Meta-analysis identifies four new loci associated with testicular germ cell tumor

Charles C Chung1,2,19, Peter A Kanetsky3,4,19, Zhaoming Wang1,2,19, Michelle A T Hildebrandt5,19, Roelof Koster6,19, Rolf I Skotheim7,8,19, Christian P Kratz1,18,19, Clare Turnbull9,19, Victoria K Cortessis10,19, Anne C Bakken7,8, D Timothy Bishop11, Michael B Cook1, R Loren Erickson12, Sophie D Fosså13, Kevin B Jacobs1,2, Larissa A Korde1,14, Sigrid M Kraggerud7,8, Ragnhild A Lothe7,8, Jennifer T Loud1, Nazneen Rahman9, Eila C Skinner15, Duncan C Thomas10, Xifeng Wu5, Meredith Y eager1,2, Fredrick R Schumacher10, Mark H Greene1, Stephen M Schwartz16,17, Katherine A McGlynn1, Stephen J Chanock1,20 & Katherine L Nathanson3,6,20

We conducted a meta-analysis to identify new susceptibility loci for testicular germ cell tumor (TGCT). In the discovery phase, we analyzed 931 affected individuals and 1,975 controls from 3 genome-wide association studies (GWAS). We conducted replication in 6 independent sample sets comprising 3,211 affected individuals and 7,591 controls. In the combined analysis, risk of TGCT was significantly associated with markers at four previously unreported loci: 4q22.2 in HPGDS (per-allele odds ratio \(OR = 1.19\), 95% confidence interval \(CI = 1.12–1.26\); \(P = 1.11 \times 10^{-8}\)), 7p22.3 in MAD1L1 (\(OR = 1.21\), 95% CI = 1.14–1.29; \(P = 5.59 \times 10^{-9}\)), 16q22.3 in RFWD3 (\(OR = 1.26\), 95% CI = 1.18–1.34; \(P = 5.15 \times 10^{-13}\)) and 17q22 (\(rs9905704\): \(OR = 1.27\), 95% CI = 1.18–1.33; \(P = 4.32 \times 10^{-13}\) and \(rs7221274\): \(OR = 1.20\), 95% CI = 1.12–1.28; \(P = 4.04 \times 10^{-8}\)), a locus that includes TEX14, RAD51C and PPM1E. These new TGCT susceptibility loci contain biologically plausible genes encoding proteins important for male germ cell development, chromosomal segregation and the DNA damage response.

In the United States, TGCT is the most common cancer in young men, with peak incidence among those aged 25 to 34 years. The incidence of TGCT has more than doubled among men of European ancestry in the United States over the past 30 years; similar increases in incidence rates have been observed in other populations of European ancestry1–3. Of note, the incidence of TGCT varies widely between populations and is much higher in individuals of European ancestry than in those of African ancestry2. Established risk factors for TGCT include family history of the disease, cryptorchidism, adult height and a prior history of TGCT; several recent studies have also implicated marijuana use4–7. First-degree relatives of affected men have consistently been shown to have greater risk of TGCT (5- to 19-fold higher for the brothers of affected men and 2- to 4-fold higher for the sons of affected men)8–11, the highest for any cancer. Furthermore, the estimated heritability of TGCT is the third highest among all cancers, with genetic effects estimated to account for 25% of susceptibility to TGCT12. These observations, coupled with findings in twin studies13–15, support a strong genetic component contributing to susceptibility to TGCT.

Despite the greatly increased relative risk of TGCT in the family members of affected men, candidate gene and linkage approaches have shown little progress in identifying specific genetic risk factors. Initially, two independent GWAS identified allele variation within KITLG at 12q22 as the strongest genetic risk factor for TGCT16,17, with a per-allele OR of greater than 3. Variants at 5p15.33 (\(TERT\); two independent loci), 5q31.3 (\(SPRY4\)), 6p21.3 (\(BAK1\)), 9p24.3 (\(DMRT1\); two independent loci) and 12p13.1 (\(ATF7IP\)) have also been associated with risk of TGCT16–21. The per-allele ORs for the identified

---

1Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), US National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland, USA. 2Cancer Genome Research Laboratory, Division of Cancer Epidemiology and Genetics, SAIC-Frederick, Frederick, Maryland, USA. 3Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. 4Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. 5Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA. 6Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. 7Department of Medicine, Translational Medicine and Human Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. 8Department of Cancer Prevention, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway. 9Department of Urology, Stanford University, Stanford, California, USA. 10Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA. 11Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St. James’s University Hospital, Leeds, UK. 12Walter Reed Army Institute of Research, Silver Spring, Maryland, USA. 13Department of Oncology, The Norwegian Radium Hospital, Oslo University Hospital, University of Oslo, Oslo, Norway. 14Division of Medical Oncology, University of Washington/Seattle Cancer Care Alliance, Seattle, Washington, USA. 15Department of Urology, Stanford University, Stanford, California, USA. 16Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington, USA. 17Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA. 18Present address: Center for Pediatrics and Adolescent Medicine, Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany. 19These authors contributed equally to this work. 20These authors jointly directed this work. Correspondence should be addressed to K.L.N. (knathans@exchange.upenn.edu).

