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

Zhaoming Wang, Katherine A McGlynn, Ewa Rajpert-De Meyts, D Timothy Bishop, Charles C Chung, Marlene D Dalgaard, Mark H Greene, Ramneek Gupta, Tom Grotmol, Trine B Haugen, Robert Karlsson, Kevin Litchfield, Nandita Mitra, Kasper Nielsen, Louise C Pyle, Stephen M Schwartz, Vésteinn Thorsson, Saran Vardhanabhuti, Fredrik Wiklund, Clare Turnbull, Stephen J Chanock, Peter A Kanetsky, Katherine L Nathanson, and the Testicular Cancer Consortium

The international Testicular Tumor Consortium (TECAC) combined five published genome-wide association studies of testicular germ cell tumor (TGCT; 3,558 cases and 13,970 controls) to identify new susceptibility loci. We conducted a fixed-effects meta-analysis, including, to our knowledge, 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 ($P < 5 \times 10^{-8}$). 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 familial 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 to the etiology of TGCT.

In Europe and the United States, TGCT is the most common cancer in young men aged 20 to 39 years. The incidence of TGCT is rising and is highest in men of northern European ancestry and lowest in men of African ancestry. Risk factors for TGCT include cryptorchidism, increased adult height, prior diagnosis, and a familial history of TGCT. 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 heritability ranges from 37% to 49% (refs. 9, 10). Despite the multiple lines of evidence demonstrating a considerable genetic component to TGCT risk, linkage and candidate gene approaches to find genes with rare, highly penetrant variants involved in TGCT etiology were unsuccessful.

In contrast, 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, with 13 loci identified after replication on the iCOGS array. One locus was identified as a candidate region. The genes mapping in or near identified susceptibility loci have highlighted 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 and combined the association test statistics for a total of 8,960,654 autosomal SNPs and 249,696 X-chromosome SNPs after excluding those with an INFO (imputation quality) score of <0.3 or a minor allele frequency (MAF) of <0.01. We conducted a fixed-effects meta-analysis for 3,558 cases and 13,970 controls (Online 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 heritability ranges from 37% to 49% (refs. 9, 10). Despite the multiple lines of evidence demonstrating a considerable genetic component to TGCT risk, linkage and candidate gene approaches to find genes with rare, highly penetrant variants involved in TGCT etiology were unsuccessful.

In contrast, 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, with 13 loci identified after replication on the iCOGS array. One locus was identified as a candidate region. The genes mapping in or near identified susceptibility loci have highlighted 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.

Received 3 April 2016; accepted 27 April 2017; published online 12 June 2017; doi:10.1038/ng.3879
Table 1 TGCT meta-analysis association results for new loci and new independent SNPs in established loci

