Published in final edited form as:
Nat Genet. 2013 June ; 45(6): . doi:10.1038/ng.2635.

Identification of nine new susceptibility loci for testicular cancer, including variants near *DAZL* and *PRDM14*

Elise Ruark[1], **Sheila Seal**[1], **Heather McDonald**[1], **Feng Zhang**[1], **Anna Elliot**[1], **KingWai Lau**[1], **Elizabeth Perdeaux**[1], **Elizabeth Rapley**[1], **Rosalind Eeles**[1,2], **Julian Peto**[3], **Zsofia Kote-Jarai**[1], **Kenneth Muir**[4], **Jeremie Nsengimana**[5], **Janet Shipley**[6], **UKTCC**[7], **D. Timothy Bishop**[5], **Michael R Stratton**[8], **Douglas F Easton**[9], **Robert A Huddart**[10], **Nazneen Rahman**[1], and **Clare Turnbull**[1]

[1]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK

[2]Royal Marsden NHS Foundation Trust, Fulham and Sutton, London and Surrey, UK

[3]Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

[4]Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK

[5]Section of Epidemiology & Biostatistics, Leeds Institute of Molecular Medicine, Leeds, LS9 7TF, UK.

[6]Molecular Cytogenetics, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK.

[7]UK Testicular Cancer Collaboration, see Supplementary Note 1.

[8]The Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK

[9]Cancer Research UK, Genetic Epidemiology Unit, Strangeways Research Laboratory, Cambridge CB1 8RN, UK.

[10]Academic Radiotherapy Unit, Institute of Cancer Research, Sutton, Surrey, SM2 5PT, UK

Abstract

Testicular germ cell tumor (TGCT) is the most common cancer in young men and is notable for its high familial risks[1,2]. To date, six loci associated with TGCT have been reported[3-7]. From GWAS

Author Contributions C.T and N.R designed the study. C.T, N.R., R.A.H. and U.K.T.C.C.. coordinated the studies that provided case samples. D.F.E., J.P., R.E, Z.K-J. and K.M. coordinated studies that provided control samples. C.T, H.M., S.S, A.E., and S.DVD coordinated sample management, selection and transfer. M.R.S., D.F.E.,D.T.B. J.N. and E.Ra. designed and executed the original case genome-wide genotyping. C.T. designed the statistical analyses. C.T. and E.Ru., conducted statistical analyses with assistance from KW.L and F.Z.. C.T. drafted the manuscript with assistance from N.R., E.Ru., J.S, J.N, D.T.B. and E.P. All authors contributed to the final paper.

Competing financial interests The authors declare no competing financial interests

URLs. Illuminus algorithm, https://www.sanger.ac.uk/resources/software/illuminus/
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/;
CaTS Power Calculator, http://www.sph.umich.edu/csg/abecasis/CaTS/;
Haploview, http://www.broadinstitute.org/haploview/haploview/;
SNAP, http://www.broadinstitute.org/mpg/snap/;
Wellcome Trust Case Control Consortium 2, http://www.wtccc.org.uk/ccc2/
COGS http://www.cogseu.org/
GenCall http://www.illumina.com/documents/products/technotes/technote_gencall_data_analysis_softwa re.pdf
NHGRI Catalog of Published Genome-Wide Association Studies, http://www.genome.gov/gwastudies/
Haploreg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php
Genevar, http://www.sanger.ac.uk/resources/software/genevar/

Europe PMC Funders Author Manuscripts

analysis of 307,291 SNPs in 986 cases and 4,946 controls, we selected for follow-up 694 SNPs, which we genotyped in a further 1,064 TGCT cases and 10,082 controls from the UK. We identified SNPs at nine new loci showing association with TGCT ($P<5\times10^{-8}$), at 1q22, 1q24.1, 3p24.3, 4q24, 5q31.1, 8q13.3, 16q12.1, 17q22 and 21q22.3, which together account for an additional 4-6% of the familial risk of TGCT. The loci include genes plausibly related to TGCT development. *PRDM14*, at 8q13.3, is essential for early germ cell specification[8] whilst *DAZL*, at 3p24.3, is required for regulation of germ cell development[9]. Furthermore, *PITX1*, at 5q31.1 regulates TERT expression, and is the third TGCT locus implicated in telomerase regulation[10].

Testicular germ cell tumor (TGCT) is the most common malignancy in men aged 15–45 years[1]. Studies in families have estimated the risk to brothers of TGCT cases to be increased 8- to 10- fold and to sons of cases 4- to 6-fold, and these risks are substantially higher than the equivalent familial relatives risk of ~2 typical for common cancers such as breast, colorectal and prostate[2]. Recent genome-wide association studies (GWAS) of TGCT have identified eight associated SNPs at six loci, which together account for >11% of the genetic risk of TGCT[3-7]. These loci lie on chromosome 12q21, encompassing *KITLG*, 5q31 (*SPRY4)*, 6p21 (*BAK1)*, 5p15 (*TERT* and *CLPTM1L)*, 12p13 (*ATF7IP*) and 9p24 (*DMRT1*)[11].

To identify further susceptibility loci for TGCT, we identified the most strongly associated 1,050 SNPs from our previous GWAS, which comprised 307,291 SNPs genotyped in 986 cases of TGCT and 4,946 controls from the UK[3,4]. These SNPs were included on a custom Illumina iSelect genotyping array (the iCOGS array). The iCOGS array includes 211,155 cancer-related SNPs selected for further evaluation in different cancers and has already yielded 80 new loci for breast, ovarian and prostate cancer[12,13]. Using the iCOGS array, we genotyped 1,094 cases of TGCT from the UK, which had not been included in the original GWAS. We compared these genotypes to those of three cancer-free UK control series, also genotyped on the iCOGS array. Following quality control exclusions, we obtained data for 694 of the 1,050 SNPs in 1,064 cases and 10,082 controls (the ‘iCOGS replication series’).

