www.impactjournals.com/oncotarget/ Oncotarget, 2018, Vol. 9, (No. 16), pp: 12630-12638

Research Paper

Validation of loci at 2q14.2 and 15q21.3 as risk factors for testicular cancer

Chey Loveday[1], Kevin Litchfield[1,2], Max Levy[1], Amy Holroyd[1], Peter Broderick[1], Zsofia Kote-Jarai[1], Alison M Dunning[3], Kenneth Muir[4,5], Julian Peto[6], Rosalind Eeles[1,7], Douglas F Easton[3,8], Darshna Dudakia[1], Nick Orr[9], Nora Pashayan[10], Alison Reid[11], Robert A Huddart[12], Richard S Houlston[1] and Clare Turnbull[1,13,14,15]

[1]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

[2]Translational Cancer Therapeutics Laboratory, The Francis Crick Institute, London, UK

[3]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

[4]Division of Health Sciences, Warwick Medical School, Warwick University, Warwick, UK

[5]Institute of Population Health, University of Manchester, Manchester, UK

[6]Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

[7]Royal Marsden NHS Foundation Trust, London, UK

[8]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[9]The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK

[10]Department of Applied Health Research, University College London, London, UK

[11]Academic Uro-oncology Unit, The Royal Marsden NHS Foundation Trust, Sutton, Surrey, UK

[12]Academic Radiotherapy Unit, Institute of Cancer Research, Sutton, Surrey, UK

[13]William Harvey Research Institute, Queen Mary University, London, UK

[14]Guys and St Thomas NHS Foundation Trust, London, UK

[15]National Cancer Registration and Analysis Service, Public Health England, London, UK

Correspondence to: Clare Turnbull, **email:** clare.turnbull@icr.ac.uk

Keywords: *testicular cancer; germ cell tumor; TGCT; GWAS; oncoarray*

Received: *July 12, 2017* ***Accepted:*** *November 15, 2017* ***Published:*** *December 07, 2017*

ABSTRACT

Testicular germ cell tumor (TGCT), the most common cancer in men aged 18 to 45 years, has a strong heritable basis. Genome-wide association studies (GWAS) have proposed single nucleotide polymorphisms (SNPs) at a number of loci influencing TGCT risk. To further evaluate the association of recently proposed risk SNPs with TGCT at 2q14.2, 3q26.2, 7q36.3, 10q26.13 and 15q21.3, we analyzed genotype data on 3,206 cases and 7,422 controls. Our analysis provides independent replication of the associations for risk SNPs at 2q14.2 (rs2713206 at $P = 3.03 \times 10^{-2}$; *P*-meta = 3.92 x 10–8; nearest gene, TFCP2L1) and rs12912292 at 15q21.3 ($P = 7.96 \times 10^{-11}$; *P*-meta = 1.55×10^{-19}; nearest gene PRTG). Case-only analyses did not reveal specific associations with TGCT histology. TFCP2L1 joins the growing list of genes located within TGCT risk loci with biologically plausible roles in developmental transcriptional regulation, further highlighting the importance of this phenomenon in TGCT oncogenesis.

INTRODUCTION

Testicular germ cell tumor (TGCT) is the most common cancer in men aged between 18 and 45, with more than 52,000 men diagnosed annually worldwide [1]. Known risk factors include a family history of the disease, a previously diagnosed germ cell tumor, subfertility, undescended testis (UDT) [2] and testicular microlithiasis [3], characterized by intratesticular calcification. Histologically, TGCT can be divided into two main

subtypes: seminomas, which resemble undifferentiated primordial germ cells and nonseminomas, which show varying degrees of differentiation toward embryonal and extraembryonal structures. Some tumors show features of both classes (mixed histology). Both subtypes are thought to arise from progenitor germ cells through a pre-invasive phase of intratubular germ cell neoplasia (ITGCN) [4].

The cancer has a strong heritable basis, which is reflected in an observed four to eight -fold familial relative risk [5–8] and, from heritability analyses utilizing familial data, it has been estimated that genetic factors contribute to nearly half of all disease risk [9]. Despite the sizable heritable component, high penetrance TGCT susceptibility variants accounting for a sizeable proportion of genetic susceptibility have not been identified. We recently described enrichment in familial TGCT using exome sequencing of rare disruptive mutations in genes relating to ciliary and microtubule functions; however, these variants account for only a minor fraction of disease heritability [10]. In contrast, interrogation via genome-wide association studies (GWAS) for common variants of more modest effect size has proved vastly more fruitful, with a total of 50 independent risk loci proposed to date [11–22].

The two most recent, contemporaneously published TGCT GWAS reported a total of 26 novel TGCT susceptibility loci, more than doubling, from 24 to 49, the number of regions identified by preceding efforts [21, 22]. In Litchfield et al., (2017) [21], we performed a new GWAS in UK TGCT cases using the OncoArray platform (UK OncoArray study, 3,206 cases, 7,422 controls). These data were combined in meta-analysis with two previously published GWAS datasets from the UK and Scandinavia (2,313 cases, 11,633 controls), followed by replication genotyping performed for the most strongly associated loci (1,801 cases, 4,027 controls). In this analysis, in total comprising 7,319 cases and 23,082 controls, we identified 19 new loci associated with TGCT susceptibility (1p36.22, 2p13.3, 4q21.1, 4q35.2, 6q25.1, 7p14.1, 8p23.1, 11q24.2, 12p11.21, 12q15, 14q22.3, 15q22.31, 15q25.2, 16p13.11, 18p11.32, 19q11, 19q13.42, 20q13.2 and 22q11.21).

