nature genetics

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

Charles C Chung[1,2,19], Peter A Kanetsky[3,4,19], Zhaoming Wang[1,2,19], Michelle A T Hildebrandt[5,19], Roelof Koster[6,19], Rolf I Skotheim[7,8,19], Christian P Kratz[1,18,19], Clare Turnbull[9,19], Victoria K Cortessis[10,19], Anne C Bakken[7,8], D Timothy Bishop[11], Michael B Cook[1], R Loren Erickson[12], Sophie D Fosså[13], Kevin B Jacobs[1,2], Larissa A Korde[1,14], Sigrid M Kraggerud[7,8], Ragnhild A Lothe[7,8], Jennifer T Loud[1], Nazneen Rahman[9], Eila C Skinner[15], Duncan C Thomas[10], Xifeng Wu[5], Meredith Yeager[1,2], Fredrick R Schumacher[10], Mark H Greene[1], Stephen M Schwartz[16,17], Katherine A McGlynn[1], Stephen J Chanock[1,20] & Katherine L Nathanson[3,6,20]

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; $P = 5.15 \times 10^{-12}$) 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; $P = 4.04 \times 10^{-9}$), 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 ancestry[1–3]. Of note, the incidence of TGCT varies widely between populations and is much higher in individuals of European ancestry than in those of African ancestry[2]. 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 use[4–7]. 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 TGCT[12]. These observations, coupled with findings in twin studies[13–15], 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 TGCT[16,17], 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 TGCT[16–21]. The per-allele ORs for the identified

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), US National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland, USA. [2]Cancer Genome Research Laboratory, Division of Cancer Epidemiology and Genetics, SAIC-Frederick, NCI-Frederick, Frederick, Maryland, USA. [3]Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [4]Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [5]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [6]Department of Medicine, Translational Medicine and Human Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [7]Department of Cancer Prevention, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway. [8]Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway. [9]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, UK. [10]Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA. [11]Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St. James's University Hospital, Leeds, UK. [12]Walter Reed Army Institute of Research, Silver Spring, Maryland, USA. [13]Department of Oncology, The Norwegian Radium Hospital, Oslo University Hospital, University of Oslo, Oslo, Norway. [14]Division of Medical Oncology, University of Washington/Seattle Cancer Care Alliance, Seattle, Washington, USA. [15]Department of Urology, Stanford University, Stanford, California, USA. [16]Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington, USA. [17]Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA. [18]Present address: Center for Pediatrics and Adolescent Medicine, Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany. [19]These authors contributed equally to this work. [20]These 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[22,23]. Many additional loci are expected to contribute to susceptibility, as has been shown for cancers of lower heritability[24]. 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 (STEED, 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 *P* values below 1×10^{-4}, of which 12 localized to 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)[18], 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 (MDA). 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 ($P < 5 \times 10^{-8}$; **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 ($P = 1.11 \times 10^{-8}$; 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[25]. Disruption of *Hpgds* leads to modification of the phenotype of $Apc^{Min/+}$ mice[26]. Seventy-one surrogate markers were highly correlated with rs17021463 in *HPGDS* ($r^2 \geq 0.8$, 1000 Genomes Project CEU data (Utah residents of Northern and Western European ancestry); **Supplementary Table 5**). Notably, rs35744894 ($r^2 = 0.87$) alters a DMRT2-binding motif (**Supplementary Table 6**); variation in *DMRT1* has been associated with risk of TGCT[19].

Fifty-three of 71 surrogate markers that were highly correlated with the rs17021463 SNP in *HPGDS* ($r^2 \geq 0.8$, 1000 Genomes Project CEU data; **Supplementary Table 5**) across a 200-kb window mapped within or near an adjacent gene, *SMARCAD1* (encoding SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily a, containing DEAD/H box 1). SMARCAD1 is a chromatin remodeler that restores silenced heterochromatin domains in dividing cells and participates in the DNA damage response[27,28]. Mice with homozygous mutations in *Smarcad1* (*Etl1*) show developmental defects, including impaired fertility[29]. Surrogate markers included one nonsynonymous substitution, rs7439869, at codon 301 ($r^2 = 0.93$, 1000 Genomes Project CEU data; c.902T>C; p.Val301Ala). Although the mutation at rs7439869 is predicted by PolyPhen-2 to be tolerated[30], it changes an OCT4- (encoded by *POUF5F1*) and SOX4-binding motif (**Supplementary Table 6**). OCT4 is a transcription factor that regulates pluripotency in a number of cell types, including primordial germ cells, and is expressed in TGCTs[31–36].

