two case-control studies that examine the effect of physical activity on thyroid cancer, the possibility of thyroid cancer lightly decreases among subjects who engaged in physical activity (OR, 0.8), furthermore, frequency seemed to have more robust association than the duration of physical activity [51]. However, in a meta-analysis, the summarized thyroid cancer risk estimation of high- compared to low-level of doing exercise from case-control and cohort studies indicated no significant relationship between regular exercise and thyroid cancer risk [52]. Although there is no definite causal linkage between physical activity and thyroid cancer risk, an increase in physical activity has been reported with the reduction in obesity and changes in hormone level [53].

Socioeconomic status. The effects of socioeconomic status (SES) on new cases of thyroid cancer are well-researched. In the study done by the Korean National Health and Nutrition Examination Survey, the medium-highest and highest household incomes have a higher risk of thyroid cancer as compared with lowest household income (OR 8.16 and 3.30, respectively) [2]. In other studies, higher SES was positively correlated with higher thyroid cancer risk; the median and high SES showed higher thyroid cancer risk (RR 2.29 and 3.67, respectively), compared to people of low SES [54]. The potential reason may be the excessive of health examination (overdiagnosis). Healthcare utilization can be affected by socioeconomic status [55].

Other risk factors. Gender difference in cancer susceptibility is one of the common conclusive findings in cancer research. For a very long period, it has been recognized that males are more likely to have cancer, and especially hematologic malignancies [56]. The gender difference in principal cancer susceptibility can be quantified by analyzing of incidence rates in males and females. In thyroid cancer (almost all diseases of the thyroid), cancer is three times less common in men than in women. In females, the incidence rate in specific age group accelerates clearly from the puberty age and peaks at 40-49 years, while the peak is at 60-69 years in male (**Figure 2**) [57]. The reasons for the disparity between men and women remain unclear. It has been assumed that reproductive, menstrual, and environmental factors can contribute to this difference. Several studies have shown that estrogen and its receptors are crucial for proliferation, migration, and invasion of thyroid cancer [58]. It exerts its growth promoting effect through a membrane-bound estrogen receptor

(ER). ERα activation seems to induce the development of thyroid cancer, while wild-type ERβ plays a protective role against thyroid cancer [59]. However, such effects or pathways have not been successfully confirmed in human studies.

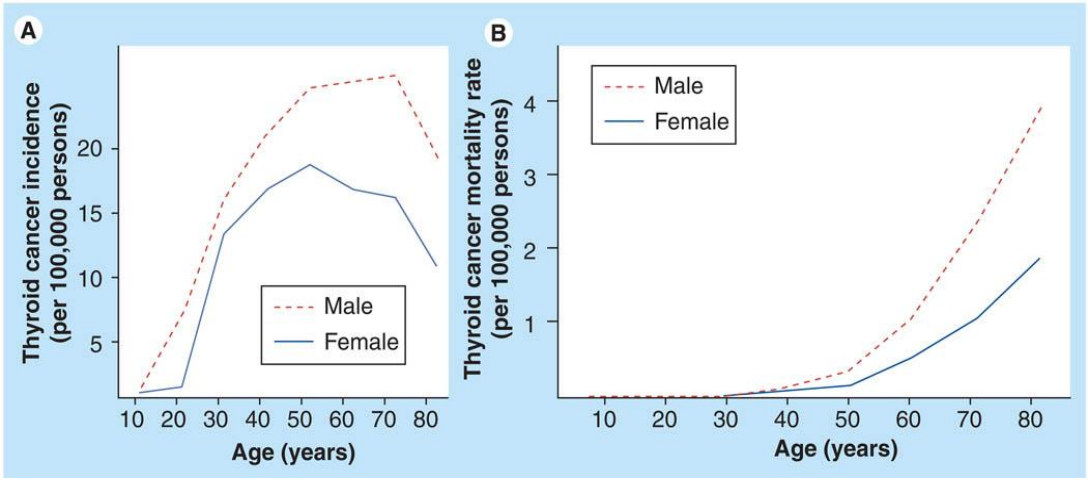

Fig 2. Thyroid cancer incidence and mortality [57]

(A) Thyroid cancer incidence by age and gender. **(B)** Thyroid cancer mortality rates by age and gender.

1.2.2 The genetic susceptibility

Various genetic factors have been associated with the progression of complex types of thyroid cancer, such as family history. Certain inherited hereditary abnormalities have been linked with the growth of different classes of thyroid cancer, and genetic impact has been indicated to be responsible for 53% of the causes of thyroid cancer [60]. Genetic factors (family history of thyroid cancer) has proven its role that having a first-degree relative (parent, sibling, or offspring) with thyroid cancer risk. The higher risk may be due to certain hereditary conditions, but heredity and family history is not known for all families. Previous studies found that a family

history of thyroid cancer in first-degree relatives was associated with increased differentiated thyroid carcinoma (DTC) risk (OR 4.1). Particularly in siblings, the risk of PTC was highest (OR 7.4) [61]. Another study between family history of thyroid cancer and the risk of PTC in French Polynesia also suggested that individuals with an affected first-degree relative had an increased risk of PTC with a 4.5-fold [62].

Recently, the advance of genomic techniques has allowed genetic studies for many types of disease. Genome-wide association studies (GWAS) has quickly developed as a widely-used tool to discover genetic factors associated with common diseases, including many types of cancers. There are several GWAS conducted in thyroid cancer in different populations. The first GWAS on DTC risk factors was found in the association of 2 variants, rs965513 on 9q.22.33 (near *FOXE1*) and rs944289 on14q13.3 (near *NKX2-1*), with DTC in two case-control groups of European descent [63]. Circulating four levels of TSH is associated SNPs, which is rs116909374 (*NKX2–1*) on 14q13.3, rs966423 (*DIRC3*) on 2q35, rs2439302 (*NRG1*) on 8p12, and rs334725 (*NFIA*) on 1p31.3 especially found in the Icelandic population; and in replication studies, including individuals from the U.S (Ohio), the Netherlands, and Spain [64]. In Asians, four SNPs rs965513 (*FOXE1*) on 9q22.33, rs944289 (*PTCSC3*) on 14q13.3, rs966423 (*DIRC3*) on 2q35 and rs2439302 (*NRG1*) on 8p12, identified by GWAS for PTC risk were established in a Han population [65]. Both rs965513 (*FOXE1*) and rs944289 (*NKX2-1*) were related to the elevated risk of sporadic Japanese PTC [66]. Previous Korean GWAS indicated that *NRG1, NKX2-1, FOXE1*, and *DIRC3* have been associated with increased DTC [60].

In a case-control study, Arg72Pro SNP of *TP53* gene was found to contribute to thyroid carcinogenesis in young people, women, non-smokers, and individuals with high TSH levels [67]. Additionally, in a previously conducted meta-analysis, *XRCC1* gene Arg280His and Arg194Trp polymorphism were identified to be irrelevant for thyroid cancer risk in the Caucasian and mixed populations. However, Arg399Gln polymorphism in *XRCC1* gene was associated with a significantly reduced risk of thyroid cancer [68]. Although previous studies suggest genotype-phenotype correlation of different effects of genetic variants on the risk of thyroid cancer, the pathology is yet unknown and these associations remain unclear, as described in a systematic review [69].

1.2.3 Genetic variants for body mass index (BMI) and obesity

Many serious diseases in the developed worldwide are associated with obesity, such as coronary heart disease, stroke, hypertension, type 2 diabetes, some type of cancer, and cardiovascular diseases [70, 71]. BMI is a single index usually utilized to distinguish a person as underweight, normal weight, overweight or obese [72]. Also, almost 40-70% of the inter-individual variability in BMI, usually used to assess obesity, has been connected to genetic factors [73]. Several GWAS have confirmed that common genetic variants contribute to obesity, with 426 findings of concrete correlation in 127 candidate genes [71].

The fat quantity and obesity-associated (*FTO*) gene was the first factor related to obesity recognized by GWAS, and the most extraordinary effect is seen in young adulthood [74]. The minor allele raises BMI by 0.39 kg/m^2 (or 1,130g in body weight)

and the probability of obesity by 1.20 times [75]. The first two SNPs that described *FTO* as an obesity susceptibility gene are also commonly associated with BMI, rs9939609 and rs9930506 [76, 77]. Several studies have reported larger SNPs to be correlated with obesity, such as rs1121980 (first intron of *FTO*) on 16q12.2, rs8050136, rs3751812, and rs7202116 [78-80]. In addition, the relationship of *FTO* with BMI from 249,796 people having European ancestry was confirmed to be rs1558902; as the common noticeably related to *FTO* marker [81]. The relationship between *FTO* SNPs and BMI was also proven in different ethnic societies, for example, East Asian and African ancestry populations [82, 83]. The biological role of *FTO* gene in body mass has been well-established. Murine models overexpressing *FTO* are distinguished by the increase of white adipose tissue and adipocyte volume. However, germline or neural knockout mice present decreased lean and lipid quantity [84]. At the lower level of cell, *FTO* has been investigated to influence lipid mass through controlling reproduction and differentiation of pre-adipocytes. The function of *FTO* in adipogenesis can be credited to its N-methyl adenosine demethylase activity and its involvement in splicing regulation [85].

Other evidence for an association in the etiology of obesity is the melanocortin 4 receptor (*MC4R*) gene that was published as the secondary relevant signal for universal obesity by the GWAS. The rs17782313 SNP near the *MC4R* gene was discovered to be connected with obesity amongst both European adolescents and adults, with an increased risk of 12% and 30%, respectively [86]. Different SNPs (rs12970134) near the *MC4R* gene were similarly proposed to improve the 12% risk of obesity amongst Europeans [87]. Consequently, several studies have examined the

correlation in diverse ethnic groups [88-90]. Furthermore, other SNPs including rs571312, rs17700144, and rs4450508, which hold in higher linkage disequilibrium (LD) with rs17782313 or rs12970134 SNPs, were additionally reviewed [91].

The beta3-adrenergic receptor gene (*ADRB3*) is known to be predominantly expressed in adipose tissue and potently stimulates lipolysis and thermogenesis. *ADRB3* is implicated in experiments in the control of lipid metabolism, from fat assimilation in the digestive tract, to triglyceride storage. Therefore, via its influence on energy releasing of obese tissue, a damage of *ADRB3* function results in obesity [92]. A meta-analysis of 31 investigations on over 9,000 patients has established an important relationship between the Trp64Arg polymorphism of the *ADRB3* gene and BMI [93]. The association amongst different community groups presented an approximately comparable magnitude, and the *ADRB3* locus has been confirmed to be one of the hereditary factors correlated with body mass. Recently, a different meta-analysis performed in Japan established that the *ADRB3* gene Trp64Arg SNPs is correlated with BMI [94].

Other genes, like *LEPR* gene including the glucocorticoid receptor gene (*GRL*), have been shown to be related to elevated BMI, improved weight gain, or obesity in several communities. Nevertheless, these findings from two up-to-date meta-analyses showed that there was no substantial proof of a relationship among these two genes and obesity [95, 96].

1.3 Use of genetic instrumental variables for assessing the causal association

1.3.1 Introduction to Mendelian randomization (MR)

"Epidemiology is the study of the distribution and determinants of health-related states or events including disease, and the application of this study to the control of diseases and other health problems" [97]. A fundamental problem in epidemiology research is that an observational relationship between two measures may indicate a causal influence of one variable on the other. If we want to answer the question: "what causes a disease?" to prevent disease etiology, or "what would be the result of a treatment?" to inform public policy, or to counsel on the impact of lifestyle choice, then we have to consider the questions of cause and effect. Unfortunately, observational studies are limited by residual confounder, reverse causality, and multiple biases that frequently make it challenging to interpret whether such an association indicates a causal association. Certainly, there is a prolonged past of observational epidemiological investigations indicating to show influential relationships linking multiple risk factors and disease that upon modern research came out to be non-causality, most reasonably reasoning to the appearance of residual confounders.

The randomized controlled trial (RCT) is the gold standard for authenticating causality among a medically related susceptibility and an outcome. This design involves dividing a set of individuals into two or more subgroups in a random way. These subgroups are each given different treatments. However, RCTs are expensive,

time-consuming, and not constantly humane or generalizable to communities outside the rigorously managed confines of the research. In addition, translating the relationship between risk factors and disease results in the observational study as a causal relationship relies on untestable and usually unreasonable hypotheses. This issue has driven to different situations where a potential factor has been broadly encouraged as a significant factor in disease prevention based on observational data, but later the proof of randomized trials did not confirm a causal explanation. For instance, vitamin C was initially found to be a strong inverse association with coronary heart disease in an observational study [98]. Nevertheless, outcomes of experimental data acquired from randomized trials presented an ineffective association with a concrete point estimation for the association [99]. More robust approaches are therefore needed for assessing causal relationships using observational data. Genetic approaches are helpful to understand causal relationship because hereditary modifications are being from birth and consequently unlikely to be confounded with environmental representatives. Mendelian randomization (MR) "is an approach that utilizes genetic variants robustly related with a modifiable exposure or biological intermediate of interest to find out the causal relationship between these variables and a medically relevant outcome being free from the effect of confounding" (see **Figure 3**). Although this originally emerged to investigate the association between adjustable exposures/biomarkers and illness, its application has extended to include utilization in molecular research, systems biology, pharmacogenomics, as well as other fields. This method is remarkable due to the rising of plenty of researches that estimates associations between molecular

intermediates (for instance, gene expression, gene methylation, metabolites, and metagenomic information), and those studies match all the similar problems of residual confounding, reverse causation as more conventional measurements [100].

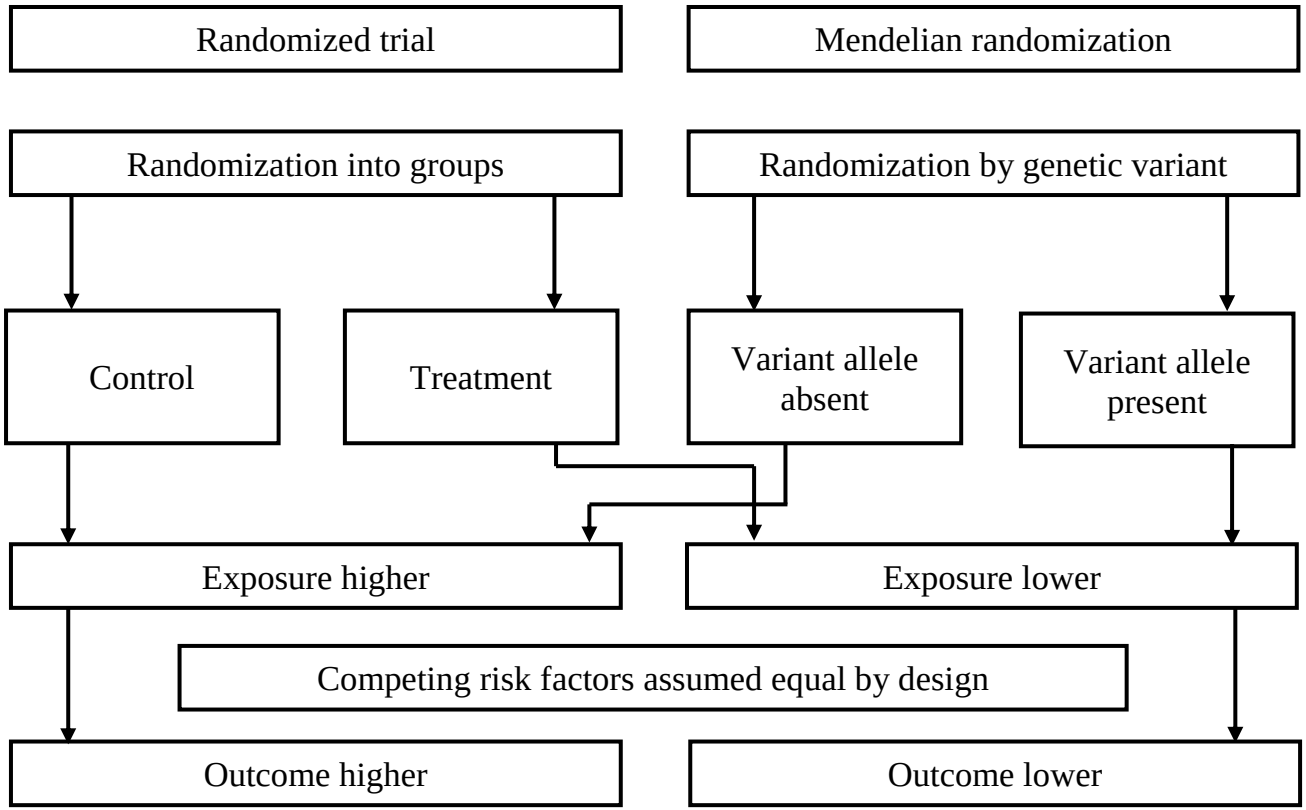

Fig 3. Comparison of a randomized controlled trial and MR

Available from: Burgess, S. and Thompson, S. (2015). "*Mendelian randomization: Methods for using genetic variations in causal estimation."* CRC Press, Taylor & Francis Group

1.3.2 The MR framework for instrumental variable analysis

The fundamental principle used in the MR method is that naturally, random allocation of alleles at conception as the basis of a natural experiment, whereby a genetic variant is used as a proxy for a clinically related risk factor for the disease. The genotype is not affected by the confounding and unchangeable by the development of outcome. Unlike, the estimation of the outcome in observational studies usually is biased by the direct measurement of the modifiable risk factor. Consequently, in the MR the majority of variance in the modifiable risk factor will be explained by the genetic variant [62]. In order to be used as an instrumental variables, such variants must satisfy three fundamental conditions: (a) the variant is associated with the exposure, (b) the variant is not associated with any confounder of the exposure-outcome association, and (c) the variant does not affect the outcome, except possibly via its association with the exposure (see **Figure 4**) [101].

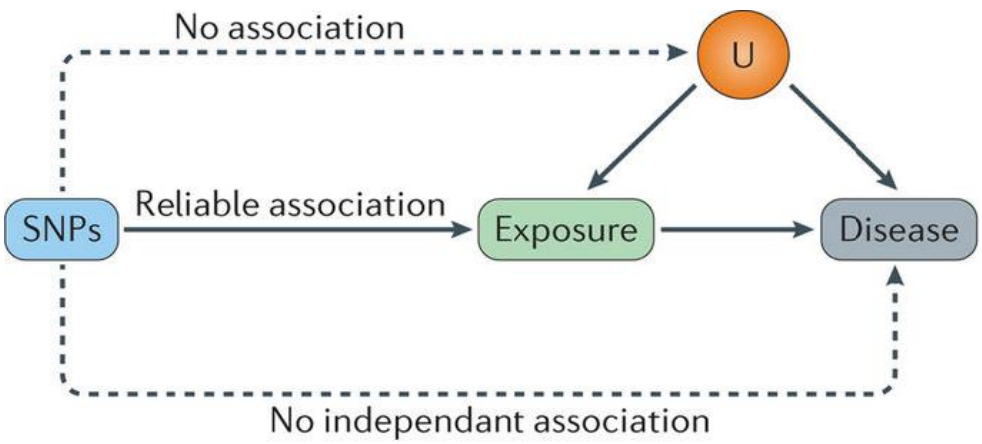

Fig 4. Instrumental variable analysis to create causal estimates by MR [101] The three main assumptions analysis are: the instrumental variable is associated with the exposure (in this example a genetic variant either in isolation or in joining with other); the instrumental variable is associated with the outcome through the studied exposure only; and there is no of other factors (U: confounders) which affect the outcome. [102].

Available from: https://www.nature.com/articles/nrcardio.2017.78

There are numerous studies utilizing those instrumental variables. The different strategies have been introduced to carry out the MR analysis and estimation of a causal effect parameter, in particular, the use of approach depending on the character of the population. The most common method used is two-stage least squares (2-SLS), and the Wald ratio estimate was most frequently applied [103].

$$\hat{\beta}_{IV} = \frac{\hat{\beta}_{Y|G}}{\hat{\beta}_{X|G}}$$

The causal effect estimates $\hat{\beta}_{IV}$ obtained from the Wald method, where $\hat{\beta}_{Y|G}$ is the regression coefficient of the regression outcome on the instrument, and $\hat{\beta}_{X|G}$ is the regression coefficient of the exposure on the instrument.

The confidence interval or standard errors for the ratio estimate can be calculated in several ways, such as Fieller method [104], delta method (Taylor series expansion) [105], and bootstrapping. There are also several other methods available for estimation of instrumental variables, which are: the ratio of coefficients methods, semi-parametric methods (generalized method of moments, continuous updating estimator, G-estimation of the structural mean), Bayesian methods, k-class estimators, and limited information maximum likelihood method [103]. In particular,

an exposure is a continuous and outcome is a continuous or binary. Those are methodologies usually used in order to conduct MR analysis.

The capacity to determine the quantity of the causality of the longstanding susceptibility on the modifiable exposure of interest was also acknowledged [106]. In addition, many candidate gene and GWAS have been reported, which now allows for the manipulation of MR studies that apply these association without recruiting new patients or designing additional research. This is indicated in the rising number of instrumental variable analysis in general and MR studies in specific (see **Figure 5**).

