HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2013 December 01.

Published in final edited form as:

Nat Genet. 2013 June ; 45(6): 680–685. doi:10.1038/ng.2634.

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

Charles C. Chung[1,2,*], **Peter A. Kanetsky**[3,4,*], **Zhaoming Wang**[1,2,*], **Michelle A. T. Hildebrandt**[5,*], **Roelof Koster**[6,*], **Rolf I. Skotheim**[7,8,*], **Christian P. Kratz**[1,*,†], **Clare Turnbull**[9,*], **Victoria K. Cortessis**[10,*], **Anne C. Bakken**[7,8], **D. Timothy Bishop**[11], **Michael B. Cook**[1], **R. Loren Erickson**[12], **Sophie D. Fossåa**[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], **Katherine A. McGlynn**[1], **Stephen J. Chanock**[1,‡], and **Katherine L. Nathanson**[3,6,‡]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA

[2]Cancer Genome Research Laboratory, Division of Cancer Epidemiology and Genetics, SAIC-Frederick Inc., 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

Corresponding Author: Katherine L. Nathanson, MD., Department of Medicine, Translational Medicine and Human Genetics, Room 351 BRB 2/3, 421 Curie Blvd, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, Phone: 215-573-9840, Fax: 215-573-6298, knathans@exchange.upenn.edu.

*The authors contributed jointly to this work.

‡These authors jointly directed this work.

†Current address - Center for Pediatrics and Adolescent Medicine, Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany

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.

Conflict of Interest

There are no conflicts of interest.

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

[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, Surrey, 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' University Hospital, Leeds, UK

[12]Walter Reed Army Institute of Research, Silver Spring, Maryland, USA

[13]Department of Oncology, Oslo University Hospital, The Norwegian Radium 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 and School of Public Health, University of Washington, Seattle, Washington, USA

Abstract

We conducted a meta-analysis to identify new loci for testicular germ cell tumor (TGCT) susceptibility. In the discovery phase, 931 affected individuals and 1,975 controls from three genome wide association studies (GWAS) were analyzed. Replication was conducted in six independent sample sets totaling 3,211 affected individuals and 7,591 controls. In the combined analysis, TGCT risk was significantly associated with markers at four novel loci: 4q22.2 in *HPGDS* (per allele odds ratio (OR) 1.19, 95%CI 1.12–1.26, $P = 1.11\times10^{-8}$); 7p22.3 in *MAD1L1* (OR 1.21, 95%CI 1.14–1.29, $P = 5.59\times10^{-9}$); 16q22.3 in *RFWD3* (OR 1.26, 95%CI 1.18–1.34, $P = 5.15\times10^{-12}$); and 17q22 (rs9905704; OR 1.27, 95%CI 1.18–1.33; $P = 4.32\times10^{-13}$, and rs7221274; OR 1.20, 95%CI 1.12–1.28 $P = 4.04\times10^{-9}$), a locus which includes *TEX14, RAD51C* and *PPM1E*. The new TGCT susceptibility loci contain biologically plausible genes encoding proteins important for male germ cell development, chromosomal segregation and DNA damage response.

In the United States, testicular germ cell tumors (TGCT) are the most common cancers in young men, with a peak incidence among those aged 25 to 34 years. The incidence of TGCT has more than doubled among white men 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 compared to African ancestry[2]. Established risk factors for TGCT include family history, cryptorchidism, adult height and prior TGCT history; several recent studies also have implicated marijuana use[4–7]. First degree relatives of affected men have been shown consistently to have an increased TGCT risk (5- to 19-fold for brothers and 2- to 4-fold for fathers)[8–11], the highest for any cancer. Further, the estimated heritability of TGCT is third among all cancers, with genetic effects estimated to account for 25% of TGCT susceptibility[12]. These observations, coupled with twin studies[13–15], support a strong genetic component contributing to TGCT susceptibility.

Despite the greatly increased relative risk of TGCT in family members, candidate gene and linkage approaches yielded little progress in identifying specific genetic risk factors. Initially, two independent genome wide association studies (GWAS) identified allele variation within *KITLG* on 12q22 as the strongest genetic risk factor for TGCT, with a per allele odds ratio (OR) greater than 3[16,17]. Variants on 5p15.33 (*TERT*- two independent loci*)*, 5q31.3 (*SPRY4)*, 6p21.3 (*BAK1)*, 9p24.3 (*DMRT1*- two independent loci), and 12p13.1 (*ATF7IP)* also have been associated with TGCT risk[16–21]. The per allele ORs for the identified TGCT susceptibility alleles are in large part higher than those identified for other cancers, which may be due, in part, to the homogeneity of the disease, as all TGCT are thought to arise from the primordial germ cell[22,23]. Multiple additional loci are expected to contribute to susceptibility as has been shown for cancers of lower heritability[24]. Combining multiple GWAS represents a step to increase power to detect additional genetic risk factors failing to reach genome-wide significance in individual studies.