Received 26 December 2012; accepted 10 April 2013; published online 12 May 2013; doi:10.1038/ng.2634
TGCT susceptibility alleles are in large part higher than those identified for other cancers, which may be owing, in part, to the homogeneity of the disease, as all TGCTs are thought to arise from primordial germ cells. Many additional loci are expected to contribute to susceptibility, as has been shown for cancers of lower heritability. Combining multiple GWAS data sets represents a strategy to increase the power to detect additional genetic risk factors that did not reach genome-wide significance in individual studies.

We performed a meta-analysis of the 340 most promising SNPs (after excluding previously reported loci) observed in the adjusted pooled analysis of the combined National Cancer Institute (NCI) scan (STED, US Servicemen's Testicular Tumor Environmental and Endocrine Determinants Study and FTCs, NCI Familial Testicular Cancer Study) with the previously reported University of Pennsylvania (UPENN) TGCT scan (Online Methods). Allelic ORs for known loci are shown in Supplementary Table 1 for the combined NCI scan. Forty SNPs from 9 loci had association values larger than 1 × 10^{-4}, of which 12 were highly correlated with the MAD1L1 gene locus (7p22.2) (details of correlation between the NCI and UPENN studies for the top 40 SNPs are shown in Supplementary Table 2). We selected the most significant SNP marker from each of nine loci, plus eight additional markers, for replication. We performed an in silico analysis of these 17 SNPs in GWAS data from the University of Southern California (USC) and the UK Testicular Cancer Collaboration (UKTCC), which was followed by genotyping in four additional TGCT case-control studies from the Fred Hutchinson Cancer Center (Adult Testicular Lifestyle and Blood Specimen (ATLAS) study), the University of Pennsylvania (Testicular Cancer in Philadelphia Area Counties (TestPAC) study), the Oslo University Hospital–The Radium Hospital, Norway (OUHRH) and the MD Anderson Cancer Center (MAD1). Details of each study are included in the Supplementary Note. The combined analysis included 4,142 TGCT cases and 9,566 controls (Supplementary Table 3). In the combined meta-analysis, we observed four new loci significantly associated with TGCT (P < 5 × 10^{-8}; Table 1 and Supplementary Table 4).

The most significantly associated SNP marker at 4q22.2, rs17021463, was located in the intron of the HPGDS gene encoding hematopoietic prostaglandin D synthase (P = 1.11 × 10^{-8}; OR = 1.19, 95% CI = 1.12–1.26) (Fig. 1a and Table 1). In mice, Hpgds is expressed in the early embryonic male gonad and seems to regulate nuclear localization of the Sox9 protein. Disruption of Hpgds leads to modification of the phenotype of ApcMin/+ mice. Seventy-one surrogate markers were highly correlated with rs17021463 in HPGDS (r^2 ≥ 0.8, 1000 Genomes Project CEU data (Utah residents of Northern and Western European ancestry); Supplementary Table 5). Notably, rs35744894 (encoding SWI/SNF-2; similarly, rs4072222 (r = 0.87, 1000 Genomes Project ancestry); Supplementary Table 5) across a 200-kb window mapped to two additional genes, Etl1 and RFWD3 (encoding Golgi glycoprotein 1) (Supplementary Table 6). The RFWD3 protein is an E3 ubiquitin ligase that positively regulates p53 expression in monocytes.

We observed a significant association with TGCT for rs4888262 at 16q23.3 (P = 5.15 × 10^{-12}; OR = 1.26, 95% CI = 1.18–1.34), which is a synonymous SNP in codon 404 (c.1212G>A; p Thr404). The RFWD3 gene (encoding ring finger WD domain 3) (Fig. 1c and Table 1). The RFWD3 protein is an E3 ubiquitin ligase that positively regulates p53 stability by forming an RFWD3-MDM2-p53 complex, thereby protecting p53 from degradation by MDM2-mediated polyubiquitination. Within the linkage disequilibrium (LD) interval were SNPs that mapped to two additional genes, GLG1 (encoding Golgi glycoprotein 1) and MLKL (encoding mixed-lineage kinase domain like), the latter of which has been recently identified as a key mediator of tumor necrosis factor (TNF)-induced necrosis, downstream of receptor-interacting protein kinase 3 (RIP3) (Fig. 1c). We note that rs3851729, which is highly correlated with rs4888262 (r^2 = 0.77, 1000 Genomes Project CEU data), maps to a highly conserved sequence in the 3’ UTR of GLG1; similarly, rs4072222 (r = 0.87, 1000 Genomes Project CEU data) maps to an intron of MLKL (Supplementary Table 5). Both susceptibility variants are cis expression quantitative trait loci (eQTLs) that influence both MLKL and RFWD3 expression in monocytes.