In the contemporaneously published study reported by Wang et al., (2017) [22], under the auspices of the TEsticular CAncer Consortium (TECAC), meta-analysis of data from five TGCT GWAS datasets (totaling 3,558 TGCT cases and 13,970 controls and inclusive of the UK/Scandinavian datasets used in Litchfield et al. 2017) identified associations at eight loci (2q14.2, 3q26.2, 4q35.2, 7q36.3, 10q26.13, 15q21.3, 15q22.31 and Xq28), two of which were also identified in the UK OncoArray study (4q35.2 and 15q22.31). Additional associations were also reported in the TECAC meta-analysis at previously established TGCT loci at 9p24.3 and 19p12.

In the present study, we sought independent evidence of replication at the five novel autosomal loci unique to the TECAC meta-analysis (2q14.2, 3q26.2, 7q36.3, 10q26.13, 15q21.3) using the UK OncoArray GWAS data.

RESULTS

The UK OncoArray GWAS includes data from 10,628 UK individuals, comprising 3,206 TGCT cases and 7,422 controls. The final number of SNPs passing quality control filters was 371, 504, which were used to impute genotypes at over 10 million SNPs. We looked for evidence of association for the five index SNPs reported in the TECAC meta-analysis using a frequentist approach under an additive model. We also performed meta-analysis combining data from the UK OncoArray GWAS and the TECAC meta-analysis using a fixed-effects model.

The strongest evidence for an association amongst the five loci in the UK OncoArray GWAS dataset was at 15q21.3 (Table 1; Figure 1). The reported index SNP from the TECAC meta-analysis, rs12912292, showed a highly significant association in the UK OncoArray dataset ($P = 7.96 \times 10^{-11}$), as did its most strongly linked directly genotyped tagging SNP (rs12899976, $r^2 > 1.0$, $D' > 1.0$, $P = 2.34 \times 10^{-11}$). Notably, SNPs in this region did show evidence of association in the meta-analysis undertaken in Litchfield et al., including rs12912292. However, due to poor phet and I^2 values associated with rs12912292, an alternative SNP (rs7175728) had been chosen for replication genotyping in 1,801 cases and 4027 controls, which failed to replicate ($P = 0.97$, OR = 0.9986). The reported index SNP at 2q14.2, rs2713206, was not significant in the UK OncoArray dataset after correcting for multiple testing (i.e. five tests), though it was significant at a nominal threshold ($P = 3.03 \times 10^{-2}$; Table 1; Figure 1). Of note, a nearby directly genotyped tagging SNP in strong LD with rs2713206 (rs2713207; $r^2 > 0.7$, $D' > 0.9$) showed a stronger level of association ($P = 9.44 \times 10^{-3}$). For the reported index SNP at 2q14.2, the point estimate for the effect size was smaller in the UK OncoArray dataset than reported in the TECAC meta-analysis, likely a reflection of "winner's curse". For the reported index SNP at both loci, genome-wide significance ($P < 5 \times 10^{-8}$) was achieved in meta-analysis of the UK OncoArray data with the constituent TECAC datasets (Table 2).

Analysis of the UK OncoArray data did not find any evidence of association with TGCT risk for the loci at 3q26.2, 7q36.3 and 10q26.13 and did not achieve genome-wide significance when combined with the TECAC data at meta-analysis (Table 1, Table 2). Directly genotyped tagging SNPs at these three regions did not show any evidence of association with TGCT risk.

Finally, we investigated whether the two SNPs showing evidence of association in the current study (rs12912292 and rs2713206) showed differences in risk allele frequency in phenotypically-defined subgroups of TGCT cases (Table 3). Neither of the two SNPs showed