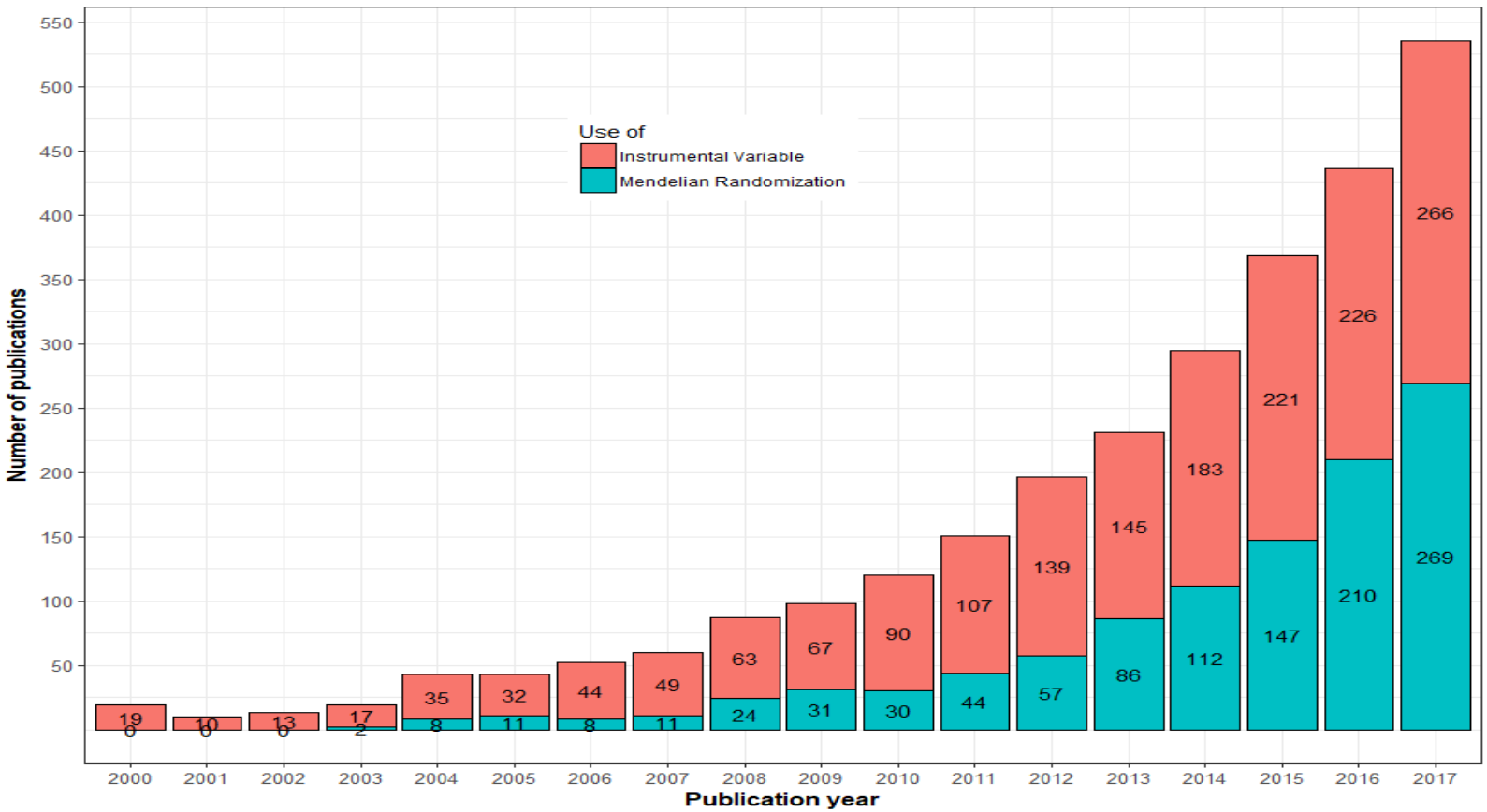

Fig 5. Use of MR and instrumental variable approaches in the literature increases over time

PubMed Search strategy (March of 2018): for MR analysis, "mendelian random*[tiab]" (medical subject headings [MeSH]; for instrumental variable analysis, "instrumental variable*[tiab]".

We searched PubMed for subjects including the phrase "Mendelian randomization" or a related term from 1 January 2003 to 31 December 2017. We excluded publications with these characteristics: (i) were discussion, reports, reviews, articles, surveys, study projects or technical articles; (ii) did not use MR (e.g. did not declare MR or a genetic IVs was used in the manual, abstract or heading and did not introduce 'Mendelian randomization' or a relevant title as a keyword); (iii) classified possible IVs for future MR researches; (iv) were principally methodological, utilizing a use of MR as a case; or (v) were reported in a health finance or economics publication preferably than a medicine publication.

After searching by the key word "Mendelian randomization" on PubMed database, 1,050 publications were obtained. After scanning for the abstracts and titles and if needed the full-text report, a further 504 reports were dismissed for analyses noted in **Figure 6**, ending in 546 available Mendelian randomization studies. The involved studies were reported between January 2003 and December 2017. A summary of the genetic variant and exposures of interest studied utilized are shown in **Table 1**. The majority of studied exposures was adiposity measures including body mass index, lipid quantity and portion body fat (130 studies), C-reactive protein (81 studies) and alcohol use (73 studies).

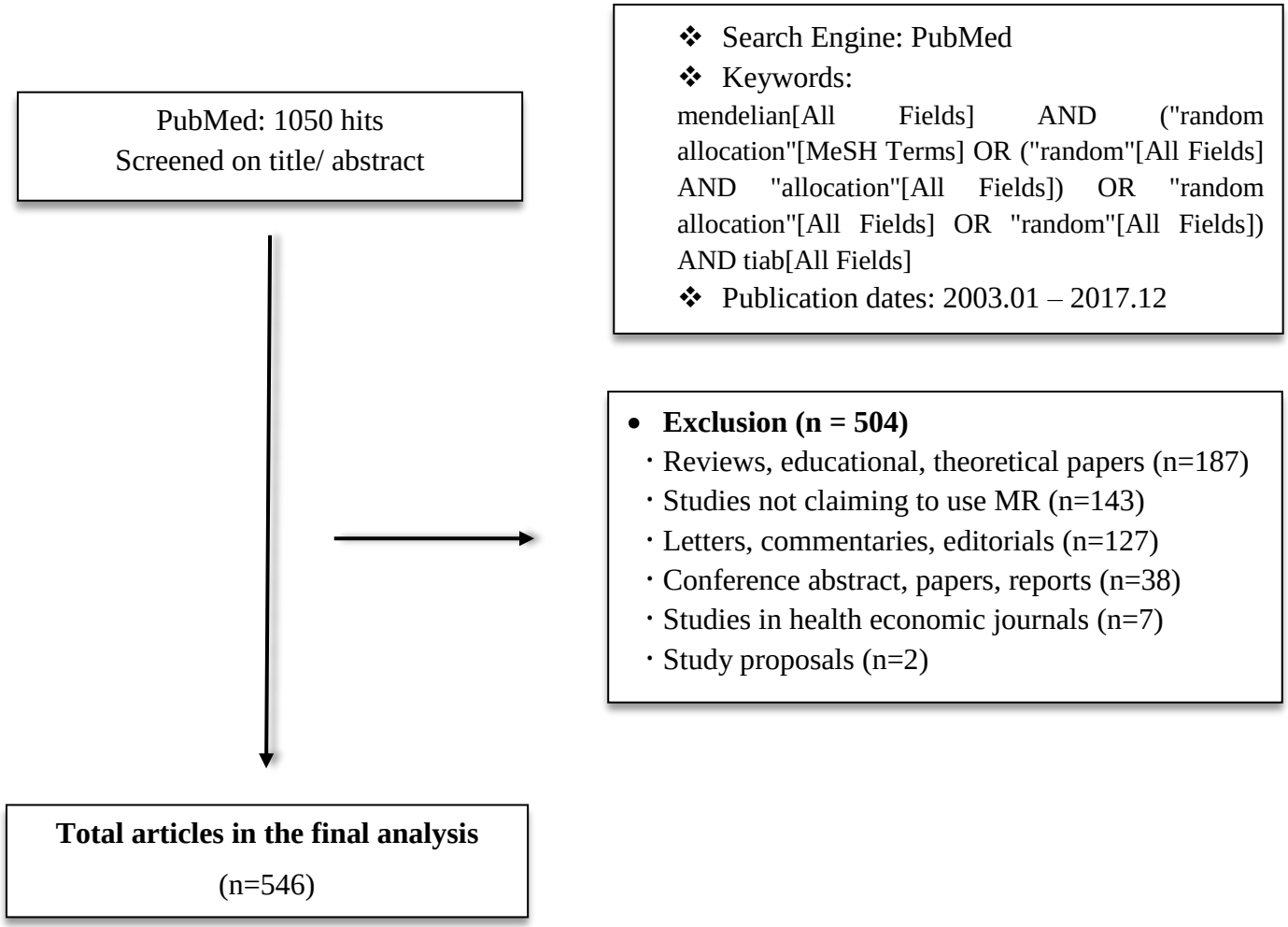

Fig 6. Flowchart of the literature search

Table 1. Exposures and genetic instruments in studies using MR analysis

Exposure	Number of studies	Genes in which variation was used as an instrument
BMI/ fat mass/ percentage body fat	130	*FTO, MC4R, TMEM18, VEGFA, genetic risk score*
C-reactive protein	81	*CRP, LEPR, HNF1A, IL6R, APOE, genetic risk score*
Alcohol use	73	*ALDH2, ADH1B, ADH1C*
Vitamin D levels	62	*GC, DHCR7/NADSYN1, CYP2R1, CYP24A1, FLG, VDR, genetic risk scores*
Homocysteine	27	*MTHFR*
"Folate metabolism"	10	*MTHFR*
LDL-cholesterol	16	*SORT1, PCSK9, LDLR, HMGCR, ABCG8, APOE, APOB, genetic risk score*
HDL-cholesterol	25	*LIPC, LIPG, ABCA1, LCAT, genetic risk score*
Total cholesterol	6	*APOE*
Remnant cholesterol	5	*APOA5, TRIB1, GCKR, genetic risk score*
Remnant cholesterol: HDL ratio	1	*LPL*
Triglycerides	5	*APOA5, genetic risk score*
Lipoprotein(a)	7	*LPA*
Lp-PLA$_2$ (activity)	4	*PLA2G7, PLA2G2A*
ApoAI	1	*APOA5-A4-C3-A1*
ApoB	1	*APOB*

Uric acid	7	*SLC2A9, ABCG2, SLC17A1, SLC22A11, SLC22A12, genetic risk score*
IL-6/ IL-6 receptor signaling	5	*IL6, IL6R*
Fetuin-A	4	*AHSG*
Adiponectin	4	*ADIPOQ*
Fibrinogen	4	*FIBA-B-G cluster, FGB*
Fasting glucose	3	*genetic risk score*
HOMA-IR	2	*GCKR, ADAMTS9, PPARG2*
Beta-cell function	2	*ADAMTS9, TCF7L2*
Non-fasting glucose	1	*GCK, G6PC2, ADCY5 DGKB, ADRA2A*
Fasting insulin	1	*INSR, IRS1*
Type 2 diabetes	2	*genetic risk score*
Type 1 diabetes	1	*genetic risk score*
Milk consumption	3	*LCT*
Iron status (ferritin/serum iron)	3	*HFE, TMPRSS6*
Bilirubin	3	*UGT1A1*
SHBG	3	*SHBG*
Testosterone	2	*SHBG, FAM9B, CYP19A1, ESR2*
Prenatal testosterone exposure	1	*Sex of co-twin*
Caffeine (intake)	2	*CYP1A2, NAT2, GSTA1*
Vitamin B-12	2	*FUT2, TCN2, CUBN*
Total transcobalamin	1	*TCN2*
Smoking	2	*CHRNA5–CHRNA3–CHRNB4 cluster*
PAI-1 levels	2	*PAI14G/5G*

Malaria infection	2	HbAS phenotype
IL-18	2	*IL18*
Macrophage migration inhibitory factor	2	*MIF*
6-propylthyouracil tasting	1	*TAS2R38*
Monocyte chemotactic protein-1	1	*CCL2*
Leukocyte telomere length	1	*genetic risk score*
Triacylglycerol	1	*genetic risk score*
sPLA$_2$-IIa	1	*PLA2G2A*
γ-glutamyl transferase	1	*GGT1*
Δ^5-desaturase and Δ^6-desaturase activity	1	*FADS1*
Monocyte CD36 expression	1	*CD36*
Factor VII	1	*F7*
Retinol-binding protein 4	1	*RBP4*
Complement factor H	1	*CFH*
Surfactant protein D	1	*SP-D*
MiR-34b	1	*Pri-miR-34b/c*
ICAM-1	1	*ICAM1*

P-selectin	1	*SELP*
CSF ApoE	1	*APOE, genetic risk score*
NT-pro-BNP	1	*BNP*
APC resistance	1	*FVL*
ACE activity	1	*ACE D/I*
Prothrombin levels	1	*F2*
Beta-carotene	1	*BCMO1*
Arsenic metabolism efficiency	1	*AS3MT*
IL-1RA	1	*IL1RN*
Inflammatory/auto-immune disease	1	*IL23R, PTPN2, PTPN22, SH2B3, IL2RA (+ 31 others)*
Ceruloplasmin	1	*CP*
Organophosphate exposure	1	*PON1*

Table 2. Available results of the association between obesity (BMI) and specific diseases using MR analysis

Outcome	Instrument	Description of results	Reference
Type 2 diabetes	30 BMI-associated variants	The potential causal association between abdominal obesity and hyperglycemia can be forced by increased insulin resistance, which is different with the possible causal association between overall obesity and insulin secretion.	[107]

Hay fever, asthma and lung function	26 BMI-associated variants and genetic risk score	Increasing BMI is causally associated with a greater prevalence of asthma and reduced lung function but not related to hay fever or biomarkers of allergy.	[108]
Cardiometabolic diseases	15 (SNPs) for childhood BMI	A genetic predisposition to childhood BMI was related to elevated risk of type 2 diabetes and coronary artery disease in adult life.	[109]
	A polygenic risk score comprising 93 SNPs associated with BMI	Higher BMI and increased risk of cardiometabolic diseases have been observed.	[110]
Intracranial aneurysms (IA) and abdominal aortic aneurysms (AAA)	97 BMI-associated variants, genetic risk score	There are potentially causal associations between BMI and risk of AAA.	[111]
Posttraumatic stress disorder (PTSD)	High-resolution polygenic score	A putative causal association between genetically determined female body shape and PTSD, which may be interfered by evolutionary mechanisms implicated in sexual behaviors among human	[112]
Alzheimer's disease	32 BMI-associated variants	Heretical and thus lifelong low BMI is not related to	[113]

		elevated risk of Alzheimer's disease in the overall population. The low BMI is not a cause for Alzheimer's disease.	
Gastric cancer	37 BMI-associated variants, genetic risk score	High BMI was associated with risen gastric cancer risk.	[114]
Venous thromboembolism (VTE)	32 BMI-associated variants	Proving for a causality between high BMI and risk of VTE, declining obesity levels will likely result in lower incidence of VTE.	[115]
Ischemic stroke subtypes	77 BMI-associated variants	Genetically predicted BMI was not significantly correlated with any ischemic stroke subtype.	[116]
Breast cancer survival	94 BMI-associated variants, genetic risk score	A causal impact of elevated BMI on decreased breast cancer survival for ER-positive breast cancer. Evidence of a causal impact of higher BMI on survival for ER-negative breast cancer individuals is inadequate.	[117]
Pediatric disease	97 BMI-associated variants, genetic risk score	Causal effects of increased BMI on susceptibility to pediatric-onset multiple sclerosis were identified.	[118, 119]

Coronary heart disease, stroke subtypes, and type 2 diabetes	97 BMI-associated variants and 49 SNPs for waist-to-hip ratio adjusted for BMI (WHRadjBMI)	Both BMI and WHRadjBMI have causal effects on CHD and type 2 diabetes mellitus. WHRadjBMI may have a more robust effect on the risk of stroke.	[120]
Atrial fibrillation (AF)	*FTO* genotype (rs1558902) and a BMI gene score comprising 39 SNPs	There was a causal relationship between BMI and incident AF.	[121]
Type 1 diabetes (T1D)	23 SNPs associated with childhood adiposity	This provides genetic support for a relationship between childhood adiposity and T1D possibility.	[122]
Lung cancer	97 BMI-associated variants	A causal effect of BMI on lung cancer risk for two of the three main histological subtypes, with confirmation of a risk development for squamous cell carcinoma, and for small cell lung cancer, but not for adenocarcinoma	[123]
Pancreatic cancer	95 BMI-associated variants	A strong causal relationship of raising BMI with the risk of pancreatic cancer was observed.	[124]

Thyroid hormone levels	32 BMI-associated variants	Higher BMI had higher Free T_3 (FT_3) but not FT_4 levels, proving that higher BMI/fat mass has a causal effect in raising Free T3 (FT3) levels.	[125]
Endometrial cancer	77 BMI-associated variants	BMI is causally related to the risk of endometrial cancer.	[126]
Multiple sclerosis (MS)	70 BMI-associated variants	Genetically high BMI is associated with risk of MS, giving evidence for a causal role for obesity in MS etiology.	[127]
Peripheral arterial disease (PAD)	14 BMI-associated variants	There was a causal relationship between obesity and PAD.	[128]
Psychiatric disorders (PD)	97 BMI-associated variants	Higher BMI may not increase the risk of bipolar disorder and schizophrenia.	[129]
Breast cancer risk	84 BMI-associated genetic variants	BMI predicted from variants is inversely correlated with the risk of both pre- and postmenopausal breast cancer.	[130]

1.4 Study objectives

The risk of multiple diseases including thyroid disease rises as body mass index (BMI) increases. Both overweight and obesity have been related to an increased risk of thyroid cancer in males and females. However, recent studies did not confirm this result. It remains to be unclear about that association. Furthermore, the observational association between BMI and thyroid cancer can be due to confounders. For example, there may be another factor decreasing BMI and causing thyroid cancer at the same time.

Mendelian randomization (MR) is a method using an instrumental variable (IVs) to indicate the causal relationship in observational research. Gene variations which are linked with environmental exposures or intermediary phenotypes can be handled as IVs to assess the influence of the exposure on a disorder outcome. Random assortment of gene variations during gametogenesis indicates that possible confounders exceed the possibility to be evenly allocated. Furthermore, the effect estimates resulting from IVs analysis is suitable to be independent of residual confounder and reverse causal relationship since the outcome cannot influence the genotypes. This study utilizes MR method in the estimation of the causal effect of BMI on the risk of thyroid cancer (**Figure 7**).

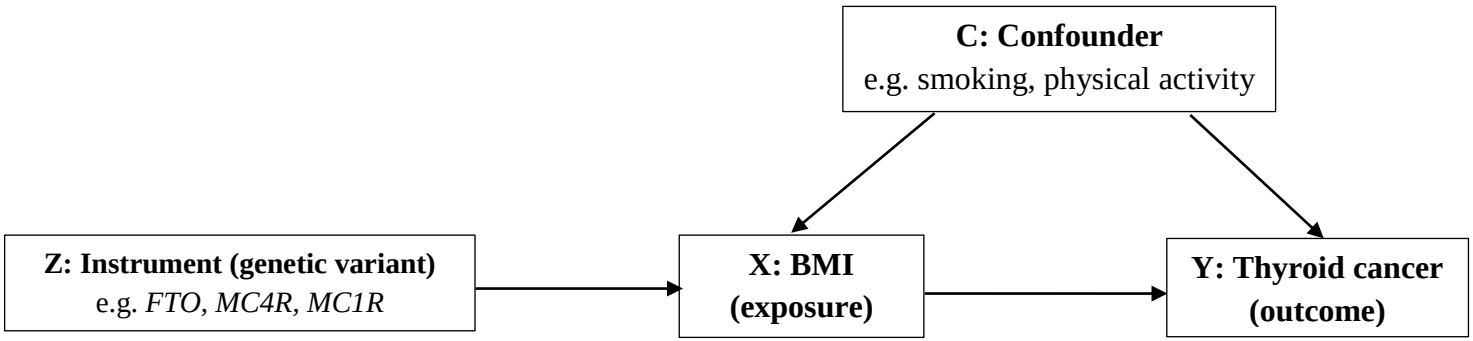

Fig 7. Directed acyclic design of instrumental variable analysis utilizing genetic variants as representatives for environmental exposures

The instrument (genetic variants: Z) associated with an exposure (BMI: X) can be utilized as substitutes to determine the result of the exposure (X) on the outcome (Thyroid cancer: Y). The three IVs conditions are symbolized by the arrows: (1) the IVs in this schematic (*FTO, MC4R, MC1R* gene variants) is robustly linked with the exposure; (2) the IVs is not associated with confounder (C); and (3) there is no alternative route that the IVs affects the outcome other than through the exposure.

2. Materials and Methods

2.1 Study population

Our study included individual participant data and genetic material from a large hospital-based case-control study by using the data from the Cancer Screenee Cohort of the National Cancer Center (NCC) in Korea from August 2002 to July 2014 and the Health Examinee (HEXA), and Rural cohort of the Korean Genome Epidemiology Study (KoGES) shared control study (**Figure 8)**. The NCC cohort study consisted of 41,109 subjects; the HEXA shared control only is one of the KoGES population-based cohorts which were launched in 2001 intending to distinguish risk factors of life-style related multiple disorders including dyslipidemia, type 2 diabetes, and hypertension. Roughly 30% of HEXA participants at the age of 40-69 years who were randomized shared the comparison group for Korean cancer and coronary artery disease (CAD) GWAS. The Rural cohort of KoGES was selected from amongst citizens aged 40-69 years of three provincial centers of Korea. During 2004 and 2008, 8,702 men and women were obtained for the baseline. Among them, 4,052 individuals without any diseases have no records of hypertension, diabetes, hyperlipidemia, heart disease, blood vessel disease of the brain, or cancer were also chosen for SNP genotyping.

The environmental and lifestyle information was collected from all of the patients who completed a self-administered questionnaire about demography and lifestyle characteristics such as age, gender, smoking, alcohol consumption, and physical exercise (**Appendix 1**).

The height and weight of each subject were measured with an Inbody 3.0 (Biospace, Seoul, Korea) body composition analyzer or X-SCAN PLUS II Body Composition Analyzer (Jawon Medical, Gyeongsan, Korea). Body mass index (BMI) was measured as weight (kg) / (height (m))^2. Ethical approval was obtained for all cohorts in the corresponding centers. All participants provided written informed consent prior to participation, and the Institutional Review Board of the NCC approved the study protocols (IRB No. NCC2016-0088) (**Appendix 2**).