We identified two highly correlated SNPs (r^2 = 0.74) in the STEED controls) at 17q22, rs9905704 (P = 4.32 × 10^{-13}; OR = 1.27, 95% CI = 1.18–1.33) and rs7221274 (P = 4.04 × 10^{-9}; OR = 1.20, 95% CI = 1.12–1.28) (Fig. 1d and Table 1). In a conditional analysis, the signal at each of these SNPs was markedly attenuated by the other, indicating a single TGCT susceptibility locus at 17q22 (Supplementary Table 7). Within this LD block are at least six plausible candidate genes: RAD51C (RAD51 homolog C), TEX14 (testis expressed 14), PPM1E (protein phosphatase, Mg2+/Mn2+ dependent, 1E), SEPT4 (septin 4), TRIM37 (tripartite motif containing 37) and SKA2 (spindle- and kinetochore-associated complex subunit 2) (Fig. 1d). Proteins encoded...
by these candidate genes, except for SKA2, have been implicated as having roles in spermatogenesis. \textit{RAD51C} is a DNA repair gene, rare mutations in which confer susceptibility to ovarian cancer. Of male \textit{Rad51c} \textit{ko/neo} mice, approximately one-third were found to be infertile, owing to impaired spermatogenesis. TEX14 is an essential component of germ cell intercellular bridges, evolutionarily structures that are conserved from invertebrates to humans that allow the clonal development of daughter cells in syncytium; targeted disruption.

### Table 1 New loci associated with TGCT through meta-analysis

| SNP  | Nearby genes | Studya | Cases | Controls | EAFc | Allelic OR (95% CI) | P value | P for heterogeneity |
|------|--------------|--------|-------|----------|------|---------------------|---------|-------------------|
| rs17021463 | **HPGDS** | NCI | 582 | 1,055 | 0.475 | 1.33 (1.14–1.55) | 2.12 × 10^{-4} |
|       |             | UPENN | 349 | 919 | 0.433 | 1.28 (1.07–1.53) | 6.67 × 10^{-3} |
|       |             | All discovery | 2,935 | 7,172 | 0.412 | 1.19 (1.07–1.35) | 7.01 × 10^{-5} |
| rs12699477 | **MAD1L1** | NCI | 582 | 1,056 | 0.412 | 1.31 (1.13–1.52) | 4.64 × 10^{-4} |
|       |             | UPENN | 349 | 919 | 0.407 | 1.37 (1.15–1.63) | 4.25 × 10^{-4} |
|       |             | All discovery | 2,832 | 8,424 | 0.428 | 1.16 (1.07–1.25) | 2.41 × 10^{-4} |
| rs4888262 | **RFWD3** | NCI | 582 | 1,056 | 0.412 | 1.31 (1.13–1.52) | 4.64 × 10^{-4} |
|       |             | UPENN | 349 | 919 | 0.407 | 1.37 (1.15–1.63) | 4.25 × 10^{-4} |
|       |             | All discovery | 1,870 | 6,422 | 0.428 | 1.16 (1.07–1.25) | 2.41 × 10^{-4} |
| rs9905704 | **TEX14** | NCI | 582 | 1,056 | 0.412 | 1.31 (1.13–1.52) | 4.64 × 10^{-4} |
|       |             | UPENN | 349 | 919 | 0.407 | 1.37 (1.15–1.63) | 4.25 × 10^{-4} |
|       |             | All discovery | 2,801 | 8,393 | 0.428 | 1.16 (1.07–1.25) | 2.41 × 10^{-4} |
| rs7221274 | **PPM1E** | NCI | 582 | 1,056 | 0.412 | 1.31 (1.13–1.52) | 4.64 × 10^{-4} |
|       |             | UPENN | 349 | 919 | 0.407 | 1.37 (1.15–1.63) | 4.25 × 10^{-4} |
|       |             | All discovery | 3,829 | 9,120 | 0.428 | 1.16 (1.07–1.25) | 2.41 × 10^{-4} |

Combined discovery results include those from initial meta-analysis of NCI and UPENN data. Combined replication results include those from meta-analysis of the other studies. aSNP genotype depicted as reference allele/effect allele. bNCI results include combined analysis results from the two GWAS STEED and FTCS performed at the NCI. cEAF values are the effect allele frequencies in controls.
of Tex14 results in male sterility in mice. TEX14 has also been implicated as an important component of kinetochores and interacts with MAD1 protein. PPM1E encodes a phosphatase that dephosphorylates and inactivates CaMK4 (calcium/calmodulin-dependent protein kinase IV), the deficiency of which causes infertility in mice. TRIM37 encodes a RING-B box–coiled-coil protein; rare mutations in this gene cause the autosomal recessive disease mulberry nanism (MIM 253250), in which adult males have testicular failure.