We tested association between each SNP and TGCT risk using a 1df trend test, adjusted for six principle components. Inflation in the test statistic, quantified using 36,879 uncorrelated non-TGCT-related SNPs, revealed evidence of only modest inflation of the test statistics (Supplementary Figure 1, λ=1.14, λ_{1000}=1.07; following adjustment for six principle components, λ=1.05, λ_{1000}=1.02). There was clear evidence of association of the six previously reported loci in the iCOGS replication series ($P_{trend}= 3.88\times10^{-9}$ to 9.31×10^{-28}), with odds ratios consistent with those previously published (Supplementary Table 1).

In a combined fixed-effects meta-analysis of the GWAS and the adjusted iCOGS replication for the remaining 640 SNPs outside of these six loci, we identified association of $P<5\times10^{-8}$ for a further 18 SNPs at nine loci: 1q22, 1q24.1, 3p24.3, 4q24, 5q31.1, 8q13.3, 16q12.1, 17q22 and 21q22.3. Following multiple logistic regression to test for independence of effects, there remained one SNP at each of the nine loci showing association with TGCT (Table 1). The associations at 1q24.1 and 17q22 have been replicated concurrently in other independent TGCT series[14,15].

For each SNP, we examined for evidence of departure from a log-additive (multiplicative) model: only rs2072499 at 1q22 showed evidence for departure from the log-additive model, indicating a recessive effect (heterozygote OR (OR_{het}): 1.00 (0.87-1.16), homozygote OR (OR_{hom}): 1.53 (1.27-1.84), $P_{2df}=5.18\times10^{-6}$, Supplementary Table 2). This observation was supported by genotypic (2df) analysis of the GWAS data for rs2072499 (OR_{het} 1.09 (0.94-1.26) OR_{hom} 1.64 (1.24-2.01), P_{2df} 1.25 $\times10^{-6}$). We found no evidence for statistical interaction between any pair of these new nine SNPs or with the previous eight SNPs; for

every combination of SNPs, the combined risk was consistent with the product of the individual risks.

We investigated whether the newly identified loci were associated with different risks in subgroups of TGCT cases characterized by specific phenotypic characteristics (Supplementary Tables 3,4,5). Consistent with our findings at the previous six loci, none of the newly identified loci showed a significant difference in effect when comparing cases of the two TGCT histological subtypes, seminoma and nonseminoma. Likewise, there was no clear evidence of differential association by age of diagnosis, presence of testicular maldescent, family history of TGCT or unilateral versus bilateral disease, although power to detect differences in effects was limited on account of the case distribution.

Encyclopedia of DNA Elements (ENCODE) data indicated that rs2072499, rs3790672, 3805663, rs7010162, rs9905704 and rs2839186 may have effects on transcription on account of evidence of location within a DNase I hypersensitivity site, local modification of histones indicating promoter/enhancer activity, influence on binding of transcription factors and/or effect on a regulatory motif. Functional annotations from the ENCODE data for all correlated variants within each of the nine LD block (r2>0.8), are shown in Supplementary Table 6. We then investigated whether the genotype at these nine SNPs was associated with differential expression of genes lying within 500kB of the SNP: there were no compelling patterns evident from expression data available from fibroblast, skin, adipose and lymphoblastoid cells (Supplementary Table 7). Further examination in testicular germ cells both for evidence of transcriptional regulation and of gene expression would be of interest.

Evidence supports TGCT developing from primordial germ cell or gonocytes, with tumor initiation described *in utero* through a pre-invasive stage termed intra germ cell neoplasia unclassified (IGCNU) or carcinoma *in situ* (CIS)[16]. It is therefore noteworthy that a number of the genes identified at loci identified in this study relate to pathways of early differentiation and development of the testicular germ cell[11] (Table 1).

rs10510452 lies in an LD block of 82kb at 3p24.3 which contains only one gene, *DAZL* (deleted in azoospermia-like). *DAZL* encodes an RNA-binding protein which has been shown from expression studies in human primordial germ cells to have a central role in early differentiation of primordial germ cells[17]. *Dazl* knockout mice are infertile, with differentiation of the germ cells halted at the A spermatogonia level [9] and expression of *DAZL* is testis-specific [18,19]. Sequence analysis indicates that the DAZ cluster on the Y-chromosome arose during primate evolution via transposition of the *DAZL* gene[20]. Of note, *DAZ* has been implicated as the critical gene within the Y-chromosome AZF (azospermia factor) region, so-named as deletions of this region cause non-obstructive spermatogenetic failure (azoospermia) in humans[21].

rs7010162 lies in an LD block of 61kb at 8q13.3 in which there is one gene, *PRDM14*, which is required for both the reacquisition of potential pluripotency and the genomewide epigenetic reprogramming necessary for primordial germ cell specification[8,22,23]. *PRDM14* encodes a transcriptional regulator which controls expression of key pluripotency genes such as *POU5F1* (*OCT4*), *NANOG* and *SOX2,* which are highly expressed in IGCNU and TGCT[24,25]. Whilst normal PRDM14 expression is restricted to germ cell and stem cell lineages, amplification and/or over-expression of *PRDM14* have been reported in a number of different cancer types, including TGCT[26-29].

rs3805663 is located in a 51kb LD block at 5q31.1 which contains two genes: *CATSPER3* and *PITX1*. *CATSPER3* (Cation-Channel, Sperm-associated 3*)* is one of a family of genes which together form a functional hetero-tetrameric cation channel, exclusively expressed in the testis and essential for hyperactivated motility of the differentiated germ cell