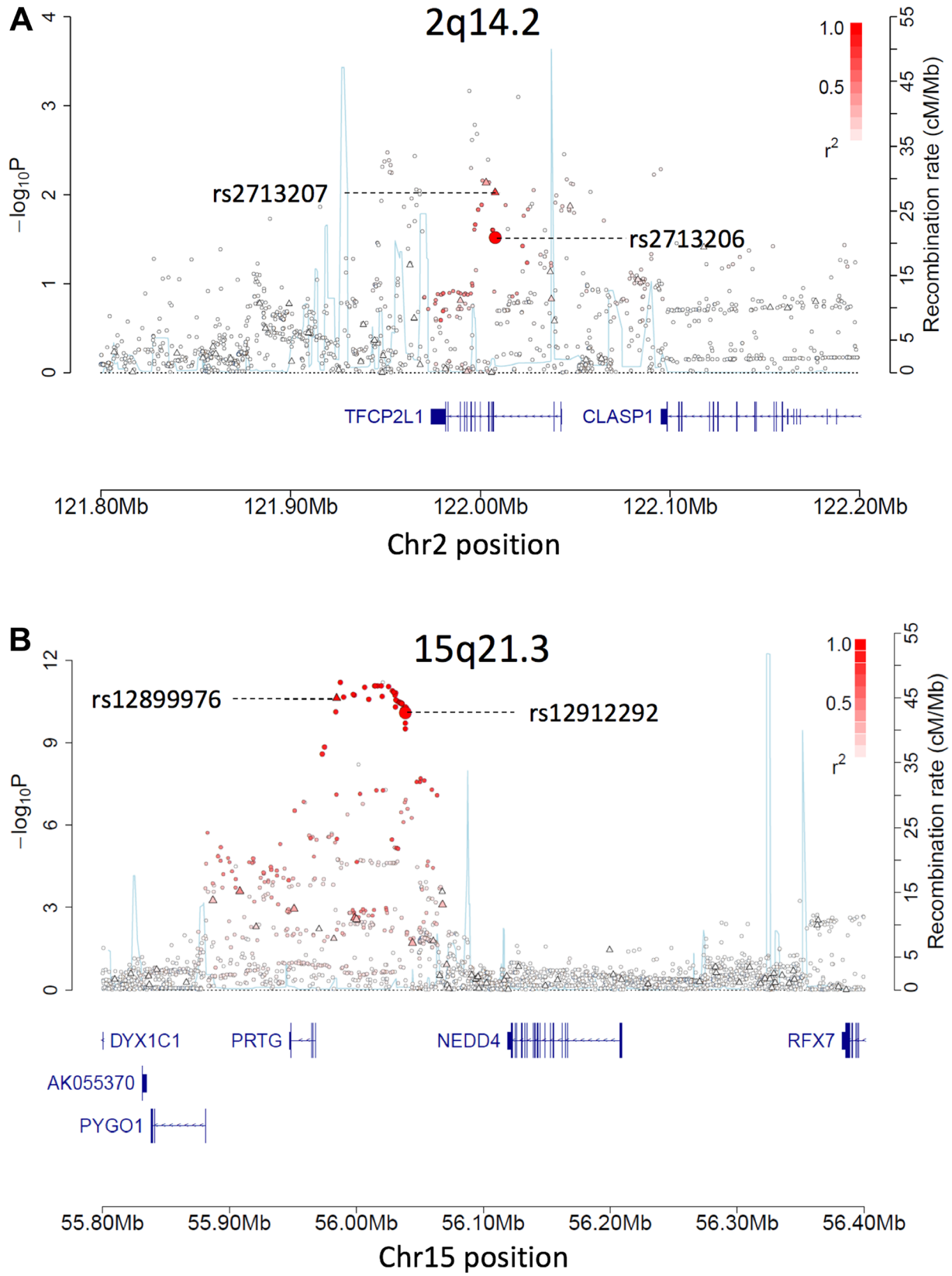

Figure 1: (**A**–**B**) Regional plots for loci 2q14.2 (A) and 15q21.3 (B) based on the UK OncoArray GWAS data. Triangles indicate directly genotyped SNPs while circles indicate imputed SNPs, with their position on the y axis indicating their -$\log_{10}$ association *P* values and their position on the x axis indicating their SNP build 37 coordinates. The intensity of red shading indicates the strength of linkage disequilibrium (LD) with the index SNP (enlarged circle). Recombination rates are plotted on the z axis (light blue). The Index SNP at each locus and its mostly strongly linked tagging SNP are labelled.

Table 1: UK OncoArray GWAS data for index SNPs identified in the TECAC meta-analysis and their strongest linked tagging SNP

Locus	Nearest gene	SNP	Type	r2	D′	Position (HG19)	INFO	Alleles	Case RAF	Control RAF	OR (95% CI)	*P*-value
2q14.2	*TFCP2L1*	rs2713206	Imputed			122007941	0.97	C/**T**	0.16	0.15	1.09 (1.01–1.18)	3.03E-02
		rs2713207	Genotyped	0.7	0.9	122007858	1.00	G/**A**	0.17	0.15	1.11 (1.03–1.20)	9.44E-03
3q26.2	*GPR160*	rs3755605	Imputed			169756119	0.99	C/**T**	0.40	0.39	1.05 (0.99–1.11)	1.32E-01
		rs7651441	Genotyped	0.6	0.8	169738278	1.00	C/**T**	0.37	0.35	1.08 (1.01–1.15)	1.81E-02
7q36.3	*NCAPG2*	rs11769858	Imputed			158501492	0.94	T/**C**	0.32	0.32	0.99 (0.92–1.05)	6.47E-01
		rs2290393	Genotyped	0.6	0.9	158438186	1.00	G/**A**	0.39	0.38	0.97 (0.91–1.03)	2.81E-01
10q26.13	*LHPP*	rs61408740	Imputed			126274612	0.99	C/**G**	0.03	0.03	0.97 (0.81–1.16)	6.72E-01
		rs1006535	Genotyped	0.2	0.8	126277624	1.00	C/**T**	0.07	0.07	0.97 (0.86–1.09)	5.98E-01
15q21.3	*PRTG*	rs12912292	Imputed			56038707	0.97	G/**A**	0.58	0.53	1.21 (1.14–1.29)	7.96E-11
		rs12899976	Genotyped	1.0	1.0	55984439	1.00	**A**/C	0.57	0.53	1.22 (1.15–1.30)	2.34E-11

RAF, risk allele frequency. Risk allele in bold and underlined. INFO, information score indicating certainty of imputation (0-1). Reported *P*-values are from a test for trend using SNPTEST.

a significant difference in frequency between cases with seminoma (n = 1,120) compared to nonseminoma/mixed histology (n = 643), cases with testicular maldescent (n = 308) compared to those with normal descent (n = 2,837), cases with a family history of TGCT (n = 53) compared to those without (n = 3,122) or cases with unilateral (n = 3,028) compared to bilateral (n = 78) disease.