**Figure 1** Regional association plots, recombination and LD structure for the four new TGCT susceptibility regions at 4q22.2, 7p22.3, 16q22.3 and 17q22. (a–d) Regional plots of association results, recombination hotspots and LD for the 4q22.2-22.3 (a), 7p22.3 (b), 16q23.1 (c) and 17q22-23.1 (d) TGCT susceptibility loci. Top, combined meta-analysis results are shown as red diamonds with rs numbers labeled; results for the NCI scan are shown in gray. For each plot, $-\log_{10} P$ values (left y axis) of the SNPs are shown according to their chromosomal positions (x axis). The line graph shows likelihood-ratio statistics (right y axis) for recombination hotspots by SequenceDhot software, and the 5 different colors represent 5 tests of 100 controls from NCI without resampling. Bottom, LD structure based on data from NCI controls ($n = 1,188$) was visualized by snp.plotter software. The physical locations of each region are based on NCBI Build 37 of the human genome.
Three SNPs (rs8077332, rs116527213 and rs9898048) mapped within TRIM37 and were in perfect LD with rs7221274 ($r^2 = 1$, 1000 Genomes Project CEU data; *Supplementary Table 5*); all are cis eQTLs affecting RAD51C expression in monocytes. Thus, fine-mapping and functional studies will be required to elucidate the biological basis of the association signal in this interval at 17q22.

In our meta-analysis of GWAS, we have identified four new susceptibility loci for TGCT at 4q22, 7q22, 16q22.3 and 17q22. In total, ten loci now have been conclusively associated with TGCT susceptibility. The four newly identified susceptibility alleles account for 2% of the risk to the brothers of men with TGCT and 3% of the risk to the sons of men with TGCT, increasing the cumulative total of 12 susceptibility alleles (2 susceptibility alleles from *TERT*-CLPTM1L1 (5p15) and two from the *DMRT1* locus (9p24)) to 14% and 21% of the risk to brothers and sons, respectively. On the basis of the high heritability of TGCT, more than 100 additional susceptibility loci are expected to be discovered. Notably, the allelic ORs associated with these newly discovered loci are in the range of 1.2 to 1.3, continuing the trend of identifying loci with higher ORs for TGCT than for other cancer types.

Each locus harbors biologically plausible candidate genes, implicating several pathways—notably, spermatogenesis and male germ cell development (*HPGDS, SMARCAD1, SEPT4, TEX14, RAD51C, PPM1E* and *TRIM37*), chromosomal segregation (*MAD1L1, TEX14, and SKA2*) and the DNA damage response (*SMARCAD1, RFWD3* and *RAD51C*). None of the four newly identified loci have previously been implicated in TGCT of other cancers, further supporting the idea that there are distinct pathways and regions implicated in TGCT susceptibility; however, rare mutations in *RAD51C* have been implicated in ovarian cancer susceptibility. Susceptibility to TGCT is particularly unique in that many of the associated genes affect male germ cell development and differentiation, thus emphasizing the potential detrimental effect that inherited variation in this developmental process can have on the tumorigenic potential of primordial germ cells. This study is the first to our knowledge to implicate variation within genes involved in chromosomal segregation in cancer susceptibility. Karyotypes in TGCT are unique among cancers in that nearly all carry the same chromosomal aberration, a gain of 12p, most often in the form of an isochromosome, which is considered essential for tumor development. Variation in these genes may lead to chromosomal instability and facilitate the development of aneuploidy. We also identified numerous potential regulatory SNPs, suggesting that newly identified associations might be mediated by plausible candidate genes within each locus, which warrant further fine-mapping and functional studies to elucidate the biological bases of the TGCT susceptibility regions. Studies of the genetic basis of TGCT continue to provide new insights into this unique disease with high heritability.

**METHODS**

Methods and any associated references are available in the online version of the paper.

---

**ACKNOWLEDGMENTS**

The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations indicate endorsement by the US Government. We thank C. Berg and P. Prorok, Division of Cancer Prevention, NCI, the screening center investigators and the staff of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, T. Riley and staff at Information Management Services, Inc., and B. O’Brien and staff at Westat, Inc., for their contributions to the PLCO Cancer Screening Trial. We thank S. Ciossek, M. McDermoth and K. D’ Andrea for expert assistance in conducting TestPAC. We thank L. Kolonel and L. Le Marchand for providing access to the Multiethnic Cohort aggressive prostate cancer scan, as well as J.P. Lewinger, M. Pike, D.J. Van Den Berg and K. Siegmund for technical and scientific contributions to the parent study at USC.