We identified a locus at 7p22.3 harboring the *MAD1L1* gene (mitotic arrest deficient like 1), which encodes MAD1. The most significantly associated SNP without study heterogeneity, rs12699477, was located in intron 17 of this gene ($P = 5.59 \times 10^{-9}$; OR = 1.21, 95% CI = 1.14–1.29) (**Fig. 1b**). Of note, the risk-conferring allele (C) at rs12699477 is more prevalent in populations of European ancestry (29%) than in those of African ancestry (8%) in 1000 Genomes Project data[37]. MAD1 is a spindle assembly checkpoint protein that delays the onset of anaphase in the mitotic cell cycle until all sister chromatids achieve proper alignment and microtubule attachment, thereby preventing aneuploidy and maintaining genomic stability[38].

Of the 35 SNPs that were highly correlated with rs12699477 in *MAD1L1* ($r^2 \geq 0.7$, 1000 Genomes Project CEU data; **Supplementary Table 5**), rs1801368 is a missense variant at codon 558 (c.1673G>A; p.Arg588His) that is located in the region encoding the second leucine-zipper domain of MAD1L1. The p.Arg558His alteration has been reported to be associated with lung cancer risk[39] and may lead to reduced binding of MAD2 to MAD1, resulting in decreased proficiency in enforcing mitotic arrest[40]. We observed additional statistically significant associations with TGCT for neighboring SNPs in the *MAD1L1* region, including for rs10275045 ($P = 3.78 \times 10^{-10}$; OR = 1.20, 95% CI = 1.13–1.27) and rs3778991 ($P = 6.73 \times 10^{-10}$; OR = 1.21, 95% CI = 1.14–1.28). However, both showed significant heterogeneity between studies (**Supplementary Table 4**). The r^2 values between our strongest signal at rs12699477 and these markers were 0.66 and 0.50, respectively, in the STEED controls. Conditional analysis showed a marked attenuation of the two significantly associated neighboring SNPs, supporting the idea that there is a single TGCT susceptibility locus across the *MAD1L1* gene at 7q22.3 (**Supplementary Table 7**).

We observed a significant association with TGCT for rs4888262 at 16q22.3 ($P = 5.15 \times 10^{-12}$; OR = 1.26, 95% CI = 1.18–1.34), which is a synonymous SNP in codon 404 (c.1212G>A; p.Thr404) of 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[41,42]. 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)[43,44] (**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 ($r^2 = 0.87$, 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[45].