| Chromosome Position (bp) | Gene | Study | INFO | Controls | Cases | Ref. allele | EAF Case | OR CI | p | p_het | p_maf | pHet | I^2 |
|--------------------------|------|-------|------|----------|-------|-------------|----------|------|---|-------|-------|------|-----|
| 2q14.2                   | TFCP2L | 1555 | C T | 0.15 | 0.20 | 1.45 | 1.18–1.79 | 4.19 × 10^{-4} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | MAP2K1 | 6887 | C T | 0.16 | 0.20 | 1.32 | 1.06–1.64 | 1.25 × 10^{-2} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | TFCP2L | 6887 | C T | 0.14 | 0.17 | 1.25 | 1.09–1.44 | 1.74 × 10^{-3} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.17 | 0.21 | 1.41 | 0.98–2.03 | 6.39 × 10^{-2} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,968 | 3,555 | 1.26 | 1.16–1.36 | 1.68 × 10^{-8} | 0.39 | 3.6 |
| 3q26.2                   | GPR160 | 6887 | C T | 0.41 | 0.44 | 1.14 | 0.98–1.33 | 9.31 × 10^{-2} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.40 | 0.44 | 1.24 | 1.12–1.37 | 2.08 × 10^{-5} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,968 | 3,555 | 1.19 | 1.12–1.26 | 3.87 × 10^{-9} | 0.67 | 0.0 |
| 4q35.2                   | ZFP42  | 6887 | C T | 0.66 | 0.63 | 0.84 | 0.72–0.98 | 2.71 × 10^{-2} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.68 | 0.63 | 0.77 | 0.70–0.86 | 1.18 × 10^{-6} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,968 | 3,555 | 0.84 | 0.79–0.89 | 3.13 × 10^{-8} | 0.02 | 67.4 |
| 7q36.3                   | NACAPG2| 6887 | C T | 0.67 | 0.65 | 0.89 | 0.76–1.06 | 1.86 × 10^{-1} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.68 | 0.64 | 0.82 | 0.62–1.08 | 1.59 × 10^{-1} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,968 | 3,555 | 0.84 | 0.79–0.89 | 2.38 × 10^{-8} | 0.92 | 0.0 |
| 9p24.3*                  | DMRT1  | 6887 | C T | 0.69 | 0.65 | 0.84 | 0.76–0.94 | 1.51 × 10^{-3} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.64 | 0.60 | 0.81 | 0.68–0.96 | 1.69 × 10^{-2} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,968 | 3,555 | 0.84 | 0.79–0.89 | 3.13 × 10^{-8} | 0.02 | 67.4 |
| 10q26.13                 | LHPP   | 6887 | C T | 0.70 | 0.66 | 0.82 | 0.73–0.91 | 3.89 × 10^{-4} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.69 | 0.63 | 0.82 | 0.62–1.08 | 1.59 × 10^{-1} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,968 | 3,555 | 0.84 | 0.79–0.89 | 2.38 × 10^{-8} | 0.92 | 0.0 |
| 15q21.3                  | PRTG   | 6887 | C T | 0.51 | 0.55 | 1.25 | 1.07–1.46 | 4.76 × 10^{-3} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.50 | 0.55 | 1.25 | 1.07–1.46 | 4.76 × 10^{-3} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,970 | 3,556 | 1.65 | 1.38–1.96 | 1.75 × 10^{-8} | 0.95 | 0.0 |
| 15q22.31                 | MAP2K1 | 6887 | C T | 0.26 | 0.30 | 1.27 | 1.07–1.50 | 5.59 × 10^{-3} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.26 | 0.30 | 1.27 | 1.07–1.50 | 5.59 × 10^{-3} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,970 | 3,556 | 1.65 | 1.38–1.96 | 1.75 × 10^{-8} | 0.95 | 0.0 |
| 19p12*                   | ZNF728 | 6887 | C T | 0.15 | 0.16 | 0.69 | 0.55–0.85 | 6.64 × 10^{-4} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.26 | 0.28 | 1.18 | 1.05–1.31 | 3.89 × 10^{-3} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,970 | 3,556 | 1.23 | 1.16–1.32 | 1.10 × 10^{-10} | 0.70 | 0.0 |
| 19p12*                   | ZNF726 | 6887 | C T | 0.21 | 0.22 | 1.04 | 0.85–1.28 | 6.77 × 10^{-1} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Denmark | 363 | C T | 0.20 | 0.22 | 1.04 | 0.85–1.28 | 6.77 × 10^{-1} |
|                         |      |      |     |          |       |              |          |      |   |       |       |      |     |
|                         | Combined | 13,968 | 3,555 | 1.29 | 1.20–1.39 | 2.40 × 10^{-11} | 0.23 | 29.2 | (continued)
significance and an additional four new 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) for these 12 loci using a prior probability of 0.0001 and an odds ratio (OR) of 1.2 (Supplementary Table 2). Two loci, rs61408740 and rs17336718, failed to surpass a BFDP threshold of 0.05, likely because of their low MAFs (0.023 and 0.053, respectively).

Prior reports identified 27 independent SNPs in 25 distinct regions and the gr/gr deletion associated with TGCT susceptibility (Fig. 1). In our current study, 19 of these SNPs (in 17 loci) reached genome-wide significance (Table 2). Eight previously reported susceptibility markers were not identified in our meta-analysis, which was 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 our findings in context, we calculated the BFDP for these 27 loci using a prior probability of 0.10, which assumes that 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 new loci but still may be too conservative for the reidentification of previously reported susceptibility markers. Only one locus among all 27, rs11705932, failed to surpass the BFDP threshold of 0.05.

![Figure 1](image-url)

**Figure 1** All identified SNP markers associated with TGCT susceptibility to date. In the ideograms, red dots and red rs number annotation indicate SNPs identified and described in the current study ($P \leq 5 \times 10^{-8}$), blue dots and blue rs number annotation represent previously identified SNP markers achieving genome-wide significance ($P \leq 5 \times 10^{-8}$) in the current study, and gray dots and gray rs number annotation represent previously identified SNPs that failed to achieve genome-wide significance in this study ($P > 5 \times 10^{-8}$).