A portion of this work was supported by the Intramural Research Program of the NCI and by support services contract HHSN261200653004C with Westat, Inc. The Penn GWAS (UPENN) and replication effort for the TestPAC study were supported by the Abramson Cancer Center at the University of Pennsylvania and US National Institutes of Health grant R01CA114478 to P.A.K. and K.L.N. The replication effort for the ATLAS study was supported by US National Institutes of Health grant R01CA085914 to S.M.S. The analysis of the USC GWAS controls was supported by the Multiethnic Cohort Study (NCI U01-CA98758). The analyses of the USC GWAS testicular cases and the Familial Study were supported by the California Cancer Research Program (99-00505V-10260 and 03-00174RS-30021) and was to grant (R01CA88343) to V.K.C. and a Whittier Foundation award to the Norris Comprehensive Cancer Center. The study at MD Anderson was supported by the Center for Translational and Public Health Genomics of the Duncan Family Institute for Cancer Prevention and Risk Assessment and by an MD Anderson Senior Research Trust Fellowship to W.X. The UK testicular cancer study was supported by the Institute of Cancer Research, Cancer Research UK and the Wellcome Trust and made use of control data generated by the Wellcome Trust Case Control Consortium 2 (WTCCC2). Support was provided by the Norwegian Cancer Society to R.A.L. and R.I.S., by Health Region South-Eastern Norway to R.A.L. and S.D.F., and by the Norwegian ExtraFoundation for Health and Rehabilitation to S.D.F.

**AUTHOR CONTRIBUTIONS**

S.J.C. and K.L.N. supervised the overall study. P.A.K., M.A.T.H., C.P.K., V.K.C., A.C.B., D.T.B., M.B.C., R.L.E., S.D.F, , I.A.K., S.M.K., N.R., E.C.S., X.W., M.H.G., S.M.S., K.A.M. and K.L.N. contributed to recruitment, study and data management. C.C.C., P.A.K., Z.W., M.A.T.H., R.K., R.I.S., C.T., K.B.J., R.A.L., J.T.L., D.C.T, M.Y. and F.R.S. contributed to genotyping or association analysis of individual studies. C.C.C., Z.W. and R.K. carried out the meta-analysis and the additional reported ENCODE analyses. C.C.C. and K.L.N. prepared the manuscript, together with P.A.K., R.K. and S.J.C., and all authors reviewed and contributed to the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

---

1. Rosen, A., Jayram, G., Drazier, M. & Eggener, S.E. Global trends in testicular cancer incidence and mortality. *Eur. Urol.* **60**, 374–379 (2011).
2. Howlader, N. et al. SEER Cancer Statistics Review, 1975–2009 (Vintage 2009 Populations) based on November 2011 SEER data submission, posted to the SEER web site, 2012 (http://seer.cancer.gov/csr/1975_2009_pops09) (2012).
3. Stang, A. et al. Gonadal and extragonadal germ cell tumours in the United States, 1973–2007. *Int. J. Androl.* **35**, 616–625 (2012).
4. McGlynn, K.A. & Trabert, B. Adolescent and adult risk factors for testicular cancer. *Nature Genet.* **45**, 537–543 (2012).
5. Broman, K. et al. Testicular, other genital, and breast cancers in first-degree relatives of testicular cancer patients and controls. *Cancer Epidemiol. Biomarkers Prev.* **13**, 1316–1324 (2004).
6. Chia, V.M. et al. Risk of cancer in first- and second-degree relatives of testicular germ cell tumors and controls. *Int. J. Cancer* **124**, 952–957 (2009).
7. Heimdal, K. et al. Risk of cancer in relatives of testicular cancer patients. *Br. J. Cancer* **73**, 970–973 (1996).
8. Sonneveld, D.J. et al. Familial testicular cancer in a single-centre population. *Eur. J. Cancer* **35**, 1368–1373 (1999).
12. Czene, K., Lichtenstein, P. & Hemminki, K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int. J. Cancer* 99, 260–266 (2002).

13. Neale, R.E., Carriere, P., Murphy, M.F. & Baade, P.D. Testicular cancer in twins: a meta-analysis. *Br. J. Cancer* 98, 171–173 (2008).

14. Swedlow, A.J., De Stavola, B.L., Snartwick, M.A. & Macconochie, N.E. Risks of breast and testicular cancers in young adult twins in England and Wales: evidence on prenatal and genetic aetiology. *Lancet* 350, 1723–1728 (1997).