DISCUSSION

In summary, we present independent evidence supporting associations between loci at 2q14.2 and 15q21.3 and susceptibility to TGCT.

rs2713206 at 2q14.2 localizes to the intron of *TFCP2L1* in an LD block of ~50 kb. TFCP2L1, a member of the CP2 family of transcription factors, is a component of a complex transcriptional network involved in the establishment and maintenance of pluripotency in embryonic stem cells. *TFCP2L1* is highly expressed in primordial germ cells during embryogenesis [23] and is downregulated during transition of fetal gonocytes into spermatogonia [24]. *TFCP2L1* is not expressed in normal adult testes, though it is in intratubular germ cell neoplasia unclassified (ITGCN, formerly known as carcinoma *in situ*, CIS) [24], a non-invasive precursor lesion from which TGCT is widely accepted to originate. As previously reported, there are eQTL variants in LD with the index SNP. The TGCT risk allele is associated with reduced *TCFP2L1* expression, supporting transcriptional regulation of this gene as the functional mechanism through which the association may be mediated [22].

rs12912292 at 15q21.3 resides in a 130 kb region of LD that only contains *PRTG* (protogenin), which encodes an immunoglobulin superfamily transmembrane protein expressed in the developing nervous system [25]. rs12912292 displays strong eQTL effects for *PRTG* in muscle-skeletal (GTEx data, P = 1.9e-13) and thyroid (P = 5.1e-12) tissues; there is, however, no evidence for association of rs12912292 with expression of *PRTG* in either normal testes or TGCT [22].

Our data did not provide evidence supporting association with TGCT risk for SNPs at three of the loci analyzed (rs3755605 at 3q26.2, rs11769858 at 7q36.3 and rs61408740 at 10q26.13). The absence of demonstrable association for these loci in the UK OncoArray dataset could be due to power, sampling error and winner's curse. Variable LD between genotyped marker and causal variants and/or their frequency, hidden population substructures not accounted for by principal component analysis, differences in effect modifiers and technical artefacts induced by the use of different genotyping platforms or quality-control criteria may also be contributory. More detailed and/or larger studies are required to further explore the observed differences.

We investigated whether the loci at 2q14.2 and 15q21.3 are associated with different risks in subgroups of TGCT cases characterized by specific phenotypic characteristics. Neither locus showed a significant difference in effect for the subtypes analyzed. For bilaterality, maldescent and family history, there was very limited power to detect a difference because of the small numbers examined. However, analysis of seminoma compared to nonseminoma was better powered, with > 95% power to detect effect difference of ≥ 1.5 fold; absence of difference of effects is consistent with observations for SNPs identified in earlier GWAS [11, 13] suggesting that, despite their distinct histological and biological features, these two subclasses of TGCT share a common biological pathway of oncogenesis.

TFCP2L1 joins the growing list of candidate genes within TGCT risk loci linked to developmental transcriptional regulation, a key disease mechanism

Table 2: Meta-analysis of the UK OncoArray GWAS and the constituent TECAC datasets

