ORIGINAL ARTICLE

Thyroid cancer susceptibility polymorphisms: confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24

Angela M Jones,[1] Kimberley M Howarth,[1] Lynn Martin,[1] Maggie Gorman,[1] Radu Mihai,[2] Laura Moss,[3] Adam Auton,[1] Catherine Lemon,[4] Hisham Mehanna,[5] Hosahalli Mohan,[6] Susan E M Clarke,[6] Jonathan Wadsley,[7] Elena Macias,[8] Andrew Coatesworth,[9] Matthew Beasley,[10] Tom Roques,[11] Craig Martin,[11] Paul Ryan,[12] Georgina Gerrard,[13] Danielle Power,[14] Caroline Bremmer,[15] The TCUKIN Consortium,* Ian Tomlinson,[1,16] Luis G Carvajal-Carmona[1]

► Additional materials are published online only. To view these files please visit the journal online (http://jmg.bmj.com/content/49/3.toc).

For numbered affiliations see end of article.

Correspondence to
Dr Luis G Carvajal-Carmona, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; luis@well.ox.ac.uk

*A full list of the TCUKIN collaborators is listed in appendix 1.

Received 24 October 2011
Revised 5 December 2011
Accepted 22 December 2011
Published Online First
25 January 2012

ABSTRACT

Five single nucleotide polymorphisms (SNPs) associated with thyroid cancer (TC) risk have been reported: rs2910164 (5q24); rs6983267 (8q24); rs965513 and rs1867277 (9q22); and rs944289 (14q13). Most of these associations have not been replicated in independent populations and the combined effects of the SNPs on risk have not been examined. This study genotyped the five TC SNPs in 781 patients recruited through the TCUKIN study. Genotype data from 6122 controls were obtained from the CORGI and Wellcome Trust Case-Control Consortium studies. Significant associations were detected between TC and rs965513A ($p=6.35\times10^{-34}$), rs1867277A ($p=5.90\times10^{-24}$), rs944289T ($p=6.95\times10^{-7}$), and rs6983267G ($p=0.016$). rs6983267 was most strongly associated under a recessive model ($P_{GG\ vs\ GT\ +\ TT}=0.004$), in contrast to the association of this SNP with other cancer types. However, no evidence was found of an association between rs2910164 and disease under any risk model ($p>0.7$). The rs1867277 association remained significant ($p=0.008$) after accounting for genotypes at the nearby rs965513 ($p=2.3\times10^{-13}$) and these SNPs did not tag a single high risk haplotype. The four validated TC SNPs accounted for a relatively large proportion (~11%) of the sibling relative risk of TC, principally owing to the large effect size of rs965513 (OR 1.74).

INTRODUCTION

Thyroid cancer (TC) is the most common endocrine malignancy and a complex disease with a largely unknown aetiology.[1] TC is characterised by one of the strongest familial relative risks in cancer. First degree relatives of TC patients are up to 8.6 times more likely to develop TC than the general population.[2] Most of the genetic variation associated with TC remains uncharacterised, and it is likely to be explained by variants of moderate or low penetrance.

A number of recent studies have identified single nucleotide polymorphisms (SNPs) associated with TC risk on chromosomes 5q24, 8q24, 9q22, and 14q13.[3–6] Two of these SNPs, rs965513 (9q22) and rs944289 (14q13), were found through a multi-stage, genome-wide association (GWA) study in the Icelandic population.[3] Subsequent replication of association was found in smaller sample sets from Ohio, USA and Spain.

The other three TC SNPs were discovered through candidate gene or SNP approaches.[4–6] rs2910164 (5q24) was chosen because it lay within pre-miR-146a, a microRNA upregulated in TC. An association with TC was found in samples from Poland, Finland and Ohio, and the C allele at rs2910164 was found to decrease levels of pre- and mature miR-146A.[4] rs6983267 (8q24) is associated with the risk of several cancers, including those of the prostate, colon and ovary, and was assessed as a TC SNP for this reason. It showed a borderline significant association with TC in the Polish population.[6] rs1867277 was studied because it lies in the 5′ UTR of *FOXE1* (or *Thyroid Transcription Factor 2*), a key gene involved in thyroid organogenesis.[7] rs1867277 and rs965513 are in moderate pairwise linkage disequilibrium (LD) in Europeans ($r^2=0.39$, $D'=0.73$, http://www.1000genomes.org/). rs1867277 was strongly associated with TC risk in Spanish and Italian cohorts.[5] None of the three candidate SNP associations has been replicated in independent studies.

The aim of this study was to examine these five TC SNPs in a relatively large UK case–control sample set, to validate or refute their associations with TC in this population, and to estimate the proportion of the familial risk of TC for which they account.

PATIENTS AND METHODS

Study samples

We recruited 781 white UK patients of northern European origin with histologically confirmed non-medullary TC through the Thyroid Cancer Genetic investigation in the UK (TCUKIN) study. In addition to obtaining standard clinicopathological information from medical records and a questionnaire completed by each patient, the

participants donated a blood sample which was used to isolate genomic DNA. The Southampton and South West Hampshire Research Ethics Committee (A) approved the TCUKIN protocol.

SNP genotyping and control genotype data

We genotyped the TCUKIN samples at the five SNPs (rs2910164, rs6983267, rs965513, rs1867277, rs944289) using the KAspar system. Probes used to genotype these polymorphisms are shown in supplementary table 1. For comparison, we used available genotype data from 5193 population controls belonging to the National Blood Donor Service (NBS) and the 1958 Birth Cohort (BC58) and 929 cancer-free controls from our COloRectal Gene Identification study (CORGI). The NBS and BC58 samples had been genotyped with Illumina 1.2M (Hap1.2M) arrays as part of the Wellcome Trust Case Control Consortium 2,[8] and the CORGI controls had been genotyped with Illumina Hap550 arrays (Hap550, N=932) as part of our ongoing studies in colorectal cancer genetics.[9] Two of the five SNPs (rs2910164 and rs1867277) were not included on the Illumina 1.2M and Hap550 arrays, but had excellent proxy markers that facilitated their imputation. rs2910164 is perfectly tagged ($D'=1$, $r^2=1$) by rs2961920, a marker present on both SNP arrays; and rs1867277 is perfectly tagged on the Hap550s by rs1443434 and on the Hap1.2Ms by rs1867278, rs1443434, rs1443435, and rs12348691. These genotypes were imputed using IMPUTE2[10]; all markers had proper_info scores >0.8 and imputation call rates >0.95, suggesting excellent imputation. In order to confirm this excellent tagging/imputation, 94 CORGI controls were genotyped for the TC SNPs using KAspar, resulting in no discordant genotypes for each SNP, whether typed or imputed.