(spermatozoa); accordingly *Catsper*3−/−male mice are completely infertile [30-32]. An equally plausible candidate gene at this locus is *PITX1*, (paired-like homeodomain transcription factor 1), which binds to specific *PITX1*-binding sites in the promoter region of *TERT*, thus regulating expression of telomerase[10]. Thus, along with the previously identified TGCT-associated loci at 5p15 (*TERT*) and 12p13 (*ATF7IP,* which regulates expression of *TERT*), the signal at 5q31.1 may also be mediated via effects on telomerase regulation. Whilst the 5p15 (*TERT)* region is notable for its associations with multiple cancers[33-39], to date there have been no associations reported with other cancers for the loci at 5q31.1 (*PITX1*) or 12p13 (*ATF7IP).*

rs9905704 lies at 17q22 in an LD block of 813kB in which there are several genes including *RAD51C* and *TEX14* (testis expressed 14). *TEX14* is a protein kinase highly expressed in the human male germ cell [40,41]. Male Tex14 −/− mice are infertile whilst the female knockouts are normally fertile[42]. Cellular studies have demonstrated that TEX14 regulates kinetochoremicrotubule assembly and spindle assembly checkpoint in the germ cells of the testis and requires recruitment of key proteins such CENPE, a kinesin-like motor protein which associates with the centromere [41-43]. It is thus noteworthy that *CENPE* (Centromere-associated protein E) is located within the 214kb TGCT-associated LD block at 4q24. Furthermore, *PMF1*, one of the two genes located in the TGCT-associated region at 1q22, is part of the MIS12 complex which is required for normal chromosome alignment, segregation and kinetochore formation during mitosis[44]. Therefore, regulation of microtubule assembly may be the mechanism linking the signals at 17q22, 4q24 and 1q22.

In summary, we have identified nine new loci for TGCT, which brings the total number of TGCT-associated loci identified to date to fifteen. These fifteen loci have provided considerable new insights into TGC tumourigenesis, implicating genes involved in germ cell specification and/or differentiation (*DAZL, PRDM14*) including the KIT-KITL signalling pathway (*KITL, SPRY4, BAK1*) and genes involved in sex-determination (*DMRT1*), microtubule assembly (*TEX14, CENPE, PMF1*) and telomerase regulation (*TERT*, *ATF7IP, PITX1).* The nine new susceptibility alleles account for ~4% of the excess familial risk to brothers and ~6% to sons of men with TGCT: this brings the cumulative totals to ~15% and ~22% respectively. The power to detect these loci in this combined analysis was modest (29-84%): accordingly there are likely to be several additional loci for TGCT of equivalent or lesser effects, which may be identifiable via still larger follow-up studies of these data and meta-analyses with other GWAS data. Moreover, imputation and fine-scale association mapping of the loci may uncover more of the missing heritability, either through identifying more strongly associated variants or by identifying additional signals, including rare variants, at these loci. Integration of these genetic variants together with non-genetic TGCT risk factors into risk models may enable clinically useful risk profiling for TGCT in unaffected men; which in turn may provide a rational basis for application of strategies for screening and targeted prevention of TGCT.

Methods

Samples

See Supplementary Note

GWAS analysis

Cases were genotyped on the Illumina HumanCNV370-Duo bead array and controls were genotyped on the Illumina Infinium 1.2M array at the Wellcome Trust Sanger Institute (see [3] and [4] for full methods). The genotypes were re-called since the analysis reported in [4], such that SNPs on both sex chromosomes were included in the current analysis. We used data on

314,861 SNPs that were successfully genotyped on both arrays. We excluded individuals: with low call rate (<95%), with abnormal autosomal heterozygosity or with >10% non-European ancestry (based on multi-dimensional scaling). We filtered out all SNPs with (i) minor allele frequency <1%, (ii) a call rate of <95% in cases or controls or (iii) minor allele frequency of 1–5% and a call rate of <99% or (iv) deviation from Hardy-Weinberg equilibrium (10^{-12} in controls and 10^{-5} in cases). 307,291 SNPs passed the QC filters and we put forward for replication the 1050 SNPs with the lowest P-value ($P<1.3 \times 10^{-3}$), based on the 1-d.f. Cochran-Armitage trend test.

iCOGS Genotyping

Genotyping was conducted using a custom Illumina Infinium array (iCOGS array) comprising 211,155 SNPs selected across multiple consortia within the COGS (Collaborative Oncological Gene-environment Study, see [12] for details of SNP inclusion). Of the 1,050 SNPs submitted from our analysis, 740 attained an Illumina design score of ≥0.8, and were included on the array. Genotyping was performed at Genome Quebec, one of the seven genotyping centres providing genotyping for the COGS collaboration. Genotypes were called using Illumina's proprietary GenCall algorithm. Pan-COGS initiatives were employed across experiments to optimize calling algorithms for the iCOGS array: firstly, initial calling employed a cluster file generated using samples from Hapmap2; secondly, a cluster file based on 3,018 individuals selected across genotyping centers, across participating consortia and across ethnicities, combined with 380 samples of European, Asian or African ancestry genotyped as part of the Hapmap and 1000 genomes project was applied to call the genotypes for the subsequent samples. Thirdly, the performance of GenCall was compared to two other calling algorithms: Illuminus54 and GenoSNP55: all three algorithms were >99% concordant in their calling for 91% of the SNPs on the array. Manual inspection of a sample of the discrepant SNPs indicated that the GenCall calls were almost invariably superior (because Illuminus and GenoSNP frequently attempted to call SNPs that clustered poorly) [12].

Quality Control

Using the full SNP set of 211,155 SNPs on the iCOGS array, we applied quality control exclusions as follows to subjects: i) subjects with overall call rate <95% or low or high heterozygosity ($P<10^{-6}$) (5 cases), ii) using identity-by-state estimates based on 37,046 uncorrelated SNPs, we identified "cryptic" duplicates and related samples and the sample with the lower call rate was excluded (7 cases), iii) we identified ethnic outliers by multi-dimensional scaling by combining the iCOGS data with the three Hapmap2 populations using 37,046 uncorrelated markers and removed individuals with >10% non-Western European ancestry (18 cases). We included 1,064 cases and 10,082 controls in the final analysis.