15. Braun, M.M., Ahlbom, A., Floderus, B., Brinton, L.A. & Hoover, R.N. Effect of twinning on incidence of cancer of the testis, breast, and other sites (Sweden). *Cancer Causes Control* 6, 519–524 (1995).

16. Dalgaard, M.D., et al. Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. *Nat. Genet.* 41, 811–815 (2009).

17. Rapley, E.A., et al. A genome-wide association study of testicular germ cell tumor. *Nat. Genet.* 41, 807–810 (2009).

18. Turnbull, C., et al. Variants near DMRT1, TERT and AFF7 are associated with testicular germ cell cancer. *Nat. Genet.* 42, 604–607 (2010).

19. Kanetsky, P.A., et al. Effect of genome-wide association studies and implications for future discoveries. *Hum. Mol. Genet.* 20, 3109–3117 (2011).

20. Ferlin, A., et al. Variants in KITLG predispose to testicular germ cell cancer independently from spermatogenic function. *Endocr. Relat. Cancer* 19, 101–108 (2012).

21. Daiguard, M.D., et al. A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation. *J. Med. Genet.* 49, 58–65 (2012).

22. Rajpert-De Meyts, E. & Hansen, C.E. From gonocytes to testicular cancer: the role of impaired gonadal development. *Ann. NY Acad. Sci.* 1120, 168–180 (2007).

23. Chung, C.C. & Charoock, S.J. Current status of genome-wide association studies in testicular germ cell cancer. *Nat. Rev. Urol.* 8, 59–75 (2011).

24. Park, J.H., et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* 42, 570–575 (2010).

25. Moroni, B., et al. Hematopoietic progestin antagonist D synthe (H-Pgds) is expressed in the early embryonic gonad and participates to the initial nuclear translocation of the SOX9 protein. *Dev. Dyn.* 240, 2335–2343 (2011).

26. Park, J.M., et al. Hematopoietic progestin antagonist D synthe suppresses intestinal adenomas in ApcMin mice. *Cancer Res.* 67, 881–889 (2007).

27. Rowbotham, S.P., et al. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol. Cell* 42, 285–296 (2011).

28. Costelloe, T., et al. The yeast Fun30 and human SMARCAD1 chromatin remodelers promote DNA end resection. *Nature* 489, 581–584 (2012).

29. Schoor, M., et al. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol. Cell* 42, 285–296 (2011).

30. Costelloe, T., et al. The yeast Fun30 and human SMARCAD1 chromatin remodelers promote DNA end resection. *Nature* 489, 581–584 (2012).

31. Schoor, M., et al. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol. Cell* 42, 285–296 (2011).

32. Costelloe, T., et al. The yeast Fun30 and human SMARCAD1 chromatin remodelers promote DNA end resection. *Nature* 489, 581–584 (2012).

33. Gidekel, S., et al. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol. Cell* 42, 285–296 (2011).

34. Costelloe, T., et al. The yeast Fun30 and human SMARCAD1 chromatin remodelers promote DNA end resection. *Nature* 489, 581–584 (2012).

35. Schoor, M., et al. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol. Cell* 42, 285–296 (2011).

36. Gidekel, S., et al. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol. Cell* 42, 285–296 (2011).

37. Altmuuler, D.M., et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58 (2010).

38. Kim, S., et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58 (2010).

39. Kim, S., et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58 (2010).

40. Kim, S., et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58 (2010).

41. Fu, X., et al. RFWD3-Mdm2 ubiquitin ligase complex positively regulates p53 stability in response to DNA damage. *Proc. Natl. Acad. Sci. USA* 107, 4579–4584 (2010).

42. Liu, S., et al. RING finger and WD repeat domain 3 (RFWD3) associates with replication protein A (RPA) and facilitates RPA-mediated DNA damage response. *J. Biol. Chem.* 286, 22314–22322 (2011).

43. Zhao, J., et al. Mixed lineage kinase domain-like is a key receptor interacting protein 3 downstream component of TNF-induced necrosis. *Proc. Natl. Acad. Sci. USA* 109, 5322–5327 (2012).

44. Sun, L., et al. Mixed lineage kinase domain-like protein mediates necrosis signaling downstream of RIP3 kinase. *Cell* 148, 213–227 (2012).

45. Zeller, T., et al. Genetics and beyond—the transcriptome of human monocyties and disease susceptibility. *PLoS ONE* 5, e10693 (2010).

46. Ihara, M., et al. Cortical organization by the septin cytoskeleton is essential for structural and mechanical integrity of mammalian spermatozoa. *Dev. Cell* 8, 343–352 (2005).