Quality control and statistical analysis

General genotyping quality control assessment was as previously described.[9] Duplicate samples were used to check genotyping quality and 100% concordance was found. Samples with multiple missing genotypes were eliminated from the analyses (N=14). All five SNPs passed our quality control thresholds including call rates >95% and Hardy-Weinberg equilibrium p values >0.05.[11]

Association statistics were obtained on per allele, genotypic and haplotype bases using logistic regression models implemented in PLINK, R, and SNPTEST.[12 13] Haplotype analyses were carried out with HAPLOVIEW[14] and PLINK. Allelic count association meta-analyses, using the Mantel-Haenszel method, were carried out in STATA. We used the IMPUTE2 software[10] formally to generate rs2910164 and rs1867277 genotypes in the control population, although the fact that perfect proxies were used rendered this task of very limited utility.[10] To test for independence between SNPs, we used conditional logistic regression models. The proportion of the familial relative risk explained by the polymorphisms investigated in the study was estimated using the method reported by Houlston *et al*.[15]

RESULTS

Single SNP analyses

Three of the five SNPs examined showed a significant association with TC risk in the UK population (table 1). The strongest associations, on a per allele basis, were observed for the 9q22 SNPs rs965513A ($p=6.35\times10^{-34}$, OR=1.99, 95% CI 1.77 to 2.21) and rs1867277A ($p=5.90\times10^{-24}$, OR=1.75, 95% CI 1.57 to 1.95). The association at rs944289T on 14q13 was also convincingly replicated ($p=6.95\times10^{-7}$, OR=1.33, 95% CI 1.18 to 1.48). For rs6983267G, we also found a nominally significant

Table 1 Association statistics for thyroid cancer risk and genetic variants at chromosomes 5q24, 8q24, 9q22 and 14q13

SNP, genotypes and risk allele	Frequency (%)		ORs for genotype or per allele overall (95% CI)	p Value
	Cases	Controls		
rs2910164				
GG	436 (0.578)	3540 (0.584)	Reference	
CG	271 (0.367)	2179 (0.360)	1.032 (0.876 to 1.214)	0.728
CC	41 (0.054)	339 (0.056)	0.987 (0.682 to 1.384)	0.985
Risk allele (C)	359 (0.238)	2857 (0.236)	1.013 (0.893 to 1.148)	0.845
rs6983267				
TT	164 (0.218)	1441 (0.236)	Reference	
GT	346 (0.461)	3012 (0.493)	1.010 (0.827 to 1.236)	0.960
GG	241 (0.321)	1662 (0.272)	1.274 (1.027 to 1.583)	0.026
Risk allele (G)	674 (0.449)	5894 (0.518)	1.140 (1.025 to 1.268)	0.016
rs965513				
GG	187 (0.249)	2748 (0.449)	Reference	
AG	394 (0.525)	2729 (0.446)	2.121 (1.763 to 2.559)	9.08×10^{-17}
AA	170 (0.226)	643 (0.105)	3.883 (3.081 to 4.893)	1.30×10^{-30}
Risk allele (A)	734 (0.489)	4015 (0.328)	1.981 (1.774 to 2.212)	6.35×10^{-34}
rs1867277				
GG	159 (0.211)	2290 (0.376)	Reference	
AG	398 (0.529)	2879 (0.473)	1.991 (1.638 to 2.428)	3.99×10^{-13}
AA	196 (0.260)	918 (0.151)	3.074 (2.446 to 3.864)	5.97×10^{-23}
Risk allele (A)	790 (0.525)	4715 (0.387)	1.749 (1.569 to 1.950)	5.90×10^{-24}
rs944289				
CC	87 (0.116)	1003 (0.164)	Reference	
CT	332 (0.441)	2924 (0.478)	1.309 (1.019 to 1.582)	0.033
TT	334 (0.444)	2193 (0.358)	1.755 (1.365 to 2.276)	4.36×10^{-6}
Risk allele (T)	1000 (0.664)	7310 (0.597)	1.330 (1.188 to 1.489)	6.95×10^{-7}

SNP, single nucleotide polymorphism.

Table 2 Thyroid cancer risk associated with different haplotypes, defined by rs965513 and rs1867277 alleles, at chromosome 9q22

Haplotype	rs965513	rs1867277	Frequency		OR	p Value
			Cases	Controls		
1	G	G	0.413	0.561	Reference	
2	A	A	0.429	0.276	2.139 (1.902 to 2.407)	2.19×10^{-35}
3	A	G	0.097	0.111	1.189 (0.978 to 1.485)	0.077
4	G	A	0.061	0.052	1.612 (1.264 to 2.037)	0.0001

Haplotype frequencies were estimated using Haploview (http://www.haploview.org/). Only samples with full data at both loci were used for the analyses (761 cases and 6085 controls).

association with TC risk (p=0.016, OR=1.14, 95% CI 1.02 to 1.27, equivalent false discovery rate=0.020). However, the association between TC and rs2910164 was not replicated ($P_{allelic}$=0.85, OR=1.01, 95% CI 0.89 to 1.14). To test if the strength of these associations were similar in cases with different histological types, we carried out cases-only interaction analyses. We found no differences between cases with papillary and follicular histology (p>0.25 for all markers, data not shown), suggesting that these associations were not restricted to any particular histological type of TC.

Are there multiple risk alleles on chromosome 9q22?

The 9q22 variants associated with TC risk, rs965513 and rs1867277, were originally reported by independent and non-overlapping GWA and candidate gene studies.[3 5] Other studies have not considered whether these SNPs represent independent signals of association. We performed unconditional logistic regression analyses incorporating rs1867277 and rs965513 genotypes as variables and sex as a covariate in the model. Both the rs965513 ($p=2.34\times10^{-13}$, OR=1.74) and rs1867277 (p=0.008, OR=1.21) association signals decreased but remained significant. We then reconstructed haplotypes at these two loci and estimated the ORs associated with having each one of three possible risk haplotypes (haplotype 2=rs965513A-rs1867277A, haplotype 3=rs965513G-rs1867277A, and haplotype 4=rs965513A-rs1867277G, table 2) compared with the non-risk haplotype (haplotype 1=rs965513G-rs1867277G). As expected, carrying the haplotype with both risk alleles (haplotype 2) increased disease risk significantly ($p=2.19\times10^{-36}$, OR=2.09). Carrying haplotypes with either one risk allele at rs965513 (haplotype 3) or at rs1867277 (haplotype 4) also increased risk, although the association signal for the haplotype 3 was weaker ($P_{haplotype\ 3}$=0.07, $OR_{haplotype\ 3}$=1.19 and $P_{haplotype\ 4}$=0.0001, $OR_{haplotype\ 4}$=1.61, table 2).

We also estimated the risk associated with 'diplotypes' at each the two 9q22 loci. Table 3 shows the genotype frequencies at the two SNPs and the ORs associated with the nine possible diplotypes. Individuals with the four risk alleles at both loci (~7.4% of the general population) had a 4.45-fold higher risk than non-carriers (~31.5% of the population), with the other diplotypes having intermediate risk levels, principally dependent on rs965513 (table 3).

These analyses showed that the two SNPs did not simply and efficiently tag a single high-risk haplotype on 9q22. However, they did not distinguish between the existence of multiple independent risk alleles at 9q22 and a third 'causal' variant tagged in complex fashion by both rs1867277 and rs965513. To undertake a limited examination of the latter possibility, we searched for SNPs in high/moderate LD ($r^2>0.5$) with both rs1867277 and rs965513 in the most recent release of the 1000 Genomes Project (phase 1, interim release, 11 May 2011, n=762 European samples). We identified four such SNPs (rs10124220, rs7848973, rs6478413, and rs1443432, supplementary table 2, supplementary figure 1). However, none of these polymorphisms lay at a site with evidence of functional importance (data not shown). We found no evidence for a role of non-synonymous variants within any of the seven nearby 9q22 genes (supplementary figure 1).