We applied quality control exclusions as follows to SNPs: i) discrepant calls in more than 2% of duplicate samples across COGS consortia, ii) call rate <95%, MAF<1%, call rate <99% if MAF=1-5%, iii) deviation from Hardy-Weinberg ($P<10^{-5}$ in controls, $P<10^{-12}$ in cases). Following quality control exclusions applied individually to case and control data, we included genotypes from 694 SNPs in subsequent analyses. We checked genotype intensity cluster plots manually for SNPs in each new region for which a genome-wide significant association was achieved.

Statistical Analysis

For the 694 SNPs in the TGCT iCOGS replication experiment, we estimated the per-allele odds ratio (OR), 1df, using logistic regression, adjusting for principal components as covariates. We generated eight principle components and examined reduction in the

inflation factor using iterative combinations of principle components. Inclusion of principle components 1-6 achieved optimal reduction in inflation with no further reduction achieved by incorporating additional principle components into the regression model. We found no significant differences in the point estimates of effect size on performing sensitivity analysis between the control groups used in the TGCT iCOGS replication experiment.

The inflation factor (λ) was derived by dividing the median of the lowest 90% of the 1-d.f. statistics by the 45% percentile of a 1-d.f. $\chi 2$ distribution (0.357), utilizing 36,879 uncorrelated SNPs selected to not include TGCT-associated SNPs. The inflation was converted to an equivalent inflation for a study with 1000 cases and 1000 controls ($\lambda_{1,000}$):

$$\lambda_{1,000}=1+500(1/\mathrm{N}_{\mathrm{cases}}+1/\mathrm{N}_{\mathrm{controls}})^{\star}(\lambda-1)$$

We excluded from further analysis the 54 SNPs lying within the six previously identified TGCT-associated loci, defining these LD blocks via the Oxford recombination rates. We included the 640 remaining SNPs in the replication analysis.

Because the inflation factor in the GWAS data had been modest (λ = 1.078, λ_{1000} = 1.045), we did not adjust the analysis of the GWAS data. We obtained overall significance levels by combining the estimates from the GWAS (unadjusted) and the iCOGS replication (adjusted for principle components 1-6) using a fixed effects meta-analysis, to derive a 1df test, calculating an I^2 statistic to evaluate heterogeneity between the GWAS and replication. We used a threshold of 5×10^{-8} to denote genome-wide significance. Of note we had previously examined the loci at 1q24.1 and 4q24 in our first GWAS replication experiment[3]: whilst there had been evidence of replication of these signals via Taqman genotyping of 565 cases and 1,758 controls, these associations had failed to achieve genome-wide significance.

We also computed genotype-specific ORs (2 df) for the iCOGS replication, adjusted for principle components 1-6 (Supplementary Table 2). We assessed each SNP for dose response ($P_{\mathrm{hetergeneity}}$) in the replication series by comparing 1-d.f. and 2-d.f. logistic regression models,both adjusted for principle components 1-6, using a likelihood ratio test (evaluated using a significance threshold of $P_{\mathrm{hetergeneity}} < 0.006$ to account for 9 tests). We examined for statistical interaction between TGCT predisposition SNPs (nine new SNPs and the previously published eight SNPs) by evaluating the effect of adding an interaction term to the regression model, adjusted for stage, using a likelihood ratio test (using a significance threshold of $P < 0.0003$ to account for 144 tests).

We assessed for modification of the odds ratios by covariate phenotype and risk factors using a case-only analysis incorporating both GWAS and iCOGS replication data (unadjusted). We evaluated the effect of age on SNP genotype via polytomous regression: we divided age at diagnosis into 5 ordinal categories and fitted maximum-likelihood multinomial logistic models, executed using the mlogit command in Stata12 (using a significance threshold of $P < 0.006$ to account for 9 tests, Supplementary Tables 3, 4, 5).

We identified variants from within each LD block reported within the 1000 genomes project (r2>0.8 and <200kB from sentinel SNP) and used HaploReg and ENCODE data to apply functional annotations relevant to the regulation of transcription: (i) whether the variant lies in a region in which modification of histone proteins is suggestive of enhancer and other regulatory activity (*H3K4Me1* and *H3K27A* histone modification) or promoter activity (*H3K4Me3* histone modification), (ii) whether the variant lies in a region where the chromatin is hypersensitive to cutting by the DNase enzyme (suggestive of regulatory region), (iii) whether the variant lies in a region of binding of transcription factor proteins

(as assayed by chromatin immunoprecipitation with antibodies specific to the transcription factor followed by sequencing of the precipitated DNA (ChIP-seq)), (iv) whether the variant affects a specific regulatory motif, as evaluated from position weighted matrices assembled from TRANSFAC, JASPAR and protein-binding microarray experiments (Supplementary Table 6)[45,46].

We investigated associations of the nine TGCT-associated SNPs with gene expression using GENEVAR, which includes gene expression profiling and genotypic data from four datasets: (i) lymphoblastoid cell lines (LCL) from 726 HapMap individuals adipose (HapMap3), (ii) LCL and skin collected from 856 healthy female twins of the MuTHER resource (Muther pilot Twin 1), (iii) adipose, LCL and skin derived from a subset of 160 MuTHER healthy female twins (Muther pilotTwin2) and (iv) fibroblast, LCL and T-cells derived from umbilical cords of 75 Geneva GenCord individuals (Gencord). Spearman Rank Correlation between normalized gene expression levels and the count of one of the alleles of the SNP (0, 1 or 2) was analysed with significance assessed by permutation (10,000 permutations). Expression of genes within 500kB of the sentinel SNP was examined (using a significance threshold of $P < 0.00001$ to account for 440 tests, Supplementary Table 7).