47. Kessel, H., et al. The Sept4 septin locus is required for sperm terminal differentiation in mice. *Dev. Cell* 8, 353–364 (2005).

48. Greenbaum, M.P., et al. TEX14 is essential for intercellular bridges and fertility in male mice. *Proc. Natl. Acad. Sci. USA* 103, 4982–4987 (2006).

49. Kuznetsos, S., et al. RAD51C deficiency in mice results in early prophase I arrest in males and sister chromatid separation at metaphase II in females. *J. Cell Biol.* 158, 185–192 (2007).

50. Wu, J.Y., et al. Spermogenesis and exchange of basic nuclear proteins are impaired in male germ cells lacking Camk4. *Nat. Genet.* 25, 448–452 (2000).

51. Karberg, S., et al. Testicular failure and male infertility in the monogenic Mulibrey nanism disorder. *Nat. Genet.* 42, 410–414 (2010).

52. Meindl, A., et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat. Genet.* 42, 410–414 (2010).

53. Lovelad, C., et al. Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nat. Genet.* 44, 475–476 (2012).

54. Mondal, G., et al. RAD51C deficiency in mice results in early prophase I arrest in males and sister chromatid separation at metaphase II in females. *J. Cell Biol.* 158, 185–192 (2007).

55. Ishida, A., et al. Interaction of RAD51C with the MRE11-RAD50-NBS1 complex and its implications for the regulation of homologous recombination. *EMBO J.* 26, 3028–3039 (2007).

56. Avila, A., et al. Gene encoding a new RING-B-box-Coiled-coil protein is mutated in mulibrey nanism. *Nat. Genet.* 25, 298–301 (2000).

57. Karberg, S., et al. Testicular failure and male infertility in the monogenic Mulibrey nanism disorder. *J. Clin. Endocrinol. Metab.* 96, 3399–3407 (2011).

58. Atkin, N.B., et al. Specific chromosome changes (12p12) in testicular tumors? *Lancet* 2, 1349 (1982).

59. Rodriguez, E., et al. Cytogenetic analysis of 124 prospectively ascertained male subjects with testicular tumors. *Cytogenet. Cell Genet.* 62, 251–258 (1993).

60. Skotheim, R.I., et al. Novel genomic aberrations in testicular germ cell tumors by array-CGH, and associated gene expression changes. *Cell Oncol.* 28, 315–326 (2006).
ONLINE METHODS

Studies. Detailed characteristics of the study populations are given in both the Supplementary Note and in Supplementary Table 3. Subjects used in the current study are all of European descent, and data from each study were collected and analyzed in accordance with local ethical permissions and informed consent. Three studies (STEED, FTCS and UPENN) were included in the discovery meta-analysis, and six studies contributed to replication by de novo genotyping (TestPAC, ATLAS, OUHRR and MDA) or in silico look-up in existing data (UKTCC and USC).

Genotyping and quality control. Genotype quality control metrics for the reported GWAS (UPENN and UKTCC) were previously described18,19. Genotype quality control metrics for STEED, FTCS and USC are described in the Supplementary Note61. The OUHRR and MDA studies were genotyped using 5′ exonuclease assay (TaqMan) and the ABI Prism 7900HT sequence detection system, all according to the manufacturer’s instructions, across several genotyping centers. Primers and probes were supplied directly by Applied Biosystems as Assays-By-Design. Technical validation was performed in HapMap samples (n = 270), with greater than 99% genotype concordance. TestPAC and ATLAS studies conducted genotyping using the iPLEX MassARRAY platform (Sequenom), following the manufacturer’s protocol. Assays at all genotyping centers included at least four negative controls and 2–5% duplicates on each plate. A standard quality control protocol was implemented, requiring samples to have SNP call rates of >95% and no deviation from Hardy-Weinberg equilibrium in controls at P < 0.00001, that <2% discordance between genotypes in duplicate had to be fulfilled and that cluster plots for SNPs that were close to failing any of the quality control criteria were re-examined centrally.

Statistical analysis. We analyzed two genome-wide scans from NCI (STEED and FTCS) as a combined data set using a logistic regression model for trend effect adjusted for age, study and additionally for one eigenvector (there was only one with P < 0.05) to account for population stratification in this population of European descent. From the top 500 SNPs by trend P values from the NCI scan, excluding previously reported ones, we selected 340 SNPs on the basis of the availability of surrogates (r2 > 0.6) in the previous TGCT GWAS from the University of Pennsylvania. Because SNP content differs between the Illumina and Affymetrix platforms, we paired the best-correlated surrogate in the other array with each marker to perform a discovery meta-analysis (111 SNPs, direct match; 229 SNPs, surrogate match). From the discovery meta-analysis, we selected 17 of 40 SNPs with association P < 1 × 10−5 for follow-up in the remaining studies. In silico follow-up was carried out in the USC and UKTCC scans, whereas additional genotyping was carried out in the TestPAC, ATLAS, OUHRR and MDA studies (Supplementary Table 3). Not all markers were available for replication efforts at all sites (Supplementary Table 4).