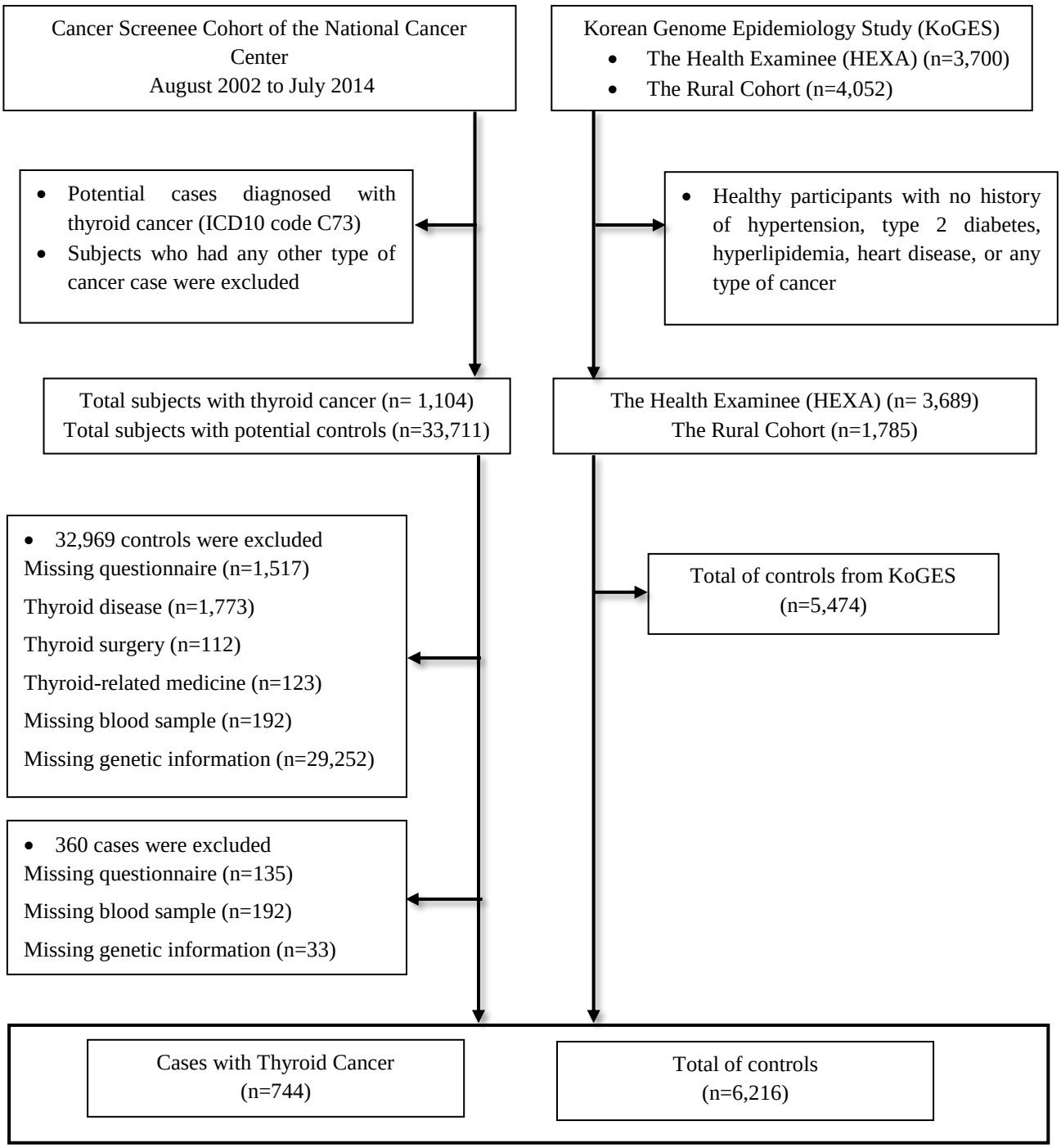

Fig 8. Flowchart of selecting study subjects in the study

2.2 Genotyping

The genotyping was performed using the Infinium OncoArray-500K BeadChip (Illumina Inc., San Diego, California) with 499,170 SNPs (275,691 genome-wide tag SNPs and 223,479 cancer-specific SNPs). The DNA samples are denatured and neutralized to prepare them for isothermal amplification. Approximately 200 ng of genomic DNA was amplified and randomly fragmented into 25 – 125-base pair fragments. Genomic DNA was subjected to initial amplification in a 40-μL reaction volume containing 20 μL genomic DNA at a concentration of 10 ng/μL. The denatured DNA is isothermally amplified overnight.

The whole-genome amplification process increases the amount of the DNA sample up to several thousand-fold without significant amplification bias. A controlled enzymatic process fragments the amplified product. The process uses endpoint fragmentation to avoid over-fragmenting the sample. After an isopropanol precipitation, centrifugation at 4°C collects the fragmented DNA. The precipitated DNA is re-suspended in hybridization buffer. Samples are implemented to a BeadChip and separated by an IntelliHyb seal (or gasket) and the loaded BeadChip is hatched overnight in the Illumina Hybridization Oven. The amplified and fragmented DNA samples anneal to locus-specific 50-mers during hybridization. Unhybridized and non-specifically hybridized DNA is removed away and the BeadChip is served for staining and extension. Single-base extension of the oligos on the BeadChip, using the captured DNA as a template, incorporates detectable labels on the BeadChip and determines the genotype call for the sample. XStain occurs in a capillary flow-through chamber. Using a laser to excite the fluorophore of the single-base extension

product on the beads, the Illumina HiScan or iScan System scans the BeadChip. The scanner records high-resolution images of the light emitted from the fluorophores. The images were analyzed using Genotyping Console™ Software (Affymetrix).

After quality controlling for monomorphic variants, Minor Allele Frequency (MAF) < 0.01, call rate < 95%, and deviation from Hardy-Weinberg Equilibrium (p-value < $1x10^{-6}$), a total of 345,675 single nucleotide polymorphisms (SNPs) were available for further analysis. The genetic variant which passed the above quality-controlling criteria was selected for this study.

2.3 Selection of BMI-associated SNPs

SNPs of the biggest GWA study of BMI were classified from reported records of the GIANT (Genetic Investigation of Anthropometric Traits) consortium (http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium) in 2015. SNPs associated with BMI at the genome-wide significance level ($p < 5\times10^{-8}$) in the population of all ancestry were involved (see **Figure 9**). We chose independent SNPs which were described as r^2 threshold of 0.001 or within 10,000 kb physical distance, based on a 1,000 Genome Project dataset.

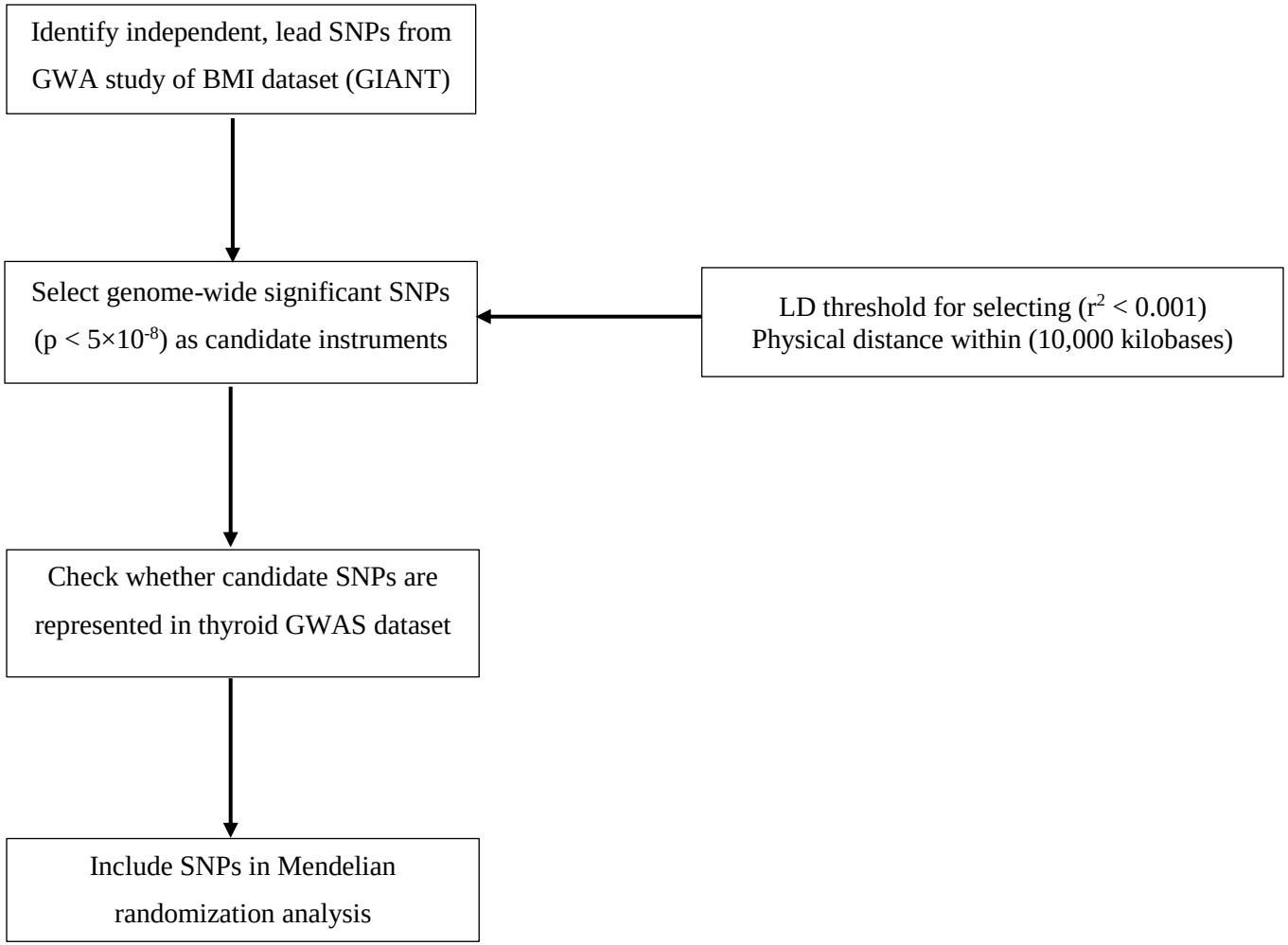

Fig 9. Procedure for selecting instruments for assessing the effect of BMI on the TC risk

2.4 Statistical analysis

Multiple IVs are aggregated into a single univariate score. This score is used as a single IV for BMI rather than multiple IVs. In MR, this is well-known as the genetic risk score. Genetic risk scores for BMI (BMI-GRS) which are used for MR are composed using external weights and determined using the following equation:

$$\text{GRS} = \sum_{i=1}^{55} \beta_i \, SNP_i$$

where β_i is the effect of the *i*th SNP for BMI reported in the GIANT GWAS and SNP_i is the dosage of the effect allele of the *i*th SNP that can be represented as random variables taking the value 0, 1, or 2.

First, to access whether weighted GRS was associated with BMI and the proportion of BMI variation explained by BMI-GRS, we used linear regression model within controls. To assess the statistical significance of the association of the instrument with BMI (BMI-GRS), an *F* statistic was calculated by the following formula, where *k* is the number of variants and n is the sample size:

$$F = \frac{R^2(n-1-k)}{(1-R^2) \times k}$$

The IV for the exposure of interest is considered to be sufficiently strong if the F-statistic is larger than 10. We then fit a logistic regression model with BMI-GRS as independent variable and thyroid cancer case-control status as an outcome.

We tested whether potential confounding factors (age, gender, drinking, smoking status, and regular exercise) were associated with observation BMI group and BMI-GRS group by using x^2 tests.

The possible causal association between BMI (X) and thyroid cancer risk (Y) was modeled using BMI-GRS as the instrumental variable. Particularly, the causal effect (β_{YX}) was estimated by using the Wald estimator [104]:

$$\beta_{YX} = \frac{\beta_{ZY}}{\beta_{ZX}}$$

Where β_{ZY} is the natural log-scale odds ratio (OR) for thyroid cancer risk associated with the instrumental variable; β_{ZX} is the regression coefficient of the instrumental variable for BMI.

The standard error for the causal effect was computed using the delta method [131]:

$$SE_{YX} = \sqrt{\left(\frac{S_{YG}}{\beta_{XG}}\right)^2 + \frac{(S_{XG}S_{YG})^2}{\beta_{XG}^4}}$$

β_{YG} and β_{XG} are standard errors of YG and XG, respectively.

To verify that violating of the third conditions of IV analysis was not influencing the estimate of the causal association which is the potential pleiotropic effect of the genetic variants, the logistic regression (logistic analysis) was used.

Analyses were implemented using PLINK (v 1.07), R (v 3.41). A two-sided p-value < 0.05 was considered statistically significant unless stated otherwise.

3. Result

3.1 Characteristics of study participants

Table 3 shows the general characteristics of the study participants which includes 744 cases and 5,216 controls. It provides baseline characteristics based on age, BMI, alcohol consumption, smoking status, and regular exercise. The mean age of study participants was 55 years for controls and 48 years for cases. In the total population, high proportion of cases was engaged in regular exercise ($p<.0001$). In male group, there were significant differences in BMI ($p<.0001$) and regular exercise ($p<.0001$). In female group, there were significant differences in BMI ($p<.0001$), alcohol consumption ($p<.0001$), smoking status ($p<.0001$), and regular exercise ($p<.0001$).

Table 3. General characteristics of the study participants

	All			Male			Female		
	Control (n=6,216)	**Case** (n=744)	**p value**	**Control** (n=2,719)	**Case** (n=229)	**p-value**	**Control** (n=3,497)	**Case** (n=515)	**p value**
Age (yrs)	54.8 ± 8.8	48.5 ± 8.5	<.0001	56.2 ± 8.9	47.8 ± 8.0	<.0001	53.7 ± 8.7	48.9 ± 8.6	<.0001
BMI (kg/m^2)	24.1 ± 3.1	24.0 ± 3.0	0.213	24.3 ± 2.9	25.6 ± 2.5	<.0001	24.0 ± 3.2	23.2 ± 2.9	<.0001
BMI (kg/m^2), n (%)			0.872			<.0001			0.001
Normal (<23)	2,323 (37.4%)	283 (38.0%)		901 (33.1%)	31 (13.5%)		1,422 (40.7%)	252 (48.9%)	
Overweight (23-24.9)	1,644 (26.4%)	199 (26.7%)		742 (27.3%)	72 (31.4%)		902 (25.8%)	127 (24.7%)	
Obese (≥25)	2,249 (36.2%)	262 (35.2%)		1,076 (39.6%)	126 (55.0%)		1,173 (33.5%)	136 (26.4%)	
Alcohol consumption			0.241			0.037			<.0001
Nondrinker	2,953 (47.5%)	336 (45.2%)		581 (21.4%)	35 (15.3%)		2,372 (67.8%)	301 (58.4%)	
Drinker	3,263 (52.5%)	408 (54.8%)		2,138 (78.6%)	194 (84.7%)		1,125 (32.2%)	214 (41.6%)	
Smoking status			0.434			0.710			<.0001
Nonsmoker	4,117 (66.2%)	504 (67.7%)		810 (29.8%)	65 (28.4%)		3,307 (94.6%)	439 (85.2%)	
Smoker	2,099 (33.8%)	240 (32.3%)		1,909 (70.2%)	164 (71.6%)		190 (5.4%)	76 (14.8%)	
Regular exercise			<.0001			<.0001			<.0001
No	3,360 (54.1%)	322 (43.3%)		1,423 (52.3%)	83 (36.2%)		1,937 (55.4%)	239 (46.4%)	
Yes	2,856 (45.9%)	422 (56.7%)		1,296 (47.7%)	146 (63.8%)		1,560 (44.6%)	276 (53.6%)	

3.2 The variants selected for the analysis

Table 4 depicts the variants chosen for the study with the nearest gene, chromosome, physical base-pair position, effect alleles and frequencies, and the coefficient value of the effect on BMI. Overall, a total of 55 SNPs was selected for analysis as IVs. These SNPs satisfied the selection procedure which was illustrated in **Figure 7**. Their functions are presented in **Supplementary Table 1**.

Table 4. The selected-variants associated with BMI

SNP	Genes	Chromosome	BP	Effect Allele	Non-Effect Allele	EA Frequency	EA beta	SE beta	p-value
rs1000940	*RABEP1*	17	5223976	G	A	0.225	0.0184	0.0033	1.81E-08
rs10132280	*STXBP6*	14	24998019	A	C	0.3333	-0.0221	0.0033	1.40E-11
rs10733682	*LMX1B*	9	128500735	A	G	0.425	0.0188	0.003	2.46E-10
rs10840100	*TRIM66*	11	8626013	G	A	0.725	0.0206	0.003	6.67E-12
rs11030104	*BDNF*	11	27641093	A	G	0.8	0.0416	0.0037	6.66E-30
rs11165643	*PTBP2*	1	96696685	C	T	0.425	-0.0221	0.003	1.43E-13
rs11191511	*CNNM2*	10	104759699	T	C	0.9417	-0.0295	0.0052	2.01E-08
rs11692326	*Intergenic*	2	207971524	T	C	0.2417	0.0194	0.0036	4.80E-08
rs12286929	*CADM1*	11	114527614	G	A	0.4333	0.0211	0.0029	5.44E-13
rs12429545	*OLFM4*	13	53000207	G	A	0.9	-0.0324	0.0044	3.15E-13
rs12940622	*RPTOR*	17	76230166	A	G	0.4583	-0.0183	0.0029	3.64E-10
rs13021737	*TMEM18*	2	622348	A	G	0.125	-0.0604	0.0039	5.44E-54
rs13025697	*Intergenic*	2	6516986	T	C	0.01695	-0.061	0.0067	7.14E-20
rs13130484	*GNPDA2*	4	44870448	C	T	0.5667	-0.0398	0.003	8.01E-41
rs13201877	*IFNGR1*	6	137717234	A	G	0.9167	-0.0236	0.0043	4.29E-08
rs13329567	*Intergenic*	15	65891421	T	C	0.2167	-0.0307	0.0035	1.53E-18

SNP	Genes	Chromosome	BP	Effect Allele	Non-Effect Allele	EA Frequency	EA beta	SE beta	p-value
rs1421085	*FTO*	16	52358455	C	T	0.45	0.0803	0.003	2.17E-158
rs1441264	*MIR548A2*	13	78478920	A	G	0.55	0.0172	0.0031	2.96E-08
rs1460676	*FIGN*	2	164275935	T	C	0.7833	-0.0209	0.0038	4.98E-08
rs14810	*KCTD15*	19	38996743	C	G	0.325	-0.0183	0.0033	1.92E-08
rs1516725	*ETV5*	3	187306698	T	C	0.0917	-0.0448	0.0044	1.39E-24
rs1528435	*UBE2E3*	2	181259207	T	C	0.5833	0.0175	0.003	4.77E-09
rs16851483	*RASA2*	3	17810167	G	T	0.9083	-0.0478	0.0075	1.85E-10
rs17094222	*HIF1AN*	10	51491638	C	T	0.2083	0.0249	0.0037	2.19E-11
rs17724992	*PGPEP1*	19	76290622	A	G	0.6917	0.0196	0.0034	7.79E-09
rs1928295	*TLR4*	9	47856286	C	T	0.425	-0.0182	0.0029	4.32E-10
rs2060604	*Intergenic*	8	42867698	T	C	0.5583	0.0203	0.003	9.46E-12
rs2112347	*HMGCR*	5	27380376	G	T	0.375	-0.0254	0.003	1.96E-17
rs2176598	*HSD17B12*	11	53873481	T	C	0.2	0.0185	0.0033	3.47E-08
rs2229616	*MC4R*	18	75037314	C	T	0.9833	0.0899	0.0113	2.07E-15
rs2365389	*FHIT*	3	15507246	C	T	0.6583	0.0195	0.003	1.35E-10
rs2820292	*NAV1*	1	5102745	A	C	0.4917	-0.0181	0.0029	5.45E-10
rs2836754	*ETS2*	21	80177445	C	T	0.65	0.0169	0.003	1.61E-08

SNP	Genes	Chromosome	BP	Effect Allele	Non-Effect Allele	EA Frequency	EA beta	SE beta	p-value
rs3800229	*FOXO3*	6	33741232	T	G	0.6917	0.0175	0.0032	4.95E-08
rs3817334	*MTCH2*	11	53980149	C	T	0.55	-0.0256	0.003	1.17E-17
rs3888190	*ATP2A1*	16	69354230	A	C	0.3583	0.0311	0.003	3.45E-25
rs4615388	*Intergenic*	6	32093039	A	T	0.25	0.0247	0.0045	4.05E-08
rs4740619	*CCDC171*	9	45545498	T	C	0.5333	0.017	0.0029	6.36E-09
rs543874	*SEC16B*	1	4409291	G	A	0.2667	0.0497	0.0037	2.29E-40
rs6091540	*ZFP64*	20	78942046	C	T	0.725	0.0185	0.0033	2.14E-08
rs6457796	*UHRF1BP1*	6	31601585	T	C	0.7417	-0.0209	0.0033	2.54E-10
rs6477694	*EPB41L4B*	9	47608600	C	T	0.3583	0.0169	0.003	1.71E-08
rs6567160	*MC4R*	18	75030739	C	T	0.2833	0.0562	0.0035	6.68E-59
rs657452	*AGBL4*	1	1459358	A	G	0.4167	0.0227	0.0031	2.12E-13
rs6804842	*RARB*	3	14452939	A	G	0.425	-0.0183	0.003	8.02E-10
rs7138803	*FAIM2*	12	57928867	G	A	0.5583	-0.032	0.003	5.12E-26
rs7243357	*GRP*	18	74999857	G	T	0.1333	-0.0219	0.0038	9.14E-09
rs7599312	*ERBB4*	2	12694008	G	A	0.7083	0.0214	0.0033	4.73E-11
rs7715256	*GALNT10*	5	29698313	G	T	0.45	0.0168	0.0029	8.85E-09
rs7903146	*TCF7L2*	10	51856937	T	C	0.25	-0.0235	0.0033	1.10E-12

SNP	Genes	Chromosome	BP	Effect Allele	Non-Effect Allele	EA Frequency	EA beta	SE beta	p-value
rs891389	*NPC1*	18	73954796	C	T	0.675	-0.0209	0.0037	1.62E-08
rs9374842	*LOC285762*	6	34052834	T	C	0.7417	0.0196	0.0034	7.20E-09
rs943005	*Intergenic*	6	32085035	T	C	0.1	0.0444	0.0038	4.52E-31
rs9540493	*MIR548X2*	13	61759583	G	A	0.55	-0.0182	0.0031	3.95E-09
rs9579083	*MTIF3*	13	60660968	G	C	0.7667	-0.0295	0.0046	1.43E-10

BMI, body mass index; BP, base pair; EA, effect allele; SE, standard error.

