Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor

Zhaoming Wang[1†], Katherine A. McGlynn[1], Ewa Rajpert-De Meyts[2], D. Timothy Bishop[3], Charles Chung[1], Marlene D. Dalgaard[2,4], Mark H. Greene[1], Ramneek Gupta[4], Tom Grotmol[5], Trine B. Haugen[6], Robert Karlsson[7], Kevin Litchfield[8], Nandita Mitra[9], Kasper Nielsen[4], Louise C. Pyle[10,11], Stephen M. Schwartz[12], Vésteinn Thorsson[13], Saran Vardhanabhuti[14], Fredrik Wiklund[7], Clare Turnbull[8,15], Stephen J. Chanock[1], Peter A. Kanetsky[16,18], Katherine L. Nathanson[10,17,18] for the Testicular Cancer Consortium

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

[2] Department of Growth and Reproduction, Copenhagen University Hospital (Rigshospitalet), Copenhagen, Denmark

[3] Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK

[4] Center of Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kongens Lyngby, Denmark

[5] Cancer Registry of Norway, Oslo, Norway

[6] Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Oslo, Norway

[7]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

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

[9] Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[10] Department of Medicine, Division of Translational Medicine and Human Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA

[11] Division of Human Genetics and Metabolism, The Children's Hospital of Philadelphia, Philadelphia 19104, PA, USA

[12] Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

[13] Institute for Systems Biology, Seattle, Washington, USA

[14] Department of Biostatistics, Harvard School of Public Health, Cambridge, Massachusetts, USA

[15] Genomics England, London, UK

[16] Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

[17] Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA

[18] These authors jointly directed this work

[†]Current address - Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

Corresponding Author:

Katherine L. Nathanson, MD

351 BRB 2/3, 421 Curie Blvd

University of Pennsylvania

Philadelphia, PA 19104

tel: 215-573-9840 | fax: 215-573-6298

e-mail: knathans@exchange.upenn.edu

Introductory Paragraph

The international TEsticular CAncer Consortium (TECAC) combined five published genome-wide association studies of testicular germ cell tumors (TGCT; 3,558 cases and 13,970 controls) to identify novel susceptibility loci. We conducted a fixed effects meta-analysis, including the first analysis of the X chromosome. Eight new loci mapping to 2q14.2, 3q26.2, 4q35.2, 7q36.3, 10q26.13, 15q21.3, 15q22.31, and Xq28 achieved genome-wide significance ($P < 5\times10^{-8}$). Most loci harbor biologically plausible candidate genes. We refined previously reported associations at 9p24.3 and 19p12 by identifying one and three additional independent SNPs, respectively. In aggregate, the 39 independent markers identified to date explain 37% of father-to-son risk, 8% of which can be attributed to the 12 new signals reported here. Our findings substantially increase the number of known TGCT susceptibility alleles, move the field closer to a comprehensive understanding of the underlying genetic architecture of TGCT, and provide further clues into the etiology of TGCT.

In Europe and the United States, testicular germ cell tumors (TGCT) are the most common cancers in young men aged 20 to 39 years[1]. The incidence of TGCT is rising, and is highest in men of Northern European and lowest in men of African ancestry[1-3]. Risk factors for TGCT include cryptorchidism, adult height, prior diagnosis and familial history of TGCT[4-8]; its heritability ranges from 37% to 49%[9,10]. Despite the multiple lines of evidence demonstrating a considerable genetic component of TGCT risk, linkage and candidate gene approaches to find rare, highly-penetrant susceptibility genes involved in TGCT etiology were unsuccessful[11].

In contrast, genome wide association studies (GWAS) of TGCT have had remarkable success in identifying susceptibility loci with strong effects. Of the 27 replicated loci, most were discovered using GWAS chip-based microarray platforms[12-18], with 13 identified after replication on the iCOGs array[19-21] and one identified as a candidate region[22]. The genes mapping at or near identified susceptibility loci have revealed several biological themes that are highly likely to be important to TGCT development, including male germ cell maturation and differentiation, KIT-MAPK signaling, DNA damage response, and chromosomal segregation.

We imputed each of five published TGCT GWAS scans[12,18,20,23], and combined the association test statistics for a total of 8,960,654 autosomal and 249,696 chromosome X single nucleotide polymorphisms (SNPs), after excluding those with INFO score < 0.3 or minor allele frequency (MAF) < 0.01. We conducted a fixed-effects meta-analysis for 3,558 cases and 13,970 controls (**Methods and Supplementary Table 1**). The genomic control factor $\lambda = 1.037$ suggests little systematic inflation from population stratification (**Supplementary Fig. 1**). We identified eight new TGCT susceptibility loci surpassing genome-wide significance, and an additional four novel independent loci in two previously established regions (9p24.3 and 19p12) (**Table 1**). Two of these loci (rs6837349 and rs12912292) showed evidence of effect measure heterogeneity ($I^2 > 0.50$) across the five sample sets. We also determined the Bayes false discovery probability

(BFDP)[24] for these 12 loci using a prior probability of 0.0001 and odds ratio of 1.2 (**Supplementary Table 2**). Two loci, rs61408740 and rs17336718, failed to surpass a BFDP < 0.05, likely because of their low minor allele frequencies (0.023 and 0.053, respectively).

Prior reports identified 27 independent SNPs, at 25 distinct regions, and the gr/gr deletion associated with TGCT susceptibility (**Fig. 1**). In our current study, 19 of the SNPs (at 17 loci) reached a level of genome-wide significance (**Table 2**). Eight previously reported susceptibility markers were not identified in our meta-analysis likely related to limited study power, staged replication of GWAS chip-based array results in published studies, and possible residual population substructure. Considering these limitations and to further place context around our findings, we calculated the BFDP for these 27 loci using a prior probability of 0.10, which assumes 10% of the previously established loci are true positives (**Supplementary Table 2**). This threshold is more liberal than one that would be used to identify novel loci, but still may be too conservative for the re-identification of previously identified susceptibility markers. Only one locus of all 27, rs11705932, failed to surpass a BFDP < 0.05.

rs12912292 is the most statistically significant novel SNP marker in our study, (OR=1.22; $P = 1.38 \times 10^{-11}$), marking a 131 kb haploblock on 15q21.3 (**Table 1** and **Supplementary Fig. 2a**). This region contains only a single gene, *PRTG*, a member of the immunoglobulin superfamily implicated in neurogenesis[25]. *PRTG* is highly expressed in thyroid, testes and uterus (**Supplementary Fig. 3a**). No variant in this region is an eQTL in either normal testes or TGCT.