Study	INFO	Case RAF	Control RAF	OR (95% CI)	*P*-value	P_{het}	I^2(%)
rs2713206 at 2q14.2							
NCI	0.92	0.20	0.15	1.45 (1.18–1.79)	4.19E-04		
UK	0.98	0.17	0.15	1.15 (1.00–1.31)	4.43E-02		
PENN	0.90	0.20	0.16	1.32 (1.06–1.64)	1.25E-02		
Norway/Sweden	0.93	0.17	0.14	1.25 (1.09–1.44)	1.74E-03		
Denmark	0.85	0.21	0.17	1.41 (0.98–2.03)	6.39E-02		
UK OncoArray	0.97	0.16	0.15	1.09 (1.01–1.18)	3.03E-02		
Combined				1.16 (1.09–1.22)	3.92E-08	0.065	51.8
rs3755605 at 3q26.2							
NCI	0.97	0.44	0.41	1.14 (0.98–1.33)	9.31E-02		
UK	0.98	0.43	0.39	1.21 (1.09–1.33)	2.25E-04		
PENN	0.97	0.43	0.41	1.08 (0.92–1.27)	3.54E-01		
Norway/Sweden	0.98	0.44	0.40	1.24 (1.12–1.37)	2.08E-05		
Denmark	0.99	0.43	0.39	1.19 (0.91–1.55)	1.94E-01		
UK OncoArray	0.99	0.40	0.39	1.05 (0.99–1.11)	1.32E-01		
Combined				1.11 (1.06–1.15)	1.10E-07	0.041	56.7
rs11769858 at 7q36.3							
NCI	0.93	0.35	0.33	1.12 (0.94–1.32)	1.86E-01		
UK	0.95	0.35	0.31	1.19 (1.06–1.32)	1.51E-03		
PENN	0.88	0.40	0.36	1.23 (1.04–1.47)	1.69E-02		
Norway/Sweden	0.91	0.34	0.30	1.22 (1.10–1.37)	3.89E-04		
Denmark	0.90	0.36	0.32	1.22 (0.93–1.61)	1.59E-01		
UK OncoArray	0.94	0.32	0.32	0.99 (0.92–1.05)	6.47E-01		
Combined				1.08 (1.03–1.13)	1.54E-04	0.004	71.0
rs61408740 at 10q26.13							
NCI	0.99	0.04	0.02	1.68 (1.09–2.60)	1.89E-02		
UK	0.95	0.04	0.03	1.64 (1.22–2.20)	1.05E-03		
PENN	0.94	0.04	0.03	1.92 (1.22–3.03)	4.92E-03		
Norway/Sweden	1.00	0.03	0.02	1.53 (1.12–2.09)	7.79E-03		
Denmark	0.96	0.03	0.02	1.61 (0.69–3.76)	2.75E-01		
UK OncoArray	0.99	0.03	0.03	0.97 (0.81–1.16)	6.72E-01		
Combined				1.15 (1.00–1.30)	1.27E-04	0.003	72.2
rs12912292 at 15q21.3							
NCI	0.95	0.55	0.51	1.25 (1.07–1.46)	4.76E-03		
UK	0.99	0.55	0.53	1.09 (0.99–1.20)	9.18E-02		
PENN	0.95	0.56	0.47	1.44 (1.23–1.70)	8.74E-06		
Norway/Sweden	0.95	0.58	0.52	1.26 (1.14–1.39)	3.42E-06		
Denmark	0.95	0.57	0.51	1.30 (1.00–1.69)	4.81E-02		
UK OncoArray	0.97	0.58	0.53	1.21 (1.14–1.29)	7.96E-11		
Combined				1.21 (1.16–1.26)	1.55E-19	0.074	50.2

RAF, risk allele frequency. Reported *P*-value derives from the fixed-effects inverse-variance method implemented in META.

Table 3: Case-only subtype analysis of UK OncoArray GWAS for the two replicated TGCT risk SNPs

Phenotype	Subcategory	rs2713206 at 2q14.2 (C/T)	rs12912292 at 15q21.3 (G/T)
Tumor Type	Seminoma (RAF)	0.17	0.58
	Nonseminoma (RAF)	0.17	0.56
	OR (95% CI)	1.00 (0.84–1.21)	0.92 (0.80–1.06)
	P-value	0.96	0.25
Bilaterality	Unilateral (RAF)	0.16	0.58
	Bilateral (RAF)	0.16	0.57
	OR (95% CI)	0.99 (0.64–1.52)	0.98 (0.71–1.35)
	P-value	0.95	0.90
Family History	Negative (RAF)	0.16	0.58
	Positive (RAF)	0.19	0.60
	OR (95% CI)	1.20 (0.73–1.96)	1.09 (0.74–1.62)
	P-value	0.46	0.64
UDT	Absent (RAF)	0.16	0.57
	Present (RAF)	0.17	0.59
	OR (95% CI)	1.09 (0.87–1.35)	1.05 (0.88–1.24)
	P-value	0.45	0.59

RAF, risk allele frequency. Reported *P*-values are from a test for trend using SNPTEST.

implicated in TGCT oncogenesis [21]. Further functional evaluation is required to explore the cellular mechanisms through which the associations are mediated. The set of 50 GWAS loci identified to date are more strongly predictive of disease than the SNP sets for cancer types such as breast, colorectal and prostate cancer despite much larger GWAS in these cancer types having identified much greater numbers of SNPs: those in the highest centile for risk estimated from the TGCT SNP set have a relative risk of > 14 compared to the population risk [21, 26]. The continued success of GWAS in TGCT provides a strong rationale for continuing studies to identify additional risk loci via these methods.

MATERIALS AND METHODS

Sample description

TGCT cases (n = 3,206) were ascertained via two UK studies: (1) a UK study of familial testicular cancer and (2) a systematic collection of UK TGCT cases. Case recruitment was via the UK Testicular Cancer Collaboration, a group of oncologists and surgeons treating TGCT in the UK (Supplementary Note). The studies were coordinated at the Institute of Cancer Research (ICR). Samples and information were obtained with full informed consent and Medical Research and Ethics Committee approval (MREC02/06/66 and 06/MRE06/41).

Control samples for the primary GWAS were all taken from within the UK. Specifically 2,976 cancer-free, male controls were recruited through two studies within the PRACTICAL Consortium (Supplementary Note): (1) the UK Genetic Prostate Cancer Study (UKGPCS) (age <65), a study conducted through the Royal Marsden NHS Foundation Trust and (2) SEARCH (Study of Epidemiology & Risk Factors in Cancer), recruited via GP practices in East Anglia (2003–2009). 4,446 cancer-free female controls from across the UK were recruited via the Breast Cancer Association Consortium (BCAC).

