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, Kimberly 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, Luis G Carvajal-Carmona

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 x 10-39), rs1867277A (p = 5.90 x 10-29), rs944289T (p = 6.95 x 10-17), and rs6983267G (p = 0.016). Rs6983267 was most strongly associated under a recessive model (GG vs GT + TT; p = 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 x 10-13) 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.7 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 Rs944289 (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 (Thyroid Transcription Factor 2), a key gene involved in thyroid organogenesis.8 Rs1867277 and rs965513 are in moderate pairwise linkage disequilibrium (LD) in Europeans (r2 = 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, 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, and the CORGI controls had been genotyped with Illumina Hap550 arrays (Hap550, N=932) as part of our ongoing studies in colorectal cancer genetics. Two of the five SNPs (rs2910164 and rs1867277) were not included on the Illumina 1.2M and Hap550 arrays, but had excellent proxy SNPs (rs2910164 and rs1867277) were not included on 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, and the CORGI controls had been genotyped with Illumina Hap550 arrays (Hap550, N=932) as part of our ongoing studies in colorectal cancer genetics. 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, r²=1) by rs2961920, a marker present on both 9q22 and 14q13 and risk allele SNP, genotypes

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 |
|-------------------------------|--------------|-----------------------------------------------|---------|
| **rs2910164**                 |              |                                               |         |
| GG                            | 438 (0.578)  | 3540 (0.584)                                  | Reference |
| CG                            | 271 (0.367)  | 2178 (0.360)                                  | 1.032 (0.876 to 1.214) | 0.728 |
| CC                            | 41 (0.054)   | 339 (0.056)                                  | 0.987 (0.862 to 1.384) | 0.985 |
| Risk allele (C)               | 359 (0.238)  | 2857 (0.236)                                  | 1.013 (0.893 to 1.148) | 0.845 |
| **rs965513**                  |              |                                               |         |
| 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)  | 5994 (0.518)                                  | 1.140 (1.025 to 1.268) | 0.016 |
| **rs1867277**                 |              |                                               |         |
| GG                            | 187 (0.249)  | 2748 (0.449)                                  | Reference |
| AG                            | 394 (0.525)  | 2729 (0.446)                                  | 2.121 (1.763 to 2.559) | 9.08×10⁻³⁷ |
| AA                            | 170 (0.226)  | 642 (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 |
| **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) | 4.36×10⁻⁴ |
| Risk allele (T)               | 1000 (0.664) | 7310 (0.597)                                  | 1.330 (1.188 to 1.489) | 6.95×10⁻⁷ |

SNP, single nucleotide polymorphism.

Quality control and statistical analysis
General genotyping quality control assessment was as previously described. 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.

Association statistics were obtained on per allele, genotypic and haplotype bases using logistic regression models implemented in PLINK, R, and SnpTest. Haplotype analyses were carried out with HAPLOVIEW and PLINK. Allelic count association meta-analyses, using the Mantel-Haenszel method, were carried out in STATA. We used the IMPUTE2 software 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. 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.
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 (Pallele = 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. 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 rs965513 (p = 2.34 × 10⁻³⁵, 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 × 10⁻³⁵, 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 (Phaplotype 3 = 0.07, ORhaplotype 3 = 1.19 and Phaplotype 4 = 0.0001, ORhaplotype 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 (r² > 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, rs6479413, 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 = 1.14, table 1), to non-carriers/heterozygotes (p = 0.009, OR = 1.27, table 1), and to non-carriers/homozygotes (p = 0.004, OR = 1.26, table 4). Wokolorczyk et al had previously found relatively weak evidence of association between TC and the rs6983267 SNP in the Polish population.

---

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

| Haplotype | rs965513 | rs1867277 | Frequency | Cases | Controls | OR (95% CI) | p Value |
|-----------|----------|-----------|-----------|-------|----------|-------------|--------|
| 1         | G        | G         | 0.413     | 0.561 | Reference |             |        |
| 2         | A        | A         | 0.428     | 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 (1.284 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).