TC is associated with variation at 8q24 under a recessive model

We found evidence that rs6983267G was associated with TC in the UK population (table 1). Interestingly, and unlike previous findings in other cancer types, rs6983267G was associated with TC risk according to a recessive model (tables 1 and 4). We found no difference in risk between non-carriers and heterozygotes (OR=1.01, 95% CI 0.83 to 1.24, p=0.921, table 1), but a significantly increased risk when homozygous carriers (GG) were compared to non-carriers (p=0.016, OR=1.140, table 1), to heterozygotes (p=0.009, OR=1.27, table 1), and to non-carriers/heterozygotes (p=0.004, OR=1.26, table 4). Wokolorczyk *et al*[6] had previously found relatively weak evidence of association between TC and the rs6983267 SNP in the Polish population

Table 3 Thyroid cancer risk associated with different genotype combinations (diplotypes) at rs965513 and rs1867277

rs965513	rs1867277	Frequency (%)		OR (95% CI)	p Value
		Cases	Controls		
GG	GG	123 (0.162)	1917 (0.315)	Reference	
	AG	60 (0.079)	744 (0.122)	1.257 (0.897 to 1.747)	0.174
	AA	7 (0.009)	73 (0.012)	1.494 (0.568 to 3.333)	0.336
AG	GG	33 (0.043)	355 (0.058)	1.449 (0.939 to 2.183)	0.071
	AG	300 (0.394)	1960 (0.322)	2.385 (1.908 to 2.995)	6.62×10^{-16}
	AA	63 (0.083)	394 (0.065)	2.491 (1.774 to 3.473)	1.10×10^{-7}
AA	GG	4 (0.005)	17 (0.003)	3.663 (0.883 to 11.465)	0.036
	AG	42 (0.055)	174 (0.029)	3.759 (2.497 to 5.580)	4.44×10^{-10}
	AA	129 (0.170)	451 (0.074)	4.455 (3.379 to 5.876)	8.50×10^{-27}

The risk alleles are rs965513A and rs1867277A.

Table 4 Evidence that the association between rs6983267 and thyroid cancer risk is best explained by a recessive model

	UK only		Meta analysis of UK and Polish studies	
Test	**OR (95% CI)**	**p Value**	**OR (95% CI)**	**p Value**
GG vs (GT + TT)	1.266 (1.071 to 1.494)	0.004	1.250 (1.089 to 1.435)	7.64×10^{-4}
GG vs GT	1.262 (1.055 to 1.509)	0.009	1.215 (1.051 to 1.404)	0.004
GT vs TT	1.010 (0.827 to 1.236)	0.960	1.087 (0.933 to 1.266)	0.142

Note that the risk allele homozygotes (GG) have significantly higher frequency compared with other genotypes combined (GT + TT) and with heterozygotes (GT). This is also the case for the GG versus TT test which is not shown. However, there is no evidence that heterozygotes are over-represented in cases (GT vs TT).

(p=0.04, OR=0.78, 95% CI 0.63 to 1.00). We carried out a meta-analysis of the Polish data and our data and found enhanced support for an association between rs6983267G and TC risk, with no evidence of inter-study heterogeneity ($P_{meta}=6.64\times10^{-4}$, per allele $OR_{meta}=1.15$, 95% CI 1.06 to 1.25, $P_{heterogeneity}=0.841$, supplementary figure 2). The meta-analysis continued to support a recessive effect of the rs6983767G allele on risk (OR >1.2, p<0.004, table 4) and found no difference in risk between heterozygous and non-carriers (OR=1.087, p=0.142, table 4).

rs2910164 at the pre-miR-146a locus is not associated with TC

Using an allelic model, our study found no evidence of an association between rs2910164 and TC risk ($P_{allelic}=0.85$, table 1). We confirmed the absence of associations between this SNP and TC in genotypic, dominant, recessive and trend models (p>0.71 for all models, supplementary table 3). The previous report of an association between rs2910164 and papillary TC risk found a highly significant association between rs2910164 heterozygosity and disease ($p=7\times10^{-7}$, OR=1.62, 95% CI 1.3 to 2.0); unusually, both homozygote genotypes were protective.[4] We tested this model in our data and failed to replicate an association between rs2910164 heterozygosity and TC risk (p=0.784; supplementary table 3). When the analyses were restricted to the cases that had histologically verified papillary TC we also failed to detect association between rs2910164 and TC (p>0.784 for all models, data not shown, supplementary table 3).

Combined effects of rs6983267, rs965513, rs1867277, and rs944289 on disease risk

We carried out case-only and case-control pairwise analyses between all four risk SNPs associated with TC and found to no evidence for SNP–SNP interaction (details not shown). We then estimated the combined effects of the four SNPs on risk. To incorporate the effects of the two 9q22 markers in the combined risk analyses, we used the estimates obtained in our diplotype analyses (table 3). Using this information, we estimated that the risk for those individuals who are homozygous at 8q24, 9q22 and 14q13—comprising ~1.7% of the UK population—is 9.96-fold higher compared with individuals who do not carry any risk allele at these loci (~1.2% of the population, table 5). The risk homozygous at the four SNPs (17 cases and 41 controls, supplementary table 5) have a 17.08-fold higher chance of having TC (95% CI 3.776 to 159.323) when compared to non-risk homozygous (two cases and 84 controls, supplementary table 5).

Contribution of 8q24, 9q22, and 14q13 variants to the familial risk of TC

We have shown that four variants at 8q24, 9q22, and 14q13 are associated with a significantly higher risk of TC in the UK population. We then determined the proportion of the sibling relative risk of TC that they explained (15). Using a TC sibling relative risk of 8.6 (2), we estimated that these four risk variants explain 10.9% of the sibling relative risk of TC. Under a conservative model that assumes the existence of a single risk variant at chromosome 9q22, these loci explain at least 6.6% of the disease heritability (supplementary table 4).

DISCUSSION

Using a relatively large sample set in a single, homogeneous European population, we confirmed associations between TC and SNPs on chromosomes 9q22 (rs965513 and rs1867277), 14q13 (rs944289), and 8q24 (rs6983267). However, we failed to replicate an association between SNP rs2910164 on 5q24 and TC risk.

The 9q22 SNPs rs965513 and rs1867277 have not previously been genotyped in the same samples. We have found that there is not a single TC risk haplotype on 9q22 that is perfectly denoted by rs965513 and rs1867277. Conversely, the association cannot be explained entirely by genotypes at only one of the two SNPs, although logistic regression analysis incorporating both SNPs did lead to a considerably reduced association signal for both SNPs, particularly rs1867277. We suggest, therefore, that rs965513 and rs1867277 tag a third variant (or variants) that is the functional variation near *FOXE1*. Perhaps contrary to the genetic data, Landa *et al*[5] showed that rs1867277A affected *FOXE1* transcript levels through the differential recruitment of the USF1/USF2 transcription factors and suggested that rs1867277 was a TC-causal SNP. Fine mapping studies at this site might benefit from the use of non-European samples. For example, the LD between rs965513 and rs1867277 is significantly weaker in populations of African ($r^2=0.01$, $D'=0.078$) or Asian ($r^2=0.00$, $D'=0.014$) ancestry (data from the 1000 Genomes Project). Finally, although challenging,[16] rare variants at *FOXE1* deserve further scrutiny.