The SNP marker rs60180747 (OR=1.23; $P = 1.10 \times 10^{-10}$) marks a 261 kb haploblock on 15q22.31 that contains several genes, including *TIPIN* (TIMELESS-interacting protein) *MAP2K1* (mitogen-activated protein kinase kinase 1), *DIS3L* (DIS3 like exosome 3'-5' exoribonuclease), *SNAPC5* (small nuclear RNA activating complex, polypeptide 5), and *LCTL* (lactose-like) (**Table**

1 and **Supplementary Fig. 2b**). Several of these proteins, particularly ZWILCH and TIPIN, have high and somewhat specific expression in testes (**Supplementary Fig. 3b-g**). TIPIN coordinates the DNA replication checkpoint by interacting with Replication protein A[26]; ZWILCH is a kinetochore protein important for proper chromatid alignment during cell division[27]. A missense mutation in *ZWILCH*, p.Ser230Gly, lies within the LD block (**Supplementary Table 3**). The LD block also contains a single eQTL to *RP11-653J6.1*, which is a long non-coding (lnc) RNA highly specific to testes (**Supplemental Fig. 3h**). *RP11-653J6.1* levels and eQTLs were not measured in TGCT Cancer Genome Atlas (TCGA). Further dissection of this signal is needed to pinpoint the candidate gene.

The SNP marker rs3755605 (OR=1.19; $P = 3.87 \times 10^{-9}$) identifies a 213 kb haploblock containing three genes, *GPR160* (G protein-coupled receptor 160), *PHC3* (polyhomeotic homolog 3), and *PRKCI* (protein kinase C, iota form) on 3q26.2 (**Table 1** and **Supplementary Fig. 2c**). Several SNPs across this block are eQTLs for *GPR160* (**Supplementary Table 3, Supplementary Fig. 4a**); the risk allele is associated with increased expression in both normal testes and TGCT (**Supplementary Fig. 5a** and **Supplementary Fig. 6a**). Several SNPs in the haploblock also are eQTLs in normal testes with *RP11-469J4.3*, a lncRNA of unknown function. *RP11-469J4.3* is highly expressed in normal testes, but lies outside the haploblock (**Supplementary Fig. 3l**, **Supplementary Fig. 4b,** and **Supplementary Fig. 5b**). *RP11-469J4.3* expression was not measured in the TCGA.

The SNP marker rs2713206 (OR = 1.26; $P = 1.68 \times 10^{-8}$) lies within a smaller LD region of only 48 kb on 2q14.2. The gene *TFCP2L1* (transcription factor CP2-like 1) overlies the entirety of the haploblock (**Table 1** and **Supplementary Fig. 2d**); SNPs in the region are eQTLs (**Supplementary Table 3**) with the risk allele associated with decreased expression of *TFCP2L1* in TGCT (**Supplementary Fig. 6b**). *TFCP2L1* is not expressed in normal adult testes

(**Supplementary Fig. 3m**) but it is highly expressed in fetal gonocytes and in germ cell neoplasia in situ, the precursor of TGCT[28,29]. *TFCP2L1* is upregulated in human primordial germ cells during embryogenesis at the time of epigenetic reprogramming[30], but downregulated during transition from fetal gonocytes into spermatogonia[29]. The SNP marker rs6837349 (OR = 0.84; $P = 3.13 \times 10^{-8}$) localizes to an intron of *ZFP42* (zinc finger protein 42) on 4q35.2, and marks a small 11 kb haploblock containing no other genes (**Table 1** and **Supplementary Fig. 2e**). The region has no eQTLs in either normal testes or TGCT, although *ZFP42* is expressed exclusively in normal testes (**Supplementary Fig. 3n)**, specifically in human spermatogonia and TGCT[29,31]. Additionally, both *ZFP42* and *TFCP2L1* are involved in embryonal stem cell pluripotency[30,32].

The SNP marker rs61408740 (OR= 1.65; $P = 1.75 \times 10^{-8}$) localizes to an intron of *LHPP* (phospholysine phosphohistidine inorganic pyrophosphate phosphatase) on 10q26.13 (**Table 1** and **Supplementary Fig. 2f**). Only two SNPs were identified with pair-wise $r^2>0.4$, one in the region of the second gene, *FAM175B* (**Supplementary Table 3**); neither are eQTLs. *LHPP* encodes an inorganic diphosphatase that functions in oxidative phosphorylation[33].

The SNP marker rs11769858 (OR = 0.84; $P = 2.38 \times 10^{-8}$) identifies an 82 kb LD block on 7q36.3 that contains a large portion of *NCAPG2* (non-SMC condensin II complex subunit G2) (**Table 1** and **Supplementary Fig. 2g**). *NCAPG2* encodes a regulatory subunit of the condensin II complex, which is highly expressed in testes (**Supplementary Fig. 3q**), and plays a role in chromosome assembly and segregation during mitosis[34].

We also identified a locus marked by SNP rs17336718 (OR = 1.41; $P = 3.84 \times 10^{-8}$) on chromosome Xq28 (**Table 1** and **Supplementary Fig. 2h**) in an intron of *TKTL1* (transketolase-like 1), which is highly expressed in normal testes. The SNP is an eQTL for TKTL1 in the TGCT

TCGA data (**Supplementary Figure 6c**). TKTL1 converts sedoheptulose to ribose and glyceraldehyde to xylulose, linking the pentose phosphate pathway to the glycolytic pathway. Interestingly, although overexpression of *TKTL1* is associated with the Warburg effect and poor prognosis in several cancer types[35-38], the risk allele is associated with lower expression.

At the previously-reported TGCT susceptibility locus *DMRT1* on chromosome 9, we identified a third independent signal, rs55873183 (OR = 1.89; $P = 2.18 \times 10^{-23}$) (**Table 1, Fig. 2a, and Supplementary Table 4a**). This intronic SNP marker has an r^2 of 0.03 and 0.06 with the two previously published SNP markers, rs7040024 and rs755383[14,17], respectively; it retained genome-wide significance in conditional analysis (**Table 1, Supplementary Table 4b, Supplementary Fig. 2i**). We also identified three additional independent signals at 19p12: rs58521262, rs34601376 and rs73019876 (**Table 1, Fig. 2b, Supplementary Tables 5a and 5b, and Supplementary Figs. 2j, 2k, 2l**). We identified a SNP, rs2194275 ($P = 9.23 \times 10^{-12}$; OR = 0.76), in moderate LD ($r^2 = 0.7$) with the previously published rs2195987 ($P = 1.21 \times 10^{-9}$; OR = 0.81), which was more statistically significant in this meta-analysis (**Supplementary Table 5b**). 19p12 contains a cluster of Krüppel-associated box zinc finger genes (KRAB-ZFPs)[39]. KRAB-ZFPs are highly and differentially expressed in germ cells, and important for the epigenetic reprogramming requisite for normal germ cell development[30]. A number of different GWA studies, including one for telomere length[40], have identified significant SNPs in this region. The 19p12 LD blocks are large: rs58521262 marks a 219 kb block, and rs73019876 marks a 184 kb block, each containing over 200 relevant SNPs and several genes. The rs73019876 haploblock is extremely eQTL rich, with *ZNF729* and *ZNF676* both eQTLs in normal testes (**Supplemental Figs. 4c** and **4d**). rs58521262 is an eQTL with *ZNF728* and *CTD-2291D10.2*, which like *ZNF729* (**Supplemental Figs. 3t**), are only expressed in normal testes (**Supplemental Fig. 3v**). Associated SNPs in the LD block also include two putative missense mutations in *ZNF728* (**Supplemental Table 3**) not predicted to be deleterious. Given the

multiple independent signals in this region, further study will be required to determine which, if any, are causally involved in TGCT.