3.3 The body mass index and genetic risk score (BMI-GRS)

The BMI-GRS was normally distributed (mean, 0.0002979; median, 0.0002974; SD, 0.001194271; minimum, -0.0045273; maximum, 0.0048982) (see **Figure 10**) and explained 1.24% of the variance in BMI ($R^2 = 0.0124$), as calculated from linear regression analyses with BMI as the outcome variable. The F-statistic for the BMI-GRS and its association with BMI was 78.4, which is large. This value means that weak instrument bias was unlikely.

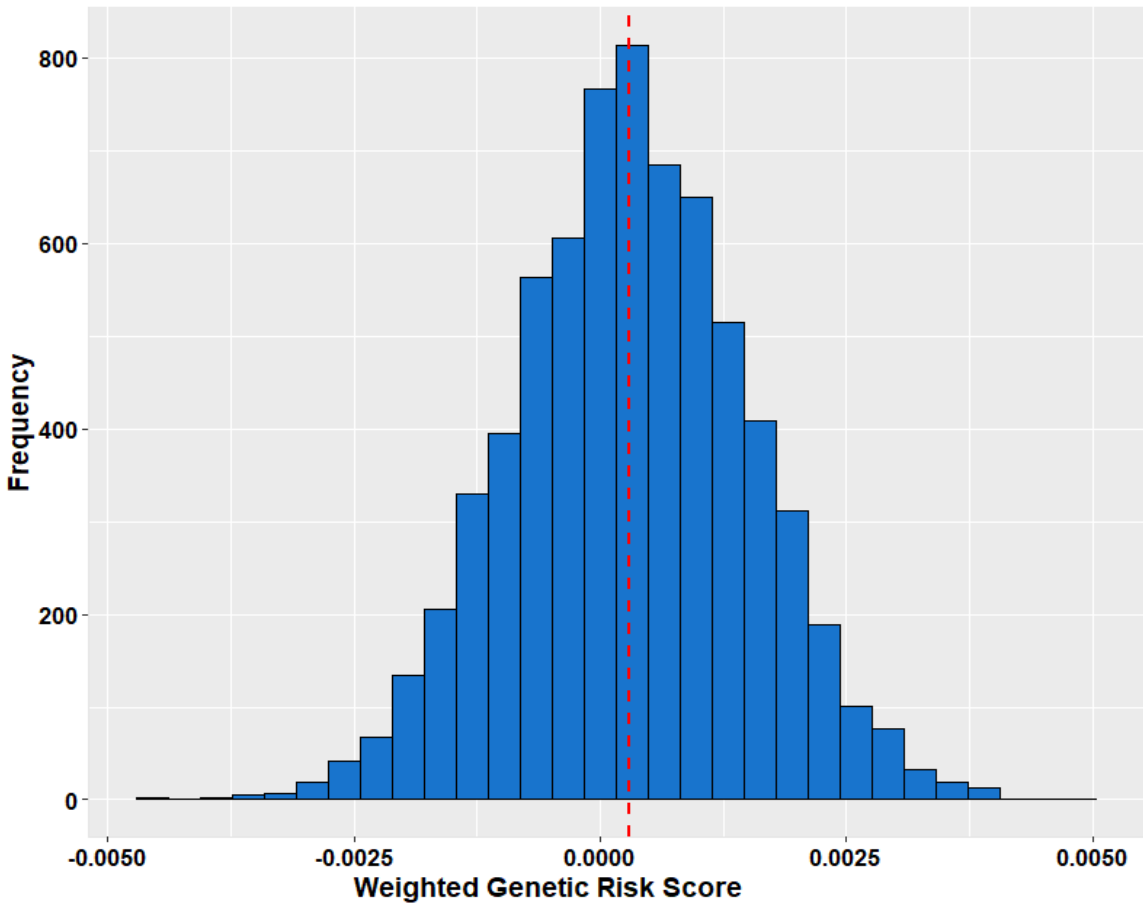

Fig 10. The distribution of weighed genetic risk score

The mean BMI-GRS was similar in thyroid cancer cases (0.0003275 ± 0.0012199) and in controls (0.00029439 ± 0.0011912) and was normally distributed (p = 0.39) by using Shapiro-Wilk's test **(Figure 11).**

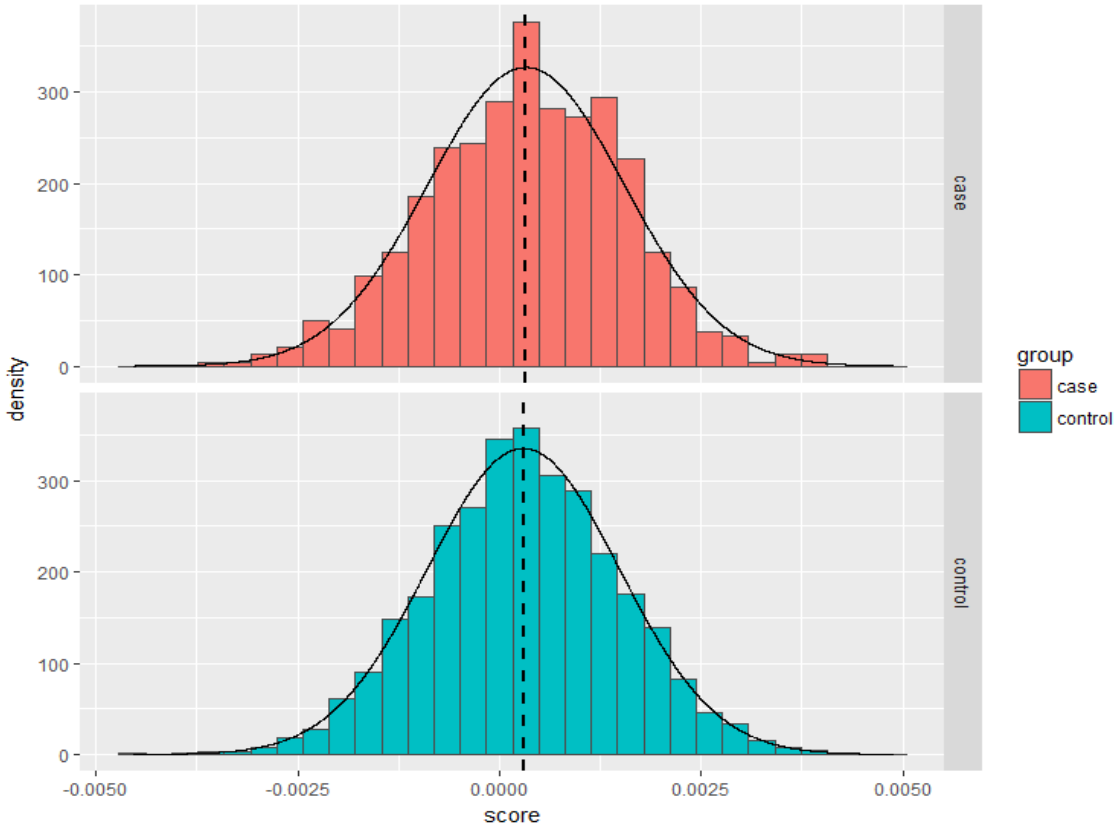

Fig 11. The distribution of weighted genetic risk score by case and control group

There was also no difference of the mean BMI-GRS in female group (0.0002908 ± 0.0011941) and in male group (0.0003075 ± 0.0011946) **(Figure 12).**

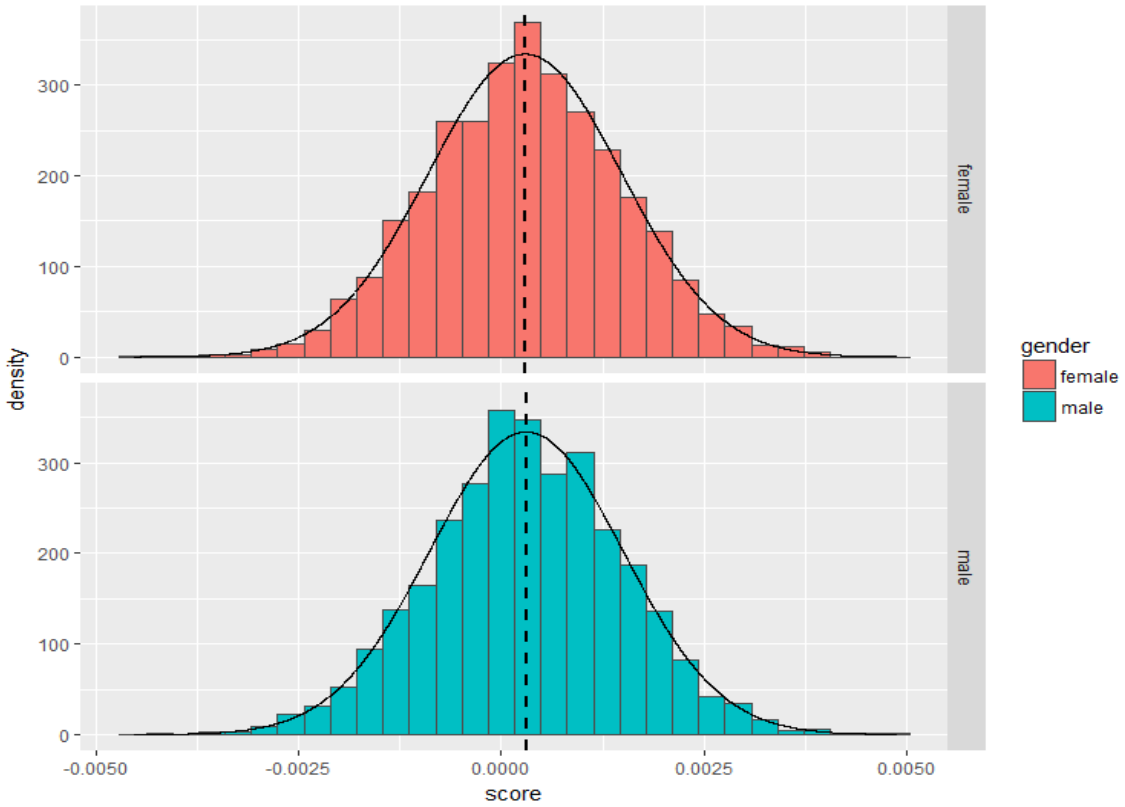

Fig 12. The distribution of weighted genetic risk score by gender

3.4 Association of potential confounders with BMI group and BMI-GRS group

The age, gender, drinking, smoking status, and regular exercise were all remarkably associated with BMI and risk of thyroid cancer and thus, being the potential confounders for the observational BMI-Thyroid cancer association. Potential confounders were dichotomized: gender (female and male); smoking status (non-smoker vs smoker); drinking status (non-drinker vs drinker); high regular exercise (yes vs no) (**Table 5**). In contrast, there was no association between BMI-GRS (IV) and age, gender, drinking, smoking status, and regular exercise.

Table 5. Association of potential confounders with BMI group and BMI-GRS group

Potential confounders	BMI Group					GRS Group				
	Group I	Group II	Group III	Group IV	p-value	Group I	Group II	Group III	Group IV	p-value
Age (Mean, SD*)	52.6 ± 9.6	54.0 ± 8.8	54.8± 8.6	55.1 ± 8.8	< .0001	53.9 ± 9.1	54.0 ± 8.9	53.8± 9.0	54.7 ± 9.1	0.014
BMI	20.4 ± 1.2	23.0 ± 0.6	24.9 ± 0.6	28.1 ± 2.0	< .0001	23.6 ± 2.9	24.0 ± 3.0	24.2 ± 3.1	24.5 ± 3.1	< .0001
Gender (%)										
Female	1,135 (65.2%)	1,054 (60.6%)	894 (51.3%)	929 (53.4%)	< .0001	1,029 (59.1%)	979 (56.3%)	999 (57.4%)	1,005 (57.8%)	0.414
Male	605 (34.8%)	685 (39.4%)	847 (48.7%)	811 (46.6%)		712 (40.9%)	760 (43.7%)	741 (42.6%)	735 (42.2%)	
Smoking (%)										
Non-smoker	1,186 (68.2%)	1,194 (68.7%)	1,117 (64.2%)	1,124 (64.6%)	0.005	1,184 (68.0%)	1,129 (64.9%)	1,146 (65.9%)	1,162 (66.8%)	0.256
Smoker	554 (31.8%)	545 (31.3%)	624 (35.8%)	616 (35.4%)		557 (32.0%)	610 (35.1%)	594 (34.1%)	578 (33.2%)	
Drinking (%)										
Non-Drinker	878 (50.5%)	824 (47.4%)	788 (45.3%)	799 (45.9%)	0.011	844 (48.5%)	803 (46.2%)	828 (47.6%)	814 (46.8%)	0.554
Drinker	862 (49.5%)	915 (52.6%)	953 (54.7%)	941 (54.1%)		897 (51.5%)	936 (53.8%)	912 (52.4%)	926 (53.2%)	
Regular exercise (%)										
No	983 (56.5%)	881 (50.7%)	875 (50.3%)	943 (54.2%)	< .0001	892 (51.2%)	953 (54.8%)	909 (52.2%)	928 (53.3%)	0.180
Yes	757 (43.5%)	858 (49.3%)	866 (49.7%)	797 (45.8%)		849 (48.8%)	786 (45.2%)	831 (47.8%)	812 (46.7%)	

* Mean ± standard deviation

In order to check the third assumption that the IVs do not affect the outcome, except possibly via its association with the exposure, the GRS was classified into four subgroups based on the quantile. We have not found any significant association between the highest quartile of GRS with the risk of thyroid cancer. The effect estimated from logistic regression in group 2 was 0.92 (95% CI 0.74-1.14; p = 0.440), group 3 was 0.96 (95% CI 0.77-1.19; p = 0.702), group 4 was 1.15 (95% CI 0.94-1.42; p = 0.180) **(Figure 13).**

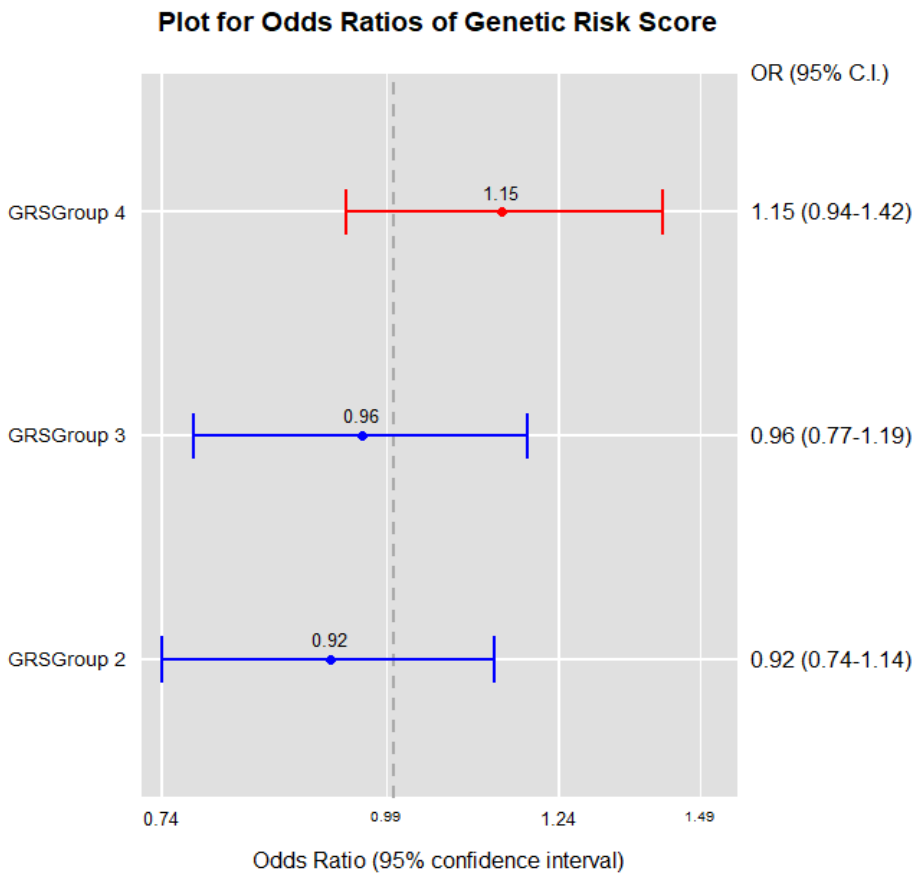

Fig 13. Odds ratios of genetic risk score and TC risk

3.5 Estimates of the effects of BMI on the risk of TC

The estimated causal odds ratio of BMI on risk of TC was 1.08 (95% CI, 0.87-1.35). The estimated gender-specific causal odds ratio of BMI on risk of thyroid cancer among male was 1.44 (95% CI, 0.75-2.75). For female, the estimated causal odds ratio of BMI on the risk of TC was 1.02 (95% CI, 0.82-1.25) (**Table 6**).

Table 6. Estimates of the effects of BMI on the risk of thyroid cancer using MR (OR per 1 kg/m^2)

All				Female				Male			
Control	**Case**	**OR [95% CI]**	**p value**	**Control**	**Case**	**OR [95% CI]**	**p value**	**Control**	**Case**	**OR [95% CI]**	**p value**
N=6,216	N=744	1.08 [0.87-1.35]	0.48	N=3,497	N=515	1.02 [0.82-1.25]	0.89	N=2,719	N=229	1.44 [0.75-2.75]	0.27

Age (years) was classified into three subgroups (younger than 40 yr, 40 to 55 yr, and older than 55 yr) and we found a significant association in 40 to 55 yr group and older than 55 yr group with the risk of thyroid cancer in both crude (OR 0.15, 95% CI 0.11-0.20 for 40 to 55 yr group; OR 0.05, 95% CI 0.04-0.07 for older than 55 yr group) and multivariable models (OR 0.14, 95% CI 0.11-0.19 for 40 to 55 yr group; OR 0.06, 95% CI 0.04-0.08 for older than 55 yr group), respectively. Female (adjusted OR 2.42; 95% CI 1.90-3.10) has a higher risk of thyroid cancer compared to male. Smoking status (adjusted OR 1.68; 95% CI 1.32-2.13) and regular exercise (adjusted OR 1.52; 95% CI 1.29-1.78) were associated with the increased risk of thyroid cancer in the total population, however there was no association between alcohol consumption and risk of thyroid cancer observed in the overall population (**Table 7**).

Table 7. Distributions of potential risk confounders and their associations with the risk for thyroid cancer

	Control (n=6,216)	Case (n=744)	Crude OR	Adjusted OR*
Age (years)				
< 40	101 (1.6%)	105 (14.1%)	1 (ref.)	1 (ref.)
40 - 55	3,031 (48.8%)	465 (62.5%)	**0.15 (0.11-0.20)**	**0.14 (0.11-0.19)**
> 55	3,084 (49.6%)	174 (23.4%)	**0.05 (0.04-0.07)**	**0.06 (0.04-0.08)**
Gender				
Male	2,719 (43.7%)	229 (30.8%)	1 (ref.)	1 (ref.)
Female	3,497 (56.3%)	515 (69.2%)	**1.75 (1.49-2.06)**	**2.42 (1.90-3.10)**
Alcohol consumption				
Nondrinker	2,953 (47.5%)	336 (45.2%)	1 (ref.)	1 (ref.)
Drinker	3,263 (52.5%)	408 (54.8%)	1.10 (0.94-1.28)	1.09 (0.91-1.30)
Smoking status				
Nonsmoker	4,117 (66.2%)	504 (67.7%)	1 (ref.)	1 (ref.)
Smoker	2,099 (33.8%)	240 (32.3%)	0.93 (0.79-1.10)	**1.68 (1.32-2.13)**
Regular exercise				
No	3,360 (54.1%)	322 (43.3%)	1 (ref.)	1 (ref.)
Yes	2,856 (45.9%)	422 (56.7%)	**1.54 (1.32-1.80)**	**1.52 (1.29-1.78)**

* Adjusted for gender, age, body mass index, alcohol consumption, smoking status, and regular exercise.

In male population, age was inversely associated with thyroid cancer risk while the regular exercise was positively associated with an increased risk of thyroid cancer (adjusted OR 1.85; 95% CI 1.38-2.50). No association was found regarding alcohol consumption and smoking status. In female population, smoking status (adjusted OR 2.43; 95% CI 1.78-3.29) and regular exercise (adjusted OR 1.41; 95% CI 1.16-1.71) were associated with increased risk of thyroid cancer. In contrast, an age for greater than 55 yr (adjusted OR 0.10; 95% CI 0.07-0.15) was inversely associated with thyroid cancer risks. No association was observed between alcohol consumption and thyroid cancer risk (**Table 8**).