We identified two highly correlated SNPs ($r^2 = 0.74$ in the STEED controls) at 17q22, rs9905704 ($P = 4.32 \times 10^{-13}$; OR = 1.27, 95% CI = 1.18–1.33) and rs7221274 ($P = 4.04 \times 10^{-9}$; 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, Mg^{2+}/Mn^{2+} 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[46–51]. *RAD51C* is a DNA repair gene, rare mutations in which confer susceptibility to ovarian cancer[52,53]. Of male *Rad51c*$^{ko/neo}$ mice, approximately one-third were found to be infertile, owing to impaired spermatogenesis[49]. 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[a]	Nearby genes	Study[b]	Cases	Controls	EAF[c]	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}	
G/T		UPENN	349	919	0.433	1.28 (1.07–1.53)	6.67×10^{-3}	
4q22.2		All discovery	931	1,974		1.31 (1.16–1.47)	6.09×10^{-6}	0.750
		UKTCC	979	4,947	0.420	1.19 (1.08–1.31)	5.66×10^{-4}	
		USC	358	258	0.428	1.09 (0.82–1.46)	0.464	
		OUHRH	798	377	0.399	1.16 (0.98–1.39)	0.092	
		TestPAC	267	575	0.417	1.14 (0.92–1.41)	0.225	
		ATLAS	297	664	0.420	1.11 (0.91–1.35)	0.292	
		MDA	236	351	0.412	1.01 (0.79–1.27)	0.960	
		All replication	2,935	7,172		1.15 (1.07–1.23)	7.01×10^{-5}	0.862
		Combined	3,866	9,146		1.19 (1.12–1.26)	1.11×10^{-8}	0.583
rs12699477	*MAD1L1*	NCI	582	1,056	0.412	1.31 (1.13–1.52)	4.64×10^{-4}	
T/C		UPENN	349	919	0.407	1.37 (1.15–1.63)	4.25×10^{-4}	
7p22.3		All discovery	931	1,975		1.34 (1.19–1.50)	7.16×10^{-7}	0.704
		UKTCC	979	4,947	0.380	1.17 (1.06–1.30)	1.34×10^{-3}	
		USC	358	258	0.352	1.32 (0.99–1.74)	0.024	
		TestPAC	266	573	0.375	1.09 (0.88–1.35)	0.440	
		ATLAS	298	671	0.373	1.08 (0.89–1.32)	0.480	
		All replication	1,901	6,449		1.16 (1.07–1.25)	2.41×10^{-4}	0.645
		Combined	2,832	8,424		1.21 (1.14–1.29)	5.59×10^{-9}	0.318
rs4888262	*RFWD3*	NCI	582	1,052	0.552	1.36 (1.17–1.59)	9.12×10^{-5}	
T/C		UPENN	349	919	0.498	1.39 (1.16–1.67)	3.12×10^{-4}	
16q22.3		All discovery	931	1,971		1.37 (1.22–1.54)	1.39×10^{-7}	0.858
		UKTCC	957	4,940	0.458	1.22 (1.10–1.34)	1.02×10^{-4}	
		USC	358	258	0.440	1.41 (1.09–1.83)	5.14×10^{-3}	
		TestPAC	260	575	0.465	1.12 (0.91–1.38)	0.2876	
		ATLAS	295	649	0.495	1.16 (0.95–1.41)	0.1328	
		All replication	1,870	6,422		1.21 (1.12–1.31)	1.62×10^{-6}	0.568
		Combined	2,801	8,393		1.26 (1.18–1.34)	5.15×10^{-12}	0.397
rs9905704	*TEX14*	NCI	582	1,054	0.719	1.37 (1.16–1.61)	1.88×10^{-4}	
G/T		UPENN	349	919	0.663	1.30 (1.08–1.56)	5.46×10^{-3}	
17q22		All discovery	931	1,973		1.33 (1.19–1.52)	3.44×10^{-6}	0.674
		UKTCC	979	4,947	0.680	1.28 (1.16–1.43)	3.65×10^{-6}	
		USC	358	258	0.649	1.35 (1.08–1.72)	0.017	
		OUHRH	802	382	0.695	1.23 (1.02–1.49)	0.028	
		TestPAC	259	575	0.669	1.14 (0.91–1.43)	0.253	
		ATLAS	300	666	0.669	1.09 (0.88–1.33)	0.403	
		MDA	234	351	0.672	1.15 (0.89–1.47)	0.273	
		All replication	2,932	7,179		1.23 (1.15–1.32)	1.37×10^{-8}	0.628
		Combined	3,863	9,152		1.27 (1.18–1.33)	4.32×10^{-13}	0.668
rs7221274	*PPM1E*	NCI	582	1,056	0.660	1.39 (1.19–1.61)	3.56×10^{-5}	
G/A		UPENN	349	919	0.598	1.19 (1.00–1.43)	0.053	
17q22		All discovery	931	1,975		1.30 (1.16–1.47)	7.79×10^{-6}	0.197
		UKTCC	979	4,947	0.620	1.23 (1.12–1.37)	3.08×10^{-5}	
		USC	358	258	0.558	1.37 (0.87–2.13)	9.43×10^{-3}	
		OUHRH	802	380	0.630	1.22 (1.02–1.47)	0.029	
		TestPAC	243	538	0.612	1.01 (0.81–1.27)	0.890	
		ATLAS	301	671	0.604	1.00 (0.83–1.22)	0.989	
		MDA	215	351	0.615	1.05 (0.83–1.33)	0.682	
		All replication	2,898	7,145		1.16 (1.09–1.25)	3.27×10^{-5}	0.228
		Combined	3,829	9,120		1.20 (1.12–1.28)	4.04×10^{-9}	0.130

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. [a]SNP genotype depicted as reference allele/effect allele. [b]NCI results include combined analysis results from the two GWAS STEED and FTCS performed at the NCI. [c]EAF values are the effect allele frequencies in controls.

of *Tex14* results in male sterility in mice[48]. TEX14 has also been implicated as an important component of kinetochores and interacts with MAD1 protein[54]. *PPM1E* encodes a phosphatase that dephosphorylates and inactivates CaMK4 (calcium/calmodulin-dependent protein kinase IV), the deficiency of which causes infertility in mice[50,55]. *TRIM37* encodes a RING–B box–coiled-coil protein; rare mutations in this gene cause the autosomal recessive disease mulibrey nanism (MIM 253250)[56], in which adult males have testicular failure[57].

