A Systematic Review and Analysis of Reporting Quality of Studies of Germline Genetic Variants Influencing Susceptibility to Nonmedullary Thyroid Cancer

1Judith E Ritchie, 2Sabapathy P Balasubramanian
1Surgical Oncology, University of Sheffield, Sheffield, United Kingdom
2Department of Endocrine Surgery, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom

ABSTRACT

Genetic susceptibility makes a significant contribution to many multifactorial diseases, such as cancer. Genetic variants have been associated with medullary thyroid cancer but their role in nonmedullary thyroid cancer (NMTC) has not been clearly characterized. Although many published reports have evaluated association between some genetic variants and NMTC, a comprehensive assessment has not been done and the quality of reporting of these studies has not been evaluated. In this study, we report the results of a systematic review of published case-control studies looking at the association of polymorphisms with the susceptibility to nonmedullary thyroid cancer and an assessment of quality of study design, implementation and interpretation.

Methods: A systematic review of the existing literature was carried out, identifying studies through a search of the Medline literature via PubMed and through scanning of references of these papers. These papers were then subject to an assessment of quality of reporting using a scoring system modified from previously published criteria.

Results: 50 studies published between 1997 and 2010 were included. These studies included 916 polymorphisms across 62 genes involved in a diverse range of cellular processes using 10,704 cases and 27,707 controls. The median quality score was 70.4% (range 52-93%). Areas of strength and weaknesses in study reporting were identified.

Conclusions: Research on the genetic susceptibility to nonmedullary thyroid cancer is sparse. Published studies are of suboptimal quality, have analyzed few variants and positive findings have not been replicated. Adherence to recently published guidelines on methodology should be encouraged. A consortium led approach involving multiple centers, including large numbers of patients in well-defined study protocols is required to investigate this subject comprehensively.

Keywords: Nonmedullary thyroid cancer, Polymorphisms, Susceptibility.

INTRODUCTION

Although relatively uncommon, thyroid carcinoma is the most common endocrine cancer encountered in clinical practice with 2,108 cases diagnosed in the UK in 2007. Furthermore, the disease burden is increasing worldwide with between 141,000 and 212,033 estimated worldwide diagnoses and 35,471 deaths reported. The disease is subclassified according to tumor histology. The majority of cases are well differentiated with over 80% papillary and 11% follicular carcinomas. Other subtypes of thyroid carcinoma are less common. Medullary carcinoma comprises around 3%, of which 18% is thought to be familial. Genetic factors in the etiology of ‘nonmedullary’ thyroid cancer have not been well characterized. This review focuses on the genetic predisposition to nonfamilial nonmedullary thyroid cancer.

Genetic variation can take several forms, the commonest being a single nucleotide change in the genomic sequence. If this change is uncommon (< 1%) and invariably associated with a distinct clinical phenotype, it is referred to as a mutation (i.e. has a high penetrance). A genetic polymorphism is a variation in gene sequence that differs from a mutation in that they occur in over 1% of the population. A vast majority is functionally neutral, but they may occasionally have a significant effect on the phenotype depending on their location in the gene and their structural or functional effect on the expressed protein. Characterizing the role of genetic polymorphisms in susceptibility or severity to thyroid cancer could potentially help risk stratification, allowing early identification of at-risk patients. Identification of genes that predispose to tumors with poor prognosis could help in targeting aggressive treatment and surveillance in selected patients.

Establishing the role of such genetic variation in multifactorial disease is difficult as small effect sizes are
expected due to their often modest genetic effect. Numerous examples of the conflicting reports on the role of genetic polymorphisms in disease exist.\(^3\)\(^4\) Reported positive associations are often later demonstrated to be ‘false-positive’ findings and the postulated reasons include inadequate power from small sample sizes, incorrect calculation of population genetic parameters and the potential for confounding by important variables, such as sex or ethnic origin. To improve the quality of methodology and reporting of ‘genetic epidemiology’ studies, guidelines have recently been published.\(^5\)\(^0\) These guidelines focus on a set of criteria that aim to address the common pitfalls in such studies.