---

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

| rs965513 | rs1867277 | Frequency (%) | Cases | Controls | OR (95% CI) | p Value |
|----------|-----------|---------------|-------|----------|-------------|--------|
| GG       | GG        | 123 (0.162)   | 1917 (0.315) | Reference |             |        |
| AG       | GG        | 60 (0.079)    | 744 (0.122)   | 1.257 (0.897 to 1.747) | 0.174 |
| AA       | GG        | 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       | AA        | 300 (0.394)   | 1960 (0.322)  | 2.385 (1.908 to 2.995) | 6.62 × 10⁻¹⁶ |
| AA       | GG        | 63 (0.083)    | 394 (0.065)   | 2.491 (1.774 to 3.473) | 1.10 × 10⁻² |
| AA       | AG        | 4 (0.005)     | 17 (0.003)    | 3.663 (0.883 to 11.465) | 0.036 |
| AA       | AA        | 42 (0.055)    | 174 (0.029)   | 3.759 (2.497 to 5.580) | 4.44 × 10⁻¹⁰ |
| AA       | GG        | 129 (0.170)   | 451 (0.074)   | 4.455 (3.379 to 5.876) | 8.50 × 10⁻²⁷ |

The risk alleles are rs965513A and rs1867277A.
We carried out a meta-analysis of the Polish data and our data and found enhanced evidence for an association between rs6983267 and TC risk, with no evidence of inter-study heterogeneity \( (P_{	ext{meta}}=6.64 \times 10^{-4}, \text{per allele OR}_{	ext{meta}}=1.15, 95\% \text{ CI 1.06 to 1.25}, \text{P}_{	ext{heterogeneity}}=0.341, \text{supplementary figure 2}) \). The meta-analysis continued to support a recessive effect of the rs6983767G allele on risk \( (\text{OR}>1.2, p=0.004, \text{table 4}) \) and found no difference in risk between heterozygous and non-carriers \( (\text{OR}=1.087, p=0.142, \text{table 4}) \).

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 \( (\text{OR}_{	ext{allelic}}=0.85, \text{table 1}) \). We confirmed the absence of associations between this SNP and TC in genotypic, dominant, recessive and trend models \( (p>0.71 \text{ for all models, \text{supplementary table 5}}) \). The previous report of an association between rs2910164 and papillary TC risk found a highly significant association between rs2910164 heterozygosity and disease \( (p=7 \times 10^{-7}, \text{OR}=1.62, 95\% \text{ CI 1.3 to 2.0}); \) unusually, both homozygote genotypes were protective.\textsuperscript{4} We tested this model in our data and failed to replicate an association between rs2910164 heterozygosity and TC risk \( (\text{OR}=0.784; \text{supplementary table 3}) \). When the analyses were restricted to the cases that had histologically verified papillary TC we also failed to detect association between rs2910164 and TC \( (p=0.784 \text{ for all models, \text{data not shown, \text{supplementary table 3}}}) \).

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 \( (\text{table 3}) \). Using this information, we estimated the risk for those individuals who are homozygous at 8q24, 9q22 and 14q13—comprising \( \sim 1.7\% \) of the UK population—is 9.96-fold higher compared with individuals who do not carry any risk allele at these loci \( (\sim 1.2\% \text{ of the population, \text{table 5}}) \). The risk homozygous at the four SNPs \( (17 \text{ cases and 41 controls, \text{supplementary table 5}}) \) have a 17.08-fold higher chance of having TC \( (95\% \text{ CI 3.776 to 159.323, \text{table 5}}) \) when compared to non-risk homozygous \( (\text{two cases and 84 controls, \text{supplementary table 5}}) \).

| Locus | Non-carriers Population frequency | Heterozygous Population frequency | Homozygous carriers Population frequency |
|-------|----------------------------------|----------------------------------|----------------------------------------|
| rs6983267* | 0.236 | NA | NA |
| 9q22\textsuperscript{†} | 0.315 | 2.385 | 0.322 |
| rs944289 | 0.164 | 1.309 | 0.478 |
| Combined | 0.012 | 3.122 | 0.210 |

\*rs69833267 heterozygous do not have increased risk of thyroid cancer \( (\text{see \text{table 1 and 4}}) \).  
\textsuperscript{†}ORs and frequencies for the 9q22 markers were obtained from the diplotype analysis presented in \text{table 3}. Population frequency from non-carriers, heterozygous and homozygous carriers is shown in \text{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 that 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 high risk alleles and those with all low risk alleles. Moreover, owing to the large size effect associated with the 9q22 SNPs, the four SNPs explain over 10% of the total sibling relative risk of TC, despite the fact that TC is one of the largest familial relative risks reported for any malignancy. It is highly plausible that future studies involving all these cases could obtain additional important common risk variants for this common disease.