We performed a meta-analysis of the most promising 340 SNPs (after excluding previously reported loci) observed in the adjusted pooled analysis of the combined 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 nine loci had *P* values below 10^{-4}, of which 12 localized to the *MAD1L1* gene locus (7p22.2) (details of correlation between NCI and UPENN study for top 40 SNPs in Supplementary Table 2). The most significant SNP marker from each of nine loci, plus eight additional markers were selected for replication (n=17). An *in silico* analysis of these 17 SNPs was performed in the GWAS data from the University of Southern California (USC) and the UK Testicular Cancer Collaboration (UKTCC)[18], followed by genotyping in four additional TGCT case-control studies from: Fred Hutchinson Cancer Center (Adult Testicular Lifestyle and Blood Specimen [ATLAS] study), University of Pennsylvania (Testicular Cancer in Philadelphia Area Counties [TestPAC] study), Oslo University Hospital-Radium Hospital, Norway (OUHRH), and 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* value < 5×10^{-8}) (Table 1; Supplementary Table 4).

The most significant 4q22.2 SNP marker, rs17021463, is located within the intron of the hematopoietic prostaglandin D synthase gene, *HPGDS* (*P* = 1.11×10^{-8}, OR 1.19, 95%CI

1.12–1.26) (Figure 1A, Table 1). In mice, *hpgds* is expressed in the early embryonic male gonad and appears 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 highly correlate with *HPGDS* rs17021463 ($r^2 \geq 0.8$, 1000 Genomes CEU data, Supplementary Table 5). Notably, rs35744894 ($r^2 = 0.87$) changes a DMRT2 binding motif (Supplementary Table 6); variation in *DMRT1* has been associated with TGCT risk[19].

Fifty-three of 71 surrogate markers that were highly correlated with *HPGDS* rs17021463 ($r^2 \geq 0.8$, 1000 Genomes CEU data, Supplementary Table 5) across a 200kb window mapped within or near an adjacent gene, *SMARCAD1* (SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily a, containing DEAD/H box 1). SMARCAD1 is a chromatin remodeler, which restores silenced heterochromatin domains in dividing cells and participates in DNA damage response[27,28]. Homozygous mutant mice display developmental defects, including impaired fertility[29]. Surrogate markers included one nonsynonymous substitution, rs7439869, at codon 301 ($r^2 = 0.93$, 1000 Genomes CEU, T>C, Val>Ala). Although it is predicted to be tolerated by PolyPhen2[30], it changes an OCT4 (*POUF5F1*) and SOX4 binding motif (Supplementary Table 6). OCT4 is a transcription factor, which regulates pluripotency in a number of cell types, including primordial germ cells, and is expressed in TGCT[31–36].

We identified a locus on 7p22.3, harboring mitotic arrest deficient-like 1 (*MAD1L1*) gene, which encodes MAD1. The most significant SNP without study heterogeneity, rs12699477, localized in intron 17 ($P = 5.59 \times 10^{-9}$, OR 1.21, 95%CI (1.14–1.29)) (Figure 1B). Of note, the risk allele (C) at rs12699477 is more prevalent in populations of European (29%) than those of African ancestry (8%) in 1000 Genomes[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].

Among the 35 SNPs that are highly correlated with *MAD1L1* rs12699477 ($r^2 \geq 0.7$, 1000 Genomes CEU data, Supplementary Table 5), rs1801368 is a missense mutation at codon 558 (G>A, Arg>His) that resides in the *MAD1L1* second leucine zipper domain. Arg558His 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 rs10275045 ($P=3.78\times10^{-10}$, OR 1.20, 95%CI (1.13–1.27)) and rs3778991 ($P=6.73\times10^{-10}$, OR 1.21, 95%CI (1.14–1.28)). However, both displayed significant study heterogeneity (Supplementary Table 4). The r^2 between our strongest signal at rs12699477 and these markers is 0.66 and 0.50, respectively, in the STEED controls. A conditional analysis resulted in a marked attenuation of the signal, supporting a single TGCT susceptibility locus across *MAD1L1* on 7q22.3 (Supplementary Table 7).

We observed a significant TGCT association with rs4888262 on 16q22.3 ($P = 5.15\times10^{-12}$, OR 1.26, 95%CI 1.18–1.34), which is a synonymous SNP in codon 404 (G>A, Thr) of the ring finger WD domain 3 (*RFWD3*) (Table 1, Figure 1C). RFWD3 is an E3 ubiquitin ligase that positively regulates p53 stability by forming a RFWD3-MDM2-p53 complex, thereby

protecting p53 from degradation by MDM2 polyubiquitination[41,42]. Within the LD interval are SNPs that map to two additional genes, the golgi glycoprotein 1 (*GLG1*) and mixed lineage kinase domain-like (*MLKL*); the latter of which has been recently identified as a key of mediator of TNF-induced necrosis, downstream of receptor interacting protein kinase 3 (RIP3)[43,44] (Figure 1C). We note that rs3851729, which is highly correlated with rs4888262 (r^2 = 0.77, 1000 Genomes CEU), maps to a highly conserved sequence in the 3′ UTR of *GLG1*; similarly, rs4072222 (r^2 = 0.87, 1000 Genomes CEU) maps to an intron of *MLKL* (Supplementary Table 5). Both susceptibility variants are *cis*-eQTLs that influence *MLKL* and *RFWD3* expression in monocytes[45].