Similar to prior reports, several of the loci identified in the current study contain biologically plausible genes implicating pathways involved in male germ cell development and pluripotency (*TFCP2L1, ZFP42*), kinetochore function (*ZWILCH*), DNA damage response (*TIPIN*), and metabolic mitochondrial function (*TKTL1* and *LHPP*). We identified eight such loci in novel genomic regions and four in previously identified regions. These additional loci bring the cumulative total of independent susceptibility alleles for TGCT to 40. Interestingly, racial differences in risk allele frequencies that parallel population-specific TGCT risk also continue to be apparent. Of the 40 identified susceptibility loci, the allele frequencies of all but one differ significantly across continents based on analysis of data from the AFR, AMR, ASN and EUR populations available in the 1000 Genomes project[41], with most comparisons having a P-value surpassing strict Bonferroni correction ($P < 0.00125$) (**Supplementary Table 6**). The 12 newly identified susceptibility alleles account for 5.3% of the genetic risk to the brothers and 8.0% of risk to the sons of TGCT patients, increasing estimates of heritability to 25% and 37% of the risk to brothers and sons, respectively. The newly identified TGCT susceptibility markers continue to demonstrate moderate effects with ORs that range from 1.17 to 1.89. In comparison with other cancer types, we have accounted for a high proportion of site-specific heritability with fewer loci[42-44].

Acknowledgements

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 Ms. Benita Weathers for her coordination of TECAC, Mr. John Pluta for biostatistical assistance, and Mr. Kurt D'Andrea for expert assistance with SNP genotyping. We thank Drs. Douglas R. Stewart and Jennifer T. Loud for critical support of the NCI Clinical Genetics Branch Familial Testicular Cancer Project (NCI 02-C-0178; NCT-00039598). The authors also thank all previous contributors to the GWAS analyzed in this study, including Drs. Nils Weinhold, Daniel Edsgärd, Henrik Leffers, and Profs. Anders Juul, Niels E. Skakkebæk, and Søren Brunak from the Danish study.

The Testicular Cancer Consortium is supported by the National Institutes of Health grant U01CA164947 to KLN, PAK and SMS. A portion of this work was supported by the Intramural Research Program of the National Cancer Institute and by a support services contract HHSN26120130003C with IMS, Inc. The Penn GWAS (Penn) was supported by the Abramson Cancer Center at the University of Pennsylvania and National Institute of Health grant CA114478 to KLN and PAK. The UK testicular cancer study was supported by the Institute of Cancer Research, Cancer Research UK and made use of control data generated by the Wellcome Trust Case Control Consortium (WTCCC) 2. CT is supported by the Movember foundation. KL is supported by a PhD fellowship from Cancer Research UK. LCP is supported by T32-GM008638. The contribution from the University of Leeds was funded by Cancer Research UK. Norwegian/Swedish study was supported by the Norwegian Cancer Society (grants number 418975 – 71081 – PR-2006-0387 and PK01-2007-0375); the Nordic Cancer Union (grant number S-12/07) and the Swedish Cancer Society (grant numbers 2008/708, 2010/808, 2011/484, and CAN2012/823). The Danish GWAS was supported by Villum Kann

Rasmussen Foundation, a NABIIT grant from the Danish Strategic Research Council, the Novo Nordisk Foundation, the Danish Cancer Society, and the Danish Childhood Cancer Foundation.

Testicular Cancer Consortium:

Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA
Katherine McGlynn, Mark Greene, Stephen Chanock

Department of Growth and Reproduction, Copenhagen University Hospital (Rigshospitalet), Copenhagen, Denmark
Ewa Rajpert-De Meyts

Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK
D. Timothy Bishop, Jeremie Nsengimana

Center of Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kongens Lyngby, Denmark
Ramneek Gupta

Cancer Registry of Norway, Oslo, Norway
Tom Grotmol

Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Oslo, Norway
Trine B. Haugen

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
Fredrik Wiklund, Robert Karlsson

Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK
Clare Turnbull

Department of Medicine, Division of Translational Medicine and Human Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA
Katherine L. Nathanson

Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
Stephen M. Schwartz

Genomics England, London, UK
Clare Turnbull

Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA
Peter A. Kanetsky

Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA
Katherine L. Nathanson

Department of Medicine, Unit of Andrology and Reproductive Medicine, University of Padova, Padova, Italy
Alberto Ferlin

Medical Oncology, University Medical Center Groningen, Groningen, The Netherlands
Jourik A. Gietema

Department of Clinical Preventive Medicine, Krek School of Medicine at University of Southern California, Los Angeles, CA
Victoria Cortessis

Department of Environmental Health, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA
Russ Hauser

Department of Epidemiology, Division of OVP, Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX
Michelle Hildebrandt

Radboud University Medical Centre, Radhoud Institute for Health Sciences, Nijmegen, Nijmegen, The Netherlands
Lambertus A. Kiemeney

Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
Davor Lessel, Christian Kubisch

deCode Genetics, Reykjavik, Iceland
Thorunn Rafnar

Department of Biostatistics and Epidemiology, University of Turin, Turin, Italy
Lorenzo Richiardi

Department of Molecular Oncology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway
Rolf Skotheim

Department of Epidemiology, Brown University, Providence, RI
Tongzhang Zheng

Conflict of Interest

There are no conflicts of interest.

Author Contributions

K.L.N. and P.A.K. supervised the overall study. K.A.M., E.R-D.M, D.T.B., M.D.D., M.H.G., R.G., T.G., T.B.H., K.L., K.N., S.M.S., F.W., C.T., P.A.K, and K.L.N. contributed to recruitment, study and data management. Z.W., K.A.M., E.R-D.M, D.T.B., M.D.D., M.H.G., R.G., T.G., T.B.H., R.K., K.L., N.M., K.N., S.V., F.W., C.T., S.J.C., P.A.K, and K.L.N. contributed to genotyping or association analysis of individual studies. Z.W., C.C.C., L.C.P., V.T., S.J.C, P.A.K. and K.L.N. carried out the meta-analysis and the additional bioinformatics analyses, including using GTeX and TCGA TGCT data. Z.W., P.A.K. and K.L.N. drafted the initial manuscript, and all authors reviewed and contributed to the manuscript.

Data Availability

Individual level data from the UK GWAS data has been deposited into European Genome-phenome Archive (EGA) accession number EGAS00001001302. Individual level data has been deposited into dbGAP from University of Pennsylvania (phs001307.v1.p1) and NCI (phs001303.v1.p1). The summary data from the TECAC meta-analysis also has been deposited into dbGAP (phs001349.v1.p1).