### Table 1 TGCT meta-analysis association results for new loci and new independent SNPs in established loci (Continued)

| Cytoband | Gene neighborhood | SNP | Position (bp) | Study | INFO Controls | Cases Ref. allele | Effect allele | EAF | OR | CI | P | Pnull | $\rho$ |
|----------|-------------------|-----|--------------|-------|--------------|----------------|----------------|-----|----|----|----|--------|------|
| 19p12a   | ZNF257            | rs73019876 | 22,267,849   | NCI   | 0.93         | 1,055          | T, 0.45        | G, 0.42        | 0.89 | 0.76–1.04 | 1.35 × 10^{-1} |
|          |                   |      |              | UK    | 0.96         | 4,946          | T, 0.45        | G, 0.40        | 0.83 | 0.75–0.91 | 1.51 × 10^{-4} |
|          |                   |      |              | Penn  | 0.95         | 918            | G, 0.51        | T, 0.49        | 0.91 | 0.78–1.07 | 2.59 × 10^{-1} |
|          |                   |      |              | Norway/Sweden | 0.95 | 6,687 | T, 0.43 | G, 0.41 | 0.85 | 0.77–0.94 | 1.22 × 10^{-3} |
|          |                   |      |              | Denmark | 0.94 | 363   | G, 0.44 | T, 0.36 | 0.72 | 0.55–0.93 | 1.25 × 10^{-2} |
|          |                   |      |              | Combined | 13,969 | 3,555 | 0.85 | T, 0.80–0.90 | 2.04 × 10^{-8} |
| Xq28     | TKTL1             | rs17336718 | 153,536,119  | NCI   | 0.97         | 1,056         | C, 0.05        | T, 0.09        | 1.33 | 1.07–1.64 | 8.85 × 10^{-3} |
|          |                   |      |              | UK    | 0.63         | 4,945         | C, 0.05        | T, 0.07        | 1.46 | 1.19–1.80 | 3.71 × 10^{-4} |
|          |                   |      |              | Penn  | 0.78         | 918           | C, 0.05        | T, 0.09        | 1.59 | 1.23–2.06 | 4.06 × 10^{-4} |
|          |                   |      |              | Denmark | 0.84 | 363   | C, 0.06 | T, 0.08 | 1.15 | 0.78–1.69 | 4.71 × 10^{-1} |
|          |                   |      |              | Combined | 7,282 | 2,231 | 1.41 | T, 1.25–1.59 | 3.84 × 10^{-8} |

Ref., reference; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
rs12912292 was the most significant new SNP marker in our study (OR = 1.22; P = 1.38 × 10^{-11}) and marks a 131-kb haplotype block at 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 the thyroid, testes, and uterus (Supplementary Fig. 3a). No variant in this region is an expression quantitative trait locus (eQTL) in either normal testes or TGCT.

The SNP marker rs61080747 (OR = 1.23; P = 1.10 × 10^{-10}) marks a 261-kb haplotype block at 15q22.31 that contains several genes, including TIPIN (encoding TIMELESS-interacting protein), MAP2K1 (encoding mitogen-activated protein kinase kinase 1), DIS3L (encoding DIS3-like exoribonuclease), SNACP5 (encoding small nuclear RNA activating complex, polypeptide 5), ZWILCH (encoding Wilcher kinetochore protein), and TIPIN (encoding TIMELESS-interacting protein) (Table 1 and Supplementary Fig. 2b). Several of the encoded proteins, particularly ZWILCH and TIPIN, have high and somewhat tissue-specific expression in testis (Supplementary Fig. 2b-g). TIPIN coordinates the DNA replication checkpoint by interacting with replication protein A^{26}, and ZWILCH is a kinetochore protein important for proper chromatid alignment during cell division^{27}. A missense mutation in ZWILCH, resulting in p.Ser230Gly, lies within the linkage disequilibrium (LD) block (Supplementary Table 3). The LD block also contains a single eQTL for RP11-653f6.1, which encodes a long noncoding RNA (lncRNA) with expression highly specific to testes (Supplemental Fig. 3h). RP11-653f6.1 levels and eQTLs were not assessed in The Cancer Genome Atlas (TCGA) TGCT studies. Further dissection of this signal is needed to pinpoint the likely candidate gene.

The SNP marker rs3755605 (OR = 1.19; P = 3.87 × 10^{-9}) identifies a 213-kb haplotype block containing three genes, GPR160 (encoding G-protein-coupled receptor 160), PHC3 (encoding polyhomeotic homolog 3), and PRKCI (encoding protein kinase C, iota form), at 3q26.2 (Table 1 and Supplementary Fig. 2c). Several SNPs across this block are eQTLs for GPR160 (Supplementary Table 3 and Supplementary Fig. 4a); the TGCT risk allele is associated with increased expression of GPR160 in both normal testes and TGCT (Supplementary Figs. 5a and 6a). Several SNPs in the haplotype block are also eQTLs in normal testes for RP11-469f4.3, an lncRNA of unknown function. RP11-469f4.3 is highly expressed in normal testes but lies outside of the haplotype block (Supplementary Figs. 3a, 4b, and 5b). RP11-469f4.3 expression was not measured in the TCGA study.