We identified two highly correlated SNPs (r^2 = 0.74 in the STEED controls) on 17q22, rs9905704 ($P = 4.32\times10^{-13}$, OR 1.27, 95%CI 1.18–1.33) and rs7221274 ($P = 4.04\times10^{-9}$, OR 1.20, 95%CI 1.12–1.28) (Table 1, Figure 1D). In a conditional analysis, the signal at one SNP was markedly attenuated by the other, indicating a single 17q22 TGCT susceptibility locus (Supplementary Table 7). Within this LD block are at least six plausible candidate genes: *RAD51C* (RAD51 homolog C [S. cerevisiae]), *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) (Figure 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, in which rare mutations confer susceptibly to ovarian cancer[52,53]. Of male *rad51c*$^{ko/neo}$ mice, approximately one-third were found to be infertile due to impaired spermatogenesis[49]. TEX14 is an essential component of germ cell intercellular bridges, evolutionarily conserved structures from invertebrates to humans that allows clonal development of daughter cells in syncytium; targeted disruption of *Tex14* results in male sterility in mice[48]. TEX14 also has been implicated as an important component of kinetochores (KTs) and interacts with MAD1[54]. *PPM1E* encodes a phosphatase that dephosphorylates to switch-off CaMK4 (calcium/calmodulin-dependent protein kinase IV), 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 (MUL; MIM 253250)[56], in which adult males have testicular failure[57]. Three SNPs - rs8077332, rs11652713, and rs9898048 - map within *TRIM37* and are in perfect LD with rs7221274 (r^2 = 1, 1000 Genomes CEU, Supplementary Table 5); all are *cis*-eQTL 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 on 17q22.

In our meta-analysis of GWAS studies, we have identified four new TGCT susceptibility loci at 4q22, 7q22, 16q22.3, and 17q22. In total, 10 loci now have been conclusively associated with TGCT susceptibility. The four newly identified susceptibility alleles account for 2% of the risk to the brothers and 3% of risk to the sons of TGCT patients, increasing the cumulative total of 12 susceptibility alleles (two susceptibility alleles from *TERT-CLPTM1L* [5p15] and two from *DMRT1* locus [9p24]) to 14% and 21% of the risk to brothers and sons, respectively. Based on the high heritability of TGCT, more than one hundred additional loci are expected to be discovered[24]. Notably, the allelic ORs associated with these novel loci are

in the range of 1.2 to 1.3, continuing the trend of identifying loci with higher odds ratios for TGCT than for other cancer types[23].

Interestingly, each locus harbors biologically plausible candidate genes implicating several pathways – most strikingly, spermatogenesis and male germ cell development (*HPGDS, SMARCAD1, SEPT4, TEX14, RAD51C, PPM1E, TRIM37*), chromosomal segregation (*MAD1L1, TEX14, SKA2*), and DNA damage response (*SMARCAD1, RFWD3, RAD51C*). None of the four newly identified loci have been previously implicated in GWAS of other cancers, further supporting that there are distinct pathways and regions implicated in TGCT susceptibility; however rare mutations in *RAD51C* have been implicated in ovarian cancer susceptibility[53]. TGCT susceptibility 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 the primordial germ cell.

This study is the first to implicate variation within genes involved in chromosomal segregation as associated with cancer susceptibility. TGCT karyotypes 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. Numerous potential regulatory SNPs were identified, suggesting that newly identified associations might be mediated by plausible 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 novel insights into this unique disease with high heritability.

ONLINE METHODS

Studies

Detailed characteristics of the study populations are described in both the Supplementary Note and 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 scans (UPENN and UKTCC) were previously described[18,19]. Genotype quality control metrics for STEED, FTCS, and USC are described in Supplementary Note[61]. OUHRH and MDA studies were genotyped using the 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 the HapMap samples (n=270) with greater than 99% genotype concordance. TestPAC and ATLAS studies conducted genotyping using the iPLEX mass

array platform (Sequenom, Inc.) following manufacturer's protocol. Assays at all genotyping centers included at least four negative controls and 2–5% duplicates on each plate. Standard quality control protocol was implemented; SNP call rate > 95%, no deviation from Hardy-Weinberg equilibrium in controls at $P<0.00001$, <2% discordance between genotypes in duplicate had to be fulfilled and cluster plots for SNPs that were close to failing any of the QC criteria were re-examined centrally.

Statistical analysis

Two genome-wide scans from the National Cancer Institute (STEED and FTCS) were analyzed as a combined dataset using a logistic regression model for trend effect adjusted for age, study, and additionally for one eigenvector (only one with $p < 0.05$) to account for population stratification in this European population. From the top 500 SNPs by trend *P* values from the NCI scan excluding previously reported ones, 340 SNPs were selected based on the availability of surrogates ($r^2 > 0.6$) in the previous TGCT GWAS scan from the University of Pennsylvania. Since SNP content differs between the Illumina and Affymetrix platforms, the best correlated surrogate per each marker was paired to perform a discovery meta-analysis (111 SNPs, direct match; 229 SNPs, surrogate match). From the discovery meta-analysis, 17 of 40 SNPs with *P* values < 10^{-4} were selected for follow up in the remaining studies. *In silico* follow-up was done in the USC and UKTCC scans, whereas additional genotyping was done in TestPAC, ATLAS, OUHRH and MDA studies (Supplementary Table 3). Not all markers were available for replication efforts from all sites (see Supplementary Table 4).

The meta-analysis was conducted using the suite of tools in GLU (Genotyping Library and Utilities) software, combining study-specific odds ratio (OR) estimates using a fixed effects model, which used the inverse-variance method to estimate the combined OR and its 95% confidence intervals (CIs). To assess existence of heterogeneity among studies, Cochran's Q statistic was used to calculate *P* for heterogeneity.