References

1. Trabert, B., Chen, J., Devesa, S.S., Bray, F. & McGlynn, K.A. International patterns and trends in testicular cancer incidence, overall and by histologic subtype, 1973-2007. *Andrology* **3**, 4-12 (2015).
2. Howlader, N. *et al.* SEER Cancer Statistics Review, 1975-2012. (National Cancer Institute, 2015).
3. Znaor, A., Lortet-Tieulent, J., Jemal, A. & Bray, F. International variations and trends in testicular cancer incidence and mortality. *Eur Urol* **65**, 1095-106 (2014).
4. 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**, 1316-24 (2004).
5. 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**, 952-7 (2009).
6. Heimdal, K. *et al.* Risk of cancer in relatives of testicular cancer patients. *Br J Cancer* **73**, 970-3 (1996).
7. Sonneveld, D.J. *et al.* Familial testicular cancer in a single-centre population. *Eur J Cancer* **35**, 1368-73 (1999).
8. McGlynn, K.A. & Trabert, B. Adolescent and adult risk factors for testicular cancer. *Nat Rev Urol* **9**, 339-49 (2012).
9. Litchfield, K. *et al.* Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches. *Sci Rep* **5**, 13889 (2015).
10. Mucci, L.A. *et al.* Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* **315**, 68-76 (2016).
11. Crockford, G.P. *et al.* Genome-wide linkage screen for testicular germ cell tumour susceptibility loci. *Hum Mol Genet* **15**, 443-51 (2006).
12. Chung, C.C. *et al.* Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat Genet* **45**, 680-5 (2013).

13. Kanetsky, P.A. *et al.* Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. *Nat Genet* **41**, 811-5 (2009).

14. Kanetsky, P.A. *et al.* A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. *Hum Mol Genet* **20**, 3109-17 (2011).

15. Schumacher, F.R. *et al.* Testicular germ cell tumor susceptibility associated with the UCK2 locus on chromosome 1q23. *Hum Mol Genet* **22**, 2748-53 (2013).

16. Rapley, E.A. *et al.* A genome-wide association study of testicular germ cell tumor. *Nat Genet* **41**, 807-10 (2009).

17. Turnbull, C. *et al.* Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat Genet* **42**, 604-7 (2010).

18. Kristiansen, W. *et al.* Two new loci and gene sets related to sex determination and cancer progression are associated with susceptibility to testicular germ cell tumor. *Hum Mol Genet* **24**, 4138-46 (2015).

19. Litchfield, K. *et al.* Identification of four new susceptibility loci for testicular germ cell tumour. *Nat Commun* **6**, 8690 (2015).

20. Ruark, E. *et al.* Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nat Genet* **45**, 686-9 (2013).

21. Litchfield, K. *et al.* Multi-stage genome-wide association study identifies new susceptibility locus for testicular germ cell tumour on chromosome 3q25. *Hum Mol Genet* **24**, 1169-76 (2015).

22. Nathanson, K.L. *et al.* The Y deletion gr/gr and susceptibility to testicular germ cell tumor. *Am J Hum Genet* **77**, 1034-43 (2005).

23. 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).

411 24. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic
412 epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
413 25. Wong, Y.H. *et al.* Protogenin defines a transition stage during embryonic neurogenesis
414 and prevents precocious neuronal differentiation. *J Neurosci* **30**, 4428-39 (2010).
415 26. Unsal-Kacmaz, K. *et al.* The human Tim/Tipin complex coordinates an Intra-S
416 checkpoint response to UV that slows replication fork displacement. *Mol Cell Biol* **27**, 3131-42
417 (2007).
418 27. Williams, B.C. *et al.* Zwilch, a new component of the ZW10/ROD complex required for
419 kinetochore functions. *Mol Biol Cell* **14**, 1379-91 (2003).
420 28. Almstrup, K. *et al.* Embryonic stem cell-like features of testicular carcinoma in situ
421 revealed by genome-wide gene expression profiling. *Cancer Res* **64**, 4736-43 (2004).
422 29. Sonne, S.B. *et al.* Analysis of gene expression profiles of microdissected cell populations
423 indicates that testicular carcinoma in situ is an arrested gonocyte. *Cancer Res* **69**, 5241-50
424 (2009).
425 30. Tang, W.W. *et al.* A Unique Gene Regulatory Network Resets the Human Germline
426 Epigenome for Development. *Cell* **161**, 1453-67 (2015).
427 31. Kristensen, D.M. *et al.* Presumed pluripotency markers UTF-1 and REX-1 are expressed
428 in human adult testes and germ cell neoplasms. *Hum Reprod* **23**, 775-82 (2008).
429 32. Scotland, K.B., Chen, S., Sylvester, R. & Gudas, L.J. Analysis of Rex1 (zfp42) function
430 in embryonic stem cell differentiation. *Dev Dyn* **238**, 1863-77 (2009).
431 33. Yokoi, F., Hiraishi, H. & Izuhara, K. Molecular cloning of a cDNA for the human
432 phospholysine phosphohistidine inorganic pyrophosphate phosphatase. *J Biochem* **133**, 607-14
433 (2003).
434 34. Kim, J.H. *et al.* The condensin component NCAPG2 regulates microtubule-kinetochore
435 attachment through recruitment of Polo-like kinase 1 to kinetochores. *Nat Commun* **5**, 4588
436 (2014).

35. Schwaab, J. *et al.* Expression of Transketolase like gene 1 (TKTL1) predicts disease-free survival in patients with locally advanced rectal cancer receiving neoadjuvant chemoradiotherapy. *BMC Cancer* **11**, 363 (2011).

36. Ahopelto, K., Bockelman, C., Hagstrom, J., Koskensalo, S. & Haglund, C. Transketolase-like protein 1 expression predicts poor prognosis in colorectal cancer. *Cancer Biol Ther*, 1-6 (2015).

37. Jayachandran, A. *et al.* Transketolase-like 1 ectopic expression is associated with DNA hypomethylation and induces the Warburg effect in melanoma cells. *BMC Cancer* **16**, 134 (2016).

38. Kayser, G. *et al.* Poor outcome in primary non-small cell lung cancers is predicted by transketolase TKTL1 expression. *Pathology* **43**, 719-24 (2011).

39. Eichler, E.E. *et al.* Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome Res* **8**, 791-808 (1998).

40. Mangino, M. *et al.* Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. *Hum Mol Genet* **21**, 5385-94 (2012).

41. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

42. Jiao, S. *et al.* Estimating the heritability of colorectal cancer. *Hum Mol Genet* **23**, 3898-905 (2014).

43. Mancuso, N. *et al.* The contribution of rare variation to prostate cancer heritability. *Nat Genet* **48**, 30-5 (2016).

44. Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* **47**, 373-80 (2015).

ONLINE METHODS

Studies

Detailed characteristics and genotype quality control metrics of the study populations (Denmark, NCI [STEED, FTCS], Norway/Sweden, Penn, UK) have been previously described[12,15,18,19,23]. 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.