Table 8. Distributions of potential confounders and their associations with the risk for thyroid cancer by gender

	Male				Female			
	Control (n=2,719)	**Case (n=229)**	**Crude OR**	**Adjusted OR***	**Control (n=3497)**	**Case (n=515)**	**Crude OR**	**Adjusted OR***
Age (years)								
< 40	34 (1.3%)	35 (15.3%)	1 (ref.)	1 (ref.)	67 (1.9%)	70 (13.6%)	1 (ref.)	1 (ref.)
40 – 55	1,147 (42.2%)	153 (66.8%)	**0.13 (0.08-0.21)**	**0.11 (0.06-0.19)**	1,884 (53.9%)	312 (60.6%)	**0.16 (0.11-0.23)**	**0.18 (0.13-0.26)**
> 55	1,538 (56.6%)	41 (17.9%)	**0.03 (0.01-0.05)**	**0.02 (0.01-0.04)**	1,546 (44.2%)	133 (25.8%)	**0.08 (0.06-0.12)**	**0.10 (0.07-0.15)**
Alcohol consumption								
Nondrinker	581 (21.4%)	35 (15.3%)	1 (ref.)	1 (ref.)	2,372 (67.8%)	301 (58.4%)	1 (ref.)	1 (ref.)
Drinker	2,138 (78.6%)	194 (84.7%)	**1.51 (1.05-2.22)**	1.11 (0.75-1.67)	1,125 (32.2%)	214 (41.6%)	**1.50 (1.24-1.81)**	1.10 (0.90-1.35)
Smoking status								
Nonsmoker	810 (29.8%)	65 (28.4%)	1 (ref.)	1 (ref.)	3,307 (94.6%)	439 (85.2%)	1 (ref.)	1 (ref.)
Smoker	1,909 (70.2%)	164 (71.6%)	1.07 (0.80-1.45)	1.10 (0.80-1.53)	190 (5.4%)	76 (14.8%)	**3.01 (2.26-3.99)**	**2.43 (1.78-3.29)**
Regular exercise								
No	1,423 (52.3%)	83 (36.2%)	1 (ref.)	1 (ref.)	1,937 (55.4%)	239 (46.4%)	1 (ref.)	1 (ref.)
Yes	1,296 (47.7%)	146 (63.8%)	**1.93 (1.46-2.56)**	**1.85 (1.38-2.50)**	1,560 (44.6%)	276 (53.6%)	**1.43 (1.19-1.73)**	**1.41 (1.16-1.71)**

* Adjusted for age, body mass index, alcohol consumption, smoking status, and regular exercise

BMI was classified into three subgroups (normal (<23), overweight (23-24.9), and obese (≥25)) and we found significant associations between highest quartile of BMI with the risk of thyroid cancer in multivariable models adjusted for gender, age, alcohol consumption, smoking status, and regular exercise. The BMI was significantly associated with an increased risk of thyroid cancer in total population in overweight (23-24.9) (adjusted OR 1.24; 95% CI 1.01-1.52) and obese (≥25) (adjusted OR 1.29; 95% CI 1.06-1.56) (**Table 9**).

Table 9. The association between thyroid cancer risk and BMI using multivariate logistic regression analysis

	All			
	Control (n=6,216)	Case (n=744)	Crude OR	Adjusted OR*
BMI (kg/m^2), n (%)				
Normal (<23)	2,323 (37.4%)	283 (38.0%)	1 (ref.)	1 (ref.)
Overweight (23-24.9)	1,644 (26.4%)	199 (26.7%)	0.99 (0.82-1.20)	**1.24 (1.01-1.52)**
Obese (≥25)	2,249 (36.2%)	262 (35.2%)	0.96 (0.80-1.14)	**1.29 (1.06-1.56)**

* Adjusted for age, body mass index, alcohol consumption, smoking status, and regular exercise

When stratifying by genders, BMI was significantly associated with an increased risk of thyroid cancer in both crude and multivariable in male population. In crude model, the risk of thyroid cancer is elevated in overweight (23-24.9) (adjusted OR 2.28; 95% CI 1.85-4.40) and obese (≥25) (adjusted OR 3.40; 95% CI 2.31-5.18). In an adjusted model, overweight (23-24.9) (adjusted OR 2.68; 95% CI 1.73-4.27), and obese (≥25) (adjusted OR 2.87; 95% CI 1.91-4.44) also led to the same effect. For female population, the significant association of BMI with a decreased risk of thyroid cancer was only found in crude model in overweight (23-24.9) (adjusted OR 0.79; 95% CI 0.63-1.00) and obese (≥25) (adjusted OR 0.65; 95% CI 0.52-0.82), this association did not remain after adjustment.

Table 10. The association between thyroid cancer risk and BMI using multivariate logistic regression analysis by gender.

	Male			
	Control (n=2,719)	Case (n=229)	Crude OR	Adjusted OR*
BMI (kg/m^2), n (%)				
Normal (<23)	901 (33.1%)	31 (13.5%)	1 (ref.)	1 (ref.)
Overweight (23-24.9)	742 (27.3%)	72 (31.4%)	**2.82 (1.85-4.40)**	**2.68 (1.73-4.27)**
Obese (≥25)	1,076 (39.6%)	126 (55.0%)	**3.40 (2.31-5.18)**	**2.87 (1.91-4.44)**
	Female			
	Control (n=3,497)	Case (n=515)	Crude OR	Adjusted OR*
BMI (kg/m^2), n (%)				
Normal (<23)	1,422 (40.7%)	252 (48.9%)	1 (ref.)	1 (ref.)
Overweight (23-24.9)	902 (25.8%)	127 (24.7%)	**0.79 (0.63-1.00)**	0.98 (0.77-1.25)
Obese (≥25)	1,173 (33.5%)	136 (26.4%)	**0.65 (0.52-0.82)**	0.90 (0.71-1.14)

* Adjusted for age, body mass index, alcohol consumption, smoking status, and regular exercise

4. Discussion

4.1 Summary of findings

We used the genomic variant data on BMI and thyroid cancer derived from National Cancer Center (NCC) and the Korean Genome Epidemiology Study (KoGES). In total, 744 cases from NCC and 6,216 healthy controls from both NCC and KoGES were selected. By combining a genetic variation of 55 BMI variants into robust genetic instruments and supporting the analyses with data on thyroid cancer risk, we found that BMI is not significantly related to the risk of thyroid cancer in the Korean population of both women and men. These findings propose that BMI is not a causal risk factor for thyroid cancer and yet the observational association is most possibly explained by reverse causation or confounders.

4.2 Observational studies on BMI and thyroid cancer risk

The trend of rising thyroid cancer incidence over the past few decades also coincides with the growing trend of obesity, but whether or how they are correlated is largely unknown. Prior to thyroid cancer disease, observational studies have shown contradictory results with interest to the relationship between BMI and thyroid cancer risk. A pooled analysis of prospective cohorts showed a summarized hazard ratio of 1.53 for thyroid cancer in obese men and women (BMI $\geq$ 30kg/m^2) [41]. A meta-analysis of 21 observational studies found that obesity is related to increased PTC risk and reduced medullary thyroid cancer risk [40]. Another meta-analysis

conclusively reported that, a 5 kg/m^2 increment in BMI was strongly related to the increased risk of thyroid cancer in men (RR, 1.33; 95% CI, 1.04-1.70) [132]. In the current study, it has been identified that there is a positive association between BMI and TC risk. An obese (OR, 1.29; 95% CI, 1.06-1.56) and an overweight (OR, 1.24; 95% CI, 1.01-1.52) have shown significant associations with TC risk (**Table 9**). Furthermore, a similar relationship was observed for the male (OR, 2.87; 95% CI, 1.91-4.44) (**Table 10**). However, there is an inverse association between BMI and TC risk for female even though it is not significant. Recently, a prospective cohort study from the Korean Cancer Prevention Study-II reported that obese male and female under 50 years old are better to be considered for the higher possibility of TC development [134]. It is important to note that genetic predisposition is one the of well-defined non-modifiable risk factors towards the TC risk. However, a case-control study based on genomic studies concluded that selected obesity-related genetic variants were not linked to PTC risk [133]. Some reasons can be explained by the phenomenon that obese people have an increased risk of thyroid cancer. First, elevated serum thyroid-stimulating hormone (TSH) levels are clinically associated with increased risk of malignancy in human thyroid nodules consequently, related to the late stage of the disease. Many observational studies in euthyroid subjects showed a positive correlation between BMI and serum TSH. In addition, TC patients have higher leptin levels than the healthy participants in a case-control study. Leptin was also indicated to intensify the migration of PTC cells. Furthermore, the well-known metabolic perturbation in obesity which is, insulin resistance may contribute to thyroid tumor growth. Insulin resistance can have effects with insulin straight

wrapper to insulin receptors or can stimulate insulin-like growth factors (IGFs), estrogen and other hormones such as TSH to improve the growth of thyroid carcinoma cells. Another reason of disagreement may be due to a specificity of study biases, or the difference of adjustment for confounder, and use of unusual cutoffs to determine the modifiable exposure [40]. Those controversies lead to popular issues in observational epidemiology, where associations are inclined to confounders and reverse causality. In this study, there is evidence that MR contribution to BMI is not related to the risk of thyroid cancer without the inherent limitations in observational studies. These results break fresh ground in the study of assessing the causality of the observed association between exposure and outcome.

4.3 Causal association using MR analysis

MR is an analytic approach to the use of genetic variants in non-experimental to assess causal inference between an exposure (risk factors) and an outcome (usually a disease). Research evidence together with GWAS has been published over the past years, which is easily accessible to conduct MR studies. According to our literature research, 546 studies have used MR analysis (as of December 2017). MR has been successfully applied to a wide range of epidemiological studies, causal effects of biomarkers on disease [134-138], estimating the causal effects of various behaviors [139-141]. The most typically studied exposures were adiposity measures including BMI, lipid quantity, and portion body fat (130 studies), C-reactive protein (81 studies), alcohol use (73 studies), and vitamin D levels (62 studies) (**Table 1**). Paucity of data with regard to the causal association between BMI and TC risk is one of the

limitations in these studies. However, in this study, we primarily focused on the MR analysis to estimate the causal association between BMI and TC risk.

There were several methods used for IV estimation. It has been identified that two-stage least square (2-SLS) regression and a Wald-type/ratio estimator as most commonly used methods from several other IV estimation strategies. For instance, the Wald-type/ratio estimator was used to assess the association between uric acid and ischaemic heart disease. The estimates were derived by first dividing the log of the hazard ratio for genotype-ischaemic heart disease by the genotype-exposure coefficient [136]. In another study, 2-SLS regression was used to estimate the causal effect of C-reactive protein (CRP) on blood pressure, pulse pressure, and hypertension. A point estimates was identified by the ratio of the coefficient for the regression of outcome on genotype and CRP on genotype [142]. Additionally, a study of causal effect of vitamin D on serum adiponectin used 2-SLS regression. The first stage is a linear regression of vitamin D on the instruments (genotype), which generates predicted values for vitamin D. The second stage is a linear regression of adiponectin on the predicted values [143]. In our study, we used Wald-type/ratio estimator to estimate the causal effect of BMI on the TC risk. The estimation of the magnitude of the causal effect was obtained from the linear regression coefficient of BMI on genetic variants and the logistic regression coefficient of TC risk on genetic variants, respectively. With a single IV, the 2-SLS regression estimate is the same as the Wald-type/ratio estimator. With multiple IVs, the 2-SLS regression can be considered as a weighted average of the Wald-type/ratio estimator calculated using the instruments one at the time. Several other methods of IV estimation were used for

IV estimation. As IV probit regression was used in associations effect of circulating CRP and cancer risk [144], generalized method of moments was used to obtain estimates of the causal association between BMI and blood pressure [145]. Also, a study based on the assessing causal effect between plasma CRP and chronic obstructive pulmonary disease considered generalized least squares regression [146]. Lastly, the limited information likelihood was used to estimate causal association of plasma HDL cholesterol with myocardial infarction [147]. Based on the complexity of methods, the simplest way to estimate the causal effect of the exposure on the outcome is via the Wald-type/ratio estimator, which was used in our study to estimate the causal effect of BMI on the TC risk.

With regard to the types of a genetic instrument used, there were three main types of genetic instrument such as single SNP, multiple SNPs and genetic risk score (GRS) or allele score. The study on causal relationship between adiposity and cardiometabolic used only the adiposity-associated variant rs9939609 at the *FTO* locus as an IV [148]. A study examining the association between *aldehyde dehydrogenase 2* (*ALDH2)* and risk of hypertension and level of blood pressure used a common polymorphism in *ALDH2* as an instrument [139]. In another study on the causal association of homocysteine level with the development of type 2 diabetes (T2B), the MR coefficient was estimated using *MTHFR* 677C > T as instrument [149]. Additionally, the single genetic variant rs10455872 has been used as an instrument to test whether the association between lipoprotein(a) levels and T2B is causal [150]. When we use single SNP as IV, the median bias of the Wald-type/ratio estimator or 2-SLS regression is insignificant for all but the weakest of IVs [151].

Similarly, when we use instrument independently which can explain variability in the exposure, use of multiple instruments can improve the precision of IV estimates. Firstly, the population stratification occurs when a sample is composed of a mixture of populations. Some genetic variants are potential candidates for using as an IVs in MR studies could have been influenced by population stratification. Secondly, pleiotropy refers to a single gene having multiple biological functions. In the context of MR analyses, SNPs in or near genes with pleiotropic effects that directly or indirectly influence the outcome other than through the exposure of interest violate the IV assumptions. Lastly, linkage disequilibrium (LD) is the correlation between allelic states at different loci on the same chromosome when assessed within a population. If SNPs are also in LD with a variant that affects the outcome of interest via a pathway that does not include the exposure of interest, the IV assumptions will be violated. Comparing IV estimates based on multiple genetic variants with independent effects on the exposure of interest provides an additional way to identify bias resulting from these issues. If IV estimates from different variants are similar, it is less plausible that LD or pleiotropy are present. [152]. Taking into consideration the studies which used multiple SNPs, one study found that causal roles of vitamin B_{12} and transcobalamin in prostate cancer. The set of SNPs in the B_{12}-related gene (*MTR, MTRR, FUT2, TCN2, TCN1, CUBN,* and *MUT*) were used as an IVs to estimate the association of vitamin B_{12} with prostate cancer [153]. Considering a study based on leukocyte telomere length and T2B risk in postmenopausal women, the IVs were selected separately for each racial/ethnic and in total six instruments for whites, four for blacks, seven for Hispanics, and two for Asians [154]. Moreover, the

rs2282679 and common *filaggrin* (*FLG*-a gene encoding a protein which provides skin hydration and photo protection) gene were used to examines the causal effect of vitamin D on serum adiponectin [143]. Furthermore, in the study examining the association of testosterone and cardiovascular risk factors (blood pressure, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and fasting glucose); three testosterone-related SNPs (rs10046, rs1008805, and rs1256031) were selected for IVs analysis [155].

In our study, we conducted to assess causal effects of BMI and TC risk, for the identification of multiple SNPs as IVs. We took advantage of published results from a GWAS of obesity (GIANT study) and identified 55 independent SNPs strongly associated with obesity. They selected SNPs based on genome-wide significance ($p<5\times10^{-8}$) (**Supplementary Material**). Genotypic effects on phenotypes are typically small, thus MR analysis requires very large sample sizes to obtain adequate power. When multiple instruments are used in the Wald estimator, the resulting IV estimate can be viewed as the efficient linear combination of the separate IV estimates. Provided that each instrument is valid, use of multiple instruments will increase the precision of the IV estimate compared with the separate IV estimates. However, inclusion of instruments that explain only a small proportion of the variability in the phenotype can increase finite sample bias of IV estimates.

The arising problem from including multiple IVs in an analysis is a weak instrument. A weak instrument is defined as an IV for which the statistical evidence of association with the exposure is not strong. The F-statistic from the regression of the exposure on the IV is usually cited as a measure of the strength of an instrument.

The value of F-statistic less than 10 is often taken to indicate a weak instrument [103]. Genetic risk score (GRS) or allele score is a single summarizing multiple genetic variant associated with a risk factor (exposure). Using a GRS as a single IV rather than each genetic variant as a separate IV can resolve the problem in IV estimation resulting from weak instrument. GRS has been constructed for many studies, including study influence blood pressure and cardiovascular disease risk; GRS was constructed based on 29 SNPs that significantly associated with Hg systolic and Hg diastolic [156]. Also, the combining of five SNPs rs780094 (*GCKR*), rs560887 (*G6PC2*), rs4607517 (*GCK*), rs13266634 (*SLC30A8*), and rs10830963 (*MTNR1B*) of fasting glucose to find the causal association between circulating glucose and carotid intima-media thickness [157]. Additionally, the study to examine the causal association of lipoprotein(a) (Lp(a)) with early atherosclerosis used GRS based on 10 Lp(a)-related SNPs as an instrument for Lp(a) levels [158]. In our study, a genetic score for BMI (BMI-GRS) was used to estimate the causal effect of BMI on TC risk. The BMI-GRS was computed from 55 BMI-associated SNPs and explained 1.24% of the variance in BMI. The F-value of 78 from the first state regression suggests that the GRS was a robust instrument for BMI. Although GRS can address a problem in weak instrument, more complex issues such as analyses with multiple risk factors are still unsolved using GRS and require the use of alternative methods.

4.4 Causal association between BMI and common diseases using MR analysis

Globally, obesity has raised as a major public health concern. Around 40-70% of inter-individual variability in BMI, commonly used to assess obesity, has been attributed to genetic factor [73]. It is a potential risk factor for several popular diseases that usually continues in uncertainty whether the risk factor is causal or whether the relationships are just prone to the effect of confounding. MR is a contemporary method that can be used to prove a causal relationship between a modifiable exposure and a disease (outcome) using an IV in observational studies. There were several studies to examine the causal associations between BMI and a variety of human general disorders using MR analysis.

Recently, a study which assesses the causal association of BMI and common diseases indicated that people whose BMI are 1 SD (SD=3.98 for BMI in European men) above the population mean will have 3.29 times increases the risk of T2D compared with the population prevalence [159]. Another study reported that increasing BMI is causally related to higher prevalence of asthma (OR, 1.009; 95% CI, 1.004-1.013), but not with hay fever (OR, 0.998; 95% CI, 0.994-1.002) or allergic sensitization (OR, 0.999; 95% CI, 0.986-1.012). For lung function, BMI was associated with decrease in forced expiratory volume in one-second (β, -0.0012; 95% CI, -0.0019-0.0006) and forced vital capacity (β, -0.0022; 95% CI, -0.0031-0.0014) [108]. In the study regarding breast cancer risk using MR analysis observed that an inverse association between BMI and breast cancer risk (OR per 5 kg/m^2, 0.65; 95%

CI, 0.56-0.75). The associations were similar in both premenopausal (OR per 5 kg/ m^2, 0.44; 95% CI, 0.31-0.62) and postmenopausal breast cancer (OR per 5 kg/m^2, 0.57; 95% CI, 0.46-0.71) [130]. Carreras-Torres *et al.* have shown a causal role of BMI on lung cancer risk for two of major histological subtypes [123]. It concluded that there are positive associations between BMI and small cell carcinoma (OR, 1.52; 95% CI, 1.15-2.00) and squamous cell carcinoma lung cancer (OR, 1.20; 95% CI, 1.01-1.43) [123]. Moreover, Holmes *et al* study observed causal effects of BMI on a range of cardiometabolic traits [160]. A 1 kg/m^2 genetically elevated BMI increased fasting glucose (0.18 mmol/l; 95% CI, 0.12-0.24), fasting insulin (8.5%; 95% CI, 5.9-11.1), interleukin-6 (7.0%; 95% CI, 4.0-10.1), and systolic blood pressure (0.70 mmHg; 95% CI, 0.24-1.16) and reduced high-density lipoprotein cholesterol (-0.02 mmol/l; 95% CI, -0.03 to -0.01) and low-density lipoprotein cholesterol (LDL-C, -0.04 mmol/l; 95% CI, -0.07 to -0.01). However, the causal effect of BMI on coronary heart disease risk is uncertain [160]. In addition, the study on causal association of obesity and multiple sclerosis showed 1 standard deviation increase in genetically determined BMI (kg/m^2) has increased odds of multiple sclerosis by 41% (OR, 1.41; 95% CI, 1.20-1.66) [127].

In our study, the causal estimate of BMI and TC risk results showed evidence that genetically influenced BMI was not causally associated with increased risk of TC (OR per 1 kg/m^2, 1.08; 95% CI, 0.87-1.35). In gender-specific IV analyses higher BMI was not associated with higher risk of TC among women (OR per 1 kg/m^2, 1.02; 95% CI, 0.82-1.25). For men, genetically influenced BMI was not associated with TC risk (OR per 1 kg/m^2, 1.02; 95% CI, 0.75-2.75). It is clear that the concept of MR

approach can be considered as an alternative for RCT where the findings make robust conclusions rather than other observational studies. For instance, several MR studies concluded that the BMI had risk effects on T2D (OR 3.29), hypertensive disease (OR 1.85), and cardiovascular disease (OR 1.30) which have been also confirmed by RCT [159, 161].

The most frequently used genetic instrument types in MR analysis are GRS and multiple SNPs or multiple instruments in which BMI is considered as an exposure (As described in the previous section). The GRS are used in MR for reasons of simplicity, increased power, and avoidance of weak instrument bias. Therefore, there were many studies using GRS for obesity to obtain causal inference. Skaaby *et al*. used GRS which was created using 26 BMI-associated SNPs to examine the causal effect of BMI on asthma, hay fever, allergic sensitization [108]. Also, in another study which focused on association of BMI and gastric cancer risk, weighted GRS (wGRS) was generated from 37 BMI-associated genetic variants as an IV [114]. Additionally, the causal role of obesity was examined in the development of depression using GRS comprised of 31 SNPs which were previously identified as robust genetic markers of body weight [162]. In this study, we developed wGRS, including 55 risk SNPs from the largest GWAS on BMI [73]. We aimed to investigate a causal role for elevated BMI and TC risk by using MR analysis. In our results, 55 SNPs were selected to construct a BMI-GRS from the largest meta-analysis of BMI-GWAS published to date [73]. To date, the number of BMI-associated SNPs represent a large, statistically significant, portion of the explained variance of observed BMI is quite substantial,

and F-statistic is considerably high. Nevertheless, the IV created in our study is sufficiently strong for conducting MR analyses.

4.5 Strengths and weaknesses

There are two main steps in order to conduct MR analysis, which is an exploration of the main three conditions and estimation of the causal effect of phenotypes and outcome. We tested these conditions using Chi-square test, linear regression, and logistic regression and found no evidence of violation of these conditions.