Recombination hotspots were identified in the vicinity of the novel 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 five unique sets of 100 control samples drawn from STEED. PHASE v2.1 program was used to calculate background recombination rates[64,65] and LD heatmap was visualized in r^2 using 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_t is the minor allele frequency 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 minor allele frequencies and odds ratio estimates from SNP association analyses were used to estimate relative risks. This formula assumes the effects of all SNPs in the set are multiplicative. The proportion of familial risk attributable to a set of SNPs was calculated as $\frac{\log\lambda}{\log\lambda_0}$, where λ_0 is the familial relative risk estimated from TGCT epidemiological studies (λ_0 =4 for affected father, λ_0 =8 for affected brother)[68].

Genomic annotation

Genomic annotation on high LD surrogates ($r^2 \geq 0.8$, 1000 Genomes CEU) of 5 SNPs (rs17021463, rs12699477, rs4888262, rs9905704, and rs7221274) from the four TGCT susceptibility loci identified in the current study was conducted using ENCODE tools – HaploReg[69] and RegulomeDB[70] (Supplementary Table 5). rs12699477 did not have surrogates with $r^2 \geq 0.8$ threshold, thus we lowered the threshold to 0.7 for surrogates, and then conducted annotation. All surrogates were queried in RegulomeDB browser to cross examine predicted regulatory DNA elements such as regions of DNase hypersensitivity, binding sites of transcription factors, and promoter regions that have been biochemically characterized to regulation transcription. Summaries of each SNP analysis by RegulomeDB browser expressed in scores are added to Supplementary Table 5. To predict potential regulatory SNPs, we assessed SNPs that meet one of the following criteria - 1) conserved (GERP and/or Siphy); 2) present in a promoter or DNase hypersensitivity region; or 3) 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 motif, and are annotated further with the motif of interest and their log-odds (LOD) motif score for the specific SNP of interest in Supplementary Table 6. Two SNPs in 3′-UTR regions were evaluated using SNP Function Prediction for changes in miRNA binding sites and are included in Supplementary Table 6.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services nor does the mention of trade names, commercial products or organization indicate endorsement by the U.S. Government. The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, NCI, the screening center investigators and staff of the PLCO Cancer Screening Trial, Mr. Thomas Riley and staff at Information Management Services, Inc., and Ms. Barbara O'Brien and staff at Westat, Inc. for their contributions to the PLCO Cancer Screening Trial. The authors thank Stephanie Ciosek, Monique McDermoth and Kurt D'Andrea for expert assistance with the conduct of TestPAC. The authors thank Laurence Kolonel and Loic Le Marchand for providing access to the Multi-Ethnic Cohort aggressive prostate cancer scan, as well as Juan Pablo Lewinger, Malcolm Pike, David J. Van Den Berg, and Kimberly 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 National Cancer Institute and by support services contract HHSN261200655004C with Westat, Inc. The Penn GWAS (UPenn) and replication effort for the TestPAC study was supported by the Abramson Cancer Center at the University of Pennsylvania and National Institute of Health grant R01CA114478 to Drs. Kanetsky and Nathanson. Replication effort for the ATLAS study was supported by the National Institute of Health grant R01CA085914 to Dr. Schwartz. USC GWAS controls were supported by the Multiethnic Cohort Study (NCI U01-CA98758). USC GWAS testicular cases and Familial Study were supported by the California Cancer Research Program (99-00505V-10260, 03-00174VRS-30021) and National Cancer Institute (R01CA102042) grants to Dr. Cortessis 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 MD Anderson Senior Research Trust Fellowship to Dr. Wu. 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 (WTCCC) 2. Support was provided by the Norwegian Cancer Society to Drs. Lothe and Skotheim, Health Region South-Eastern Norway to Drs. Lothe and Fosså and the Norwegian ExtraFoundation for Health and Rehabilitation to Dr. Fosså.