GWAS

Genotyping was conducted using a custom Infinium OncoArray-500K BeadChip (OncoArray) from Illumina (Illumina, San Diego, CA, USA), comprising a 250K SNP genome-wide backbone and 250K SNP custom content selected across multiple consortia within COGS (Collaborative Oncological Gene-environment Study). OncoArray genotyping was conducted in accordance with the manufacturer's recommendations by the Edinburgh Clinical Research Facility, Wellcome Trust CRF, Western General Hospital, Edinburgh EH4 2XU.

OncoArray data was filtered as follows: we excluded individuals with low call rate (< 95%), with abnormal autosomal heterozygosity (> 3 SD above the mean) or with > 10% non-European ancestry (based on multi-dimensional scaling); we excluded SNPs with minor allele frequency < 1%, a call rate of < 95% in cases or

controls or with a minor allele frequency of 1–5% and a call rate of < 99%; we excluded SNPs deviating from Hardy-Weinberg equilibrium ($P > 10^{-12}$ in controls and 10^{-5} in cases). The final number of SNPs passing quality control filters was 371,504. These data are deposited at European Genome–phenome Archive [EGA] under accession code EGAS00001001836.

Imputation

Genome-wide imputation was performed for all GWAS datasets. The 1000 genomes phase 1 data (Sept-13 release) was used as a reference panel, with haplotypes pre-phased using SHAPEIT2 [27]. Imputation was performed using IMPUTE2 software [28].

Statistical analyses

OncoArray data tests of association between imputed SNPs and TGCT was performed under an additive genetic model in SNPTESTv2.5 [29], adjusting for principal components. Inflation in the test statistics was observed at only modest levels, $\lambda_{1000} = 1.03$ [21]. The inflation factor λ was based on the 90% least-significant SNPs [30]. The adequacy of the case-control matching and possibility of differential genotyping of cases and controls were formally evaluated using Q-Q plots of test statistics [21]. Population ancestry structure for the cohort was assessed through visualization of the first two principle components [21]; stable ancestral clustering was observed. Unadjusted case-only analyses on binary phenotypic characteristics were performed under an additive genetic model in SNPTESTv2.5, arbitrarily assigning one subdivision for a given phenotype to control status. Meta-analyses were performed using the fixed-effects inverse-variance method based on the β estimates and standard errors from each study using META v1.7 [31]. Cochran's Q-statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity was calculated [32]. Regional plots were generated using visPIG software [33]. Power calculations were performed using the methods described by Skol et al. 2006 [34], implemented via the web interface at http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html.

Author contributions

C.T., K.L. and R.S.H. designed the study. Case samples were recruited by A.R., R.A.H. and through UKTCC. R.E., A.M.D., K.M., J.P., Z.K.-J., N.P. and D.F.E. supplied OncoArray control data. N.O. administrated genotyping of OncoArray case samples. D.D. coordinated all case sample administration and tracking. K.L., M.L., A.H. and P.B. prepared samples for genotyping experiments. C.T., R.S.H. and K.L. designed bioinformatics and statistical analyses. C.L., K.L. and M.L. performed bioinformatics and statistical analyses. C.L. drafted the manuscript with assistance from C.T. and R.S.H. All authors reviewed and contributed to the manuscript.

ACKNOWLEDGMENTS

We thank the patients with TGCT and the clinicians involved in their care for participation in this study. We thank the patients and all clinicians forming part of the UK Testicular Cancer Collaboration (UKTCC) for their participation in this study. This study would not have been possible without the contributions of M.K. Bolla (Breast Cancer Association Consortium (BCAC)), Q. Wang (BCAC), K. Michailido (BCAC), J. Dennis (BCAC), P. Hall (Collaborative Oncological Gene-environment Study (COGS)); D.F. Easton (BCAC), A. Berchuck (Ovarian Cancer Association Consortium), R. Eeles (PRACTICAL), G. Chenevix-Trench (The Consortium of Investigators of Modifiers of BRCA1/2), J. Dennis, P. Pharoah, A. Dunning, K. Muir, J. Peto, A. Lee and E. Dicks. We also thank the following for their contributions to this project: J. Simard, P. Kraft, C. Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory; and K.F. Doheny and the staff of the Center for Inherited Disease Research (CIDR) genotyping facility. The results published here are based in part on data generated by the TCGA Research Network. This study makes use of data generated by the Wellcome Trust Case Control Consortium 2 (WTCCC2). We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

GRANT SUPPORT

We acknowledge National Health Service funding to the National Institute for Health Research Biomedical Research Centre. We thank the UK Genetics of Prostate Cancer Study (UKGPCS) study teams for the recruitment of the UKGPCS controls. Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) (U19 CA 148537 for Elucidating Loci Involved in Prostate cancer Susceptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008I). Additional analytical support was provided by NIH NCI U01 CA188392. The PRACTICAL consortium was supported by Cancer Research UK Grants C5047/A7357, C1287/A10118,