Our study validated the association between TC and rs6983267 and extended the range of cancer types associated with this variant. Interestingly, however, we found that rs6983267G is associated with TC risk in a recessive fashion; all other rs6983267 cancer associations follow an allelic dosage model. Recessive cancer predisposition SNPs have rarely been

Table 5 Estimates to genotype relative risk at rs6983267, 9q22, and rs944289

	Non-carriers	Heterozygous		Homozygous carriers	
Locus	**Population frequency**	**OR**	**Population frequency**	**OR**	**Population frequency**
rs6983267*	0.236	NA	NA	1.274	0.272
9q22†	0.315	2.385	0.322	4.455	0.074
rs944289	0.164	1.309	0.478	1.755	0.358
Combined	0.012	3.122	0.210	9.961	0.017

*rs69833267 heterozygous do not have increased risk of thyroid cancer (see tables 1 and 4).
†ORs and frequencies for the 9q22 markers were obtained from the diplotype analysis presented in table 3. Population frequency from non-carriers, heterozygous and homozygous carriers is shown in table 3.

found. In part, this probably reflects suboptimal power in GWA studies and our finding emphasises the role of candidate SNP analyses across cancer types. Given that rs6983267 itself may be functional in predisposition to colorectal and other cancers,[17] one possibility is that the true, recessive functional variation in TC is not rs6983267, but an SNP in strong LD with it.

The first association between TC and common genetic variants was found at the pre-miR-146a locus (4), a micro-RNA that is upregulated in thyroid tumours.[18] There was no good evidence in our study of an association between rs2910164 and TC risk under all models tested, including the heterozygous model that showed the disease association in the study of Jazdewski *et al*.[4] It is notable that deviations from Hardy-Weinberg equilibrium were present in the case genotypes of Jazdewski *et al*; it is not clear whether this was the result or the cause of the heterozygote association with TC risk. Other possible explanations for the differences between Jazdewski *et al*'s study and our own include chance, systematic differences between cases and controls (whether related to ascertainment or technical issues) and population specific effects in either study.

The four validated TC risk SNPs explain an approximately 10-fold differential risk between those with all high risk alleles and those with all low risk alleles. Moreover, owing to the large effect size associated with the 9q22 SNPs, the four SNPs explain over 10% of the total sibling relative risk of TC, despite the fact that TC has one of the largest familial relative risks reported for any malignancy. It is highly plausible that future studies involving only a few thousand cases and controls could identify additional important common risk variants for this common disease.

Author affiliations
[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
[2]Department of Endocrine Surgery, John Radcliffe Hospital, Oxford, UK
[3]Velindre Cancer Centre, Cardiff, UK
[4]Mount Vernon Hospital, Northwood, UK
[5]Institute of Head and Neck Studies and Education, University Hospitals of Coventry and Warwickshire, Walsgrave, Coventry, UK
[6]Guys and St Thomas' NHS Foundation Trust and King's College London, London, UK
[7]Weston Park Hospital, Sheffield, UK
[8]Kent and Canterbury Hospital, Canterbury, UK
[9]York Hospital, York, UK
[10]Bristol Hematology and Oncology Centre, Bristol, UK
[11]Norfolk and Norwich University Hospital NHS Trust, Norwich, UK
[12]Medway Maritime Hospital, Gillingham, UK
[13]St. James University Hospital, Leeds, UK
[14]St Mary's Hospital, London, UK
[15]Newcross Hospital, Wolverhampton, UK
[16]NIHR Comprehensive Biomedical Research Centre, University of Oxford, Oxford, UK

Acknowledgements We are grateful to all of the individuals who participated in the TCUKIN study. We are also grateful to the National Cancer Research Network and to the National Cancer Research Institute's Thyroid Cancer Subgroup for supporting the TCUKIN study. We acknowledge the help of the Wellcome Trust Case-Control Consortium in making control data publicly available.

Funding Cancer Research UK provided principal funding for this study. LGCC and IT receive funding from the FP7 CHIBCHA Consortium. The Wellcome Trust Centre for Human Genetics is funded by the Wellcome Trust (Grant number; 075491/Z/04).

Correction notice This paper has been corrected since it was first published online. The corresponding author's name should read Luis G Carvajal-Carmona.

Competing interests None.

Patient consent Obtained.

Ethics approval Ethics approval was provided by the Southampton and South West Hampshire Research Ethics Committee (A).

Contributors LGCC and IT are the leaders of TCUKIN and designed the study, performed the main statistical analyses and co-wrote the manuscript. AMJ, KMH, LM and MG carried out the experiments and coordinated sample collection. AA carried out some of the statistical analyses. CL, H Mehanna, H Mohan, SEMC, JW, EM, AC, MB, TR, CM, PR, GG, DP, and CB are members of the TCUKIN Consortium, coordinated sample collection, and summarised the clinical data. All authors contributed to the final version of the manuscript.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The data presented in the manuscript are available on request.

REFERENCES

1. **Kondo T,** Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nat Rev Cancer* 2006;**6**:292—306.
2. **Goldgar DE,** Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 1994;**86**:1600—8.
3. **Gudmundsson J,** Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M, Magnusdottir DN, Matthiasdottir S, Stacey SN, Skarphedinsson OB, Helgadottir H, Li W, Nagy R, Aguillo E, Faure E, Prats E, Saez B, Martinez M, Eyjolfsson GI, Bjornsdottir US, Holm H, Kristjansson K, Frigge ML, Kristvinsson H, Gulcher JR, Jonsson T, Rafnar T, Hjartarsson H, Mayordomo JI, de la Chapelle A, Hrafnkelsson J, Thorsteinsdottir U, Kong A, Stefansson K. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet* 2009;**41**:460—4.
4. **Jazdzewski K,** Liyanarachchi S, Swierniak M, Pachucki J, Ringel MD, Jarzab B, de la Chapelle A. Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. *Proc Natl Acad Sci U S A* 2009;**106**:1502—5.
5. **Ruiz-Llorente S,** Montero-Conde C, Milne RL, Moya CM, Cebrian A, Leton R, Cascon A, Mercadillo F, Landa I, Borrego S, Perez de Nanclares G, Alvarez-Escola C, Diaz-Perez JA, Carracedo A, Urioste M, Gonzalez-Neira A, Benitez J, Santisteban P, Dopazo J, Ponder BA, Robledo M. Association study of 69 genes in the ret pathway identifies low-penetrance loci in sporadic medullary thyroid carcinoma. *Cancer Res* 2007;**67**:9561—7.
6. **Wokolorczyk D,** Gliniewicz B, Sikorski A, Zlowocka E, Masojc B, Debniak T, Matyjasik J, Mierzejewski M, Medrek K, Oszutowska D, Suchy J, Gronwald J, Teodorczyk U, Huzarski T, Byrski T, Jakubowska A, Gorski B, van de Wetering T, Walczak S, Narod SA, Lubinski J, Cybulski C. A range of cancers is associated with the rs6983267 marker on chromosome 8. *Cancer Res* 2008;**68**:9982—6.
7. **Klonisch T,** Hoang-Vu C, Hombach-Klonisch S. Thyroid stem cells and cancer. *Thyroid* 2009;**19**:1303—15.
8. **Craddock N,** Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczyk K, Bowes J, Brand OJ, Braund PS, Bredin F, Breen G, Brown MJ, Bruce IN, Bull J, Burren OS, Burton J, Byrnes J, Caesar S, Clee CM, Coffey AJ, Connell JM, Cooper JD, Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Ebbs B, Eccles D, Edkins S, Edwards C, Elliot A, Emery P, Evans DM, Evans G, Eyre S, Farmer A, Ferrier IN, Feuk L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Freathy RM, Gibbs P, Gilbert P, Gokumen O, Gordon-Smith K, Gray E, Green E, Groves CJ, Grozeva D, Gwilliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassanali N, Hebaishi H, Hines S, Hinks A, Hitman GA, Hocking L, Howard E, Howard P, Howson JM, Hughes D, Hunt S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jolley JD, Jones IR, Jones LA, Kirov G, Langford CF, Lango-Allen H, Lathrop GM, Lee J, Lee KL, Lees C, Lewis K, Lindgren CM, Maisuria-Armer M, Maller J, Mansfield J, Martin P, Massey DC, McArdle WL, McGuffin P, McLay KE, Mentzer A, Mimmack ML, Morgan AE, Morris AP, Mowat C, Myers S, Newman W, Nimmo ER, O'Donovan MC, Onipinla A, Onyiah I, Ovington NR, Owen MJ, Palin K, Parnell K, Pernet D, Perry JR, Phillips A, Pinto D, Prescott NJ, Prokopenko I, Quail MA, Rafelt S, Rayner NW, Redon R, Reid DM, Renwick A, Ring SM, Robertson N, Russell E, St Clair D, Sambrook JG, Sanderson JD, Schuilenburg H, Scott CE, Scott R, Seal S, Shaw-Hawkins S, Shields BM, Simmonds MJ, Smyth DJ, Somaskantharajah E, Spanova K, Steer S, Stephens J, Stevens HE, Stone MA, Su Z, Symmons DP, Thompson JR, Thomson W, Travers ME, Turnbull C, Valsesia A, Walker M, Walker NM, Wallace C, Warren-Perry M, Watkins NA, Webster J, Weedon MN, Wilson AG, Woodburn M, Wordsworth BP, Young AH, Zeggini E, Carter NP, Frayling TM, Lee C, McVean G, Munroe PB, Palotie A, Sawcer SJ, Scherer SW, Strachan DP, Tyler-Smith C, Brown MA, Burton PR, Caulfield MJ, Compston A, Farrall M, Gough SC, Hall AS, Hattersley AT, Hill AV, Mathew CG, Pembrey M, Satsangi J, Stratton MR, Worthington J, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand W, Parkes M, Rahman N, Todd JA, Samani NJ, Donnelly P; Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010;**464**:713—20.
9. **Tomlinson I,** Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier JB, Houlston R. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 2007;**39**:984—8.
10. **Howie BN,** Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;**5**:e1000529.