A review of the studies on the ‘genetic predisposition to nonmedullary thyroid cancer’ study has summarized the results of findings of 25 studies published until May 2008.\(^1\) This article highlights the problems discussed above and limitations of the use of the candidate gene approach as opposed to genome-wide association studies that have the potential to more reliably identify potentially significant genetic variants.\(^2\)\(^3\)\(^4\) However, a systematic assessment of the quality of the studies in this review and subsequent similar studies (published since May 2008) has not been carried out.

The aim of this study was two-fold: To carry out a systematic review of published case-control studies looking at the association of polymorphisms with the susceptibility to nonmedullary thyroid cancer and assess the quality of study design, implementation and interpretation.

**METHODS**

The Medline database was searched via PubMed on the 10th of February 2010. The detailed search strategy was as follows: ((“thyroid neoplasms” [TIAB] not Medline [SB]) or “thyroid neoplasms” [MeSH Terms] or thyroid cancer [Text Word]) and (“genetic polymorphism” [Text Word] or “polymorphism, genetic” [MeSH Terms] or polymorphism [Text Word]) All original articles that evaluated the role of genetic variants in the susceptibility and/or severity of thyroid cancer using a case-control approach were included. Studies carried out on ‘medullary thyroid cancer’ alone, ‘benign thyroid disease’ alone, studies without a control group and studies carried out on ‘tumor DNA’ samples only were excluded. Studies were reviewed and scored on two independent occasions by the first author (JER).

A total of 10 quality issues were identified from previous papers and studies carrying out quality assessments in genetic epidemiology studies in other diseases.\(^13\),\(^28\) Each included study was scored from 1 to 3 on each of the 10 criteria based on definitions agreed by the authors at the start of the study (see Table 1 for the definition of the scores). The overall study score was expressed as a percentage of the total possible score for the paper. Where a criterion was not relevant to a particular paper, no score was given and three points were deducted from the total score before calculation of the percentage score.

**RESULTS**

The search strategy initially identified 463 published case-control studies, of which 50 fulfilled the selection criteria (Fig. 1). These were published between 1997 and 2010. These studied 916 separate polymorphisms in 62 genes in 10,704 cases and 27,707 controls. One study did not describe the 92 genes and 786 SNPs studied, and one did not name the MC1R genetic variants it studied. The genes evaluated are involved in a diverse range of cellular processes (Table 2). Only 20 polymorphisms were studied more than once. Thirty-six studies reported positive findings identifying 45 significant polymorphisms in 39 candidate genes. These candidate genes found to be of interest are highlighted in Table 2. However, only three of these positive findings have been replicated in other studies. These are p53 gene (Arg/Pro codon 72) gene, XRCC3 C18067T and XRCC3 (Arg399Gln) gene.

Median quality score was 70.37% (range 51.85-93.33). A total of 40 studies reported significant results, of which 6 reported on ‘p’ values only; 39 reported on odds ratios and 1 reported on relative risk.

The results are summarized for each of the quality criterion examined. Percentage breakdown of scores for each criterion is summarized in Figure 2.