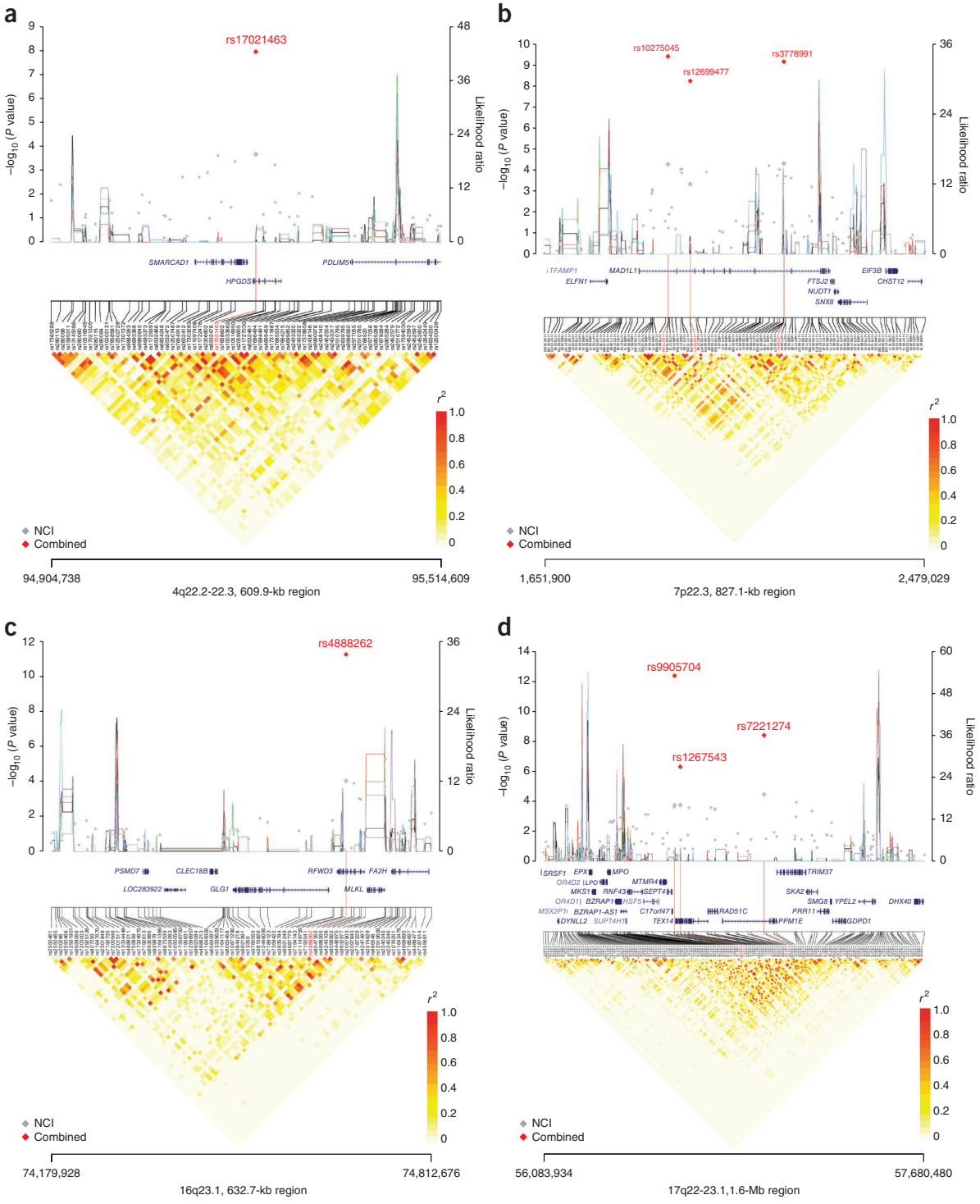

Figure 1 Regional association plots, recombination plots and LD structure for the four new TGCT susceptibility regions at 4q22.2, 7p22.3, 16q22.3 and 17q22. (**a**–**d**) Regional plots of association results, recombination hotspots and LD for the 4q22.2-22.3 (**a**), 7p22.3 (**b**), 16q23.1 (**c**) 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 SequenceLDhot 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, rs11652713 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[45]. 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-CLPTM1L* (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[24]. 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[23].