**Author affiliations** 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 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, JM, 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, Ezrat 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; 86:1600–16.

3. **Gudmundsson J, Sulem P, Budvardsson DF, Jonasson JG, Sigurdsson A, Bertholdsson JT, He H, Blondal T, Geller F, Jakobsson M, Magnusdottir DN, Matthiassdottir S, Stacey SN, Skaphdottir DS, Helgadottir H, Li W, Nagy R, Agullo E, Faune E, Prate E, Saeed B, Martinz M, Eyjolfsson JG, Bjorsdottir US, Holm H, Kristjansson K, Frigge ML, Krvinsins H, Guelcher J, Jonsson T, Rafnar T, Hjartarson H, Mayerodmo JI, de la Chapelle A, Hafknecht J, Thorsteinsson U, Kong A, Stafansson K. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations.** Nat Genet 2009; 41:461–4.

4. **Jazdzewski K, Lymanarchachi S, Swiemiat M, Pachuck J, Ringel MD, Zarab J, de la Chapelle A. Polymorphic mature micro-RNAs from passenger strand of pre-miR-146a contribute to thyroid cancer.** Proc Natl Acad Sci USA 2009; 106:1502–5.

5. **Ruiz-Llorente S, Monterro-Condé C, Milne RL, Moya CM, Cebrian A, Leton R, Cascon A, Mercadalco F, Landa I, Burrigo S, Perez de Nancalagas G, Alvarez-Escuda C, Diaz-Perez JA, Carracedo A, Urosez M, Gonzalez-Neira A, Benitez J, Santisteban P, Dopazo J, Ponder BA, Robledo M. Association study of 69 genes in the net pathway identifies low-penetrance loci in sporadic medullary thyroid carcinoma.** Cancer Res 2007; 67:9561–7.

6. **Wolkorczyk D, Glinievska B, Sikorski A, Zbrowicka E, Masoj D, Bebniak T, Matusjesci J, Mierzejewska M, Medrek K, Dzusowska B, Suchy J, Gronwald J, Teodorczyk U, Huzarski T, Bryski J, 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; 68:982–6.