**Selection of Controls**

An ideal control cohort should be an unbiased and representative sample from the population at risk of developing the disease in question. Attempts should be made to ensure that cases and controls are similar (or matched) with regards to age, gender and ethnicity. Small risks of bias can be magnified in larger studies and have been found to correlate with greater error rates with rarer alleles.\(^2\)\(^8\) Only 24 studies (48%) qualified for a score of 3, as defined in Table 1.\(^2\)\(^7\)\(^8\)\(^23\)\(^39\)\(^42\)\(^48\)\(^51\)\(^53\)\(^54\)\(^56\)\(^58\)\(^61\)\(^64\)\(^66\)\(^68\)\(^71\) Thirty-two (64%) explicitly stated participants’ ethnic origins\(^2\)\(^4\)\(^7\)\(^8\)\(^12\)\(^15\)\(^19\)\(^20\)\(^23\)\(^26\)\(^27\)\(^29\)\(^30\)\(^32\)\(^33\)\(^35\)\(^36\)\(^38\)\(^42\)\(^43\)\(^45\)\(^49\)\(^52\)\(^53\)\(^54\)\(^57\)\(^59\)\(^61\)\(^64\)\(^66\)\(^68\)\(^71\) and 13 (26%) either did not mention ethnicity or failed to state any details.\(^5\)\(^6\)\(^10\)\(^16\)\(^17\)\(^25\)\(^27\)\(^37\)\(^39\)\(^41\)\(^48\)\(^51\)\(^56\)\(^67\)\(^68\)\(^71\) thereby leaving readers to assume ethnicity based on the geographical location of the study. Six studies\(^2\)\(^5\)\(^27\)\(^35\)\(^36\)\(^38\) classified subjects by color rather than ethnicity. In addition, of the 22 studies that matched cases to controls,\(^2\)\(^4\)\(^7\)\(^8\)\(^23\)\(^26\)\(^27\)\(^33\)\(^39\)\(^42\)\(^48\)\(^51\)\(^53\)\(^54\)\(^56\)\(^60\)\(^61\)\(^64\)\(^68\)\(^71\) nine (41%) did so by ethnicity.\(^8\)\(^23\)\(^33\)\(^42\)\(^53\)\(^54\)\(^61\)\(^66\)\(^67\)

A healthy control group is essential to avoid recruitment bias due to ‘disease association’. Twenty-seven studies (54%) stated the use of healthy controls.\(^4\)\(^6\)\(^8\)\(^15\)\(^17\)\(^19\)\(^20\)\(^23\)\(^26\)\(^30\)\(^35\)\(^38\)\(^41\)\(^43\)\(^45\)\(^47\)\(^49\)\(^56\)\(^60\)\(^64\)\(^68\)\(^71\) Controls were not healthy in seven (12%) studies,\(^10\)\(^23\)\(^39\)\(^48\)\(^54\)\(^66\)\(^68\)\(^71\) and included recruits suffering from a range of nonmalignant diseases. Fourteen (28%) failed to report on the health of their controls,\(^2\)\(^3\)\(^7\)\(^12\)\(^20\)\(^31\)\(^37\)\(^36\)\(^51\)\(^53\)\(^57\)\(^59\)\(^61\)\(^65\) whilst two studies simply described their control cohorts as ‘cancer-free’.\(^3\)\(^7\)\(^36\) Seven (14%) studies recruited controls from the general population increasing the likelihood of sampling an unbiased cohort more representative of the source population.\(^8\)\(^32\)\(^42\)\(^45\)\(^65\)\(^67\)\(^71\)

The remainder is subject to recruitment bias:
17 (34%) recruited from hospital visitors or outpatients with nonmalignant pathology, 16 (32%) from blood donors and volunteers, and 10 (20%) studies failed to report the source of recruitment at all. Hardy-Weinberg Equilibrium

Hardy-Weinberg equilibrium (HWE) is a well-established principle in population genetics. It assumes stability of allele frequencies at the genetic locus within a predominantly healthy population, against which the allele frequencies of the study.