C1287/A16563, C5047/A3354, C5047/A10692, C8197/A16565, and C16913/A6135; the European Commission's Seventh Framework Programme grant agreement 223175 (HEALTH-F2-2009-223175) (D.F.E., R.E. and Z.K.-J.); and the NIH Cancer Post-Cancer GWAS initiative grant 1 U19 CA 148537-01 (the GAME-ON initiative). We thank the following for funding support: the Institute of Cancer Research and the Everyman Campaign, the Prostate Cancer Research Foundation, Prostate Research Campaign UK (now Prostate Action), the Orchid Cancer Appeal, the National Cancer Research Network UK and the National Cancer Research Institute (NCRI) UK. We are grateful for NIHR funding to the Biomedical Research Centre at the Institute of Cancer Research and the Royal Marsden NHS Foundation Trust. This study was supported by the Movember Foundation and the Institute of Cancer Research. K.L. was supported by a PhD fellowship from Cancer Research UK. R.S.H. and P.B. are supported by Cancer Research UK (C1298/A8362 Bobby Moore Fund for Cancer Research UK).

REFERENCES

1. Le Cornet C, Lortet-Tieulent J, Forman D, Beranger R, Flechon A, Fervers B, Schuz J, Bray F. Testicular cancer incidence to rise by 25% by 2025 in Europe? Model-based predictions in 40 countries using population-based registry data. Eur J Cancer. 2014; 50:831–9. https://doi.org/10.1016/j.ejca.2013.11.035.
2. Lutke Holzik MF, Rapley EA, Hoekstra HJ, Sleijfer DT, Nolte IM, Sijmons RH. Genetic predisposition to testicular germ-cell tumours. Lancet Oncol. 2004; 5:363–71. https://doi.org/10.1016/S1470-2045(04)01493-7.
3. Rashid HH, Cos LR, Weinberg E, Messing EM. Testicular microlithiasis: a review and its association with testicular cancer. Urol Oncol. 2004; 22:285–9. https://doi.org/10.1016/S1078-1439(03)00177-7.
4. Skakkebaek NE. Possible carcinoma-in-situ of the testis. Lancet. 1972; 2:516–7. https://doi.org/10.1016/S0140-6736(72)91909-5.
5. Swerdlow AJ, De Stavola BL, Swanwick MA, Maconochie NE. Risks of breast and testicular cancers in young adult twins in England and Wales: evidence on prenatal and genetic aetiology. Lancet. 1997; 350:1723–8. https://doi.org/10.1016/S0140-6736(97)05526-8.
6. Hemminki K, Li X. Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. Br J Cancer. 2004; 90:1765–70. https://doi.org/10.1038/sj.bjc.6601714.
7. McGlynn KA, Devesa SS, Graubard BI, Castle PE. Increasing incidence of testicular germ cell tumors among black men in the United States. J Clin Oncol. 2005; 23:5757–61. https://doi.org/10.1200/JCO.2005.08.227.
8. Kharazmi E, Hemminki K, Pukkala E, Sundquist K, Tryggvadottir L, Tretli S, Olsen JH, Fallah M. Cancer Risk in Relatives of Testicular Cancer Patients by Histology Type and Age at Diagnosis: A Joint Study from Five Nordic Countries. Eur Urol. 2015; 68:283–9. https://doi.org/10.1016/j.eururo.2014.12.031.
9. Litchfield K, Thomsen H, Mitchell JS, Sundquist J, Houlston RS, Hemminki K, Turnbull C. Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches. Sci Rep. 2015; 5:13889. https://doi.org/10.1038/srep13889.
10. Litchfield K, Levy M, Dudakia D, Proszek P, Shipley C, Basten S, Rapley E, Bishop DT, Reid A, Huddart R, Broderick P, Castro DG, O'Connor S, et al. Rare disruptive mutations in ciliary function genes contribute to testicular cancer susceptibility. Nat Commun. 2016; 7:13840. https://doi.org/10.1038/ncomms13840.
11. Rapley EA, Turnbull C, Al Olama AA, Dermitzakis ET, Linger R, Huddart RA, Renwick A, Hughes D, Hines S, Seal S, Morrison J, Nsengimana J, Deloukas P, et al. A genome-wide association study of testicular germ cell tumor. Nat Genet. 2009; 41:807–10. https://doi.org/10.1038/ng.394.
12. Kanetsky PA, Mitra N, Vardhanabhuti S, Li M, Vaughn DJ, Letrero R, Ciosek SL, Doody DR, Smith LM, Weaver J, Albano A, Chen C, Starr JR, et al. Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. Nat Genet. 2009; 41:811–5. https://doi.org/10.1038/ng.393.
13. Turnbull C, Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, Ricketts M, Linger R, Nsengimana J, Deloukas P, Huddart RA, Bishop DT, Easton DF, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat Genet. 2010; 42:604–7. https://doi.org/10.1038/ng.607.
14. Kanetsky PA, Mitra N, Vardhanabhuti S, Vaughn DJ, Li M, Ciosek SL, Letrero R, D'Andrea K, Vaddi M, Doody DR, Weaver J, Chen C, Starr JR, et al. A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. Hum Mol Genet. 2011; 20:3109–17. https://doi.org/10.1093/hmg/ddr207.
15. Turnbull C, Rahman N. Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour. Int J Androl. 2011; 34:e86–96; discussion e-7. https://doi.org/10.1111/j.1365-2605.2011.01162.x.
16. Chung CC, Kanetsky PA, Wang Z, Hildebrandt MA, Koster R, Skotheim RI, Kratz CP, Turnbull C, Cortessis VK, Bakken AC, Bishop DT, Cook MB, Erickson RL, et al. Meta-analysis identifies four new loci associated with testicular germ cell tumor. Nat Genet. 2013; 45:680–5. https://doi.org/10.1038/ng.2634.
17. Ruark E, Seal S, McDonald H, Zhang F, Elliot A, Lau K, Perdeaux E, Rapley E, Eeles R, Peto J, Kote-Jarai Z, Muir K, Nsengimana J, et al. Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. Nat Genet. 2013; 45:686–9. https://doi.org/10.1038/ng.2635.
18. Kristiansen W, Karlsson R, Rounge TB, Whitington T, Andreassen BK, Magnusson PK, Fossa SD, Adami HO,