Genotype imputation

Genotype imputation was conducted by each center following a similar protocol. SNPs with a call rate < 95% or Hardy-Weinberg proportion test *P*-value < 0.000001 or MAF < 1% were removed prior to imputation. Imputation was conducted using IMPUTE2 software version 2.2.2 and version 3 of the 1,000 Genomes Project Phase 1 data as the reference set. First, the genomic coordinates were lifted over from NCBI human genome build 36 to build 37 using the UCSC lift over tool. Second, the strand of the inference data was aligned with the 1,000 Genomes data informed by allele state comparison or allele frequency matching for A/T and G/C SNPs. A pre-phasing strategy with SHAPEIT software version 1 was adopted to improve the imputation performance. The phased haplotypes from SHAPEIT were imported directly into the IMPUTE2 program. We applied sliding windows of 4Mb with 250kb as an overlapping buffer and generated 744 segments for imputing autosomes. For Chromosome X, the pseudoautosomal region 1 (PAR1), PAR2, and the remaining region, which was split into 37 sections, were imputed separately. We excluded imputed loci with INFO score < 0.3 or MAF < 0.01 from further association analysis. Further, we acknowledge the limitations of imputation, including that the accuracy of imputation depends on the linkage disequilibrium between markers in the reference panel and markers to be imputed, and that the quality of imputation across scans differs because of imperfect population matching of data sets to the reference panel.

488

489 ***Statistical analysis***

490 Within each data set, a test for trend was performed for each SNP using SNPTEST software
491 version 2.2 or 2.5. Fixed-effects meta-analysis was used to combine individual within-study
492 association estimates from five imputed GWAS scans. Genetic effect heterogeneity across
493 studies was assessed by using I^2 and P-value calculated from the Cochran's Q statistic. To
494 refine the association signals of each risk region, we first performed LD pruning using pairwise
495 $R^2 > 0.3$ and then conducted the conditional association analyses to estimate the independent
496 effect of each SNP by simultaneously including all specified SNPs from the same region and
497 with their unconditional *P*-values $< 5 \times 10^{-8}$ into the same logistic regression model.

498

499 ***Heritability analysis***

500 To evaluate the familial risk explained by the new loci identified in our study, we estimated the
501 contribution of each SNP based on the formula $h^2_{SNP} = \beta^2 \times 2f(1-f)$, where β is the log per allele
502 odds ratio and f is the risk allele frequency[45]. We calculated the proportion of familial risk
503 explained by dividing the summed contribution of all h^2_{SNP} by the total heritability, which was
504 derived from the log relative risk (RR), where RR = 4 for affected father and RR = 8 for affected
505 brothers[46].

506

507 ***In silico bioinformatics analysis***

508 We used HaploReg v4.1 and RegulomeDB v1.1 to explore potential non-coding functional
509 annotation within the ENCODE database in the genomic region surrounding our SNPs of
510 interest, with particular attention to annotations in induced pluripotent stem cells (IPSC) and
511 embryonic stem cells (ESC), as we considered these tissue types as best proxies for TGCT.
512 Specifically, we interrogated the linkage disequilibrium (LD) block of neighboring SNPs in a
513 haploblock defined as pair-wise $r^2 > 0.4$ with the index SNP (**Supplementary Table 4**). We also

searched the GTEx v6 database to determine whether the haploblock SNPs were implicated as eQTLs in their sample of 157 normal adult testis tissues with available genotype. Of note, the normal testis contains an abundance of stromal cells (i.e., Sertoli and Leydig cells), so may not be an exact surrogate for germ cells, and in particular the primordial germ cells from which TGCT is believed to develop. Finally, we assessed our 12 novel susceptibility loci for eQTLs among the 128 cases of TGCT with linked genotype data available in The Cancer Genome Atlas.

For the correlation between genotype and expression data from the TGCT TCGA data set, the genotype data was downloaded from the NCI's Genomic Data Commons (https://gdc.nci.nih.gov/). Data was converted to PLINK v1.07[47] format. Subjects were screened for discordant sex, insufficient genotype call rate (>0.05), and excessive heterozygosity (> +/- 3 SD from the mean). SNPs were screened for MAF (>0.01), Hardy-Weinberg equilibrium violations ($p<0.0001$), and missingness (>0.01). All quality control steps were performed in PLINK. A total of five subjects were removed (all for heterozygosity violations), leaving 145 valid for analysis. 1000 Genomes Phase 3[41] was used as the reference set. Alignment to the reference set and haplotype estimation was performed using Shapeit v2[48], and additional SNPs were imputed using Impute2[49]. Imputed SNPs with an info score <0.4 were discarded. For the 12 SNPs of interest, the risk allele was calculated as the allele with increased odds of TGCT (OR>1). For each subject, the zygosity with respect to the risk allele was calculated, and genotypes were tabulated.

All available TCGA TGCT data were retreived from the TCGA Data Coordinating Center and processed through the TCGA pipeline at the TCGA Genome Data Analysis Center at the Institute for Systems Biology. Gene expression matrices were generated for 133 primary tumor samples using available (TCGA Level 3) gene expression values from RNA sequencing,

expressed as RSEM values[50]. Imputed genotypes for all novel SNPs reported in this paper were related to gene expression, for the 128 cases with both genotype and gene expression levels available. Assocations were tested using a linear regression model (using the *lm* function in R).

Technical validation of imputed SNPs

To technically validate our imputation findings, we optimized TaqMan assays (Applied Biosystems) for 12 loci based on the standard pipeline at the Cancer Genomics Research Laboratory at National Cancer Institute (**Supplementary Table 7)**. For six loci that failed initial TaqMan assay design, LD surrogate SNPs were used. We randomly selected about 1000 samples previously scanned in one of three GWAS (~300 each from NCI, Penn and Norway/Sweden) for TaqMan genotyping. For the imputed probabilistic genotypes, a threshold of 0.80 was applied to derive the discrete genotypes. The average concordance rates are 0.98, 0.97 and 0.93 for NCI, Penn and Norway/Sweden respectively (**Supplementary Table 7**)**.**

References ON LINE METHODS

45. Park, J.H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* **42**, 570-5 (2010).
46. Hemminki, K. & Li, X. Familial risk in testicular cancer as a clue to a heritable and environmental etiology. *Br J Cancer* **90**, 1765-1770 (2004).
47. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
48. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).

563 49. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation
564 method for the next generation of genome-wide association studies. *PLoS Genet* **5**,
565 e1000529 (2009).
566 50. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with
567 or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

568

URLs

GLU module, http://code.google.com/p/glu-genetics/;

GTEx portal, http://www.gtexportal.org/home/;

HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php;

IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html;

RegulomeDB, http://regulome.stanford.edu/;

SHAPEIT, http://www.shapeit.fr/;

SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html;

The Human Protein Atlas, http://www.proteinatlas.org/;

UCSC lift over tool, http://hgdownload.cse.ucsc.edu/downloads.html

Table 1. TGCT meta-analysis association results for novel loci and new independent SNPs in established loci