Table 1: Definition of scores allocated to each ‘quality’ criterion/item

<table>
<thead>
<tr>
<th>Item</th>
<th>Score 1</th>
<th>Score 2</th>
<th>Score 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td>No explicit description; different ethnicity</td>
<td>Some categories met</td>
<td>Ethnicity explicitly stated and same as control group; healthy cohort; source of recruitment clearly stated</td>
</tr>
<tr>
<td>HWE</td>
<td>Not mentioned; inappropriate calculation</td>
<td>Mentioned but no detail, such as test used or p value. Not clear to verify</td>
<td>Estimated and actual genotype frequencies compared using Chi-square goodness of fit and p-value reported. Significant result should have explanation for probable cause for disequilibrium</td>
</tr>
<tr>
<td>Case Group</td>
<td>No definition/explanation</td>
<td>Some details included</td>
<td>Clear explanation of types of cancers included and/or exclusion/inclusion criteria to allow replication with unambiguous case selection</td>
</tr>
<tr>
<td>Primers/probes</td>
<td>Not published or referenced</td>
<td>Partly published/referenced</td>
<td>Full/adequate publishing/referencing of sequences for all polymorphisms studied</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>No comment/description/referencing</td>
<td>1 of 2 categories completely/partly filled</td>
<td>Genotyping procedure for all polymorphisms studied either clearly described or adequately referenced. Validation (2nd assay/repeats/internal controls) carried out in at least a small sample</td>
</tr>
<tr>
<td>Blinding</td>
<td>No mention/description</td>
<td>–</td>
<td>Clearly described phenotype blinding during genetic analysis</td>
</tr>
<tr>
<td>Sample size/power</td>
<td>No mention</td>
<td>Fulfilled in part</td>
<td>Clear prospective/retrospective power calculations; clear description of parameters; population statistics referenced/obtained from appropriate source; inadequately powered studies discussed limitations</td>
</tr>
<tr>
<td>Statistics</td>
<td>No description</td>
<td>Inadequate description</td>
<td>Well-defined tests of significance. Odds ratios (p-values) with confidence intervals calculated</td>
</tr>
<tr>
<td>Corrections for multiple testing</td>
<td>No mention of corrections/limitation</td>
<td>Explained in part</td>
<td>Methods of correction, e.g. Bonferroni adequately described. Results without corrections described as exploratory and limitations emphasized. Only valid for studies &gt;1 polymorphism</td>
</tr>
<tr>
<td>Independent replication</td>
<td>No mention of prior reports or need for validation</td>
<td>Explained in part</td>
<td>Clear reference to prior positive result OR study used 2 cohorts for testing and validation OR need for replication of findings and preliminary nature of association clearly emphasized</td>
</tr>
</tbody>
</table>

Table 2: Functional classification of the genes studied in this review. Those genes in which significant polymorphisms have been identified are highlighted in bold

<table>
<thead>
<tr>
<th>Classification</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA repair</td>
<td>BRCA1, RAD51, BRCA2, RAD18, RAD52, ERCC2, XRCCL1, XRCCL3, XRCCL4, XRCCL7, EMSY, APEX, BRIP1, ZNF350, ADPRT, APE1, ATM, MTF1</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>CHK2, p53</td>
</tr>
<tr>
<td>Gene expression</td>
<td>Pre-miR-146a and 146b</td>
</tr>
<tr>
<td>Immunity</td>
<td>HLADR11, Interleukin10, Interleukin 6, CCR5</td>
</tr>
<tr>
<td>Phase 1 xenobiotic metabolisers</td>
<td>CYP1A1, CYP2D6, NAT2<em>5, NAT</em>6, NAT2<em>7, NAT2</em>14</td>
</tr>
<tr>
<td>Phase 2 xenobiotic metabolisers</td>
<td>GSTM, GSTT, GSTD, GSTP, GPX3</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>Fas, Fasl, (cytotoxic T lymphocytes &amp; NK cells)</td>
</tr>
<tr>
<td>Proto-oncogenes</td>
<td>RET and its coreceptors GFRA2 and GFRA1, L-lyc</td>
</tr>
<tr>
<td>Thyroid physiology</td>
<td>Thyroglobulin gene, THRA1, TSHR, NR1A1a</td>
</tr>
<tr>
<td>Angiogenesis and vasculogenesis</td>
<td>VEGF</td>
</tr>
<tr>
<td>Cell motility</td>
<td>PAK1</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>CAPN5, P2Xr receptor, EPAC, TGFB1, PTPRJ</td>
</tr>
<tr>
<td>Receptors, function undetermined</td>
<td>GNB3, oestrogen receptors</td>
</tr>
<tr>
<td>Underdetermined function</td>
<td>Chromosome 1p 12-13 loci, chromosome 8q24</td>
</tr>
<tr>
<td>Receptors, known function</td>
<td>Melanocortin receptor, Vitamin D receptor</td>
</tr>
<tr>
<td>Transcription factors</td>
<td>FOXE1 gene</td>
</tr>
<tr>
<td>Tumor suppressor gene</td>
<td>ARLTS1</td>
</tr>
</tbody>
</table>