References

1. Rosen A, Jayram G, Drazer M, Eggener SE. Global trends in testicular cancer incidence and mortality. Eur Urol. 2011; 60:374–9. [PubMed: 21612857]
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. National Cancer Institute; Bethesda MD: 2012.
3. Stang A, et al. Gonadal and extragonadal germ cell tumours in the United States, 1973–2007. Int J Androl. 2012; 35:616–25. [PubMed: 22320869]
4. McGlynn KA, Trabert B. Adolescent and adult risk factors for testicular cancer. Nat Rev Urol. 2012; 9:339–49. [PubMed: 22508459]
5. Daling JR, et al. Association of marijuana use and the incidence of testicular germ cell tumors. Cancer. 2009; 115:1215–23. [PubMed: 19204904]
6. Trabert B, Sigurdson AJ, Sweeney AM, Strom SS, McGlynn KA. Marijuana use and testicular germ cell tumors. Cancer. 2011; 117:848–53. [PubMed: 20925043]
7. Lacson JC, 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. 2012; 118:5374083.
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. 2004; 13:1316–24. [PubMed: 15298952]
9. Chia VM, et al. Risk of cancer in first- and second-degree relatives of testicular germ cell tumor cases and controls. Int J Cancer. 2009; 124:952–7. [PubMed: 19035442]
10. Heimdal K, et al. Risk of cancer in relatives of testicular cancer patients. Br J Cancer. 1996; 73:970–3. [PubMed: 8611417]
11. Sonneveld DJ, et al. Familial testicular cancer in a single-centre population. Eur J Cancer. 1999; 35:1368–73. [PubMed: 10658529]
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. 2002; 99:260–6. [PubMed: 11979442]
13. Neale RE, Carriere P, Murphy MF, Baade PD. Testicular cancer in twins: a meta-analysis. Br J Cancer. 2008; 98:171–3. [PubMed: 18071360]
14. Swerdlow AJ, De Stavola BL, Swanwick MA, Maconochie NE. Risks of breast and testicular cancers in young adult twins in England and Wales: evidence on prenatal and genetic aetiology. Lancet. 1997; 350:1723–8. [PubMed: 9413462]
15. Braun MM, Ahlbom A, Floderus B, Brinton LA, Hoover RN. Effect of twinship on incidence of cancer of the testis, breast, and other sites (Sweden). Cancer Causes Control. 1995; 6:519–24. [PubMed: 8580300]
16. Kanetsky PA, et al. Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. Nat Genet. 2009; 41:811–5. [PubMed: 19483682]
17. Rapley EA, et al. A genome-wide association study of testicular germ cell tumor. Nat Genet. 2009; 41:807–10. [PubMed: 19483681]
18. Turnbull C, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat Genet. 2010; 42:604–7. [PubMed: 20543847]
19. 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]
20. Ferlin A, et al. Variants in KITLG predispose to testicular germ cell cancer independently from spermatogenic function. Endocr Relat Cancer. 2012; 19:101–8. [PubMed: 22194441]

21. Dalgaard MD, et al. A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation. J Med Genet. 2012; 49:58–65. [PubMed: 22140272]
22. Rajpert-de Meyts E, Hoei-Hansen CE. From gonocytes to testicular cancer: the role of impaired gonadal development. Ann N Y Acad Sci. 2007; 1120:168–80. [PubMed: 18184914]
23. Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. Hum Genet. 2011; 130:59–78. [PubMed: 21678065]
24. Park JH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet. 2010; 42:570–5. [PubMed: 20562874]
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. 2011; 240:2335–43. [PubMed: 21887724]
26. Park JM, et al. Hematopoietic prostaglandin D synthase suppresses intestinal adenomas in ApcMin/+ mice. Cancer Res. 2007; 67:881–9. [PubMed: 17283118]
27. Rowbotham SP, et al. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. Mol Cell. 2011; 42:285–96. [PubMed: 21549307]
28. Costelloe T, et al. The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection. Nature. 2012; 489:581–4. [PubMed: 22960744]
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. 1999; 85:73–83. [PubMed: 10415348]
30. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–9. [PubMed: 20354512]
31. Gidekel S, Pizov G, Bergman Y, Pikarsky E. Oct-3/4 is a dose-dependent oncogenic fate determinant. Cancer Cell. 2003; 4:361–70. [PubMed: 14667503]
32. Tsai CC, Su PF, Huang YF, Yew TL, Hung SC. Oct4 and Nanog directly regulate Dnmt1 to maintain self-renewal and undifferentiated state in mesenchymal stem cells. Mol Cell. 2012; 47:169–82. [PubMed: 22795133]
33. Cheng L, et al. OCT4: biological functions and clinical applications as a marker of germ cell neoplasia. J Pathol. 2007; 211:1–9. [PubMed: 17117392]
34. Clark AT. The stem cell identity of testicular cancer. Stem Cell Rev. 2007; 3:49–59. [PubMed: 17873381]
35. Koster R, et al. Cytoplasmic p21 expression levels determine cisplatin resistance in human testicular cancer. J Clin Invest. 2010; 120:3594–605. [PubMed: 20811155]
36. Skotheim RI, et al. Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal development. Cancer Res. 2005; 65:5588–98. [PubMed: 15994931]
37. Altshuler DM, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010; 467:52–8. [PubMed: 20811451]
38. Kim S, Sun H, Tomchick DR, Yu H, Luo X. Structure of human Mad1 C-terminal domain reveals its involvement in kinetochore targeting. Proc Natl Acad Sci U S A. 2012; 109:6549–54. [PubMed: 22493223]
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. 2010; 47:616–22. [PubMed: 20516147]
40. Iwanaga Y, Kasai T, Kibler K, Jeang KT. 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. 2002; 277:31005–13. [PubMed: 12042300]
41. Fu X, et al. RFWD3-Mdm2 ubiquitin ligase complex positively regulates p53 stability in response to DNA damage. Proc Natl Acad Sci U S A. 2010; 107:4579–84. [PubMed: 20173098]
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. 2011; 286:22314–22. [PubMed: 21558276]