Cytoband	Gene Neighborhood	SNP	Position	Study	Info	Controls	Cases	Reference allele	Effect allele	Effect allele frequency		OR	CI	P	P_{het}	I^2
										Control	Case					
2q14.2	*TFCP2L1*	rs2713206	122007941	NCI	0.92	1055	581	C	T	0.15	0.20	1.45	(1.18-1.79)	4.19E-04		
				UK	0.98	4945	985	C	T	0.15	0.17	1.15	(1.00-1.31)	4.43E-02		
				PENN	0.90	918	480	C	T	0.16	0.20	1.32	(1.06-1.64)	1.25E-02		
				Norway/Sweden	0.93	6687	1326	C	T	0.14	0.17	1.25	(1.09-1.44)	1.74E-03		
				Denmark	0.85	363	183	C	T	0.17	0.21	1.41	(0.98-2.03)	6.39E-02		
				Combined		13968	3555					1.26	(1.16-1.36)	1.68E-08	0.39	3.6
3q26.2	*GPR160*	rs3755605	169756119	NCI	0.97	1055	581	C	T	0.41	0.44	1.14	(0.98-1.33)	9.31E-02		
				UK	0.98	4946	985	C	T	0.39	0.43	1.21	(1.09-1.33)	2.25E-04		
				PENN	0.97	918	480	C	T	0.41	0.43	1.08	(0.92-1.27)	3.54E-01		
				Norway/Sweden	0.98	6687	1326	C	T	0.40	0.44	1.24	(1.12-1.37)	2.08E-05		
				Denmark	0.99	363	183	C	T	0.39	0.43	1.19	(0.91-1.55)	1.94E-01		
				Combined		13969	3555					1.19	(1.12-1.26)	3.87E-09	0.67	0.0
4q35.2	*ZFP42*	rs6837349	188921355	NCI	0.99	1055	581	G	T	0.66	0.63	0.84	(0.72-0.98)	2.71E-02		
				UK	0.99	4945	985	G	T	0.66	0.62	0.83	(0.75-0.92)	3.98E-04		
				PENN	0.69	918	480	G	T	0.66	0.68	1.15	(0.94-1.40)	1.80E-01		
				Norway/Sweden	0.99	6687	1326	G	T	0.68	0.63	0.77	(0.70-0.86)	1.18E-06		
				Denmark	0.74	363	183	G	T	0.66	0.65	0.96	(0.70-1.32)	8.00E-01		
				Combined		13968	3555					0.84	(0.79-0.89)	3.13E-08	0.02	67.4
7q36.3	*NCAPG2*	rs11769858	158501492	NCI	0.93	1055	581	T	C	0.67	0.65	0.89	(0.76-1.06)	1.86E-01		
				UK	0.95	4945	985	T	C	0.69	0.65	0.84	(0.76-0.94)	1.51E-03		
				PENN	0.88	918	480	T	C	0.64	0.60	0.81	(0.68-0.96)	1.69E-02		
				Norway/Sweden	0.91	6687	1326	T	C	0.70	0.66	0.82	(0.73-0.91)	3.89E-04		
				Denmark	0.90	363	183	T	C	0.68	0.64	0.82	(0.62-1.08)	1.59E-01		
				Combined		13968	3555					0.84	(0.79-0.89)	2.38E-08	0.92	0.0
9p24.3*	*DMRT1*	rs55873183	878563	NCI	0.73	1055	581	A	G	0.06	0.07	1.34	(0.93-1.93)	1.14E-01		
				UK	0.82	4945	985	A	G	0.06	0.09	1.90	(1.52-2.38)	1.54E-08		
				PENN	0.73	918	480	A	G	0.06	0.09	1.97	(1.36-2.84)	3.05E-04		
				Norway/Sweden	0.80	6687	1326	A	G	0.07	0.11	2.08	(1.70-2.54)	1.21E-12		
				Denmark	0.81	363	183	A	G	0.07	0.11	1.84	(1.11-3.03)	1.75E-02		
				Combined		13968	3555					1.89	(1.67-2.14)	2.18E-23	0.36	8.1
10q26.13	*LHPP*	rs61408740	126274612	NCI	0.99	1056	582	C	G	0.02	0.04	1.68	(1.09-2.60)	1.89E-02		
				UK	0.95	4945	985	C	G	0.03	0.04	1.64	(1.22-2.20)	1.05E-03		
				PENN	0.94	919	480	C	G	0.03	0.04	1.92	(1.22-3.03)	4.92E-03		
				Norway/Sweden	1.00	6687	1326	C	G	0.02	0.03	1.53	(1.12-2.09)	7.79E-03		
				Denmark	0.96	363	183	C	G	0.02	0.03	1.61	(0.69-3.76)	2.75E-01		
				Combined		13970	3556					1.65	(1.38-1.96)	1.75E-08	0.95	0.0
15q21.3	*PRTG*	rs12912292	56038707	NCI	0.95	1055	581	G	A	0.51	0.55	1.25	(1.07-1.46)	4.76E-03		
				UK	0.99	4946	985	G	A	0.53	0.55	1.09	(0.99-1.20)	9.18E-02		
				PENN	0.95	918	480	G	A	0.47	0.56	1.44	(1.23-1.70)	8.74E-06		
				Norway/Sweden	0.95	6687	1326	G	A	0.52	0.58	1.26	(1.14-1.39)	3.42E-06		
				Denmark	0.95	363	183	G	A	0.51	0.57	1.30	(1.00-1.69)	4.81E-02		
				Combined		13969	3555					1.22	(1.15-1.29)	1.38E-11	0.03	61.4
15q22.31	*MAP2K1*, *TIPIN*	rs60180747	66663261	NCI	0.99	1055	582	A	C	0.26	0.30	1.27	(1.07-1.50)	5.59E-03		
				UK	0.99	4946	986	A	C	0.26	0.30	1.23	(1.10-1.37)	2.34E-04		
				PENN	0.99	919	481	A	C	0.27	0.34	1.37	(1.15-1.63)	4.53E-04		
				Norway/Sweden	1.00	6687	1326	A	C	0.26	0.28	1.18	(1.05-1.31)	3.89E-03		
				Denmark	1.00	363	183	A	C	0.27	0.31	1.22	(0.93-1.60)	1.59E-01		
				Combined		13970	3558					1.23	(1.16-1.32)	1.10E-10	0.70	0.0
19p12*	ZNF728	rs58521262	23205184	NCI	0.99	1056	581	G	A	0.151	0.106	0.69	(0.55-0.85)	6.64E-04		
				UK	1.00	4946	985	G	A	0.140	0.106	0.74	(0.64-0.86)	4.83E-05		
				PENN	0.98	919	481	G	A	0.154	0.123	0.78	(0.63-0.98)	2.95E-02		
				Norway/Sweden	0.99	6687	1326	G	A	0.173	0.136	0.74	(0.65-0.84)	4.90E-06		