groups can be contrasted. Deviation occurs where genotype and allele frequencies in study control groups significantly differ from the general population. It is a statistical means by which the control group’s representation of the population genetics can be assessed. It thereby requires information on the genotype and allele frequencies of the population in which the study is taking place, which should ideally be published or referenced alongside the statistical tests used. Deviation is denoted by a significant p-value, and is more likely to occur where there is selection bias.

Only 29 studies (58%) clearly reported calculating the HWE;2,4,8,12,16,17,23,30,35,37,39,41,42,45,47-49,51,54,59-61,64-66 of these 15 reported the p-value for the calculations 3,4,12,16,17,23,30,35,37,39,41,42,45,47-49,51,54,59-61,64-66 and only 10 reported detailed calculations and allele frequencies.3,4,12,16,17,35,37,36,41,51 Seven studies reported deviations from HWE.2,4,29,30,45,64,65 Twenty-one (41%) studies made no mention of carrying out the Hardy-Weinberg equilibrium.5,7,10,15,19,20,25-27,33,38,43,46,53,56,57,67,68,71

Selection of Cases

Selection bias can arise during selection of both cases and controls. The characteristics of the cases should be reported in sufficient detail (such as the subtypes of cancer included and the inclusion and exclusion criteria), so as to allow study replication without ambiguity in case selection. Sixteen studies (32%) reported inclusion and exclusion criteria,2,12,16,23,35,37,36,38,45,46,54,56,59,66,68,71 32 (64%) reported some detail2-8,10,15,19,20,25-27,29,30,32,33,39,42,43,47-49,51-53,57,60,61,64,65,67 and only two (4%) gave no definition or explanation.17,41 Five studies (10%) failed to report the types of thyroid cancers included in the study.15,43,53,60,71

Primer Sequence

Clarity of the exact primer sequences used enables repetition of the experiments by other investigators and increases confidence in the methodology used. Of the total number of studies, seven (14%) failed to either fully publish or adequately reference the primer sequences used for all polymorphisms studied.7,12,32,45,49,60,63
Experimental Reproducibility

Experimental protocols must aim to minimize the risk of misclassification of genotypes under study by incorporating quality control, particularly as genotyping error rates have been estimated between 1 and 3%. Quality control procedures include reanalysis of random samples, use of a second genotyping method or sequencing. Thirty-one studies (62%) used quality control measures. Experimental methods used were not adequately described or referenced in 11 studies. Seven (22%) had 51 to 175 cases; 14 (28%) had 176 to 300 and progressively greater numbers required to detect progressively smaller risks. Only nine studies (18%) reported carrying out statistical power calculations. Of the remaining 41 studies, only six reported limitations consequent to lack of power. Underpowered studies risk overestimating true effect, therefore positive results from preliminary studies need replication in an independent and larger study. Very little, however, has been done in terms of replicating findings. Sample sizes in the studies included in this review were largely inadequate. Nine (18%) studies had less than 50 cases recruited; 21 (42%) had 51 to 175 cases; 14 (28%) had 176 to 300 and 6 (12%) had over 300.

Phenotype Blinding

Investigators determining the genotype should ideally be blinded to outcome, and those analyzing outcomes and exposures blinded to genotype. Only three studies reported blinding to phenotype during analysis.