Turnbull C, Haugen TB, Grotmol T, Wiklund F. Two new loci and gene sets related to sex determination and cancer progression are associated with susceptibility to testicular germ cell tumor. Hum Mol Genet. 2015; 24:4138–46. https://doi.org/10.1093/hmg/ddv129.

19. Litchfield K, Sultana R, Renwick A, Dudakia D, Seal S, Ramsay E, Powell S, Elliott A, Warren-Perry M, Eeles R, Peto J, Kote-Jarai Z, Muir K, et al. Multi-stage genome-wide association study identifies new susceptibility locus for testicular germ cell tumour on chromosome 3q25. Hum Mol Genet. 2015; 24:1169–76. https://doi.org/10.1093/hmg/ddu511.

20. Litchfield K, Holroyd A, Lloyd A, Broderick P, Nsengimana J, Eeles R, Easton DF, Dudakia D, Bishop DT, Reid A, Huddart RA, Grotmol T, Wiklund F, et al. Identification of four new susceptibility loci for testicular germ cell tumour. Nat Commun. 2015; 6:8690. https://doi.org/10.1038/ncomms9690.

21. Litchfield K, Levy M, Orlando G, Loveday C, Law PJ, Migliorini G, Holroyd A, Broderick P, Karlsson R, Haugen TB, Kristiansen W, Nsengimana J, Fenwick K, et al. Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor. Nat Genet. 2017. https://doi.org/10.1038/ng.3896.

22. Wang Z, McGlynn KA, Rajpert-De Meyts E, Bishop DT, Chung CC, Dalgaard MD, Greene MH, Gupta R, Grotmol T, Haugen TB, Karlsson R, Litchfield K, Mitra N, et al. Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor. Nat Genet. 2017. https://doi.org/10.1038/ng.3879.

23. Tang WW, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell. 2015; 161:1453–67. https://doi.org/10.1016/j.cell.2015.04.053.

24. Sonne SB, Almstrup K, Dalgaard M, Juncker AS, Edsgard D, Ruban L, Harrison NJ, Schwager C, Abdollahi A, Huber PE, Brunak S, Gjerdrum LM, Moore HD, et al. Analysis of gene expression profiles of microdissected cell populations indicates that testicular carcinoma *in situ* is an arrested gonocyte. Cancer Res. 2009; 69:5241–50. https://doi.org/10.1158/0008-5472.CAN-08-4554.

25. Wong YH, Lu AC, Wang YC, Cheng HC, Chang C, Chen PH, Yu JY, Fann MJ. Protogenin defines a transition stage during embryonic neurogenesis and prevents precocious neuronal differentiation. J Neurosci. 2010; 30:4428–39. https://doi.org/10.1523/JNEUROSCI.0473-10.2010.

26. Litchfield K, Mitchell JS, Shipley J, Huddart R, Rajpert-De Meyts E, Skakkebaek NE, Houlston RS, Turnbull C. Polygenic susceptibility to testicular cancer: implications for personalised health care. Br J Cancer. 2015; 113:1512–8. https://doi.org/10.1038/bjc.2015.334.

27. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2011; 9:179–81. https://doi.org/10.1038/nmeth.1785.

28. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012; 44:955–9. https://doi.org/10.1038/ng.2354.

29. 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. https://doi.org/10.1038/ng2088.

30. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JM, Faham M, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet. 2005; 37:1243–6. https://doi.org/10.1038/ng1653.

31. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, Berrettini W, Knouff CW, Yuan X, Waeber G, Vollenweider P, Preisig M, Wareham NJ, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet. 2010; 42:436–40. https://doi.org/10.1038/ng.572.

32. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002; 21:1539–58. https://doi.org/10.1002/sim.1186.

33. Scales M, Jager R, Migliorini G, Houlston RS, Henrion MY. visPIG--a web tool for producing multi-region, multi-track, multi-scale plots of genetic data. PLoS One. 2014; 9:e107497. https://doi.org/10.1371/journal.pone.0107497.

34. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006; 38:209–13. https://doi.org/10.1038/ng1706.