LD matrices between SNPs reported in HapMap were based on Data Release 27/phaseII + III (Feb 2009) on NCBI B36 assembly, dbSNP b126, viewed using Haploview software (v4.2) and plotted using SNAP. LD blocks were evaluated using the HapMap recombination rates (cM/Mb) and defined using the Oxford recombination hotspots[47]. GWAS associations for other diseases were evaluated using the NHGRI (National Human Genome Research Institute) Catalog of Published Genome-Wide Association Studies. All genomic references are based on NCBI Build 36. Analyses were performed using R (v2.6), Stata12 (State College) and PLINK (v1.07) software.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank subjects with TGCT and the clinicians involved in their care for participation in this study. We also thank M. Warren Perry, D. Dudakia, J. Pugh, R. Linger, J. Marke, D. Hughes and D. Pernet for recruitment of subjects and database entry for the TGCT collections and the UKGPCS study teams for recruitment of the UKGPCS controls. We acknowledge National Health Service funding to the National Institute for Health Research Biomedical Research Centre. The COGS research initiative, leading to these results has received support from the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175), Cancer Research UK (Grants C1287/A10710 and C5047/A7357) and Prostate Action. This study makes use of data generated by the Wellcome Trust Case Control Consortium (WTCCC) 2. A full list of the investigators who contributed to the generation of the data is available from the WTCCC website. D.F.E. is a Principal Research Fellow of Cancer Research UK. The initial GWAS study was supported by the Institute of Cancer Research, Cancer Research UK and the Wellcome Trust. This study was supported by the Institute of Cancer Research and Movember.

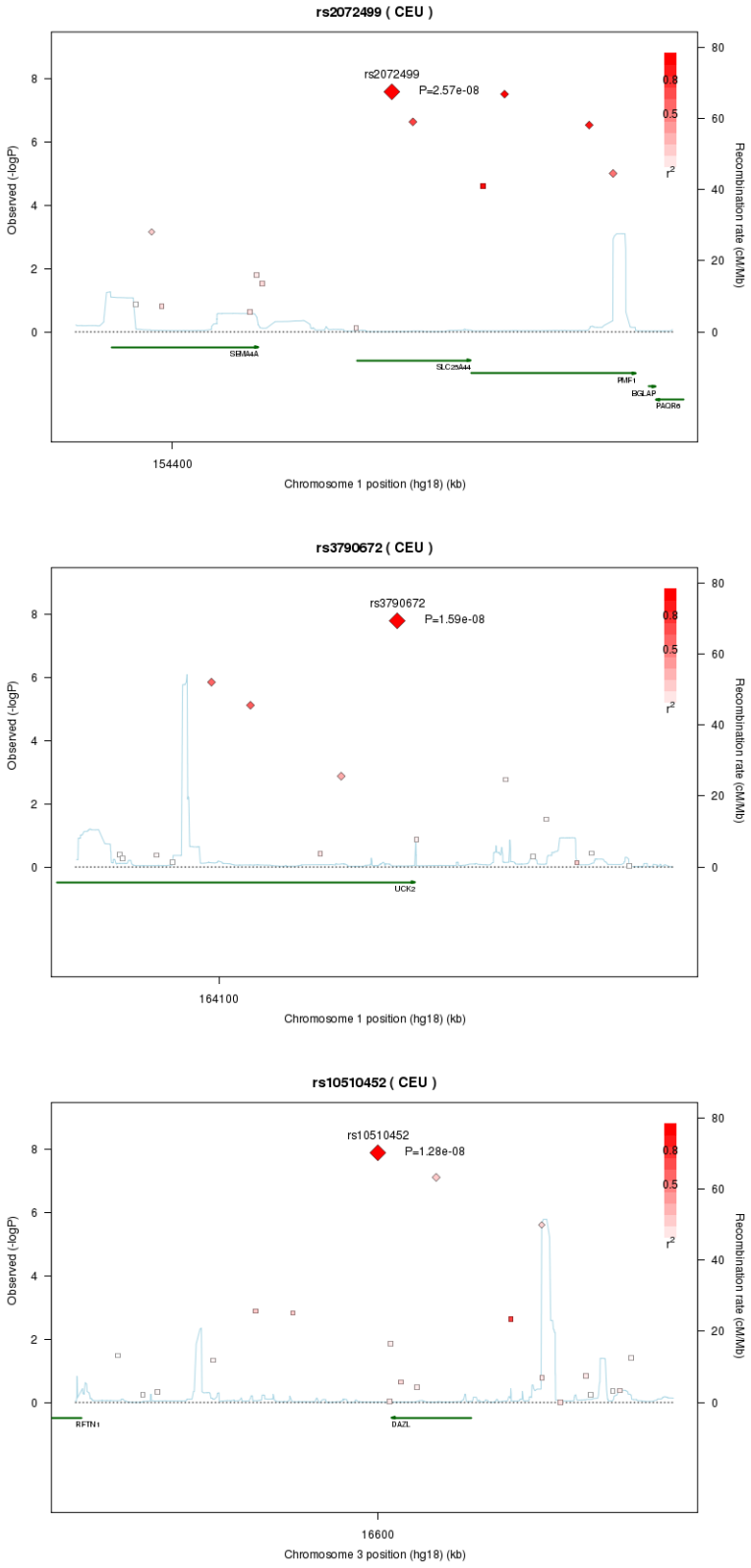
References

1. Bray F, Ferlay J, Devesa SS, McGlynn KA, Moller H. Interpreting the international trends in testicular seminoma and nonseminoma incidence. Nat.Clin.Pract.Urol. 2006; 3:532–543. [PubMed: 17031378]
2. Hemminki K, Li X. Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. Br.J.Cancer. 2004; 90:1765–1770. [PubMed: 15208620]
3. Rapley EA, et al. A genome-wide association study of testicular germ cell tumor. Nat.Genet. 2009; 41:807–810. [PubMed: 19483681]