The meta-analysis was conducted using the suite of tools in GLU (Genotyping Library and Utilities) software, combining study-specific OR estimates using a fixed-effects model, which used the inverse-variance method to estimate combined ORs and their 95% CIs. To assess the existence of heterogeneity between studies, we used Cochran’s Q statistic to calculate P values for heterogeneity.

Recombination hotspots were identified in the vicinity of the newly discovered TGCT-associated loci using SequenceLDhot62, a program that uses the approximate marginal likelihood method63 and calculates likelihood-ratio statistics at a set of possible hotspots. We tested 5 unique sets of 100 control samples drawn from STEED. The PHASE v2.1 program was used to calculate background recombination rates64,65, and LD heatmaps were visualized in r2 using the snp.plotter program65.

The relative risk attributable to a set of SNPs (λ) was estimated using the following formula67:

$$\lambda = \frac{\prod_{i=1}^{n} (p_i q_i + q_i p_i)^2}{\prod_{i=1}^{n} (p_i r_i + q_i r_i)^2}$$

where qi is the minor allele frequency (MAF) of SNP i and pi = 1 − qi. SNP-specific risks for rare homozygotes, heterozygotes and common homozygotes are denoted by rui, rij and rui, respectively. The NCI controls (n = 1,140) were used to estimate MAFs, and OR estimates from SNP association analyses were used to estimate relative risks. This formula assumes that the effects of all SNPs in the set are multiplicative. The proportion of familial risk attributable to a set of SNPs was calculated as log(λ)/log(λ0), where λ0 is the familial relative risk estimated from TGCT epidemiological studies (λ0 = 4 with an affected father, λ0 = 8 with an affected brother)68.

Genomic annotation. Genomic annotation on high-LD surrogates (r2 ≥ 0.8, 1000 Genomes Project CEU data) of five SNPs (rs17021463, rs12699477, rs4888262, rs9905704 and rs7221274) from the four TGCT susceptibility loci identified in the current study was conducted using the Encyclopedia of DNA Elements (ENCODE) tools HaploReg69 and RegulomeDB70 (Supplementary Table 5). rs12699477 did not have surrogates that met the requirement of r2 ≥ 0.8; thus, we lowered the threshold to 0.7 for surrogates and then conducted annotation. All surrogates were queried in the RegulomeDB browser to cross-examine predicted regulatory DNA elements such as regions of DNase I hypersensitivity, binding sites of transcription factors and promoter regions that have been biochemically characterized to regulate transcription. Summaries of each SNP analysis by the RegulomeDB browser expressed in scores are shown in Supplementary Table 5. To predict potential regulatory SNPs, we assessed SNPs that met one of the following criteria: (i) conserved (GERP and/or Siphy); (ii) present in a promoter, enhancer or DNase I hypersensitivity region; or (iii) predicted to have a cis eQTL or having a RegulomeDB score of ≤3. Twenty-nine SNPs that passed one of these criteria also changed a transcription factor binding motif and are annotated further with the transcription factor motif of interest and the logarithm of odds (LOD) motif score for the specific SNP of interest (Supplementary Table 6). Two SNPs in 3′ UTRs were evaluated using SNP Function Prediction for changes in microRNA-binding sites and are included in Supplementary Table 6.

61. Schumacher, F.R. et al. Testicular germ cell tumor susceptibility associated with the UCK2 locus on chromosome 1q23. Nat. Genet. 40, D930–D934 (2012).
62. Fearnhead, P. SequenceLDhot: detecting recombination hotspots. Bioinformatics 22, 3061–3066 (2006).
63. Fearnhead, P. & Donnelly, P. Approximate likelihood methods for estimating local recombination rates. J. R. Stat. Soc. B 64, 657–680 (2002).
64. Crawford, D.C. et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat. Genet. 36, 700–706 (2004).
65. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165, 2213–2233 (2003).
66. Luna, A. & Nicodemus, K.K. SNP plotter: an R-based SNPhaplotype association and linkage disequilibrium plotting package. Bioinformatics 23, 774–776 (2007).
67. Cox, A. et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat. Genet. 39, 352–358 (2007).
68. Hemminki, K. & Li, X. Familial risk in testicular cancer as a clue to a heritable risk. Nat. Genet. 39, 352–358 (2007).
69. Crawford, D.C. et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat. Genet. 36, 700–706 (2004).
70. Ward, L.D. & Kelis, M. Haploloop: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40, D939–D934 (2012).