				Denmark	0.99	363	183	G	A	0.183	0.137	0.71	(0.51-1.00)	5.10E-02		
				Combined		13971	3556					0.74	(0.68-0.80)	4.87E-14	0.95	0.0
19p12*	ZNF726	rs34601376	24050828	NCI	0.84	1055	581	A	T	0.216	0.220	1.04	(0.85-1.28)	6.77E-01		
				UK	0.92	4945	985	A	T	0.202	0.242	1.29	(1.14-1.46)	5.08E-05		
				PENN	0.83	918	480	A	T	0.191	0.229	1.32	(1.07-1.64)	1.02E-02		
				Norway/Sweden	0.88	6687	1326	A	T	0.195	0.238	1.39	(1.22-1.58)	4.19E-07		
				Denmark	0.69	363	183	A	T	0.146	0.168	1.32	(0.85-2.05)	2.23E-01		
				Combined		13968	3555					1.29	(1.20-1.39)	2.40E-11	0.23	29.2
19p12*	*ZNF257*	rs73019876	22267849	NCI	0.93	1055	581	T	G	0.45	0.42	0.89	(0.76-1.04)	1.35E-01		
				UK	0.96	4946	985	T	G	0.45	0.40	0.83	(0.75-0.91)	1.51E-04		
				PENN	0.95	918	480	T	G	0.51	0.49	0.91	(0.78-1.07)	2.59E-01		
				Norway/Sweden	0.95	6687	1326	T	G	0.43	0.41	0.85	(0.77-0.94)	1.22E-03		
				Denmark	0.94	363	183	T	G	0.44	0.36	0.72	(0.55-0.93)	1.25E-02		
				Combined		13969	3555					0.85	(0.80-0.90)	2.04E-08	0.54	0.0
Xq28	*TKTL1*	rs17336718	153536119	NCI	0.97	1056	582	C	T	0.05	0.09	1.33	(1.07-1.64)	8.85E-03		
				UK	0.63	4945	986	C	T	0.05	0.07	1.46	(1.19-1.80)	3.71E-04		
				PENN	0.78	918	480	C	T	0.05	0.09	1.59	(1.23-2.06)	4.06E-04		
				Denmark	0.84	363	183	C	T	0.06	0.08	1.15	(0.78-1.69)	4.71E-01		
				Combined		7282	2231					1.41	(1.25-1.59)	3.84E-08	0.51	0.0

*New independent SNPs in established loci

585 **Table 2. TGCT meta-analysis association results for previously published susceptibility loci**

Cytoband	Gene Neighborhood	SNP	Position	OR	CI	P	P_{het}	I^2
1q22	*KIAA0446 SLC25A44*	rs2072499	156169610	1.20	(1.13-1.27)	1.63E-09	0.78	0.0
1q24.1	*UCK2*	rs3790672	165873392	1.27	(1.20-1.35)	2.15E-14	0.84	0.0
3p24.3	*DAZL*	rs10510452	16625048	0.82	(0.77-0.87)	3.36E-10	0.76	0.0
3q23*	*TFDP2 DKFZp434G222*	rs11705932	141818850	0.88	(0.82-0.94)	3.51E-04	0.90	0.0
3q25.31		rs1510272	156300724	0.83	(0.78-0.88)	7.15E-09	0.45	0.0
4q22.3*	*SMARCAD1 HPGDS*	rs17021463	95224812	0.87	(0.82-0.92)	1.36E-06	0.06	56.5
4q24	*CENPE CENPE variant protein*	rs2720460	104054686	0.78	(0.74-0.83)	9.88E-17	0.92	0.0
5p15.33	*TERT hTERT*	rs2736100	1286516	1.29	(1.22-1.37)	7.69E-20	0.41	0.0
5p15.33	*CLPTM1L*	rs4635969	1308552	1.46	(1.37-1.57)	2.83E-27	0.15	41.0
5q31.1*	*CATSPER3 PITX1 AK026965*	rs3805663	134366200	0.88	(0.83-0.93)	9.16E-06	0.36	8.2
5q31.3	*SPRY4*	rs4624820	141681788	1.51	(1.42-1.59)	2.59E-46	0.51	0.0
6p21.31	*BAK1 AY383626 C6orf227*	rs210138	33542538	1.55	(1.44-1.66)	2.51E-34	0.66	0.0
7p22.3	*MAD1L1*	rs12699477	1968953	1.21	(1.14-1.28)	2.24E-10	0.13	43.7
8q13.3*	*PRDM14*	rs7010162	70976505	0.86	(0.81-0.91)	1.42E-07	0.33	13.5
9p24.3	*DMRT1*	rs7040024**	845516	0.67	(0.62-0.71)	1.21E-32	0.04	59.3
9p24.3	*DMRT1*	rs755383**	863635	1.49	(1.41-1.58)	6.52E-41	0.52	0.0
11q14.1*	*GAB2*	rs7107174	77997936	1.19	(1.10-1.29)	6.35E-06	0.36	8.7
12p13.1	*ATF7IP PLBD1*	rs2900333	14653867	0.85	(0.80-0.90)	2.71E-08	0.20	33.6
12q21.32	*KITLG*	rs3782181	88953561	2.02	(1.88-2.18)	1.32E-76	0.90	0.0
16p13.13*	*BCAR4 CATX-11 RSL1D1*	rs4561483	11920037	0.86	(0.81-0.91)	4.19E-07	0.45	0.0
16q12.1	*HEATR3 AF086132*	rs8046148	50142944	1.24	(1.15-1.33)	3.15E-09	0.21	32.2
16q23.1	*RFWD3*	rs4888262	74670458	0.83	(0.78-0.88)	5.65E-11	0.08	52.7
16q24.2*	*ZFPM1*	rs55637647	88549264	1.18	(1.11-1.26)	1.33E-07	0.40	1.4
17q12	*HNF1B*	rs7501939	36101156	1.26	(1.19-1.34)	1.27E-14	0.42	0.0
17q22	*TEX14*	rs9905704	56632543	1.27	(1.19-1.35)	1.99E-14	0.68	0.0
19p12	*AK125686*	rs2195987	24149545	1.23	(1.15-1.32)	1.21E-09	0.89	0.0
21q22.3*	*MCM3APAS MCM3AP*	rs2839186	47690068	1.13	(1.07-1.20)	2.00E-05	0.02	67.1

* Indicates sub genome-wide statistical significance.