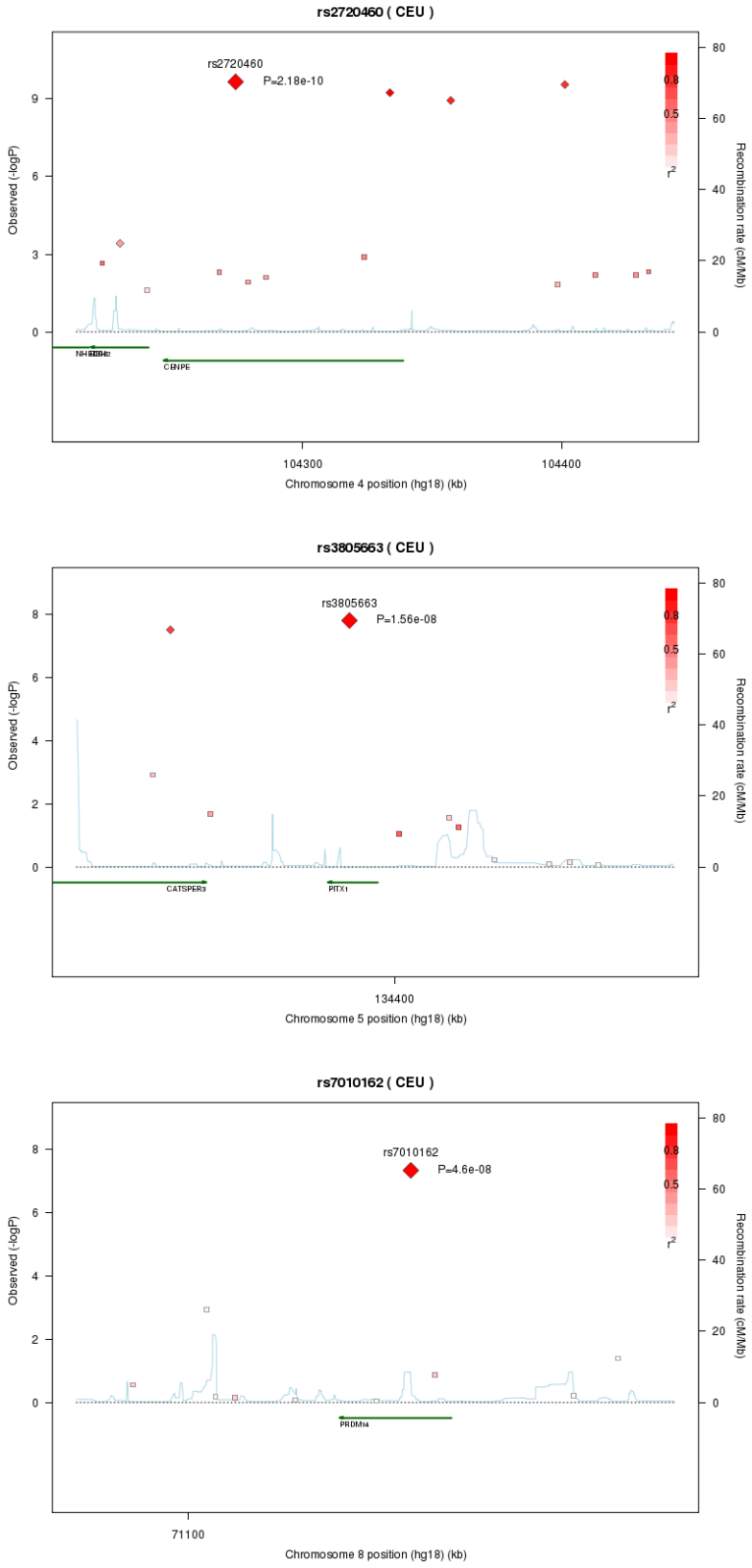
4. Turnbull C, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat.Genet. 2010; 42:604–607. [PubMed: 20543847]
5. Turnbull C, Rahman N. Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour. Int J Androl. 2011; 34:e86–96. discussion e96-7. [PubMed: 21623831]
6. Kanetsky PA, et al. A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. Hum Mol Genet. 2011; 20:3109–17. [PubMed: 21551455]
7. Kanetsky PA, et al. Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. Nat.Genet. 2009; 41:811–815. [PubMed: 19483682]
8. Yamaji M, et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet. 2008; 40:1016–22. [PubMed: 18622394]
9. Schrans-Stassen BH, Saunders PT, Cooke HJ, de Rooij DG. Nature of the spermatogenic arrest in Dazl −/− mice. Biol Reprod. 2001; 65:771–6. [PubMed: 11514340]
10. Qi DL, et al. Identification of PITX1 as a TERT suppressor gene located on human chromosome 5. Mol Cell Biol. 2011; 31:1624–36. [PubMed: 21300782]
11. Gilbert D, Rapley E, Shipley J. Testicular germ cell tumours: predisposition genes and the male germ cell niche. Nat Rev Cancer. 2011; 11:278–88. [PubMed: 21412254]
12. Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nature Genetics. 2013 (In Press).
13. Eeles R, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. 2013 (In press).
14. Schumacher FR, et al. Testicular germ cell tumor susceptibility associated with the UCK2 locus on chromosome 1q23. Hum Mol Genet. 2013
15. Chung CC, et al. Meta-analysis identifies four new loci for testicular germ cell tumor. Nat Genet. 2013
16. Skakkebaek NE. Possible carcinoma-in-situ of the testis. Lancet. 1972; 2:516–517. [PubMed: 4115573]
17. Kee K, Angeles VT, Flores M, Nguyen HN, Reijo Pera RA. Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. Nature. 2009; 462:222–5. [PubMed: 19865085]
18. Shan Z, et al. A SPGY copy homologous to the mouse gene Dazla and the Drosophila gene boule is autosomal and expressed only in the human male gonad. Hum Mol Genet. 1996; 5:2005–11. [PubMed: 8968755]
19. Yen PH, Chai NN, Salido EC. The human autosomal gene DAZLA: testis specificity and a candidate for male infertility. Hum Mol Genet. 1996; 5:2013–7. [PubMed: 8968756]
20. Saxena R, et al. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat Genet. 1996; 14:292–9. [PubMed: 8896558]
21. Reijo R, et al. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet. 1995; 10:383–93. [PubMed: 7670487]
22. Kurimoto K, Yamaji M, Seki Y, Saitou M. Specification of the germ cell lineage in mice: a process orchestrated by the PR-domain proteins, Blimp1 and Prdm14. Cell Cycle. 2008; 7:3514–8. [PubMed: 19001867]
23. Ohinata Y, et al. A signaling principle for the specification of the germ cell lineage in mice. Cell. 2009; 137:571–84. [PubMed: 19410550]
24. Chia NY, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature. 2010; 468:316–20. [PubMed: 20953172]
25. Tsuneyoshi N, et al. PRDM14 suppresses expression of differentiation marker genes in human embryonic stem cells. Biochem Biophys Res Commun. 2008; 367:899–905. [PubMed: 18194669]
26. Dettman EJ, Justice MJ. The zinc finger SET domain gene Prdm14 is overexpressed in lymphoblastic lymphomas with retroviral insertions at Evi32. PLoS One. 2008; 3:e3823. [PubMed: 19043588]