Each locus harbors biologically plausible candidate genes, implicating several pathways—most 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 GWAS 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[53]. 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[58–60]. 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.

URLs. GLU, http://code.google.com/p/glu-genetics/; SequenceLDhot, http://www.maths.lancs.ac.uk/~fearnhea/Hotspot/; snp.plotter, http://cbdb.nimh.nih.gov/~kristin/snp.plotter.html; PHASE v2.1, http://www.stat.washington.edu/stephens/phase/download.html; InPower, http://dceg.cancer.gov/bb/tools/inpower/readme; HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php; RegulomeDB, http://regulome.stanford.edu/.

METHODS

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

Note: Supplementary information is 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. Ciosek, 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 HHSN261200655004C 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 03-00174VRS-30021) and by an NCI grant (R01CA102042) 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 X.W. 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., L.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.

1. Rosen, A., Jayram, G., Drazer, 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. *Nat. Rev. Urol.* **9**, 339–349 (2012).
5. Daling, J.R. *et al.* Association of marijuana use and the incidence of testicular germ cell tumors. *Cancer* **115**, 1215–1223 (2009).
6. Trabert, B., Sigurdson, A.J., Sweeney, A.M., Strom, S.S. & McGlynn, K.A. Marijuana use and testicular germ cell tumors. *Cancer* **117**, 848–853 (2011).
7. Lacson, J.C. *et al.* Population-based case-control study of recreational drug use and testis cancer risk confirms an association between marijuana use and nonseminoma risk. *Cancer* **118**, 5374–5383 (2012).
8. Bromen, 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).
9. Chia, V.M. *et al.* Risk of cancer in first- and second-degree relatives of testicular germ cell tumor cases and controls. *Int. J. Cancer* **124**, 952–957 (2009).
10. Heimdal, K. *et al.* Risk of cancer in relatives of testicular cancer patients. *Br. J. Cancer* **73**, 970–973 (1996).
11. 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. Swerdlow, A.J., De Stavola, B.L., Swanwick, M.A. & Maconochie, 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 twinship on incidence of cancer of the testis, breast, and other sites (Sweden). *Cancer Causes Control* **6**, 519–524 (1995).
16. Kanetsky, P.A. *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 *ATF7IP* are associated with testicular germ cell cancer. *Nat. Genet.* **42**, 604–607 (2010).
19. Kanetsky, P.A. *et al.* A second independent locus within *DMRT1* is associated with testicular germ cell tumor susceptibility. *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. Dalgaard, 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. & Hoei-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. & Chanock, S.J. Current status of genome-wide association studies in cancer. *Hum. Genet.* **130**, 59–78 (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. Moniot, B. *et al.* Hematopoietic prostaglandin D synthase (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 prostaglandin D synthase suppresses intestinal adenomas in $Apc^{Min/+}$ 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 remodellers promote DNA end resection. *Nature* **489**, 581–584 (2012).
29. Schoor, M., Schuster-Gossler, K., Roopenian, D. & Gossler, A. Skeletal dysplasias, growth retardation, reduced postnatal survival, and impaired fertility in mice lacking the SNF2/SWI2 family member ETL1. *Mech. Dev.* **85**, 73–83 (1999).
30. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
31. Gidekel, S., Pizov, G., Bergman, Y. & Pikarsky, E. Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell* **4**, 361–370 (2003).
32. Tsai, C.C., Su, P.F., Huang, Y.F., Yew, T.L. & Hung, S.C. Oct4 and Nanog directly regulate Dnmt1 to maintain self-renewal and undifferentiated state in mesenchymal stem cells. *Mol. Cell* **47**, 169–182 (2012).
33. Cheng, L. *et al.* OCT4: biological functions and clinical applications as a marker of germ cell neoplasia. *J. Pathol.* **211**, 1–9 (2007).
34. Clark, A.T. The stem cell identity of testicular cancer. *Stem Cell Rev.* **3**, 49–59 (2007).
35. Koster, R. *et al.* Cytoplasmic p21 expression levels determine cisplatin resistance in human testicular cancer. *J. Clin. Invest.* **120**, 3594–3605 (2010).
36. Skotheim, R.I. *et al.* Differentiation of human embryonal carcinomas *in vitro* and *in vivo* reveals expression profiles relevant to normal development. *Cancer Res.* **65**, 5588–5598 (2005).
37. Altshuler, D.M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
38. Kim, S., Sun, H., Tomchick, D.R., Yu, H. & Luo, X. Structure of human Mad1 C-terminal domain reveals its involvement in kinetochore targeting. *Proc. Natl. Acad. Sci. USA* **109**, 6549–6554 (2012).
39. Guo, Y. *et al.* Functional evaluation of missense variations in the human *MAD1L1* and *MAD2L1* genes and their impact on susceptibility to lung cancer. *J. Med. Genet.* **47**, 616–622 (2010).
40. Iwanaga, Y., Kasai, T., Kibler, K. & Jeang, K.T. Characterization of regions in hsMAD1 needed for binding hsMAD2. A polymorphic change in an hsMAD1 leucine zipper affects MAD1-MAD2 interaction and spindle checkpoint function. *J. Biol. Chem.* **277**, 31005–31013 (2002).
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 monocytes 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. Kissel, 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. Kuznetsov, 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.* **176**, 581–592 (2007).
50. Wu, J.Y. *et al.* Spermiogenesis and exchange of basic nuclear proteins are impaired in male germ cells lacking Camk4. *Nat. Genet.* **25**, 448–452 (2000).
51. Karlberg, S., Tiitinen, A. & Lipsanen-Nyman, M. Failure of sexual maturation in Mulibrey nanism. *N. Engl. J. Med.* **351**, 2559–2560 (2004).
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. Loveday, C. *et al.* Germline *RAD51C* mutations confer susceptibility to ovarian cancer. *Nat. Genet.* **44**, 475–476 (2012).
54. Mondal, G., Ohashi, A., Yang, L., Rowley, M. & Couch, F.J. Tex14, a Plk1-regulated protein, is required for kinetochore-microtubule attachment and regulation of the spindle assembly checkpoint. *Mol. Cell* **45**, 680–695 (2012).
55. Ishida, A., Sueyoshi, N., Shigeri, Y. & Kameshita, I. Negative regulation of multifunctional Ca^{2+}/calmodulin-dependent protein kinases: physiological and pharmacological significance of protein phosphatases. *Br. J. Pharmacol.* **154**, 729–740 (2008).
56. Avela, K. *et al.* Gene encoding a new RING–B-box–Coiled-coil protein is mutated in mulibrey nanism. *Nat. Genet.* **25**, 298–301 (2000).
57. Karlberg, 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. & Baker, M.C. Specific chromosome change i(12p) in testicular tumors? *Lancet* **2**, 1349 (1982).
59. Rodriguez, E. *et al.* Cytogenetic analysis of 124 prospectively ascertained male germ cell tumors. *Cancer Res.* **52**, 2285–2291 (1992).
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, OUHRH 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 described[18,19]. Genotype quality control metrics for STEED, FTCS and USC are described in the **Supplementary Note**[61]. The OUHRH 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 ($r^2 > 0.6$) 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 $P < 1 \times 10^{-4}$ 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, OUHRH 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 SequenceLDhot[62], a program that uses the approximate marginal likelihood method[63] 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 rates[64,65], and LD heatmaps were visualized in r^2 using the snp.plotter program[66].