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 RA, Paglino V, Robson S, Vulcovic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini J, St Cyr L, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Atwood J, Auran A, Ball SB, Balmforth AJ, Barnett GC, Barroso I, Barton A, Bennett AJ, Bhatkar S, Blasko K, Bowes J, Brand OH, Braun F, Breen F, Brown MJ, Bruce IM, Bun J, Burren OS, Burton J, Bylles E, Caesar S, Clee CM, Caffey A, Connell JM, Cooper JD, Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Elsb E, Eccles D, Edkins S, Edwards C, Elliot A, Emery F, Evans DM, Evans G, Eyre S, Farmer A, Ferrell JW, Fear L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Feathby J, Ggib P, Gilbert P, Gokumon O, Gordon-Smith K, Gray E, Green,E, Groves CJ, Grozeva D, Gwiliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassansai N, Hebahsi H, Hines S, Hinks A, Hitman GA, Hocking J, Howard E, Howard P, Howson JM, Hughes D, Hunt S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jollett JD, Jones IR, Jones IA, Kins M, Langford CF, Lango Allen H, Lathrop GM, Lee J, Lee KL, Lees D, Lindgren CM, Maisa-Muirr-Armer M, Mallier J, Mansfield J, Martin P, Massey DC, McArilde VL, Mcgillipn G, McClay K, Mentzer A, Mimmack ML, Morgan AE, Morris AF, Mowat C, Myers S, Newman W, Nimmo ER, O’Donovan MC, Oatifn A, Onyah J, Ovington NR, Owen MJ, Pain K, Parnell K, Pernet D, Perry J, Phillips A, Pinto D, Precont CJ, Prokopenko I, Quali MA, Refet S, Rangé NW, Reden R, Reid DM, Renwick A, Ring SM, Robertson N, Russel E, St Clair D, Sambrook JG, Sanderson JD, Schultuben H, Scott CE, Scott R, Seal S, Shaw-Hawkins S, Sheds BM, Simmonds MJ, Smyth DJ, Samaskantiarrajah E, Spanova K, Steer S, Stephens J, Stevens HE, Stone MA, Suz S, Symons DP, Thompson JR, Thompson W, Travis ME, Turnbull C, Valasesia A, Walker M, Walker NM, Wallace C, Warren-Perry M, Watkins NA, Webster J, Weeend MN, Wilson AG, Woodburn M, Wordsworth BP, Young AH, Zeggini E, Carter NP, Frayling TM, Lee C, McVeigan G, Munroe PB, Falotie A, Sawyer SJ, Scherer SW, Stochian DP, Tyler-Cromwell S, Brown MA, Burton FR, Cairfild MJ, Conston A, Farall M, Gough SC, Hall AS, Houlston R, Ivimey CL, James NA, Jones R, Laird NF, Le Couteux D, Lithi M, Lohmander C, Pembrey M, Satsangi J, Stratton MR, Whorton J, Deloukas P, Duncanson A, Kwiakotsinki DP, McCarthy MI, Duwehand 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, Caravajal-Carmo L, Broderick P, Kerr Z, Spain S, Peneger 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, Malik R, Logan R, Knezevic J, Johnstone E, Sieber O, Sierie D, Gray R, Thomas H, Pett J, Cazier JB, Howell S. 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, Dobkins S, Tenesa A, Farrington S, Prendergast J, Theodoratou E, Barnetson R, Cantu D, Newcomb P, Hopper JL, Jenkins MA, Gaillanger 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;81:59–75.

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. Haploviev: 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, Lube L, 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, Nittymaki 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, Topel J, Schreiber S, Volzke H, Chang-Claude J, Hoffmeister M, Brenner H, Zanke BW, Montpetit A, Hudson TJ, Gaillanger S, Campbell H, Dunlop MG. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 2010;42:1426–35.

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, Halkki O, Vanharanta S, Kivioja T, Bjoerklund M, Wei G, Yan J, Nittymaki I, Mecklin JP, Jauvinen H, Ristimaki A, Di Bernardo M, East P, Carvajal-Carmona L, Houlston RS, Tomlinson I, Pain K, Ukkonen E, Karhu A, Taijale J, Aaltonen LA. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signalling. *Nat Genet* 2009;41:865–90.

18. He H, Jazdzewski K, Li W, Livaranarachi S, Nagy R, Volinsky S, Calvin GA, Liu CG, Franssila K, Suster S, Klos RT, Croce CM, de la Chapelle A. The role of microRNA genes in papillary thyroid carcinoma. *Proc Nat 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 6PD, UK; Dr Andrew Goodman, Royal Devon & Exeter Hospital, Exeter, EX2 5DQ, UK; Dr Radu Mihai, John Radcliffe Hospital, Oxford OX3 9DU, UK; Dr James Gildersleeve, 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 Brummer, Newcross Hospital, Wolverhampton WV10 0QP, UK; Dr Georgina Gerhard, 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 Haematology 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; Dr Andrew Couttsworth, York Hospital, York YO31 8HW, UK; Dr Hamish Courtney, Royal Victoria Hospital, Belfast BT12 6BA, UK; Dr Stephen Whithaker and Dr Katie Wood, Royal Surrey County Hospital, Guildford GU2 7XX, UK; Dr James McCaul, Bradford Royal Infirmary, Bradford BD9 6JU, 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 6DY, 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 9QD, UK; Mr Edward Babu, Hillingdon Hospital, Uxbridge UB8 3NN, UK; Dr Narayanani Srinan, Royal Shrewsbury Hospital, Shrewsbury SY3 8QX, UK; Mr Simon Ellenbogen, Tameside General Hospital, Ashton-under-Lyne OL6 9RW, UK; Dr Paul Ryan, Midway Maritime Hospital, Gillingham ME7 5NY, UK; Dr Arshad Jamil, University Hospital North Staffs, Stoke on Trent ST4 6GQ, UK.