Our study has several strengths. Firstly, our study was conducted on a largest homogenous population. We took that advantage by using the summary statistic from largest genetic studies of obesity. To our knowledge, this is the first MR study of the association of BMI and TC risk and was undertaken in a large sample size. Moreover, the individual thyroid cancer data in Korea from both NCC and KoGES. It has more advantages than using the single dataset for analysis. Secondly, the number of SNPs is 55, which is quite large and means we extracted huge genetic information for the study. One important assumption in a successful MR study is that genetic instruments must be strongly associated with the exposure (here is BMI). Our results have shown that the F-statistic of the first-stage (linear regression of combining information of 55 SNPs) was large (78), which means a weak instrument bias was unlikely (a sufficiently strong instrument with F-statistic greater than 10). Additionally, there is a considerably high statistical power due to large sample size in the current study.

Last but not least, our study used powerful consortia dataset without measuring the parameter. It turns out saving both the time and cost.

The weakness of our study could be due to use of combined data from the different studies. The problems came up because we could not figure out whether the relationship between BMI and TC risk is linear or not. The literature research showed that the relationship between BMI and health problems was non-linear correlations. Our MR analysis is assumed that BMI with a continuous variable was linear, which is not examined with summary data. Another limitation of our study is that it does not provide an explanation of mechanisms by which BMI is causally associated with TC.

Finally, our study has provided best genetic evidence to date supplementing a causal association between BMI and TC risk. It is not conceivable based on the existing data to completely oppose the hypothesis.

5. Conclusion

In conclusion, combining significant genetic variation of BMI genes into strong genetic instruments and extending these investigations with data on TC risk of 55 BMI variants from National Cancer Center (NCC) and the Korean Genome Epidemiology Study (KoGES), we found that BMI is not related to the risk of TC in the Korean population.

Although BMI is not a potentially modifiable risk factor for TC, obesity is associated with a variety of health issues that are of clinical and public health significance and should be taken into account. The strong evidence showed that being overweight or obese enhances the risk for more than 10 types of cancer such as breast

cancer, kidney cancer, esophageal, and colorectal cancer. Our findings of the causal effects of BMI and TC risk could have a significant impact on several academic fields including medical research, the pharmaceutical industry, and public health. It can be suggested that further studies to replicate this association would be required considering insight into mechanisms before giving some recommendation to the general population.

Supplementary Material

Selecting of BMI variants

Selecting process was initiated by using the summary statistics dataset from the website of the GIANT consortium. (https://www.broadinstitute.org/collaboration/giant/index.php/Main_Page). We used the combined gender, All Ancestries dataset file.

In order to choose the index SNPs were associated with BMI, the significance level of the genome-wide is $5x10^{-8}$ was set for a cut-off value. For making sure that each of SNPs is independent of each other we were used the two criteria which are Linkage Disequilibrium (LD) with the R^2 threshold of 0.001 and physical distance threshold within 10,000 kilobases. All of the selected index SNPs were ranked based on p-value ($p=5x10^{-8}$) from smallest to largest. The reference dataset for checking independent of index SNPs was used "1000 Genomes Project" (http://www.1000genomes.org/). The officially PLINK website provides the base code for selecting the SNPs (http://zzz.bwh.harvard.edu/plink/clump.shtml)

The selecting parameters were as follows:

Significance thresholds for index SNPs (p-value)	5×10^{-8}
Secondary significance threshold for clumped SNPs (p-value)	5×10^{-8}
LD threshold for clumping (R^2)	0.001
Physical distance threshold for clumping (kilobases)	10000

Supplementary Table 1. List of selected SNPs, nearest gene and their function

SNPs	Genes	Function of nearest gene
rs1000940	*RABEP1*	Protein homodimerization activity and growth factor activity
rs10132280	*STXBP6*	Phosphatidylinositol-4,5-bisphosphate binding and GTP-Rho binding
rs10733682	*LMX1B*	Transcription factor activity, sequence-specific DNA binding and sequence-specific DNA binding
rs10840100	*TRIM66*	Protein homodimerization activity and chromatin binding
rs11030104	*BDNF*	Receptor binding and neurotrophin TRKB receptor binding
rs11165643	*PTBP2*	Nucleic acid binding and RNA binding
rs11191511	*CNNM2*	Adenyl nucleotide binding
rs11692326	*None*	Intergenic
rs12286929	*CADM1*	Protein homodimerization activity and PDZ domain binding
rs12429545	*OLFM4*	Protein homodimerization activity and cadherin binding
rs12940622	*RPTOR*	Protein complex binding
rs13021737	*TMEM18*	No further details
rs13025697	*None*	Intergenic
rs13130484	*GNPDA2*	Hydrolase activity and glucosamine-6-phosphate deaminase activity
rs13201877	*IFNGR1*	Cytokine receptor activity and interferon-gamma receptor activity.
rs13329567	*None*	Intergenic
rs1421085	*FTO*	Errous iron binding and oxidative RNA demethylase activity.
rs1441264	*MIR548A2*	RNA gene
rs1460676	*FIGN*	Protein C-terminus binding and microtubule-severing ATPase activity
rs14810	*KCTD15*	Protein coding
rs1516725	*ETV5*	Transcription factor activity, sequence-specific DNA binding and transcription regulatory region DNA binding

rs1528435	*UBE2E3*	Ligase activity and acid-amino acid ligase activity.
rs16851483	*RASA2*	GTPase activator activity
rs17094222	*HIF1AN*	Protein homodimerization activity and oxidoreductase activity
rs17724992	*PGPEP1*	Cysteine-type peptidase activity and pyroglutamyl-peptidase activity.
rs1928295	*TLR4*	Receptor activity and lipopolysaccharide binding
rs2060604	*None*	Intergenic
rs2112347	*HMGCR*	Protein homodimerization activity and NADP binding
rs2176598	*HSD17B12*	Oxidoreductase activity and collagen binding
rs2229616	*MC4R*	G-protein coupled receptor activity and peptide hormone binding
rs2365389	*FHIT*	Identical protein binding and hydrolase activity.
rs2820292	*NAV1*	Protein coding
rs2836754	*ETS2*	Transcription factor activity, sequence-specific DNA binding and RNA polymerase II core promoter proximal region sequence-specific DNA binding
rs3800229	*FOXO3*	Transcription factor activity, sequence-specific DNA binding and protein kinase binding
rs3817334	*MTCH2*	Protein coding
rs3888190	*ATP2A1*	Calcium ion binding and nucleotide binding.
rs4615388	*None*	Intergenic
rs4740619	*CCDC171*	Transcription factor activity, sequence-specific DNA binding and signal transducer activity.
rs543874	*SEC16B*	Protein coding
rs6091540	*ZFP64*	Nucleic acid binding
rs6457796	*UHRF1BP1*	Identical protein binding and histone deacetylase binding.
rs6477694	*EPB41L4B*	Structural constituent of cytoskeleton and cytoskeletal protein binding.
rs6567160	*MC4R*	G-protein coupled receptor activity and peptide hormone binding
rs657452	*AGBL4*	Tubulin binding and metallocarboxypeptidase activity

rs6804842	*RARB*	Transcription factor activity, sequence-specific DNA binding and protein complex binding.
rs7138803	*FAIM2*	Protein coding
rs7243357	*GRP*	Receptor binding and neuropeptide hormone activity.
rs7599312	*ERBB4*	Protein homodimerization activity and protein kinase activity.
rs7715256	*GALNT10*	Carbohydrate binding and polypeptide N-acetylgalactosaminyltransferase activity
rs7903146	*TCF7L2*	Transcription factor activity, sequence-specific DNA binding and chromatin binding
rs891389	*NPC1*	Receptor activity and cholesterol binding
rs9374842	*LOC285762*	RNA gene
rs943005	*None*	Intergenic
rs9540493	*MIR548X2*	RNA gene
rs9579083	*MTIF3*	Translation initiation factor activity and ribosomal small subunit binding.

Variants are listed here with the nearest genes and putative function of that gene listed by http://www.genecards.org/, https://www.infino.me/, https://www.selfdecode.com/, and https://www.ncbi.nlm.nih.gov/snp/. Just because a gene is near to a given variant, it does not mean that the variant exerts its effect on thyroid cancer via BMI because of that gene.

Appendix

Appendix 1. Structured General Questionnaires for Screenee Cohort

검진자 코호트 설문지

우리나라 사망원인의 1위는 암입니다. 전체 사망자 중 약 27%가 암으로 인한 사망자이며, 매년 많은 수의 새로운 암 환자가 발생하고 있습니다. 암은 발생 이후에 치료하는 것 보다는 발생 이전에 미리 예방하는 것이 환자와 가족의 고통 및 부담을 줄일 수 있는 방법입니다. 암 예방을 위해서는 그 원인을 알고 이를 관리하는 것이 중요한데, 현재 국립암센터에서는 이러한 암 예방을 위한 진료, 연구, 암 관리사업 등이 진행 중에 있습니다. 이와 관련하여 본 센터에서는 건강검진을 하기 위해서 내원하시는 분들을 대상으로 심층적인 건강 설문조사를 수행하여 암 검진 진료 시 도움이 되는 기초 자료로 활용하고, 향후 암 발생 여부를 조사함으로써 암과 위험요인 간의 원인적 연관성 관계를 밝히고자 합니다.

본 설문의 내용은 별도로 작성하신 동의서 내용에 따라 활용되며 개인 신상에 관한 정보는 철저히 비밀이 유지됨을 알려드립니다. 설문지에 포함된 각 문항들을 자세히 검토하신 후 정확하게 답변해 주셔서, 귀하의 건강상태를 정확히 파악하고 암 위험요인을 규명하는데 중요한 자료로 사용될 수 있도록 협조 부탁드립니다.

등 록 번 호		방 문 일 자	
성 명		주민등록번호	-
휴 대 전 화		실 제 생 년 월 일	____년 ___월 ___일 (양력 ○ 음력 ○)
전 화 번 호 (실제 거주지)	직장 ( ) - 집 ( ) -		
주 소 (실제거주지)	(우편번호 : -)		
조 사 일 시			

국립암센터 암예방검진센터
전화:031-920-1313

Ⅰ. 건강검진 관련

1. 지금까지 병, 의원에서 다음과 같은 검사나 시술을 받은 경험이 있으시면 해당항목에 표시해 주십시오.

검진여부		검사명	검진이유		평생동안 받은 총 횟수	검진결과		처음 검사 받은 연도	마지막검사 받은 연도
있다	없다		이상증상	정기검진		정상	정상 이외의 결과		
○	○	위내시경	○	○	_____회	○	○위염 ○궤양 ○식도염 ○암 ○용종 ○기타( )		
○	○	위조영술 (상부위장관조영)	○	○	_____회	○			
○	○	대장내시경	○	○	_____회	○			
○	○	대장조영술	○	○	_____회	○			
○	○	대변잠혈검사	○	○	_____회	○			
○	○	직장진찰	○	○	_____회	○			
○	○	복부 초음파	○	○	_____회	○			
○	○	알파태아단백검사	○	○	_____회	○			
○	○	전립선특이항원 (PSA)	○	○	_____회	○			
○	○	전립선생검	○	○	_____회	○			
○	○	전산화단층촬영 (CT)	○	○	_____회	○	촬영부위		
○	○	자기공명영상촬영 (MRI)	○	○	_____회	○	촬영부위		
○	○	양전자단층촬영 (PET)	○	○	_____회	○	촬영부위		
○	○	갑상선혈액검사	○	○	_____회	○	○기능저하 ○기능항진 ○기타( )		
○	○	갑상선초음파	○	○	_____회	○	○결절 ○암 ○기타( )		
○	○	유방진찰	○	○	_____회	○			
○	○	유방촬영술	○	○	_____회	○			
○	○	유방초음파	○	○	_____회	○			
○	○	자궁경부암검사	○	○	_____회	○			
○	○	산부인과초음파	○	○	_____회	○			
○	○	산부인과골반진찰	○	○	_____회	○			
○	○	인유두종바이러스 (HPV)	○	○	_____회	○	○비정상(양성) ○결과 모름 ○이전 양성에서 음성으로 됨		

Ⅱ. 과거병력 관련

2. 혹시 귀하께서는 병원에서 암을 진단받으신 적이 있습니까?

○ 있다 ○ 없다 (3번 문항으로)

(2-1) 있으셨다면 어떤 암인지 기입해주시고, 처음 진단받은 연도를 표시해 주십시오.
(여러 개의 암일 경우는 모두 적어 주세요)

암종류	진단연도	치료받은 방법
		○ 악물오법 ○ 수술오법 ○ 면역(유전자)오법 ○ 골수이식 ○ 내시경적 절제술 ○ 방사선오법 ○ 대체오법 ○ 기타( ) ○ 없음
		○ 악물오법 ○ 수술오법 ○ 면역(유전자)오법 ○ 골수이식 ○ 내시경적 절제술 ○ 방사선오법 ○ 대체오법 ○ 기타( ) ○ 없음

3. 과거 암 이외에 ***의사도부터 진단받으셨던*** 질병 및 검사결과 이상이 있습니까?

○ 있다 ○ 없다 (4번 문항으로)

(3-1) 있으셨다면 다음 중 해당 질병을 아래 표를 참고하여 표시해주시고, 처음 진단받은 연도와 현재 치료여부를 모두 표시해 주십시오.

있음	질병명	완치	현재 치료 중	치료 받은 적 있으나 현재 치료하지 않음	치료 받은 적 없음	처음 진단연도
○	고혈압	○	○	○	○	
○	당뇨병	○	○	○	○	
○	고지혈증	○	○	○	○	
○	심근경색	○	○	○	○	
○	협심증	○	○	○	○	
○	부정맥, 판막질환	○	○	○	○	
○	뇌졸중(중풍)	○	○	○	○	
○	위염	○	○	○	○	
○	위궤양	○	○	○	○	
○	십이지장궤양	○	○	○	○	
○	식도염	○	○	○	○	
○	바렛식도	○	○	○	○	
○	과민성 장증후군	○	○	○	○	
○	대장용종(폴립)	○	○	○	○	
○	베쳇병, 염증성 장질환	○	○	○	○	
○	치질	○	○	○	○	
○	간기능이상	○	○	○	○	
○	지방간	○	○	○	○	

있음	질병명	완치	현재 치료 중	치료 받은 적 있으나 현재 치료하지 않음	치료 받은 적 없음	처음 진단연도
○	B형 간염	○	○	○	○	
○	C형 간염	○	○	○	○	
○	간경화	○	○	○	○	
○	담석	○	○	○	○	
○	담낭염	○	○	○	○	
○	알러지(아토피)	○	○	○	○	
○	천식, 기관지염	○	○	○	○	
○	폐결핵	○	○	○	○	
○	갑상선기능저하증	○	○	○	○	
○	갑상선기능항진증	○	○	○	○	
○	갑상선 결절	○	○	○	○	
○	신장/요로결석	○	○	○	○	
○	전립선비대증	○	○	○	○	
○	골관절염	○	○	○	○	
○	류마티스 관절염	○	○	○	○	
○	허리/목디스크	○	○	○	○	
○	골다공증	○	○	○	○	
○	외상사고	○	○	○	○	
○	파킨슨병	○	○	○	○	
○	루푸스	○	○	○	○	
○	중증근무력증	○	○	○	○	
○	우울증	○	○	○	○	
○	기타( )	○	○	○	○	

Ⅲ. 수술력 관련

4. 귀하께서는 과거에 수술이나 치료에 관련된 시술을 받으신 적이 있습니까?

○ 있다　　○ 없다 (5번 문항으로)

(4-1) 있으셨다면 어떤 수술을 받으셨습니까? 해당부위에 모두 표시해 주십시오.

수술부위	수술/시술명	해당연도	수술부위	수술/시술명	해당연도
○ 심장			○ 대장		
○ 뇌			○ 맹장(충수돌기)		
○ 위(내시경시술 포함)			○ 치질		
○ 간			○ 자궁(소파수술제외)		
○ 담낭			○ 난소		
○ 신장/방광			○ 유방(조직검사포함)		
○ 췌장			○ 기타 ( )		

Ⅳ. 약물 복용력 관련

5. ***최근 2년간*** 다음의 약물이나 영양제를 정기적으로 복용하고 계십니까?

○ 예 (만약 '예'라면 아래에 자세히 기입해 주십시오)
○ 아니오
○ 모름

종류	총 복용량					총 복용기간
	주 1-3알	주 4-6알	매일 1알	매일 2알	매일 3알	
혈압약	○	○	○	○	○	____년 ____개월
심장약	○	○	○	○	○	____년 ____개월
아스피린	○	○	○	○	○	____년 ____개월
항혈소판제제	○	○	○	○	○	____년 ____개월
항응고제	○	○	○	○	○	____년 ____개월
당뇨약	○	○	○	○	○	____년 ____개월
고지혈증약	○	○	○	○	○	____년 ____개월
소화제	○	○	○	○	○	____년 ____개월
궤양치료제	○	○	○	○	○	____년 ____개월
골다공증치료제	○	○	○	○	○	____년 ____개월
항우울제	○	○	○	○	○	____년 ____개월
진정수면제	○	○	○	○	○	____년 ____개월
진통소염제	○	○	○	○	○	____년 ____개월
기타영양제/보조식품	○	○	○	○	○	____년 ____개월
진통제	○	○	○	○	○	____년 ____개월
칼슘제	○	○	○	○	○	____년 ____개월
종합비타민제 (이름: _____)	○	○	○	○	○	____년 ____개월
단일비타민제 (이름: _____)	○	○	○	○	○	____년 ____개월
기타1 (_____)	○	○	○	○	○	____년 ____개월
기타2 (_____)	○	○	○	○	○	____년 ____개월

종류	총 복용기간	종류	총 복용기간
한약	____년 ____개월	민간요법	____년 ____개월

(5-1) 귀하께서 그 이외의 건강 보조식품으로서 규칙적으로 복용하시고 있는 것을 아래에서 표시해 주십시오.

○ 철분제 ○ 클로렐라 ○ 스피루리나 ○ 효모
○ 글루코사민 ○ 감마리놀렌산 ○ 레시틴 ○ 키토산
○ 홍삼 ○ 오메가-3 ○ DHA/EPA

V. 가족력 관련

6. 가족이나 친척 중에 암을 진단 받으신 분이 계십니까?

○ 있다 (있으시다면 누가, 무슨 암을 진단 받았는지 모두 기입해 주십시오)
○ 없다

암명	본인과의 관계								
	부	모	남자형제	여자형제	자녀(아들)	자녀(딸)	사촌	할아버지	할머니
	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○

7. 귀하의 직계가족께서 의사로부터 암 이외의 다음과 같은 질환을 진단 받으신 분이 있습니까?

○ 있다 (있으시다면 아래에 자세히 기입해 주십시오)
○ 없다
○ 모름

질병명	본인과의 관계					
	부	모	남자형제	여자형제	자녀(아들)	자녀(딸)
	○	○	○	○	○	○
	○	○	○	○	○	○
	○	○	○	○	○	○
	○	○	○	○	○	○
	○	○	○	○	○	○

<비고> 암 이외의 기타 질환 종류			
01. 고혈압	12. 바렛식도	23. 담낭염	34. 허리/목디스크
02. 당뇨병	13. 과민성 장증후군	24. 알러지(아토피)	35. 골다공증
03. 고지혈증	14. 대장용종(폴립)	25. 천식, 기관지염	36. 외상사고
04. 심근경색	15. 베쳇병, 염증성장질환	26. 폐결핵	37. 파킨슨병
05. 협심증	16. 치질	27. 갑상선기능저하증	38. 루푸스
06. 부정맥, 판막질환	17. 간기능이상	28. 갑상선기능항진증	39. 중증근무력증
07. 뇌졸중(중풍)	18. 지방간	29. 갑상선 결절	40. 우울증
08. 위염	19. B형 간염	30. 신장/오로결석	41. 기타 ( )
09. 위궤양	20. C형 간염	31. 전립선비대증	
10. 십이지장궤양	21. 간경화	32. 골관절염	
11. 식도염	22. 담석	33. 류마티스관절염	

VI. 간염 및 헬리코박터균 감염 여부

8. B형 간염 보균자이신 경우에만 대답해 주십시오.

(8-1) 항바이러스제제를 복용한 적이 있습니까?

- ○ 현재 복용 중이다
- ○ 과거에 복용한 적이 있다
- ○ 치료받은 적 없다

(8-2) 귀하께서는 B형 간염에 대해 항바이러스 치료를 받으신 적이 있습니까? 있다면 어떤 종류의 약입니까?

- ○ 없다
- ○ 있다 - 인터페론 (기간 :)
- ○ 있다 - 제픽스 (기간 :)
- ○ 있다 - 헵세라 (기간 :)
- ○ 있다 - 그 외 (기간 :)
- ○ 잘 모름

(8-3) 귀하께서 간염 항원 보유자라면 현재 간에 대한 정기적인 검사를 받고 계십니까?

정기검사 여부	검사방법(모두 선택)
○ 6개월마다 시행	○ 간초음파
○ 1년마다 시행	○ 혈청 알파태아단백
○ 1~2년마다 시행	○ 간기능 혈액검사
○ 불규칙함	○ 기타
○ 거의 따로 하지 않음	

9. 귀하는 과거에 헬리코박터균의 검사를 받으신 적이 있으십니까?

○ 예 ○ 아니오 ○ 모름

(9-1) "예"인 경우 검사 결과는 어떠하였습니까?

- ○ 양성(균이 있음)
- ○ 음성(균이 없음)
- ○ 모름

(9-2) 헬리코박터균 검사가 양성이었다면 치료(제균치료)를 받으신 적이 있습니까?

- ○ 치료받은 적 없음
- ○ 치료 받고 음성 판정 받음
- ○ 치료받았으나 음성으로 되었는지 확인하지 않음
- ○ 잘 모름

10. 귀하께서는 과거에 수혈(피주사)을 받으신 적이 있습니까? 있으시면, 수혈을 받으신 횟수와 처음 수혈 당시 나이를 표시해 주십시오.