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 U S A. 2012; 109:5322–7. [PubMed: 22421439]
44. Sun L, et al. Mixed lineage kinase domain-like protein mediates necrosis signaling downstream of RIP3 kinase. Cell. 2012; 148:213–27. [PubMed: 22265413]
45. Zeller T, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One. 2010; 5:e10693. [PubMed: 20502693]
46. Ihara M, et al. Cortical organization by the septin cytoskeleton is essential for structural and mechanical integrity of mammalian spermatozoa. Dev Cell. 2005; 8:343–52. [PubMed: 15737930]
47. Kissel H, et al. The Sept4 septin locus is required for sperm terminal differentiation in mice. Dev Cell. 2005; 8:353–64. [PubMed: 15737931]
48. 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]
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. 2007; 176:581–92. [PubMed: 17312021]
50. Wu JY, et al. Spermiogenesis and exchange of basic nuclear proteins are impaired in male germ cells lacking Camk4. Nat Genet. 2000; 25:448–52. [PubMed: 10932193]
51. Karlberg S, Tiitinen A, Lipsanen-Nyman M. Failure of sexual maturation in Mulibrey nanism. N Engl J Med. 2004; 351:2559–60. [PubMed: 15590968]
52. Meindl A, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet. 2010; 42:410–4. [PubMed: 20400964]
53. Loveday C, et al. Germline RAD51C mutations confer susceptibility to ovarian cancer. Nat Genet. 2012; 44:475–6. author reply 476. [PubMed: 22538716]
54. 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]
55. Ishida A, Sueyoshi N, Shigeri Y, Kameshita I. Negative regulation of multifunctional Ca2+/calmodulin-dependent protein kinases: physiological and pharmacological significance of protein phosphatases. Br J Pharmacol. 2008; 154:729–40. [PubMed: 18454172]
56. Avela K, et al. Gene encoding a new RING-B-box-Coiled-coil protein is mutated in mulibrey nanism. Nat Genet. 2000; 25:298–301. [PubMed: 10888877]
57. Karlberg S, et al. Testicular failure and male infertility in the monogenic Mulibrey nanism disorder. J Clin Endocrinol Metab. 2011; 96:3399–407. [PubMed: 21865362]
58. Atkin NB, Baker MC. Specific chromosome change i(12p) in testicular tumors? Lancet. 1982; 2:1349. [PubMed: 6128640]
59. Rodriguez E, et al. Cytogenetic analysis of 124 prospectively ascertained male germ cell tumors. Cancer Res. 1992; 52:2285–91. [PubMed: 1313741]
60. Skotheim RI, et al. Novel genomic aberrations in testicular germ cell tumors by array-CGH, and associated gene expression changes. Cellular Oncology. 2006; 28:315–26. [PubMed: 17167184]
61. Schumacher FR, et al. Testicular germ cell tumor susceptibility associated with the *UCK2* locus on chromosome 1q23. Hum Mol Genet. (In press).
62. Fearnhead P. SequenceLDhot: detecting recombination hotspots. Bioinformatics. 2006; 22:3061–6. [PubMed: 17060358]
63. Fearnhead P, Donnelly P. Approximate likelihood methods for estimating local recombination rates. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2002; 64:657–680.
64. Crawford DC, et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat Genet. 2004; 36:700–6. [PubMed: 15184900]
65. Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics. 2003; 165:2213–33. [PubMed: 14704198]
66. Luna A, Nicodemus KK. snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. Bioinformatics. 2007; 23:774–6. [PubMed: 17234637]

67. Cox A, et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet. 2007; 39:352–8. [PubMed: 17293864]

68. Hemminki K, Li X. Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. Br J Cancer. 2004; 90:1765–70. [PubMed: 15208620]

69. 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]

70. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012; 22:1790–7. [PubMed: 22955989]