27. Dettman EJ, et al. Prdm14 initiates lymphoblastic leukemia after expanding a population of cells resembling common lymphoid progenitors. Oncogene. 2011; 30:2859–73. [PubMed: 21339739]
28. Liu B, Zhang S, Hui L, Qiu X, Cui Z. [Relationship between the expression of PRDM14 in non-small cell lung cancer and the clinicopathologic characteristics]. Zhongguo Fei Ai Za Zhi. 2010; 13:867–72. [PubMed: 20840815]
29. Nishikawa N, et al. Gene amplification and overexpression of PRDM14 in breast cancers. Cancer Res. 2007; 67:9649–57. [PubMed: 17942894]
30. Jin JL, et al. Catsper3 and catsper4 encode two cation channel-like proteins exclusively expressed in the testis. Biol Reprod. 2005; 73:1235–42. [PubMed: 16107607]
31. Jin J, et al. Catsper3 and Catsper4 are essential for sperm hyperactivated motility and male fertility in the mouse. Biol Reprod. 2007; 77:37–44. [PubMed: 17344468]
32. Lobley A, Pierron V, Reynolds L, Allen L, Michalovich D. Identification of human and mouse CatSper3 and CatSper4 genes: characterisation of a common interaction domain and evidence for expression in testis. Reprod Biol Endocrinol. 2003; 1:53. [PubMed: 12932298]
33. Shete S, et al. Genome-wide association study identifies five susceptibility loci for glioma. Nat.Genet. 2009; 41:899–904. [PubMed: 19578367]
34. Law MH, et al. Meta-analysis combining new and existing data sets confirms that the TERT-CLPTM1L locus influences melanoma risk. J Invest Dermatol. 2012; 132:485–7. [PubMed: 21993562]
35. Stacey SN, et al. New common variants affecting susceptibility to basal cell carcinoma. Nat.Genet. 2009; 41:909–914. [PubMed: 19578363]
36. Haiman CA, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat Genet. 2011; 43:1210–4. [PubMed: 22037553]
37. Landi MT, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am.J.Hum.Genet. 2009; 85:679–691. [PubMed: 19836008]
38. Petersen GM, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat.Genet. 2010
39. Rafnar T, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat.Genet. 2009; 41:221–227. [PubMed: 19151717]
40. Wang PJ, McCarrey JR, Yang F, Page DC. An abundance of X-linked genes expressed in spermatogonia. Nat Genet. 2001; 27:422–6. [PubMed: 11279525]
41. Wu MH, et al. Sequence and expression of testis-expressed gene 14 (Tex14): a gene encoding a protein kinase preferentially expressed during spermatogenesis. Gene Expr Patterns. 2003; 3:231–6. [PubMed: 12711554]
42. Greenbaum MP, et al. TEX14 is essential for intercellular bridges and fertility in male mice. Proc Natl Acad Sci U S A. 2006; 103:4982–7. [PubMed: 16549803]
43. Mondal G, Ohashi A, Yang L, Rowley M, Couch FJ. Tex14, a Plk1-regulated protein, is required for kinetochore-microtubule attachment and regulation of the spindle assembly checkpoint. Mol Cell. 2012; 45:680–95. [PubMed: 22405274]
44. Obuse C, et al. A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. Nat Cell Biol. 2004; 6:1135–41. [PubMed: 15502821]
45. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–9. [PubMed: 21441907]
46. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–4. [PubMed: 22064851]
47. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. Science. 2005; 310:321–324. [PubMed: 16224025]