- ○ 있다 (총 _____회, 만 _____세)
- ○ 없다
- ○ 모름

11. 귀하께서는 과거에 침을 맞아보신 적이 있습니까?

○ 있다 ○ 없다 ○ 모름

VII. 흡연, 음주관련

12. 귀하는 다음 보기 중 어디에 해당하십니까?

- ○ 현재 흡연 한다.
- ○ 과거 흡연했으나 금연한지 1년 이내이다.
- ○ 과거 흡연했으나 금연한지 1년이 넘었다.
- ○ 담배를 피운 적이 없다.

(12-1) 피우기 시작한 나이와 지금까지 피우신 총 기간, 흡연 시 하루에 대략 피우신 양을 표시해 주십시오.

흡연시작 만 나이	만 _____세
하루에 피우는 양	_____개피
지금까지 피운 총 기간	_____년 _____개월

(12-2) 끊으셨다면, 담배를 끊으신 만 나이를 적어 주십시오. 만 _____세

13. 집안에서 다른 사람이 피우는 담배 연기를 맡는 경우가 얼마나 되십니까?

- ○ 없음
- ○ 주 1~2회
- ○ 주 3~4회
- ○ 주 5~6회
- ○ 매일
- ○ 모름

14. 술을 마신 적이 있습니까? 마신 적이 있으시다면 지금까지 총 몇 년간 드셨는지 표시해 주십시오.

○ 예, 지금도 마십니다 (총 _____년)
○ 예, 하지만 지금은 끊었습니다 (_____년 ____개월 동안 마시고, _____년 ____개월 전에 끊었다)
○ 아니오, 원래 안 마십니다

15. 술 종류별로 1년 동안에 드신 평균 횟수와 한번 드실 때의 양을 표시하여 주십시오.
(한번 드실 때의 평균 몇 잔을 드시는지 해당 잔의 cc를 참고해서 표시해 주십시오)

술종류	없다	지난 1년간 섭취한 평균횟수							한번 드실 때의 평균 잔 수
		월1회	월2~3회	주1회	주2~3회	주4~6회	매일1회	매일2회 이상	
맥주	○	○	○	○	○	○	○	○	_____맥주잔(200cc)
소주	○	○	○	○	○	○	○	○	_____소주잔(50cc)
양주	○	○	○	○	○	○	○	○	_____양주잔(30cc)
막걸리	○	○	○	○	○	○	○	○	_____막걸리잔(240cc)
포도주	○	○	○	○	○	○	○	○	_____포도주잔(90cc)
과실주	○	○	○	○	○	○	○	○	_____소주잔(50cc)
기타 ( )	○	○	○	○	○	○	○	○	_____잔(_____cc)

Ⅷ. 신체활동 관련

귀하께서는 지난 7일간 하신 모든 격렬한 활동을 생각해 보십시오. 격렬한 신체활동이란 힘들게 움직이는 활동으로서 평상보다 숨이 훨씬 더 차게 만드는 활동입니다. 한번에 적어도 10분 이상 지속한 활동만을 생각하여 응답해주시기 바랍니다.

16. 지난 7일간 무거운 물건 나르기, 달리기, 에어로빅, 빠른 속도로 자전거 타기 등과 같은 격렬한 신체 활동을 며칠간 하였습니까?

일주일에 _____일 ○ 격렬한 신체활동 없었음

(16-1) 그런 날 중 하루에 격렬한 신체활동을 하면서 보낸 시간이 보통 얼마나 됩니까?

하루 중 ____시간 _____분 ○ 모름

귀하께서는 지난 7일간 하신 모든 중간정도 신체활동을 생각해 보십시오. 중간정도 신체활동이란 중간정도 힘들게 움직이는 활동으로서 평소보다 숨이 조금 더 차게 만드는 활동입니다. 한번에 적어도 10분 이상 지속한 활동만을 생각하여 응답해주시기 바랍니다.

17. 지난 7일간 가벼운 물건 나르기, 보통 속도로 자전거 타기, 복식 테니스 등과 같은 중간정도 신체 활동을 며칠간 하였습니까? 걷기는 포함시키지 마십시오.

일주일에 _____일 ○ 중간정도 신체활동 없었음

(17-1) 그런 날 중 하루에 중간정도의 신체활동을 하며 보낸 시간이 보통 얼마나 됩니까?

하루 중 ____시간 _____분 ○ 모름

지난 7일간 걸은 시간을 생각해 보십시오. 직장이나 집에서, 교통수단을 이용할 때 걸은 것뿐만 아니라 오락활동, 스포츠, 운동, 여가 시간에 걸은 것도 포함됩니다.

18. 귀하께서 지난 7일 동안 한 번에 최소한 10분 이상 걸은 날이 며칠이나 됩니까?

일주일에 _____일 ○ 걷지 않았음

(18-1) 그런 날 중 하루에 걸으면서 보낸 시간이 보통 얼마나 됩니까?

하루 중 ____시간 _____분 ○ 모름

마지막 질문은 지난 7일간 주중에 앉아서 보낸 시간에 관한 것입니다. 여기에는 직장과 집에서 학업이나 여가시간에 앉아서 보낸 시간이 포함됩니다. 또한 책상에 앉아 있거나, 친구를 만나거나 독서할 때 앉거나, 텔레비전을 앉아서 또는 누워서 시청한 시간이 포함됩니다.

19. 지난 7일 동안 평일에 앉아서 지낸 시간이 얼마나 됩니까?

하루 중 ____시간 _____분 ○ 모름

20. 평소에 규칙적으로 하는 운동이 있습니까?

○ 아니오 ○ 예

(20-1) 있으시다면, 가장 자주 하시는 운동순으로 적어 주세요.

	운동 종류	***일주일***에 평균 몇 시간
1		____시간 ____분 ○ 모름
2		____시간 ____분 ○ 모름
3		____시간 ____분 ○ 모름

IX. 식습관

21. 일주일에 5일 이상은 하루에 3번 식사를 한다.

○ 예 ○ 아니오

22. 식사시간은 평균 10분 이상이다.

○ 예 ○ 아니오

23. 육류나 계란 중 1가지를 일주일에 5번 이상 먹는다.

○ 예 ○ 아니오

○ 육류는 탁구공 2개 분량 이하 또는 계란 1개 이하
○ 육류는 탁구공 2개 분량 이상 또는 계란 1개 이상

24. 어패류(생선, 오징어, 조개 등) 일주일에 3번 이상 먹는다.

○ 예 ○ 아니오

25. 두부나 두유를 일주일에 3번 이상 먹는다.

○ 예 ○ 아니오

26. 김치를 제외한 채소류, 해조류, 버섯 등을 매끼 먹는다.

○ 예 ○ 아니오

27. 과일은 일주일에 5일 이상 먹는다.

○ 예 ○ 아니오

28. 우유나 유제품(요구르트, 요거트, 치즈 등)을 일주일에 5일 이상 먹는다.

○ 예 ○ 아니오

29. 식사 때 국과 김치를 제외한 3가지 이상 반찬을 먹는다.

○ 예 ○ 아니오

30. 외식할 때 음식이 짜다고 느낀다.

○ 예 ○ 아니오

31. 음식을 짜게 드십니까?

○ 예(짜게) ○ 보통 ○ 아니오(싱겁게)

32. 만약 굽다 탄 고기가 있으면 그 고기를 드십니까?

○ 먹지 않는다
○ 가끔 먹는다
○ 먹을 때가 많다
○ 해당 없음(고기 안 먹음)

X. 일반적 사항 관련

33. 귀하의 키/체중에 대해서 여쭙겠습니다.

(33-1) 현재의 몸무게와 키는 어느 정도입니까?

_____kg _____cm

(33-2) 2년 전의 몸무게는?

_____kg ○ 모름

(33-3) 만 35세 무렵의 몸무게는?

_____kg ○ 모름

(33-4) 만 18세(고등학교 졸업 무렵)때는?

_____kg ○ 모름

34. 귀하의 결혼상태는 어디에 해당하십니까?

○ 미혼 ○ 기혼 ○ 별거
○ 사별 ○ 동거 ○ 이혼 ○ 기타

35. 귀하의 학력은 어떻게 되십니까?

○ 학교에 다니지 않았다
○ 초등학교 중퇴
○ 초등학교 졸업 또는 중학교 중퇴
○ 중학교 졸업 또는 고등학교 중퇴
○ 고등학교 졸업
○ 기술(전문)학교 졸업
○ 대학교 중퇴
○ 대학교 졸업
○ 대학원 이상
○ 모름

36. 귀 가정의 월 평균 총 수입은 어느 정도 입니까?

○ 100만원 미만 ○ 100－199만원
○ 200－399만원 ○ 400－699만원
○ 700만원 이상 ○ 기타( )
○ 잘 모름

37. 귀하는 현재 어떤 일에 종사하고 계십니까?

○ 의회의원, 고위임직원 및 관리직	의회의원 및 고위임원, 행정 및 경영 관리자, 일반관리자 등
○ 전문가	과학전문가, 컴퓨터관련 전문가, 공학전문가, 보건의료 전문가, 교육전문가, 행정/경영 및 재정 전문가, 법률/사회서비스 및 종교전문가, 문화/예술 및 방송관련 전문가 등
○ 기술공 및 준전문가	과학관련 기술종사자, 컴퓨터관련 준전문가, 공학관련 기술종사자, 보건의료 준전문가, 교육 준전문가, 경영 및 재정 준전문가, 사회서비스 및 종교 준전문가, 예술/연예 및 경기 준전문가, 기타 준전문가 등
○ 사무직	일반사무 관련 종사자, 고객서비스 사무 종사자 등
○ 서비스 종사자	대인서비스 관련 종사자, 조리 및 음식 서비스 종사자, 여행 및 운송 관련 종사자, 서비스 종사자 등
○ 판매 종사자	도소매 판매 종사자, 통신 판매 종사자, 모델 및 홍보 종사자, 보험 및 부동산 등
○ 농업, 임업 및 어업 숙련 종사자	농업 숙련종사자, 임업 숙련종사자, 어업 숙련종사자 등
○ 기능원 및 관련기능 종사자	추출 및 건설 기능종사자, 금속, 기계 및 관련 기능 종사자, 기계설치 및 정비 기능 종사자, 정밀기구/세공 및 수공예 종사자, 기타 기능원 및 관련기능 종사자 등
○ 장치, 기계조작 및 조립 종사자	고정기계장치 및 시스템 조작 종사자, 기계 조작원 및 관련 종사자, 조립 종사자, 운전원 및 관련 종사자 등
○ 단순노무 종사자	서비스 관련 단순노무 종사자, 농림어업 관련 단순노무 종사자, 제조관련 단순노무 종사자, 광업/건설 및 운송 관련 단순노무 종사자 등
○ 군인	
○ 주부 및 가사종사자	
○ 무직	
○ 기타( )	

※ 여성분만 응답해 주세요

38. 초경나이와 규칙적으로 생리를 시작한 나이를 만나이로 표시해 주십시오.

초경 연령	만 _____세	○ 모름 ○ 아직 생리가 없음
규칙적으로 생리를 시작한 연령	만 _____세	

****폐경되신 분들도 표시해 주십시오.***

(38-1) 생리 주기가 규칙적인 편입니까?

○ 예 (평균 _____일, 예시: 28일)
○ 아니오

(38-2) 생리하는 기간은 일정합니까?

○ 예 (평균 _____일, 예시: 7일)
○ 아니오

(38-3) 마지막 생리 시작 일을 표시해 주십시오.

_____월	_____일	○ 폐경

39. ***폐경이 되신 분만*** 대답해 주십시오.

(39-1) 폐경이 되신 연령과 폐경이 된 사유를 표시해 주십시오.

폐경 연령	폐경이 된 사유
만_____세	○ 자연폐경 (나이가 들어서) ○ 수술 (자궁/난소적출술) ○ 방사선치료 ○ 약물치료 ○ 기타 _____

(39-2) 폐경 이후 여성호르몬 약의 사용여부와 사용하신 경우에는 사용기간을 표시해 주십시오.

폐경이후 여성호르몬 약의 사용여부	사용 기간
○ 예, 지금도 사용하고 있다	____년 ____개월
○ 예, 지금은 사용하지 않는다	
○ 아니오	

40. 임신 경험이 있으신 분은 다음 해당 사항에 표시해 주십시오.

총 임신 횟수	자연유산 횟수	인공유산 횟수	사산 횟수	질식(정상) 분만횟수	제왕절개 횟수

(40-1) 첫 임신연령과 첫 만삭분만(37주~42주) 연령, 마지막 분만연령을 표시해 주십시오.

첫 임신연령	첫 만삭분만 연령	마지막 분만 연령
만 _____세	만 _____세	만 _____세

(40-2) 첫 번째 임신의 결과는 어떻게 되셨습니까?

○ 정상분만 ○ 사산(죽은 아이 분만)
○ 자궁외 임신으로 인한 유산
○ 자연유산 ○ 인공유산

(40-3) 임신 중 임신성당뇨증이나 임신성고혈압(임신 중독증)로 진단 받은 적이 있습니까?

○ 임신성당뇨증 ○ 임신성고혈압
○ 아니오 ○ 모름

41. 아기에게 직접 자신의 젖(모유)을 먹인 적이 있습니까?

○ 예 ○ 아니오 ○ 모름

(41-1) 모유수유 경험이 있으신 분만 응답해주세요.

모유수유를 한 자녀 수	총 _____명
모유수유를 시작했을 때의 나이	만 _____세
첫 번째 모유수유 기간	_____개월
총 모유수유 기간	총 _____개월

42. 먹는 피임약을 써본 적이 있습니까?

○ 아니오
○ 예, 그러나 지금은 사용하지 않는다.
○ 예, 지금 사용한다.
○ 모름

(42-1) 몇 세 때 복용을 시작하였습니까?

만 _____세

(42-2) 총 복용 기간은 얼마입니까?

_____년 _____개월 ○ 모름

43. 유방의 자가 검진을 시행 해 보신 적이 있습니까?

○ 있다 ○ 없다

Appendix 2. The Approval by the Institutional Review Board (IRB) of the National Cancer Center

국립암센터 의생명연구심의위원회 (서식 69호_V4.1)

심의결과통보서

국립암센터 의생명연구심의위원회 심의 결과를 아래와 같이 알려 드립니다.

수신	책임연구자		김정선	소속	암역학예방연구부
	의뢰자		국립암센터		
IRB 번호		NCC2016-0088			
접수번호		2016-0147-0001			
구분		신속심의			
연구과제명	국문	갑상선암 발생에 영향을 미치는 식이요인과 유전요인과의 상호관련성			
	영문	Diet-Gene Interaction in the Risk of Thyroid Cancer			
연구상세분류		보관된 검체, 인간대상 연구,인체 유래물(검체)연구(유전정보 포함), 연구단계:기타			
심의 유형		신규과제			
접수일		2016년 04월 01일			
승인일		2016년 04월 12일			
지속심의 제출주기		1년	연구위험도	Level I -최소위험 2018.12.01 이후 접수되어 승인된 신규과제부터 해당	
연구승인 유효일		2017년 04월 11일 지속심의신청서는 연구승인유효일 2개월 전 제출			
심의결과		승인			

1/8

Bibliography

1. Pellegriti, G., et al., *Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors.* J Cancer Epidemiol, 2013. **2013**: p. 965212.
2. Choi, S.W., et al., *The association between the socioeconomic status and thyroid cancer prevalence; based on the Korean National Health and Nutrition Examination Survey 2010-2011.* J Korean Med Sci, 2013. **28**(12): p. 1734-40.
3. Jung, K.W., et al., *Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2014.* Cancer Res Treat, 2017. **49**(2): p. 292-305.
4. Jung, K.W., et al., *Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2015.* Cancer Res Treat, 2018. **50**(2): p. 303-316.
5. Ahn, H.S., H.J. Kim, and H.G. Welch, *Korea's thyroid-cancer "epidemic"--screening and overdiagnosis.* N Engl J Med, 2014. **371**(19): p. 1765-7.
6. Lee, J.H. and S.W. Shin, *Overdiagnosis and screening for thyroid cancer in Korea.* Lancet, 2014. **384**(9957): p. 1848.
7. Ito, Y., et al., *Increasing incidence of thyroid cancer: controversies explored.* Nat Rev Endocrinol, 2013. **9**(3): p. 178-84.
8. Vigneri, R., P. Malandrino, and P. Vigneri, *The changing epidemiology of thyroid cancer: why is incidence increasing?* Curr Opin Oncol, 2015. **27**(1): p. 1-7.
9. Kitahara, C.M. and J.A. Sosa, *The changing incidence of thyroid cancer.* Nat Rev Endocrinol, 2016. **12**(11): p. 646-653.

10. Greco, A., et al., *Chapter 16 - Thyroid Cancer A2 - Dellaire, Graham*, in *Cancer Genomics*, J.N. Berman and R.J. Arceci, Editors. 2014, Academic Press: Boston. p. 265-280.
11. Iglesias, M.L., et al., *Radiation exposure and thyroid cancer: a review.* Arch Endocrinol Metab, 2017. **61**(2): p. 180-187.
12. Ron, E., et al., *Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. 1995.* Radiat Res, 2012. **178**(2): p. AV43-60.
13. Williams, D., *Radiation carcinogenesis: lessons from Chernobyl.* Oncogene, 2008. **27 Suppl 2**: p. S9-18.
14. Liu, Y., L. Su, and H. Xiao, *Review of Factors Related to the Thyroid Cancer Epidemic.* Int J Endocrinol, 2017. **2017**: p. 5308635.
15. Yu, X., et al., *A five-year follow-up study of goiter and thyroid nodules in three regions with different iodine intakes in China.* J Endocrinol Invest, 2008. **31**(3): p. 243-50.
16. Zimmermann, M.B. and V. Galetti, *Iodine intake as a risk factor for thyroid cancer: a comprehensive review of animal and human studies.* Thyroid Res, 2015. **8**: p. 8.
17. Tang, K.T. and C.H. Lee, *BRAF mutation in papillary thyroid carcinoma: pathogenic role and clinical implications.* J Chin Med Assoc, 2010. **73**(3): p. 113-28.
18. Guan, H., et al., *Association of high iodine intake with the T1799A BRAF mutation in papillary thyroid cancer.* J Clin Endocrinol Metab, 2009. **94**(5): p. 1612-7.
19. Kim, H.J., et al., *Iodine intake as a risk factor for BRAF mutations in papillary thyroid cancer patients from an iodine-replete area.* Eur J Nutr, 2018. **57**(2): p. 809-815.

20. Cho, Y.A. and J. Kim, *Thyroid cancer risk and smoking status: a meta-analysis.* Cancer Causes Control, 2014. **25**(9): p. 1187-95.
21. Myung, S.K., et al., *Risk Factors for Thyroid Cancer: A Hospital-Based Case-Control Study in Korean Adults.* Cancer Res Treat, 2017. **49**(1): p. 70-78.
22. Jorde, R. and J. Sundsfjord, *Serum TSH levels in smokers and non-smokers. The 5th Tromso study.* Exp Clin Endocrinol Diabetes, 2006. **114**(7): p. 343-7.
23. Derwahl, M. and D. Nicula, *Estrogen and its role in thyroid cancer.* Endocr Relat Cancer, 2014. **21**(5): p. T273-83.
24. Stansifer, K.J., et al., *Modifiable risk factors and thyroid cancer.* Otolaryngol Head Neck Surg, 2015. **152**(3): p. 432-7.
25. Kitahara, C.M., 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): p. 1615-24.
26. Pienaar, W.P., et al., *The thyrotropin releasing hormone stimulation test in alcoholism.* Alcohol Alcohol, 1995. **30**(5): p. 661-7.
27. El-Serag, H.B., H. Hampel, and F. Javadi, *The association between diabetes and hepatocellular carcinoma: a systematic review of epidemiologic evidence.* Clin Gastroenterol Hepatol, 2006. **4**(3): p. 369-80.
28. Huxley, R., et al., *Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies.* Br J Cancer, 2005. **92**(11): p. 2076-83.
29. Bao, C., et al., *Diabetes mellitus and incidence and mortality of kidney cancer: a meta-analysis.* J Diabetes Complications, 2013. **27**(4): p. 357-64.

30. Jian Gang, P., et al., *Diabetes mellitus and the risk of prostate cancer: an update and cumulative meta-analysis.* Endocr Res, 2015. **40**(1): p. 54-61.
31. Boyle, P., et al., *Diabetes and breast cancer risk: a meta-analysis.* Br J Cancer, 2012. **107**(9): p. 1608-17.
32. Johnson, J.A., et al., *Diabetes and cancer (1): evaluating the temporal relationship between type 2 diabetes and cancer incidence.* Diabetologia, 2012. **55**(6): p. 1607-18.
33. Johnson, J.A. and M. Pollak, *Insulin, glucose and the increased risk of cancer in patients with type 2 diabetes.* Diabetologia, 2010. **53**(10): p. 2086-8.
34. Yeo, Y., et al., *Diabetes mellitus and risk of thyroid cancer: a meta-analysis.* PLoS One, 2014. **9**(6): p. e98135.
35. Luo, J., et al., *Diabetes, Diabetes Treatment, and Risk of Thyroid Cancer.* J Clin Endocrinol Metab, 2016. **101**(3): p. 1243-8.
36. Kitahara, C.M., et al., *Physical activity, diabetes, and thyroid cancer risk: a pooled analysis of five prospective studies.* Cancer Causes Control, 2012. **23**(3): p. 463-471.
37. Aschebrook-Kilfoy, B., et al., *Diabetes and thyroid cancer risk in the National Institutes of Health-AARP Diet and Health Study.* Thyroid, 2011. **21**(9): p. 957-63.
38. Shih, S.R., et al., *Diabetes and thyroid cancer risk: literature review.* Exp Diabetes Res, 2012. **2012**: p. 578285.
39. Vucenik, I. and J.P. Stains, *Obesity and cancer risk: evidence, mechanisms, and recommendations.* Ann N Y Acad Sci, 2012. **1271**: p. 37-43.