11. **Carvajal-Carmona LG,** Cazier JB, Jones AM, Howarth K, Broderick P, Pittman A, Dobbins S, Tenesa A, Farrington S, Prendergast J, Theodoratou E, Barnetson R, Conti D, Newcomb P, Hopper JL, Jenkins MA, Gallinger S, Duggan DJ, Campbell H, Kerr D, Casey G, Houlston R, Dunlop M, Tomlinson I. Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. *Hum Mol Genet* 2011;**20**:2879—88.
12. **Purcell S,** Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559—75.
13. **Marchini J,** Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906—13.
14. **Barrett JC,** Fry B, Maller J, *et al*. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**:263—5.
15. **Houlston RS,** Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, Carvajal-Carmona L, Howarth K, Jaeger E, Spain SL, Walther A, Barclay E, Martin L, Gorman M, Domingo E, Teixeira AS, Kerr D, Cazier JB, Niittymaki I, Tuupanen S, Karhu A, Aaltonen LA, Tomlinson IP, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Cetnarskyj R, Porteous ME, Pharoah PD, Koessler T, Hampe J, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Chang-Claude J, Hoffmeister M, Brenner H, Zanke BW, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2008;**40**:1426—35.
16. **Carvajal-Carmona LG.** Challenges in the identification and use of rare disease-associated predisposition variants. *Curr Opin Genet Dev* 2010;**20**:5.
17. **Tuupanen S,** Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjorklund M, Wei G, Yan J, Niittymaki I, Mecklin JP, Jarvinen H, Ristimaki A, Di-Bernardo M, East P, Carvajal-Carmona L, Houlston RS, Tomlinson I, Palin K, Ukkonen E, Karhu A, Taipale J, Aaltonen LA. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 2009;**41**:885—90.
18. **He H,** Jazdzewski K, Li W, Liyanarachchi S, Nagy R, Volinia S, Calin GA, Liu CG, Franssila K, Suster S, Kloos RT, Croce CM, de la Chapelle A. The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci U S A* 2005;**102**:19075—80.

APPENDIX 1

Collaborators in the TCUKIN Study include: Dr Laura Moss, Velindre Cancer Centre, Cardiff CF14 2TL, UK; Dr Christopher Scrase, The Ipswich Hospital, Ipswich, IP4 5PD, UK; Dr Andrew Goodman, Royal Devon & Exeter Hospital, Exeter, EX2 5DW, UK; Dr Radu Mihai, John Radcliffe Hospital, Oxford OX3 9DU, UK; Dr James Gildersleve, Royal Berkshire Hospital, Reading RG1 5AN, UK; Dr Catherine Lemon, Mount Vernon Hospital, Northwood, HA6 2RN, UK; Dr Antony Robinson, Royal United Hospital, Bath BA1 3NG, UK; Dr Caroline Brammer, Newcross Hospital, Wolverhampton WV10 0QP, UK; Dr Georgina Gerrard, St. James University Hospital, Leeds LS9 7TF, UK; Professor Hisham Mehanna, Institute of Head and Neck Studies and Education, University Hospitals of Coventry and Warwickshire, Walsgrave, Coventry CV2 2DX, UK; Dr Matthew Beasley, Bristol Hematology and Oncology Centre, Bristol BS2 8ED, UK; Dr Hosahalli K Mohan and Dr Susan EM Clarke, Guys and St Thomas' NHS Foundation Trust and King's College London, London SE1 9RT, UK; Dr Kate Goodchild, Luton & Dunstable Hospital, Luton LU4 0DZ, UK; Dr Jonathan Wadsley, Weston Park Hospital, Sheffield S10 2SJ, UK; Dr Abdel Hamid, Scunthorpe General Hospital, Scunthorpe DN15 7BH, UK; Dr Danielle Power, St. Mary's Hospital, London W2 1NY, UK; Dr Elena Macias, Kent and Canterbury Hospital, Canterbury, CT1 3NG, UK; Dr Jerry Sharp, Royal Derby Hospital, Derby DE22 3NE, UK; Mr Andrew Coatsworth, York Hospital, York YO31 8HW, UK; Dr Hamish Courtney, Royal Victoria Hospital, Belfast BT12 6BA, UK; Dr Stephen Whitaker and Dr Katie Wood, Royal Surrey County Hospital, Guildford GU2 7XX, UK; Dr James McCaul, Bradford Royal Infirmary, Bradford BD9 6RJ, UK; Dr Christopher Ashford, Worcestershire Royal Hospital, Worcester WR5 1DD, UK; Dr Tom Roques and Dr Craig Martin, Norfolk and Norwich University Hospital NHS Trust, Norwich NR4 7UY, UK; Dr Vivienne Loo, Broomfield Hospital, Chelmsford CM1 7ET, UK; Dr Jennifer Marshall, Southampton General Hospital, Southampton SO16 6YD, UK; Dr Amy Roy, Derriford Hospital, Plymouth PL6 8DH, UK; Dr Joanna Simpson, The Royal Sussex County Hospital, Brighton BN2 5BE, UK; Dr Nick Rowell, Maidstone Hospital, Maidstone ME16 9QQ, UK; Mr Edward Babu, Hillingdon Hospital, Uxbridge UB8 3NN, UK; Dr Narayanan Srihari, Royal Shrewsbury Hospital, Shrewsbury SY3 8XQ, UK; Mr Simon Ellenbogen, Tameside General Hospital, Ashton-under-Lyne OL6 9RW, UK; Dr Paul Ryan, Medway Maritime Hospital, Gillingham ME7 5NY, UK; Dr Arshad Jamil, University Hospital North Staffs, Stoke on Trent ST4 6QG, UK.