The relative risk attributable to a set of SNPs (λ) was estimated using the following formula[67]

$$\lambda = \prod_{i=1}^{n} \frac{p_i(p_i r_{2i} + q_i r_{1i})^2 + q_i(p_i r_{1i} + q_i r_{0i})^2}{(p_i^2 r_{2i} + 2p_i q_i r_{1i} + q_i^2 r_{0i})^2}$$

where q_i is the minor allele frequency (MAF) of SNP i and $p_i = 1 - q_i$. SNP-specific risks for rare homozygotes, heterozygotes and common homozygotes are denoted by r_{0i}, r_{1i} and r_{2i}, 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 $\log\lambda/\log\lambda_0$, where λ_0 is the familial relative risk estimated from TGCT epidemiological studies ($\lambda_0 = 4$ with an affected father, $\lambda_0 = 8$ with an affected brother)[68].

Genomic annotation. Genomic annotation on high-LD surrogates ($r^2 \geq 0.8$, 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 HaploReg[69] and RegulomeDB[70] (**Supplementary Table 5**). rs12699477 did not have surrogates that met the requirement of $r^2 \geq 0.8$; 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. *Hum. Mol. Genet.* published online; doi:10.1093/hmg/ddt109 (5 March 2013).
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 SNP/haplotype 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 and environmental aetiology. *Br. J. Cancer* **90**, 1765–1770 (2004).
69. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
70. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).