** Pairwise r^2=0.38.

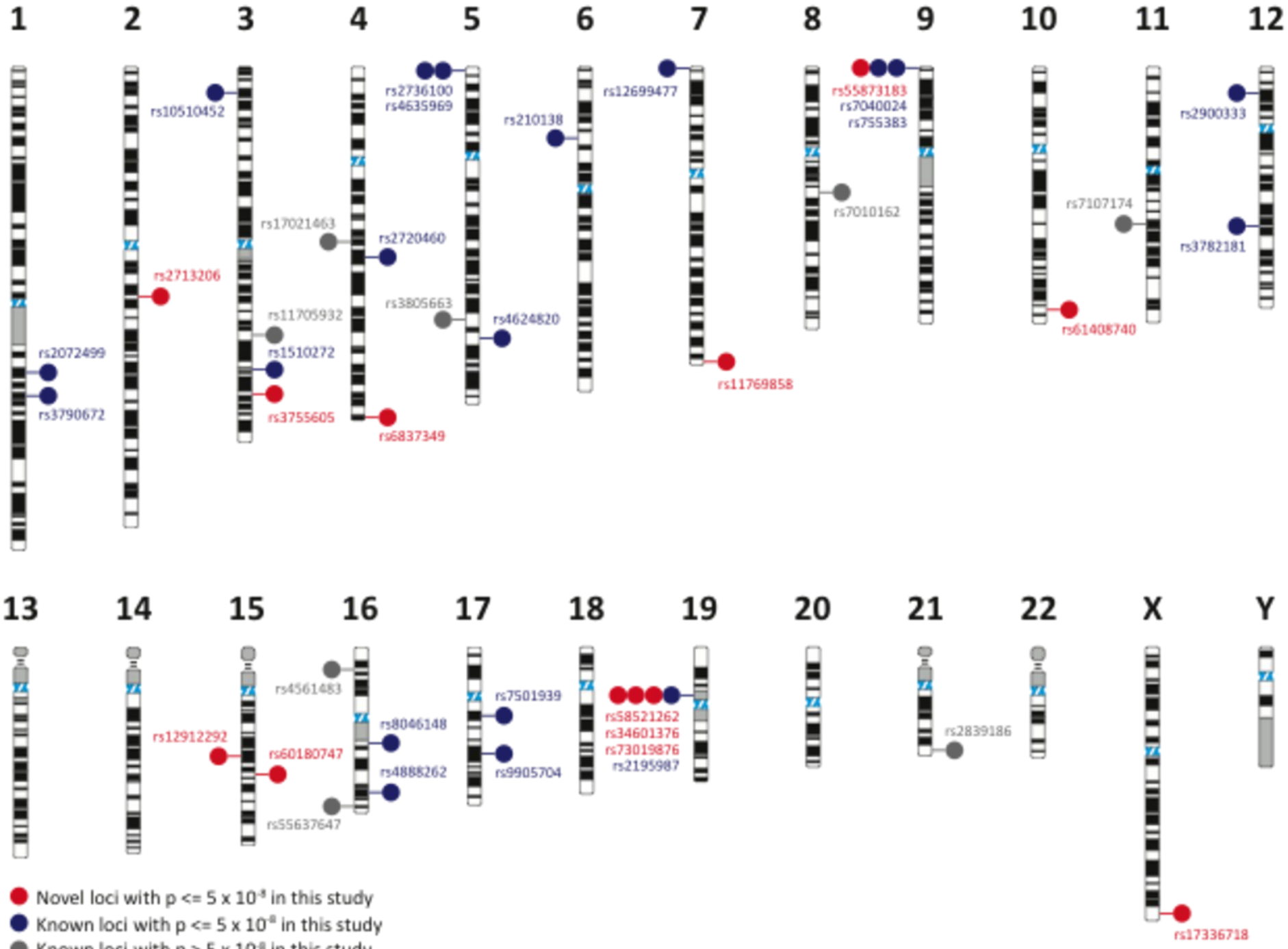
Figure Legends

Figure 1. All identified SNP markers associated with TGCT susceptibility to date

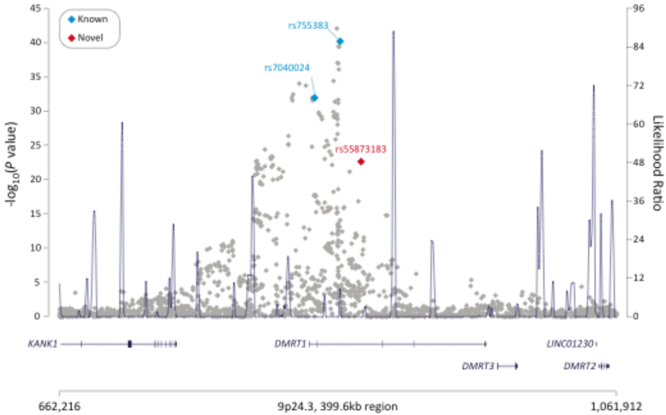
In the ideogram, red dots and red rs number annotation indicate SNPs identified and described in the current study ($P \leq 1 \times 10^{-8}$); blue dots and blue rs number annotation represent previously identified SNP markers achieving genome wide significance ($P \leq 1 \times 10^{-8}$) in the current study; and gray dots and gray rs number annotation are previously identified SNPs that fail to achieve genome wide significance in this study ($P > 1 \times 10^{-8}$).

Figure 2. Genetic association between SNP markers and TGCT risk for regions with multiple independent signals

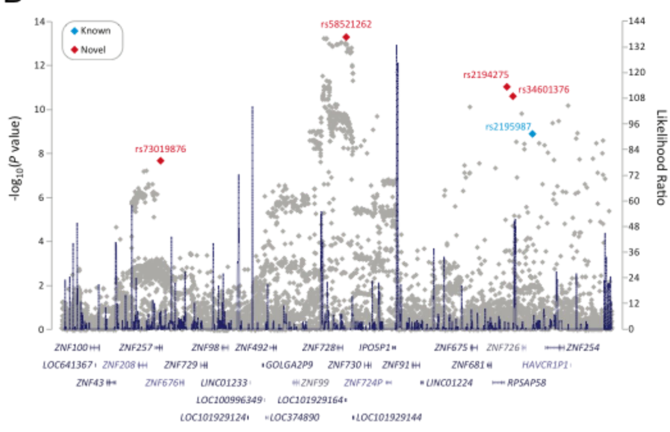
The strength of the association signals ($-\log_{10} P$-values) for individual SNPs at (a) 9p24.3 and (b) 19p12-11 are plotted on the Y-axis relative to their genomic locations (GRCh37) along the X-axis. Red diamonds are the newly identified independent SNPs, blue diamonds are previously reported SNP markers, and all other SNPs are colored gray. The line graph shows likelihood ratio statistics (right Y-axis) for recombination hotspots calculated with SequenceLDhot software using 1000 Genomes Project CEU population data. Gene annotation along the X-axis is based on NCBI RefSeq genes from the UCSC Genome Browser.


1
2
3
4
5
6
7
8
9
10
11
12
rs10510452
rs2713206
rs2072499
rs3790672
rs17021463
rs11705932
rs1510272
rs3755605
rs2736100
rs4635969
rs2720460
rs3805663
rs6837349
rs4624820
rs210138
rs12699477
rs11769858
rs55873183
rs7040024
rs755383
rs7010162
rs7107174
rs61408740
rs2900333
rs3782181
13
14
15
16
17
18
19
20
21
22
X
Y
rs12912292
rs60180747
rs55637647
rs4561483
rs8046148
rs4888262
rs7501939
rs9905704
rs58521262
rs34601376
rs73019876
rs2195987
rs2839186
rs17336718
Novel loci with p <= 5 x 10-8 in this study
Known loci with p <= 5 x 10-8 in this study
Known loci with p > 5 x 10-8 in this study

A


Known
Novel
rs755383
rs7040024
rs55873183
-log10(P value)
Likelihood Ratio
KANK1
DMRT1
LINC01230
DMRT3
DMRT2
662,216
9p24.3, 399.6kb region
1,061,912

B


Known
Novel
rs58521262
rs2194275
rs34601376
rs2195987
rs73019876
-log10(P value)
Likelihood Ratio
ZNF100
ZNF257
ZNF98
ZNF492
ZNF728
IPOSP1
ZNF675
ZNF726
ZNF254
LOC641367
ZNF208
ZNF729
GOLGA2P9
ZNF730
ZNF91
ZNF681
HAVCR1P1
ZNF43
ZNF676
LINC01233
ZNF99
ZNF724P
LINC01224
RPSAP58
LOC100996349
LOC101929164
LOC101929124
LOC374890
LOC101929144
21,768,278
19p12-p11, 2.78Mb region
24,550,666