Table **6 (web only) Multivariate analysis of prognostic factors- Europe only population**

Variable	Chi-Squared	Significance	Goodness of fit
Cardiac n=320			
eGFR ($< 60ml/min/1.73m^2$)	39.16	<0.001	0.9135
Hearing impairment [$]	Not included		
LVM index ($\geq 50\ g/m^2$)	13.26	<0.001	
Proteinuria (>300mg/24 hour)	11.36	<0.001	
Presence of vertigo	15.01	<0.001	
Presence of Abnormal ECG	4.29	<0.05	
Presence of Anhidrosis/Hypohidrosis	4.19	<0.05	
Presence of Angiokeratomas or telangectasias	*	<<0.0001*	

Renal n=179			
eGFR (< 60 ml/min/1.73m^2)	52.13	<0.001	0.9250
Male Gender	18.99	<0.001	
Proteinuria (>300mg/ 24 hours)	4.75	<0.05	
Presence of Angiokeratomas or telangiectasias		All patients had yes on this variable	
Neuro n=320			
Hearing impairment [$]	12.18	<0.001	0.3388
eGFR (< 60 ml/min/1.73m^2)	16.18	<0.001	
Presence of Vertigo	Not included		
Presence of Anhidrosis/ Hypohidrosis	6.90	<0.01	
Death n=320			
eGFR (< 60 ml/min/1.73m^2)	17.05	<0.001	0.5178
Presence of Abnormal ECG	7.65	<0.05	

Composite n=320			0.3048
eGFR (< 60, ml/min/1.73m^2)	74.51	<0.001	
Hearing impairment [$]	30.85	<0.001	
Microalbuminuria	5.67	<0.01	
LVM index (≥ 50 g/m^2)	7.26	<0.01	
Presence of anhidrosis/ hypohidrosis (Y/N)	Not included		
Proteinuria *(>300mg/24hour)	13.31	<0.001	

* Added

[$] Hearing impairment is average pure tune audiometric thresholds at 0.5, 1 & 2 kHz in one or both ears > 25 dB ISO

Figure 3 (web only)

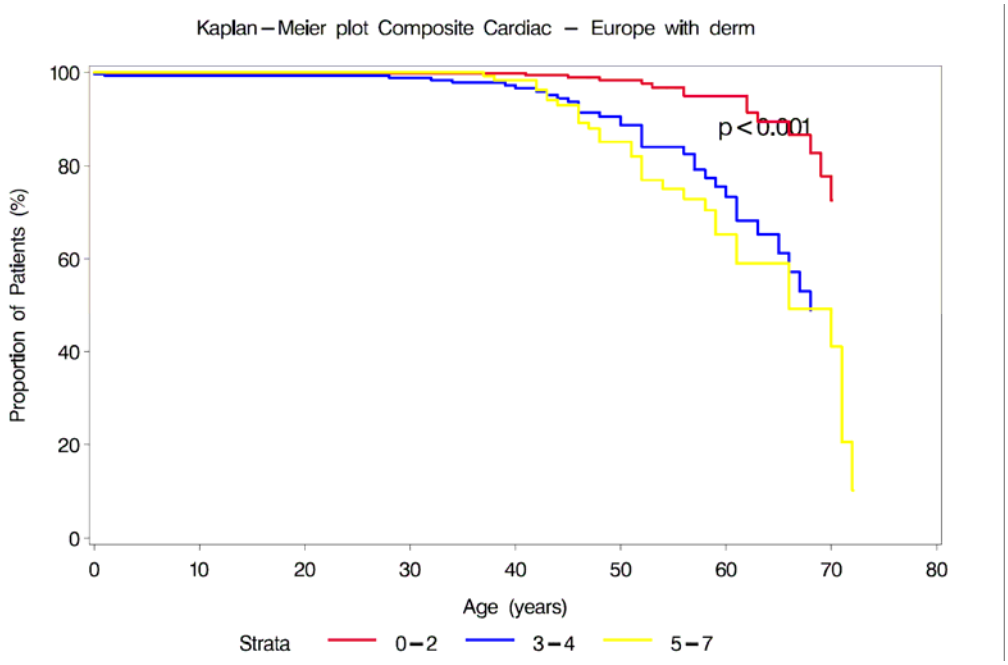
Kaplan-Meier Analysis of Predictive Scores - Europe only population