Power and Sample Size

Candidate gene polymorphisms are only expected to have a low to moderate effect on disease penetrance. The low relative risk mandates that studies be sufficiently powered in order to determine significance. Sample sizes can be calculated priori based on population frequency of the variants under study; power being generally increased by large sample sizes with progressively greater numbers required to detect progressively smaller risks. Only nine studies (18%) reported carrying out statistical power calculations. Of the remaining 41 studies, only six reported limitations consequent to lack of power. Underpowered studies risk overestimating true effect, therefore positive results from preliminary studies need replication in an independent and larger study. Very little, however, has been done in terms of replicating findings. Sample sizes in the studies included in this review were largely inadequate. Nine (18%) studies had less than 50 cases recruited; 21 (42%) had 51 to 175 cases; 14 (28%) had 176 to 300 and 6 (12%) had over 300.

Statistics

Statistical tests used were generally well described, and all reported p-values using a significance level of < 0.05. Thirty-four studies (68%) reported significant p values. Thirty-nine (68%) reporting odds ratios and one study (2%) reported on relative risk. Total of 38 studies (75%) reported on odds ratios, confidence intervals and p values.

Correction For Multiple Testing

Correction for testing for multiple polymorphisms was not applicable in 12 studies but was only clearly described in six of the remaining studies.
Consortium’ (BCAC).14 This is a group of several independent investigators working on the detection and validation of genetic polymorphisms that predispose to sporadic breast cancer (http://www.srl.cam.ac.uk/consortia/bcac/). The BCAC has adopted a genome wide association approach and identified several susceptibility variants by genotyping and analyzing hundreds of thousands of polymorphisms across the genome in thousands of patients.18 This has been followed by further testing of polymorphisms identified to be of interest in the initial phase of the project.

This study has clearly demonstrated that there is a wide variation in reporting quality across the studies with little validation of results either within or among the different studies. Good quality studies should essentially meet two standards. The first is that studies should be carried out using robust protocols that seek to eliminate sources of confounding and bias in order to produce valid, significant and replicable data. There are a number of ways by which significant levels of bias can be introduced to genetic association studies. For example, the control group needs to be reflective of the general population in which the study is carried out. Ethnicity, source and method of recruitment can introduce variability between controls and subjects, and subjects should be healthy, disease-free and of the same ethnicity as subjects, and selected in a random manner. Furthermore, statistical measures, such as calculation of ‘a priori’ sample size and ‘Hardy-Weinberg Equilibrium’ increase the validity of the results and make them more meaningful. However, reporting of these measures remains low. For instance, it has been purported that reporting of HWE is universally poor with only 20-69% of gene association studies reporting concordance with Hardy-Weinberg Equilibrium.62

The second set of standards should ensure that studies should publish in sufficient detail so as to facilitate their replication with minimal variability as well as allow thorough scrutiny for a systematic review and/or meta-analysis. Comprehensive description of statistical and scientific methodology, and resultant raw data allow transparency as they could allow calculations to be reproduced, and ultimately allow thorough scrutiny by readers and reviewers and aid performance of meta-analysis on these datasets.

Any critical appraisal of the literature using a quantitative scoring system has its limitations. The aim of using such a scoring tool was simply an attempt to quantify the problem and highlight its prevalence across the literature. This assessment of study ‘quality’ has only been done through the window of reporting. However, it has been shown that poor ‘quality of reporting’ does affect interpretation of results and an exaggeration of positive findings.55 Items in our scoring system carry similar weight without full and appropriate consideration of their individual influence on the magnitude of effect. This may lead to an unbalanced interpretation of the study design, as several minor flaws may result in worse score compared to a study with a major flaw. Despite these limitations, we believe that this provides a crude yet useful measure of problems in current literature and identifies areas of weaknesses that need to be addressed in future studies.

In conclusion, this study has shown that literature on the role of ‘genetic factors and susceptibility to nonmedullary thyroid cancer’ is sparse and the quality is suboptimal. The role of initiatives, such as the STREGA guidelines in improving the quality of such studies remains to be seen. The development of large collaborations, the adoption of modern advances in high-throughput genotyping and bio-informatics and the use of the ‘genome-wide association approach’ seem to be the way forward in ensuring that advances are made in our knowledge of low penetrance genetic susceptibility to nonmedullary thyroid cancer.

REFERENCES