Figure 1a

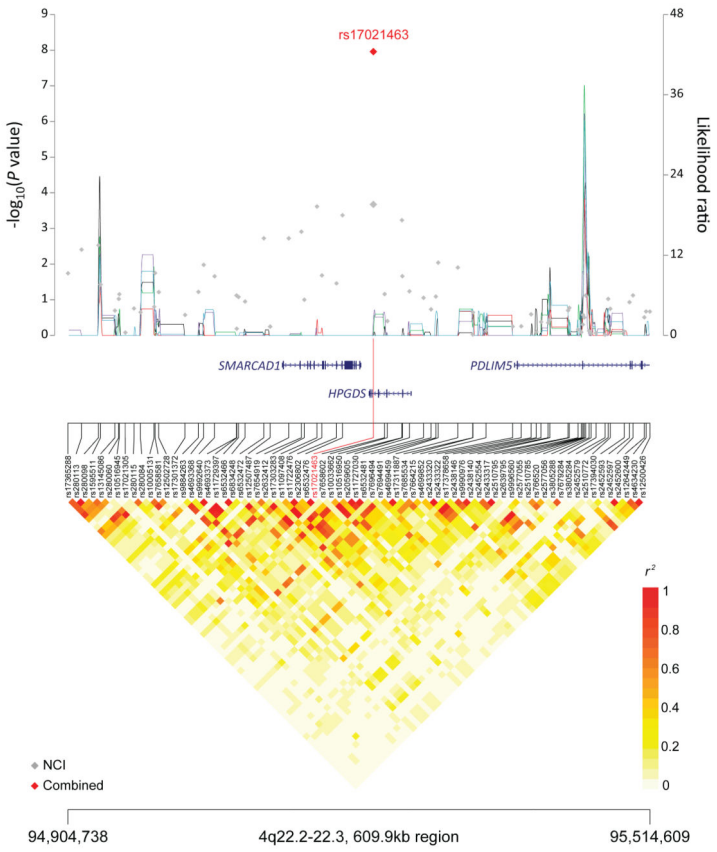

Figure 1b

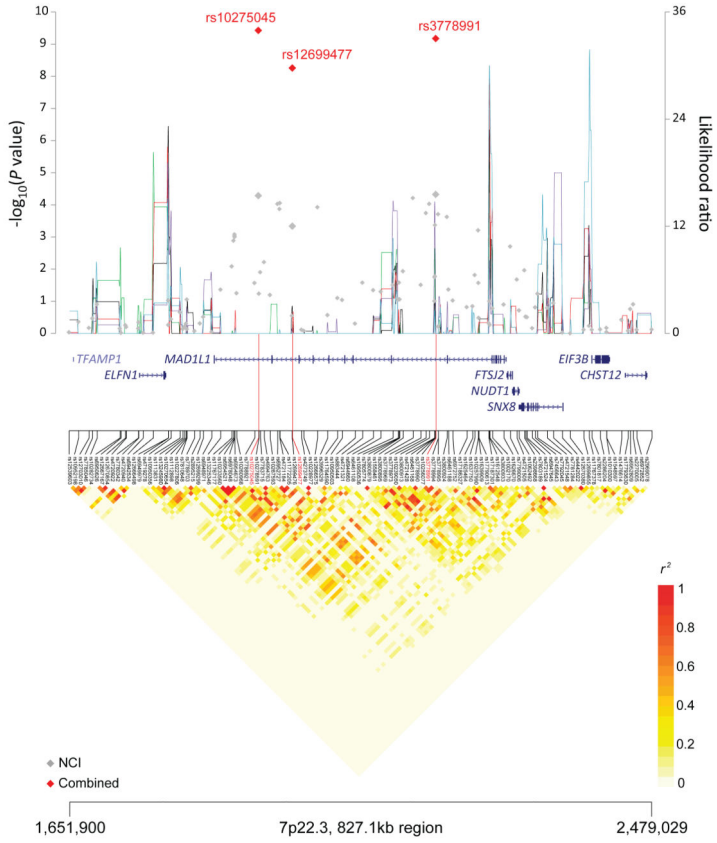

Figure 1c

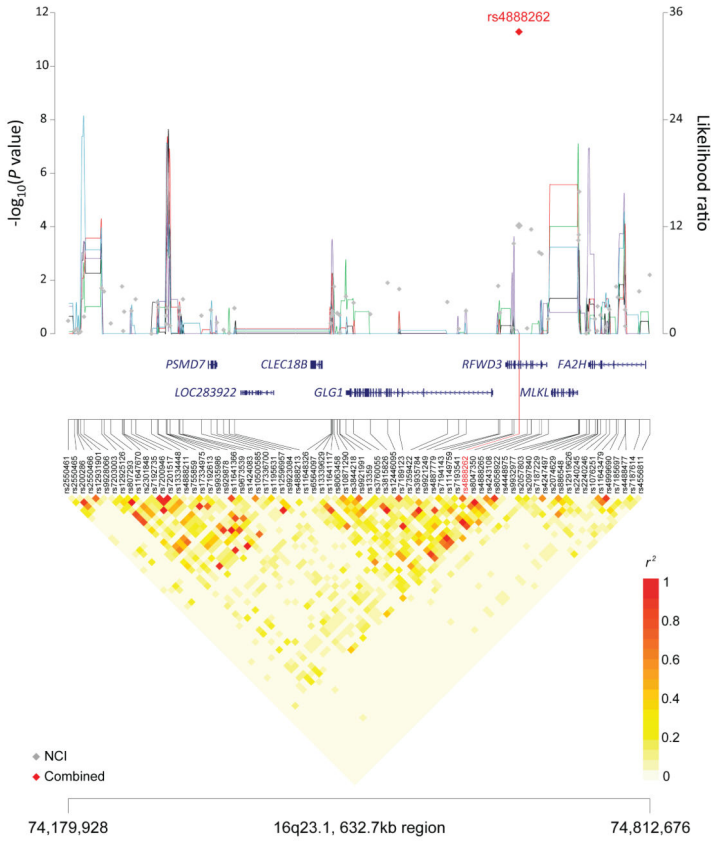

Figure 1d

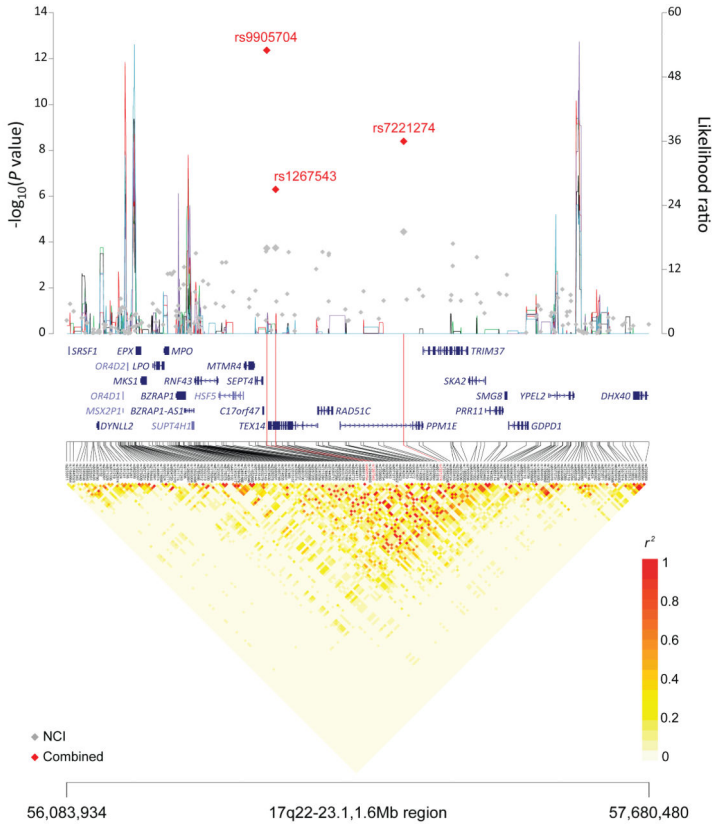

Figure 1. Recombination plot and linkage disequilibrium structure for the four new TGCT susceptibility regions at 4q22.2, 7p22.3, 16q22.3 and 17q22 (a–d)