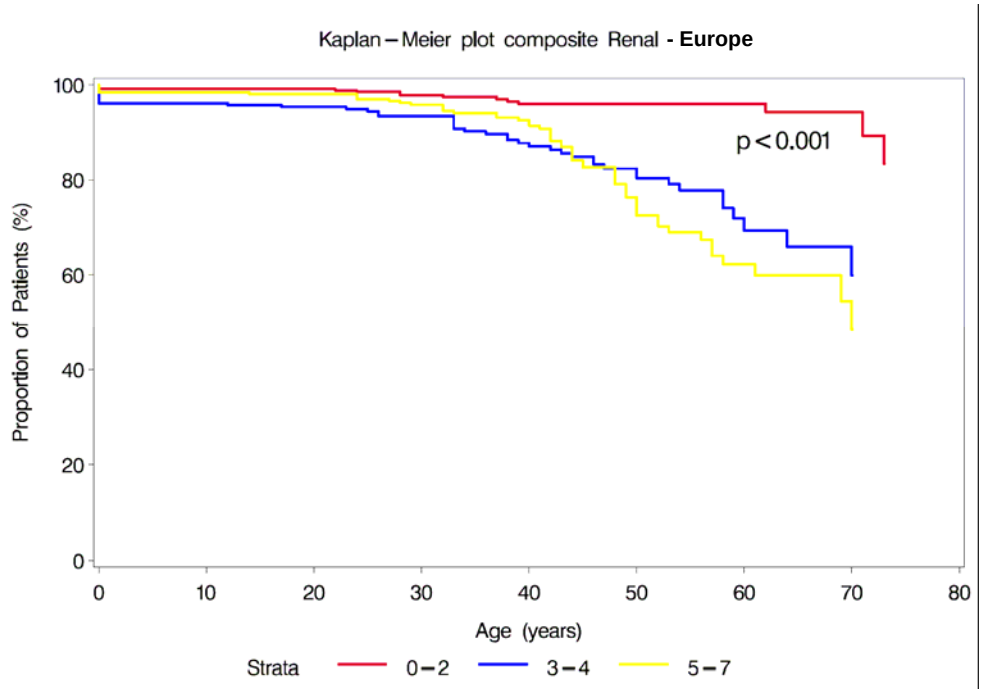
(a) Cardiac score

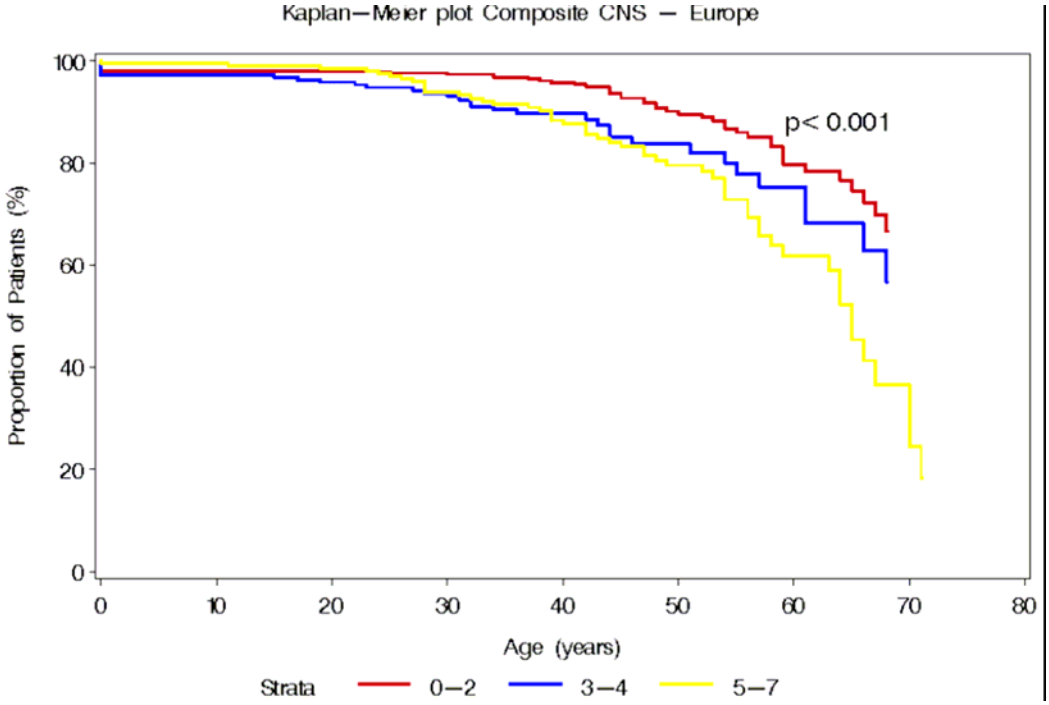
(b) Renal score

(c) Neurological

(d) Composite




Kaplan – Meier plot composite Renal - Europe
Proportion of Patients (%)
100
80
60
40
20
0
p < 0.001
0
10
20
30
40
50
60
70
80
Age (years)
Strata
0 – 2
3 – 4
5 – 7


Kaplan – Meier plot Composite CNS – Europe
Proportion of Patients (%)
100
80
60
40
20
0
p< 0.001
0
10
20
30
40
50
60
70
80
Age (years)
Strata
0 – 2
3 – 4
5 – 7

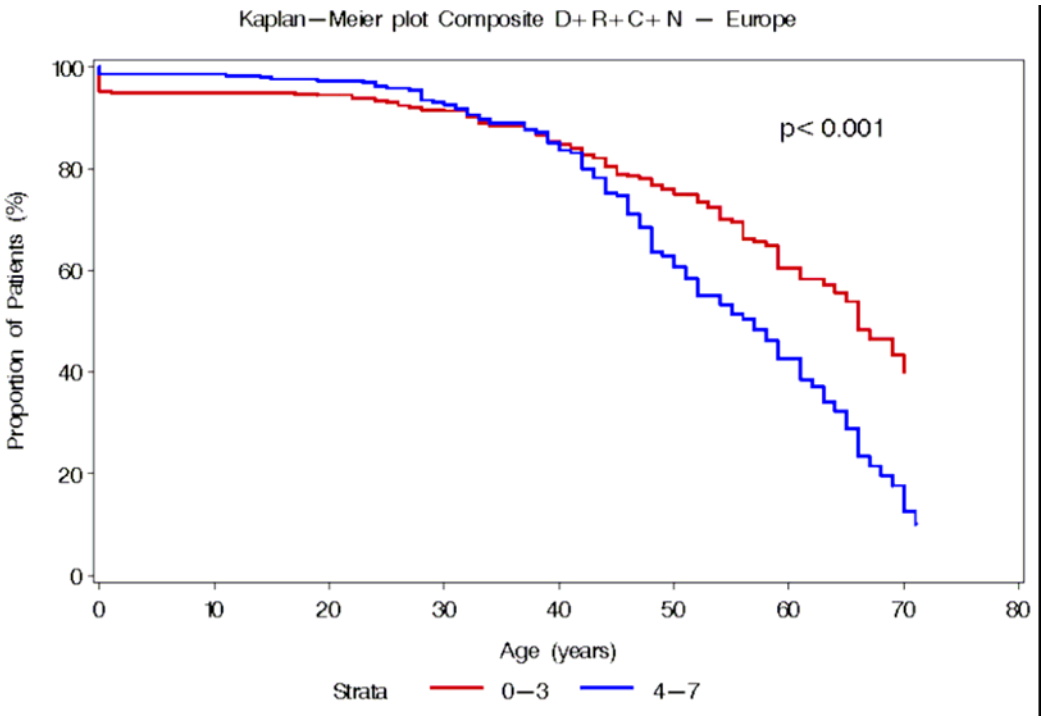


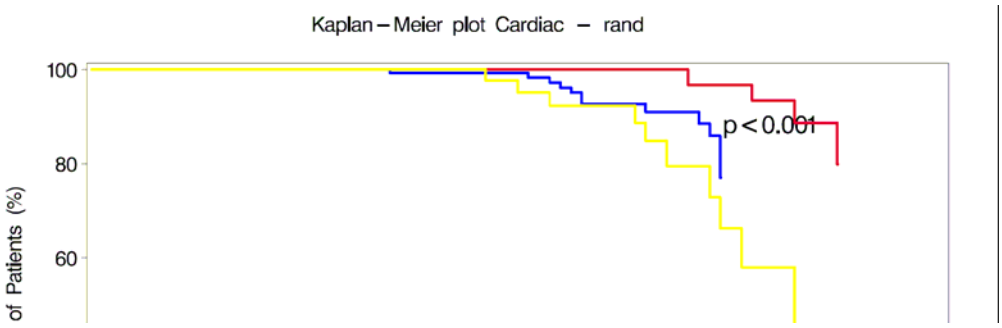
Figure 4 (web only) Kaplan-Meier Analysis of Predictive Scores- Randomly selected population

(a) Cardiac score

(b) Renal score

(c) Neurological

(d) Composite


Kaplan – Meier plot Cardiac – rand
100
80
60
of Patients (%)
p < 0.001

Supplementary Material for

Thyroid cancer susceptibility polymorphisms: confirmation of loci on chromosomes 9q22, 5q24 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24

Angela M. Jones, Kimberley M. Howarth, Lynn Martin, Maggie Gorman, Radu Mihai, Laura Moss, Adam Auton, Catherine Lemon, Hisham Mehanna, Hosahalli Mohan, Susan E.M. Clarke, Jonathan Wadsley, Elena Macias, Andrew Coatesworth, Matthew Beasley, Tom Roques, Craig Martin, Paul Ryan, Georgina Gerrard, Danielle Power, Caroline Bremmer, The TCUKIN Consortium, Ian Tomlinson and Luis G. Carvajal-Carmona

Supplementary Table 1. Kaspar probe sequences used to genotype the markers examined in this study