Europe PMC Funders Author Manuscripts



Europe PMC Funders Author Manuscripts



Europe PMC Funders Author Manuscripts

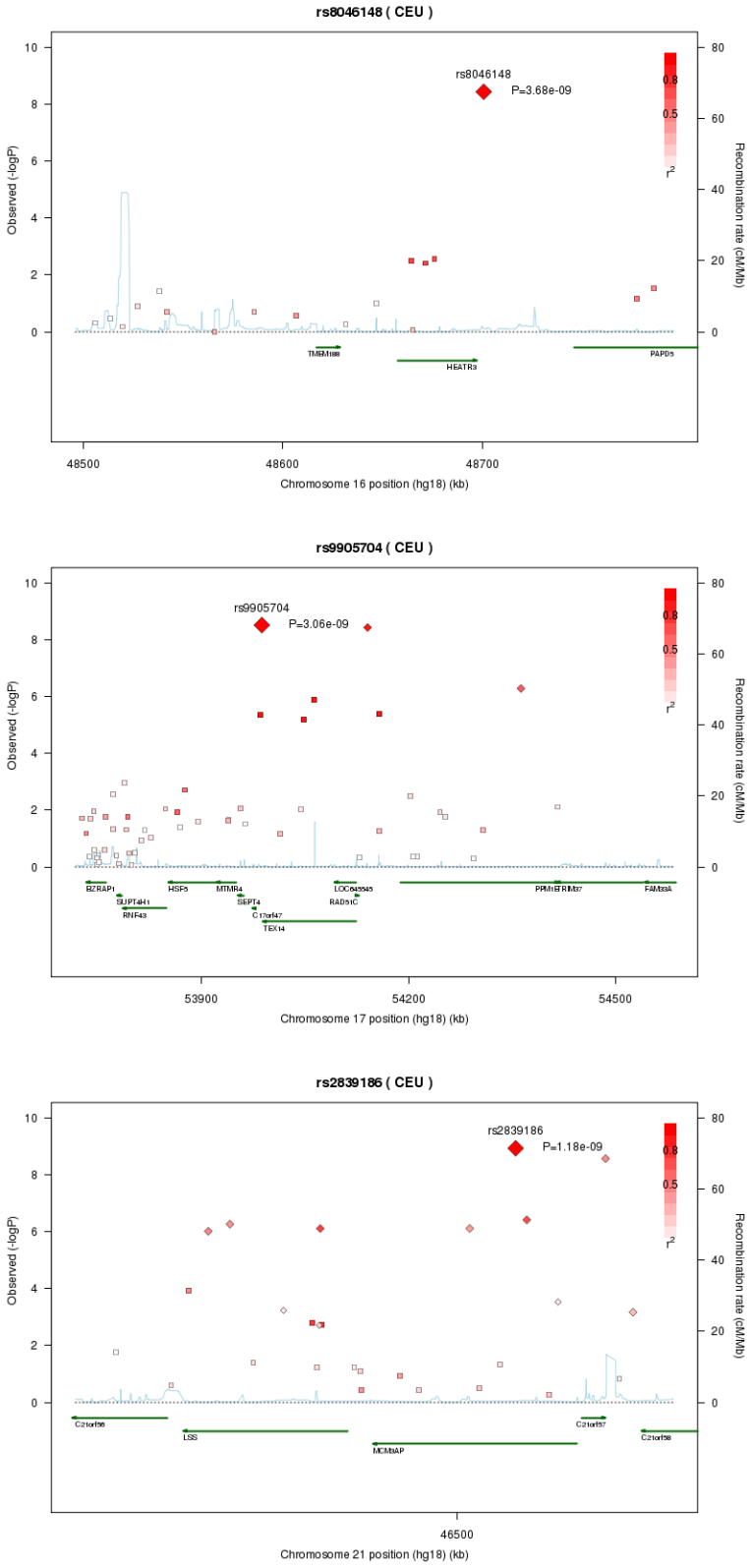


Figure 1. Regional plots of the nine new TGCT loci (a–i)

Plots show the genomic regions of association with TGCT on chromosome (a), 1q22 (b) 1q24.1, (c) 3p24.3, (d) 4q24, (e) 5q31.1, (f) 8q13.3, (g) 16q12.1, (h) 17q22 and (i) 21q22.3. Shown by **diamonds** are the $-\log_{10}$ association *P* values of SNPs in 986 cases and 4,946 controls from the GWAS, combined with 1,064 cases of TGCT and 10,082 controls from the iCOGs replication. Shown by squares are SNPs genotyped in the GWAS alone. The intensity of red shading indicates the strength of LD with the index SNP (labeled). Also shown are the SNP build 36 coordinates in kilobases (kb), recombination rates in centimorgans (cM) per megabase (Mb) (in blue) and the genes in the region (in green).

Europe PMC Funders Author Manuscripts

Table 1

Summary results for nine SNPs showing genome-wide association ($P<5\times10^{-8}$) in the combined analysis.

SNP[1]	Location[2]	Alleles[3]	RAF[4]	GWAS		iCOGS replication		Combined
				OR[5] (95% CI)	P_{trend}[6]	adjusted OR[7] (95% CI)	P_{trend}[6]	P_{meta}[8]
rs2072499	154,436,234 1q22	G/A	0.35	1.24 (1.12-1.37)	1.84×10^{-5}	1.19 (1.08-1.3)	2.98×10^{-4}	2.57×10^{-8}
rs3790672	164,140,016 1q24.1	C/T	0.28	1.26 (1.14-1.4)	1.10×10^{-5}	1.20 (1.09-1.33)	2.92×10^{-4}	1.59×10^{-8}
rs10510452	16,600,052 3p24.3	A/G	0.69	1.24 (1.12-1.39)	7.06×10^{-5}	1.24 (1.12-1.37)	4.69×10^{-5}	1.28×10^{-8}
rs2720460	104,274,135 4q24	A/G	0.61	1.28 (1.16-1.42)	2.25×10^{-6}	1.24 (1.12-1.36)	1.97×10^{-5}	2.18×10^{-10}
rs3805663	134,394,099 5q31.1	T/C	0.63	1.20 (1.08-1.33)	4.86×10^{-4}	1.25 (1.13-1.38)	7.14×10^{-6}	1.56×10^{-8}
rs7010162	71,139,059 8q13.3	G/A	0.61	1.21 (1.09-1.34)	2.03×10^{-4}	1.22 (1.11-1.34)	6.05×10^{-5}	4.60×10^{-8}
rs8046148	48,700,445 16q12.1	G/A	0.79	1.29 (1.13-1.46)	1.03×10^{-4}	1.32 (1.17-1.48)	8.75×10^{-6}	3.68×10^{-9}
rs9905704	53,987,542 17q22	T/G	0.67	1.29 (1.16-1.43)	3.74×10^{-6}	1.21 (1.1-1.34)	1.50×10^{-4}	3.06×10^{-9}
rs2839186	46,514,496 21q22.3	T/C	0.46	1.20 (1.09-1.32)	3.01×10^{-4}	1.26 (1.15-1.38)	7.43×10^{-7}	1.18×10^{-9}

[1] dbSNP rs number

[2] Build 36 position, chromosomal region

[3] Risk/non-risk associated alleles.

[4] RAF:Frequency of the risk allele

[5] OR: per allele odds ratio

[6] P_{trend}: p-value for trend, via logistic regression

[7] Adjusted OR: per allele odds ratio, adjusted for six principle components

[8] P_{meta}: P-value for fixed effects meta-analysis