40. Ma, J., et al., *Obesity and risk of thyroid cancer: evidence from a meta-analysis of 21 observational studies.* Med Sci Monit, 2015. **21**: p. 283-91.
41. Kitahara, C.M., et al., *Obesity and thyroid cancer risk among U.S. men and women: a pooled analysis of five prospective studies.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(3): p. 464-72.
42. Zhao, Z.G., et al., *Overweight, obesity and thyroid cancer risk: a meta-analysis of cohort studies.* J Int Med Res, 2012. **40**(6): p. 2041-50.
43. Iribarren, C., et al., *Cohort study of thyroid cancer in a San Francisco Bay area population.* Int J Cancer, 2001. **93**(5): p. 745-50.
44. Fiore, E., 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.* Endocr Relat Cancer, 2009. **16**(4): p. 1251-60.
45. Boelaert, K., *The association between serum TSH concentration and thyroid cancer.* Endocr Relat Cancer, 2009. **16**(4): p. 1065-72.
46. Parameswaran, R., S. Brooks, and G.P. Sadler, *Molecular pathogenesis of follicular cell derived thyroid cancers.* Int J Surg, 2010. **8**(3): p. 186-93.
47. Pazaitou-Panayiotou, K., S.A. Polyzos, and C.S. Mantzoros, *Obesity and thyroid cancer: epidemiologic associations and underlying mechanisms.* Obes Rev, 2013. **14**(12): p. 1006-22.
48. Paes, J.E., et al., *The relationship between body mass index and thyroid cancer pathology features and outcomes: a clinicopathological cohort study.* J Clin Endocrinol Metab, 2010. **95**(9): p. 4244-50.
49. Hedayati, M., et al., *Leptin: a correlated Peptide to papillary thyroid carcinoma?* J Thyroid Res, 2011. **2011**: p. 832163.

50. McTiernan, A., et al., *Physical activity and cancer etiology: associations and mechanisms.* Cancer Causes Control, 1998. **9**(5): p. 487-509.
51. Xhaard, C., et al., *Recreational Physical Activity and Differentiated Thyroid Cancer Risk: A Pooled Analysis of Two Case-Control Studies.* Eur Thyroid J, 2016. **5**(2): p. 132-8.
52. Schmid, D., et al., *Physical activity, diabetes, and risk of thyroid cancer: a systematic review and meta-analysis.* Eur J Epidemiol, 2013. **28**(12): p. 945-58.
53. Anzuini, F., A. Battistella, and A. Izzotti, *Physical activity and cancer prevention: a review of current evidence and biological mechanisms.* J Prev Med Hyg, 2011. **52**(4): p. 174-80.
54. Fei, X., et al., *Contribution of industrial density and socioeconomic status to the spatial distribution of thyroid cancer risk in Hangzhou, China.* Sci Total Environ, 2018. **613-614**: p. 679-686.
55. Reitzel, L.R., et al., *Trends in thyroid cancer incidence in Texas from 1995 to 2008 by socioeconomic status and race/ethnicity.* Thyroid, 2014. **24**(3): p. 556-67.
56. Dorak, M.T. and E. Karpuzoglu, *Gender differences in cancer susceptibility: an inadequately addressed issue.* Front Genet, 2012. **3**: p. 268.
57. Rahbari, R., L. Zhang, and E. Kebebew, *Thyroid cancer gender disparity.* Future Oncol, 2010. **6**(11): p. 1771-9.
58. Magri, F., et al., *Expression of estrogen and androgen receptors in differentiated thyroid cancer: an additional criterion to assess the patient's risk.* Endocr Relat Cancer, 2012. **19**(4): p. 463-71.
59. Lu, Y., J. Li, and J. Li, *Estrogen and thyroid diseases: an update.* Minerva Med, 2016. **107**(4): p. 239-44.

60. Son, H.Y., et al., *Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer.* Nat Commun, 2017. **8**: p. 15966.
61. Xu, L., et al., *Family history of cancer and risk of sporadic differentiated thyroid carcinoma.* Cancer, 2012. **118**(5): p. 1228-35.
62. Brindel, P., et al., *Family history of thyroid cancer and the risk of differentiated thyroid cancer in French polynesia.* Thyroid, 2010. **20**(4): p. 393-400.
63. Kohler, A., et al., *Genome-wide association study on differentiated thyroid cancer.* J Clin Endocrinol Metab, 2013. **98**(10): p. E1674-81.
64. Figlioli, G., et al., *Novel genome-wide association study-based candidate loci for differentiated thyroid cancer risk.* J Clin Endocrinol Metab, 2014. **99**(10): p. E2084-92.
65. Wang, Y.L., et al., *Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population.* J Med Genet, 2013. **50**(10): p. 689-95.
66. Matsuse, M., et al., *The FOXE1 and NKX2-1 loci are associated with susceptibility to papillary thyroid carcinoma in the Japanese population.* J Med Genet, 2011. **48**(9): p. 645-8.
67. Khan, M.S., et al., *Significant association of TP53 Arg72Pro polymorphism in susceptibility to differentiated thyroid cancer.* Cancer Biomark, 2015. **15**(4): p. 459-65.
68. Wang, C. and Z. Ai, *Association of XRCC1 polymorphisms with thyroid cancer risk.* Tumour Biol, 2014. **35**(5): p. 4791-7.
69. He, W., et al., *Key genes and pathways in thyroid cancer based on gene set enrichment analysis.* Oncol Rep, 2013. **30**(3): p. 1391-7.

70. Nguyen, D.M. and H.B. El-Serag, *The epidemiology of obesity.* Gastroenterol Clin North Am, 2010. **39**(1): p. 1-7.
71. Yang, W., T. Kelly, and J. He, *Genetic epidemiology of obesity.* Epidemiol Rev, 2007. **29**: p. 49-61.
72. WHO. *BMI Classification*. Global Database on Body Mass Index 2006; Available from: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html.
73. Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**(7538): p. 197-206.
74. Goutzelas, Y., et al., *Association analysis of FTO gene polymorphisms with obesity in Greek adults.* Gene, 2017. **613**: p. 10-13.
75. Loos, R.J. and G.S. Yeo, *The bigger picture of FTO: the first GWAS-identified obesity gene.* Nat Rev Endocrinol, 2014. **10**(1): p. 51-61.
76. Frayling, T.M., et al., *A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.* Science, 2007. **316**(5826): p. 889-94.
77. Scuteri, A., et al., *Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits.* PLoS Genet, 2007. **3**(7): p. e115.
78. Dina, C., et al., *Variation in FTO contributes to childhood obesity and severe adult obesity.* Nat Genet, 2007. **39**(6): p. 724-6.
79. Grant, S.F., et al., *Association analysis of the FTO gene with obesity in children of Caucasian and African ancestry reveals a common tagging SNP.* PLoS One, 2008. **3**(3): p. e1746.
80. Yang, J., et al., *FTO genotype is associated with phenotypic variability of body mass index.* Nature, 2012. **490**(7419): p. 267-72.

81. Speliotes, E.K., et al., *Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.* Nat Genet, 2010. **42**(11): p. 937-48.
82. Wen, W., et al., *Meta-analysis identifies common variants associated with body mass index in east Asians.* Nat Genet, 2012. **44**(3): p. 307-11.
83. Monda, K.L., et al., *A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry.* Nat Genet, 2013. **45**(6): p. 690-6.
84. Church, C., et al., *Overexpression of Fto leads to increased food intake and results in obesity.* Nat Genet, 2010. **42**(12): p. 1086-92.
85. Jiao, Y., et al., *The Fto Gene Regulates the Proliferation and Differentiation of Pre-Adipocytes in Vitro.* Nutrients, 2016. **8**(2): p. 102.
86. Loos, R.J., et al., *Common variants near MC4R are associated with fat mass, weight and risk of obesity.* Nat Genet, 2008. **40**(6): p. 768-75.
87. Thorleifsson, G., et al., *Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity.* Nat Genet, 2009. **41**(1): p. 18-24.
88. Rouskas, K., et al., *Common variants in FTO, MC4R, TMEM18, PRL, AIF1, and PCSK1 show evidence of association with adult obesity in the Greek population.* Obesity (Silver Spring), 2012. **20**(2): p. 389-95.
89. Hotta, K., et al., *Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population.* J Hum Genet, 2009. **54**(12): p. 727-31.
90. Willer, C.J., et al., *Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.* Nat Genet, 2009. **41**(1): p. 25-34.

91. Xi, B., et al., *Association between common polymorphism near the MC4R gene and obesity risk: a systematic review and meta-analysis.* PLoS One, 2012. **7**(9): p. e45731.
92. Malik, S.G., et al., *Association of beta3-adrenergic receptor (ADRB3) Trp64Arg gene polymorphism with obesity and metabolic syndrome in the Balinese: a pilot study.* BMC Res Notes, 2011. **4**: p. 167.
93. Fujisawa, T., et al., *Meta-analysis of the association of Trp64Arg polymorphism of beta 3-adrenergic receptor gene with body mass index.* J Clin Endocrinol Metab, 1998. **83**(7): p. 2441-4.
94. Kurokawa, N., et al., *Association of BMI with the beta3-adrenergic receptor gene polymorphism in Japanese: meta-analysis.* Obes Res, 2001. **9**(12): p. 741-5.
95. Marti, A., et al., *Meta-analysis on the effect of the N363S polymorphism of the glucocorticoid receptor gene (GRL) on human obesity.* BMC Med Genet, 2006. **7**: p. 50.
96. Paracchini, V., P. Pedotti, and E. Taioli, *Genetics of leptin and obesity: a HuGE review.* Am J Epidemiol, 2005. **162**(2): p. 101-14.
97. World Health Organization. *Epidemiology*. 2018 [cited 2018 31 May]; Available from: http://www.who.int/topics/epidemiology/en/.
98. Khaw, K.T., et al., *Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. European Prospective Investigation into Cancer and Nutrition.* Lancet, 2001. **357**(9257): p. 657-63.
99. Heart Protection Study Collaborative, G., *MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomised placebo-controlled trial.* Lancet, 2002. **360**(9326): p. 23-33.

100. Evans, D.M. and G. Davey Smith, *Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality.* Annu Rev Genomics Hum Genet, 2015. **16**: p. 327-50.
101. Didelez, V. and N. Sheehan, *Mendelian randomization as an instrumental variable approach to causal inference.* Stat Methods Med Res, 2007. **16**(4): p. 309-30.
102. Holmes, M.V., M. Ala-Korpela, and G.D. Smith, *Mendelian randomization in cardiometabolic disease: challenges in evaluating causality.* Nat Rev Cardiol, 2017. **14**(10): p. 577-590.
103. Burgess, S., D.S. Small, and S.G. Thompson, *A review of instrumental variable estimators for Mendelian randomization.* Stat Methods Med Res, 2017. **26**(5): p. 2333-2355.
104. Palmer, T.M., et al., *Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses.* Am J Epidemiol, 2011. **173**(12): p. 1392-403.
105. Bowden, J., et al., *Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator.* Genet Epidemiol, 2016. **40**(4): p. 304-14.
106. Smith, G.D. and S. Ebrahim, *'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?* Int J Epidemiol, 2003. **32**(1): p. 1-22.
107. Wang, T., et al., *Causal Association of Overall Obesity and Abdominal Obesity with Type 2 Diabetes: A Mendelian Randomization Analysis.* Obesity (Silver Spring), 2018.
108. Skaaby, T., et al., *Estimating the causal effect of body mass index on hay fever, asthma and lung function using Mendelian randomization.* Allergy, 2018. **73**(1): p. 153-164.

109. Geng, T., et al., *Childhood BMI and Adult Type 2 Diabetes, Coronary Artery Diseases, Chronic Kidney Disease, and Cardiometabolic Traits: A Mendelian Randomization Analysis.* Diabetes Care, 2018.
110. Lyall, D.M., et al., *Association of Body Mass Index With Cardiometabolic Disease in the UK Biobank: A Mendelian Randomization Study.* JAMA Cardiol, 2017. **2**(8): p. 882-889.
111. van 't Hof, F.N., et al., *Genetic variants associated with type 2 diabetes and adiposity and risk of intracranial and abdominal aortic aneurysms.* Eur J Hum Genet, 2017. **25**(6): p. 758-762.
112. Polimanti, R., et al., *A putative causal relationship between genetically determined female body shape and posttraumatic stress disorder.* Genome Med, 2017. **9**(1): p. 99.
113. Nordestgaard, L.T., et al., *Body Mass Index and Risk of Alzheimer's Disease: A Mendelian Randomization Study of 399,536 Individuals.* J Clin Endocrinol Metab, 2017. **102**(7): p. 2310-2320.
114. Mao, Y., et al., *Genetically predicted high body mass index is associated with increased gastric cancer risk.* Eur J Hum Genet, 2017. **25**(9): p. 1061-1066.
115. Lindstrom, S., et al., *Assessing the causal relationship between obesity and venous thromboembolism through a Mendelian Randomization study.* Hum Genet, 2017. **136**(7): p. 897-902.
116. Larsson, S.C., et al., *Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study.* Neurology, 2017. **89**(5): p. 454-460.
117. Guo, Q., et al., *Body mass index and breast cancer survival: a Mendelian randomization analysis.* Int J Epidemiol, 2017. **46**(6): p. 1814-1822.

118. Gianfrancesco, M.A., et al., *Evidence for a causal relationship between low vitamin D, high BMI, and pediatric-onset MS.* Neurology, 2017. **88**(17): p. 1623-1629.
119. Gianfrancesco, M.A., et al., *Causal Effect of Genetic Variants Associated With Body Mass Index on Multiple Sclerosis Susceptibility.* Am J Epidemiol, 2017. **185**(3): p. 162-171.
120. Dale, C.E., et al., *Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis.* Circulation, 2017. **135**(24): p. 2373-2388.
121. Chatterjee, N.A., et al., *Genetic Obesity and the Risk of Atrial Fibrillation: Causal Estimates from Mendelian Randomization.* Circulation, 2017. **135**(8): p. 741-754.
122. Censin, J.C., et al., *Childhood adiposity and risk of type 1 diabetes: A Mendelian randomization study.* PLoS Med, 2017. **14**(8): p. e1002362.
123. Carreras-Torres, R., et al., *Obesity, metabolic factors and risk of different histological types of lung cancer: A Mendelian randomization study.* PLoS One, 2017. **12**(6): p. e0177875.
124. Carreras-Torres, R., et al., *The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study.* J Natl Cancer Inst, 2017. **109**(9).
125. Taylor, P.N., et al., *Paradoxical Relationship Between Body Mass Index and Thyroid Hormone Levels: A Study Using Mendelian Randomization.* J Clin Endocrinol Metab, 2016. **101**(2): p. 730-8.
126. Painter, J.N., et al., *Genetic Risk Score Mendelian Randomization Shows that Obesity Measured as Body Mass Index, but not Waist:Hip Ratio, Is Causal for Endometrial Cancer.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(11): p. 1503-1510.

127. Mokry, L.E., et al., *Obesity and Multiple Sclerosis: A Mendelian Randomization Study.* PLoS Med, 2016. **13**(6): p. e1002053.
128. Huang, Y., et al., *Obesity and peripheral arterial disease: A Mendelian Randomization analysis.* Atherosclerosis, 2016. **247**: p. 218-24.
129. Hartwig, F.P., et al., *Body mass index and psychiatric disorders: a Mendelian randomization study.* Sci Rep, 2016. **6**: p. 32730.
130. Guo, Y., et al., *Genetically Predicted Body Mass Index and Breast Cancer Risk: Mendelian Randomization Analyses of Data from 145,000 Women of European Descent.* PLoS Med, 2016. **13**(8): p. e1002105.
131. Thomas, D.C., D.A. Lawlor, and J.R. Thompson, *Re: Estimation of bias in nongenetic observational studies using "Mendelian triangulation" by Bautista et al.* Ann Epidemiol, 2007. **17**(7): p. 511-3.
132. Renehan, A.G., et al., *Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies.* Lancet, 2008. **371**(9612): p. 569-78.
133. Kitahara, C.M., et al., *Common obesity-related genetic variants and papillary thyroid cancer risk.* Cancer Epidemiol Biomarkers Prev, 2012. **21**(12): p. 2268-71.
134. Collaboration, C.R.P.C.H.D.G., et al., *Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data.* BMJ, 2011. **342**: p. d548.
135. Pichler, I., et al., *Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study.* PLoS Med, 2013. **10**(6): p. e1001462.

136. Palmer, T.M., et al., *Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts.* BMJ, 2013. **347**: p. f4262.
137. Collaboration, I.R.G.C.E.R.F., et al., *Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies.* Lancet, 2012. **379**(9822): p. 1205-13.
138. Interleukin-6 Receptor Mendelian Randomisation Analysis, C., et al., *The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis.* Lancet, 2012. **379**(9822): p. 1214-24.
139. Chen, L., et al., *Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach.* PLoS Med, 2008. **5**(3): p. e52.
140. Bjorngaard, J.H., et al., *The causal role of smoking in anxiety and depression: a Mendelian randomization analysis of the HUNT study.* Psychol Med, 2013. **43**(4): p. 711-9.
141. Lewis, S.J., et al., *Smoking is associated with, but does not cause, depressed mood in pregnancy--a mendelian randomization study.* PLoS One, 2011. **6**(7): p. e21689.
142. Davey Smith, G., et al., *Association of C-reactive protein with blood pressure and hypertension: life course confounding and mendelian randomization tests of causality.* Arterioscler Thromb Vasc Biol, 2005. **25**(5): p. 1051-6.
143. Husemoen, L.L., et al., *Investigating the causal effect of vitamin D on serum adiponectin using a Mendelian randomization approach.* Eur J Clin Nutr, 2014. **68**(2): p. 189-95.

144. Heikkila, K., et al., *C-reactive protein-associated genetic variants and cancer risk: findings from FINRISK 1992, FINRISK 1997 and Health 2000 studies.* Eur J Cancer, 2011. **47**(3): p. 404-12.

145. Timpson, N.J., et al., *Does greater adiposity increase blood pressure and hypertension risk?: Mendelian randomization using the FTO/MC4R genotype.* Hypertension, 2009. **54**(1): p. 84-90.

146. Dahl, M., et al., *C reactive protein and chronic obstructive pulmonary disease: a Mendelian randomisation approach.* Thorax, 2011. **66**(3): p. 197-204.

147. Voight, B.F., et al., *Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study.* Lancet, 2012. **380**(9841): p. 572-80.

148. Fall, T., et al., *The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis.* PLoS Med, 2013. **10**(6): p. e1001474.

149. Huang, T., et al., *Association of homocysteine with type 2 diabetes: a meta-analysis implementing Mendelian randomization approach.* BMC Genomics, 2013. **14**: p. 867.

150. Ye, Z., et al., *The association between circulating lipoprotein(a) and type 2 diabetes: is it causal?* Diabetes, 2014. **63**(1): p. 332-342.

151. Burgess, S., F. Dudbridge, and S.G. Thompson, *Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods.* Stat Med, 2016. **35**(11): p. 1880-906.

152. Palmer, T.M., et al., *Using multiple genetic variants as instrumental variables for modifiable risk factors.* Stat Methods Med Res, 2012. **21**(3): p. 223-42.

153. Collin, S.M., et al., *The causal roles of vitamin B(12) and transcobalamin in prostate cancer: can Mendelian randomization analysis provide definitive answers?* Int J Mol Epidemiol Genet, 2011. **2**(4): p. 316-27.
154. You, N.C., et al., *A prospective study of leukocyte telomere length and risk of type 2 diabetes in postmenopausal women.* Diabetes, 2012. **61**(11): p. 2998-3004.
155. Zhao, J., et al., *Genetically predicted testosterone and cardiovascular risk factors in men: a Mendelian randomization analysis in the Guangzhou Biobank Cohort Study.* Int J Epidemiol, 2014. **43**(1): p. 140-8.
156. International Consortium for Blood Pressure Genome-Wide Association, S., et al., *Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk.* Nature, 2011. **478**(7367): p. 103-9.
157. Rasmussen-Torvik, L.J., et al., *Association of a fasting glucose genetic risk score with subclinical atherosclerosis: The Atherosclerosis Risk in Communities (ARIC) study.* Diabetes, 2011. **60**(1): p. 331-5.
158. Kivimaki, M., et al., *Conventional and Mendelian randomization analyses suggest no association between lipoprotein(a) and early atherosclerosis: the Young Finns Study.* Int J Epidemiol, 2011. **40**(2): p. 470-8.
159. Zhu, Z., et al., *Causal associations between risk factors and common diseases inferred from GWAS summary data.* Nat Commun, 2018. **9**(1): p. 224.
160. Holmes, M.V., et al., *Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis.* Am J Hum Genet, 2014. **94**(2): p. 198-208.

161. Look, A.R.G. and R.R. Wing, *Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the Look AHEAD trial.* Arch Intern Med, 2010. **170**(17): p. 1566-75.

162. Jokela, M., et al., *Body mass index and depressive symptoms: instrumental-variables regression with genetic risk score.* Genes Brain Behav, 2012. **11**(8): p. 942-8.

Acknowledgement

Firstly, I would like to express my sincere gratefulness to my advisor Professor Jeongseon Kim for the endless support of my study and related research, for her patience, motivation, and extensive experience. Her supervision encouraged me in all the time of study and writing of this thesis. I could not have imagined having a better advisor and instructor for my study. Besides my advisor, I would like to thank the rest of my thesis committee: Professor Eun Kyung Lee and Professor Yul Hwangbo, for their insightful explanations and encouragement, but also for the hard question which incented me to widen my research from various perspectives. I would also like to acknowledge Dr. Soo Ji Lee of the Department of Public Health at Seoul National University, and I am gratefully indebted to her for her very valuable comments on this thesis. I thank my fellow lab mates in for the stimulating discussions, and for all the fun we have had in the last two years. Also, I thank my friends in the following Graduate School of Cancer Science and Policy. Last but not the least, I would like to thank my family: my parents and to my sister for supporting me spiritually throughout writing this thesis and my life in general.