SNP	Probe	Sequence
rs2910164	Allele C	GAAGGTGACCAAGTTCATGCTGGTTGTGTCAGTGTCAGACCTC
	Allele G	GAAGGTCGGAGTCAACGGATTGGTTGTGTCAGTGTCAGACCTG
	Common	CGATGACAGAGATATCCCAGCTGAA
rs6983267	Allele G	GAAGGTGACCAAGTTCATGCTCATAAAAATTCTTTGTACTTTTCTCAGTGC
	Allele T	GAAGGTCGGAGTCAACGGATTCACATAAAAATTCTTTGTACTTTTCTCAGTGA
	Common	CCAGAGTTAATACCCTCATCGTCCTT
rs965513	Allele A	GAAGGTGACCAAGTTCATGCTGTGGCTGGAATGGAACAGATCAAAA
	Allele G	GAAGGTCGGAGTCAACGGATTGGCTGGAATGGAACAGATCAAAG
	Common	GTCTTTGTTAGCATTGTGAGAACAGACTA
	Allele A	GAAGGTGACCAAGTTCATGCTCCAGAGTCCAGTCCCGGTCA
	Allele G	GAAGGTCGGAGTCAACGGATTCAGAGTCCCGGTCG
	Common	GGTGCTTCTCGAGGCGGGCA
rs944289	Allele C	GAAGGTGACCAAGTTCATGCTCAATTTAATTTGGTTGAAAGATAGTCATTGC
	Allele T	GAAGGTCGGAGTCAACGGATTGCAATTTAATTTGGTTGAAAGATAGTCATTGT
	Common	GGACATTAGATTATTTTAAATTCCCAGCTA

Supplementary Table 2. 9q22 SNPs that are in high linkage disequilibrium (r^2 >0.5) with both rs1867277 and rs965513. Data from the 1000 Genomes Project Phase 1, Interim release, May 11 2011.

SNP	**Location on chromosome 9**	**r^2 with rs955513**	**r^2 with rs1877277**
rs6478413	100,582,024	0.54	0.64
rs10124220	100,583,074	0.59	0.71
rs1443432	100,583,195	0.54	0.64
rs7848973	100,588,839	0.58	0.70

Supplementary Table 3. Association between rs2910164 genotypes and thyroid cancer risk using genotypic, Cochran-Armitage trend, allelic, dominant, recessive and heterozygous disease models

Test	**Counts in all TC cases**	**Counts in All controls**	**P all cases**	**P papillary cases**
Genotypic	41/277/436	339/2179/3540	0.913	0.938
Trend	360/1156	2857/9259	0.846	0.919
Allelic	359/1149	2857/9259	0.846	0.919
Dominant	318/439	2518/3540	0.749	0.825
Recessive	41/713	339/5719	0.858	0.836
Heterozygous	277/477	2179/3879	0.708	0.784

Supplementary Table 4. Proportion of thyroid cancer heritability explained by four SNPs at chromosome 8q24, 9q22 and 14q13.

SNP	**OR per allele**[1]	**Attributable sibling relative risk**
rs6983267	1.140	0.007
rs965513	1.780 (1.981)	0.045
rs1867277	1.29- (1.749)	0.043
rs944289	1.330	0.018
Combined	3.760	0.111
No effect of rs1867277	3.995	0.066
No effect of rs965513	3.813	0.064

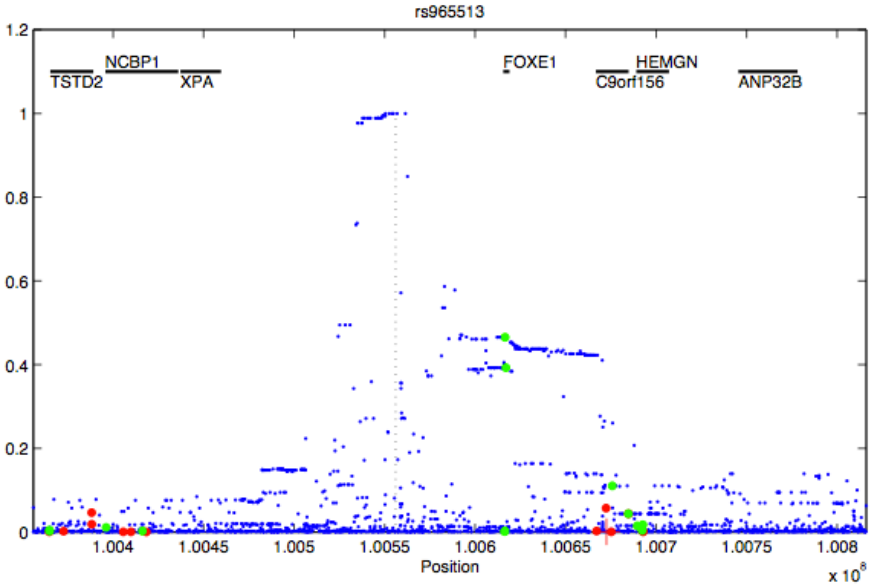
[1]. For the 9q22 markers we show their ORs obtained from the logistic regression testing when both genotypes are incorporated models and the ORs when they are analyses separately (in parenthesis)

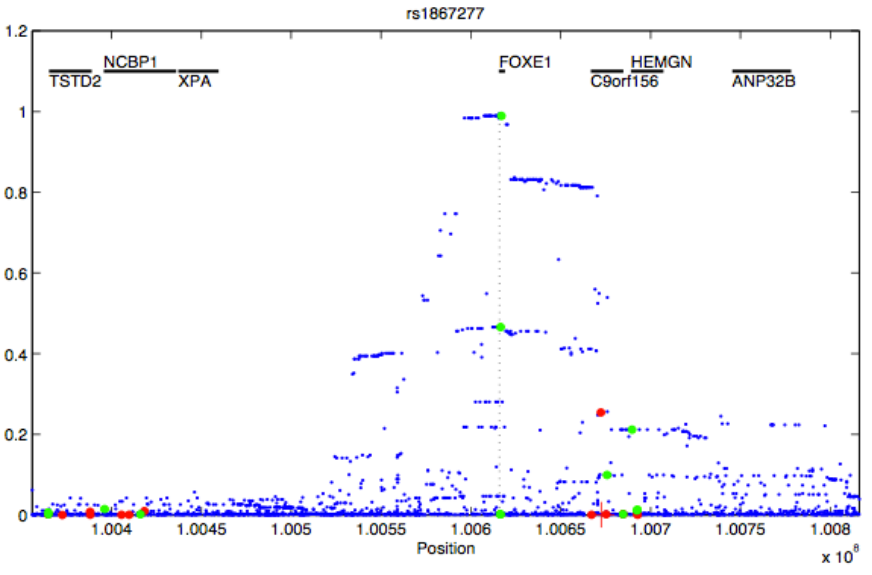
Supplementary Table 5. Number of risk alleles at rs6983267, rs965513, rs1867277 and rs944289 in thyroid cases and controls.

Number of alleles	**Cases (n=755)**	**Controls (=6076)**
0	2 (0.3%)	84 (1.4%)
1	19(2.5%)	389(6.4%)
2	61(8.1%)	1008(16.6%)
3	123(16.3%)	1439(23.7%)
4	182(24.1%)	1371(22.6%)
5	171(22.6%)	969(15.9%)
6	121 (16%)	562(9.2%)
7	59 (7.8%)	213(3.5%)
8	17(2.3%)	41(0.7%)

Supplementary Figure 1. Linkage disequilibrium plots for rs965513 (A) and rs1867277 (B). The approximate location of the seven 9q22 genes is shown on top of the figure. The vertical dotted lines in the middle of each figure indicate the position of rs965513 and rs1867277, respectively. r^2 values between either rs965513 or rs1867277 and surrounding markers are shown in the y-axis. The x-axis shows the physical positions of markers on chromosome 9q22. Synonymous variants are highlighted in green circles, non-synonymous variants are highlighted in red circles and premature stop generating variants are highlighted with red crosses.

A)





Supplementary Figure 2. Meta-analysis of associations between the rs6983267G allele and thyroid cancer