Regional plots of association results, recombination hotspots and linkage disequilibrium for the (a) 4q22.2–22.3:94,904,738–95,514,609, (b) 7p22.3:1,651,900–2,479,029, (c) 16q23.1:74,179,928–74,812,676 and (d) 17q22–23.1:56,083,934–57,680,480 TGCT susceptibility loci. (a–d) Combined meta-analysis results are shown as red diamonds with rs numbers labeled, and the NCI scan in gray. For each plot, −log10*P* values (y axis, left) of the SNPs are shown according to their chromosomal positions (x axis). Linkage disequilibrium structure based on NCI controls (n=1,188) was visualized by snp.plotter software. The line graph shows likelihood ratio statistics (y axis, right) for recombination hotspot by SequenceLDhot software and five different colors represent 5 tests of 100 controls from NCI without resampling. Physical locations of each region are based on NCBI Build 37 of the human genome.

Table 1

Meta-analysis identified novel loci associated with testicular germ cell tumor

SNP[1]	Nearby genes	Study[2]	Cases	Controls	EAF[3]	Allelic OR (95% CI)	*P* value	*P* for heterogeneity
rs17021463	*HPGDS*	NCI	582	1055	0.475	1.33 (1.14–1.55)	2.12×10^{-4}	
G|T	4q22.2	UPENN	349	919	0.433	1.28 (1.07–1.53)	6.67×10^{-3}	
		Discovery	931	1974		1.31 (1.16–1.47)	6.09×10^{-6}	0.750
		UKTCC	979	4947	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	
		Replication	2935	7172		1.15 (1.07–1.23)	7.01×10^{-5}	0.862
		Combined	3866	9146		1.19 (1.12–1.26)	1.11×10^{-8}	0.583
rs12699477	*MAD1L1*	NCI	582	1056	0.412	1.31 (1.13–1.52)	4.64×10^{-4}	
T|C	7p22.3	UPENN	349	919	0.407	1.37 (1.15–1.63)	4.25×10^{-4}	
		Discovery	931	1975		1.34 (1.19–1.50)	7.16×10^{-7}	0.704
		UKTCC	979	4947	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	
		Replication	1901	6449		1.16 (1.07–1.25)	2.41×10^{-4}	0.645
		Combined	2832	8424		1.21 (1.14–1.29)	5.59×10^{-9}	0.318
rs4888262	*RFWD3*	NCI	582	1052	0.552	1.36 (1.17–1.59)	9.12×10^{-5}	
T|C	16q22.3	UPENN	349	919	0.498	1.39 (1.16–1.67)	3.12×10^{-4}	
		Discovery	931	1971		1.37 (1.22–1.54)	1.39×10^{-7}	0.858
		UKTCC	957	4940	0.458	1.22 (1.10–1.34)	1.02×10^{-4}	

SNP[1]	Nearby genes	Study[2]	Cases	Controls	EAF[3]	Allelic OR (95% CI)	*P* value	*P* for heterogeneity
		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	
		Replication	1870	6422		1.21 (1.12–1.31)	1.62×10^{-6}	0.568
		Combined	2801	8393		1.26 (1.18–1.34)	5.15×10^{-12}	0.397
rs9905704	*TEX14*	NCI	582	1054	0.719	1.37 (1.16–1.61)	1.88×10^{-4}	
G|T	17q22	UPENN	349	919	0.663	1.30 (1.08–1.56)	5.46×10^{-3}	
		Discovery	931	1973		1.33 (1.19–1.52)	3.44×10^{-6}	0.674
		UKTCC	979	4947	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	
		Replication	2932	7179		1.23 (1.15–1.32)	1.37×10^{-8}	0.628
		Combined	3863	9152		1.27 (1.18–1.33)	4.32×10^{-13}	0.668
rs7221274	*PPM1E*	NCI	582	1056	0.660	1.39 (1.19–1.61)	3.56×10^{-5}	
G|A	17q22	UPENN	349	919	0.598	1.19 (1.00–1.43)	0.053	
		Discovery	931	1975		1.30 (1.16–1.47)	7.79×10^{-6}	0.197
		UKTCC	979	4947	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	
		Replication	2898	7145		1.16 (1.09–1.25)	3.27×10^{-5}	0.228
		Combined	3829	9120		1.20 (1.12–1.28)	4.04×10^{-9}	0.130

[1] SNP genotype depicted as reference allele|effect allele

[2] NCI depicts combined analysis results of the two GWAS scans STEED and FTCS performed at NCI Discovery depicts initial meta-analysis of NCI and UPENN and Replication depicts meta-analysis of the rest

[3] EAF depicts effect allele frequency in control