The SNP marker rs2713206 (OR = 1.26; P = 1.68 × 10^{-8}) lies within a smaller LD region of only 48 kb at 2q14.2. TFCP2L1 (encoding transcription factor CP2-like 1) overlies the entirety of the haplotype block (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 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 is downregulated during transition from fetal gonocytes into spermatogonia^{29}. The SNP marker rs6837349 (OR = 0.84; P = 3.13 × 10^{-8}) is located in an intron of ZFP42 (encoding zinc-finger protein 42) at 4q35.2 and marks a small 11-kb haplotype block 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. Additionally, both ZFP42 and TFCP2L1 are involved in embryonic stem cell pluripotency.

The SNP marker rs61408740 (OR = 1.65; P = 1.75 × 10⁻⁸) is located in an intron of LHPP (encoding phospho-lysine phosphohistidine inorganic pyrophosphate phosphatase) at 10q26.13 (Table 1 and Supplementary Fig. 2f). Only two SNPs were identified with pairwise LD (r² > 0.4 with rs61408740, one in the region of the second gene, FAM175B (Supplementary Table 3); neither of these SNPs are eQTLs. LHPP encodes an inorganic diphosphatase that functions in oxidative phosphorylation.

The SNP marker rs11769858 (OR = 0.84; P = 2.38 × 10⁻⁸) identifies an 82-kb LD block at 7q36.3 that contains a large portion of NCAPG2 (encoding non-SMC condensin II complex subunit G2) (Table 1 and Supplementary Fig. 2g). NCAPG2 encodes a regulatory subunit of the condensin II complex and is highly expressed in testes (Supplementary Fig. 3g) and plays a role in chromosome assembly and segregation during mitosis.

We also identified a locus marked by SNP rs17336718 (OR = 1.41; P = 3.84 × 10⁻⁸) at Xq28 (Table 1 and Supplementary Fig. 2h) in an intron of TKT1L (encoding transketolase-like 1), which is highly expressed in normal testes. The SNP is an eQTL for TKT1L in TCGA TGCT data (Supplementary Fig. 6c). TKT1L converts sedoheptulose to ribose and glyceraldehyde to xylulose, linking the pentose phosphate pathway to the glycolytic pathway. Interestingly, although overexpression of TKT1L is associated with the Warburg effect and poor prognosis in several cancer types, the TGCT risk allele is associated with lower expression of TKT1L.

At the previously reported TGCT susceptibility locus encompassing DMRT1 on chromosome 9, we identified a third independent signal, rs55873183 (OR = 1.89; P = 2.18 × 10⁻²³) (Table 1, Fig. 2a, and Supplementary Table 4a). This intronic SNP marker has r² values of 0.03 and 0.06 with the two previously published SNP markers, rs7040024 and rs755383, respectively; it retained genome-wide significance in conditional analysis (Table 1, Supplementary Table 4b, and Supplementary Fig. 2i). We also identified three additional independent signals at 19p12: rs58521262, rs34601376, and rs73019876 (Table 1, Fig. 2b, Supplementary Table 5a, and Supplementary Fig. 2j–l). We identified a SNP, rs2194275 (P = 9.23 × 10⁻¹²; OR = 0.76), in moderate LD (r² = 0.7) with the previously published rs2195987 (P = 1.21 × 10⁻⁹; OR = 0.81), which was more significant in this meta-analysis (Supplementary Table 5b). The region at 19p12 contains a cluster of Krüppel-associated box zinc-finger protein (KRAB-ZFP) genes.

A number of different GWAS, including one in embryonic stem cell pluripotency, have identified TGCT susceptibility markers continue to reveal significant associations, even after adjustment for ancestry. The newly identified TGCT susceptibility markers continue to demonstrate moderate effects, with ORs that range from 1.17 to 1.89.

Newly identified TGCT susceptibility markers continue to explain moderate effects, with ORs that range from 1.17 to 1.89.
METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services nor does the mention of trade names, commercial products, or organizations indicate endorsement by the US government. We thank B. Weathers for her coordination of TECAC, J. Pluta for biostatistical assistance, and K. D’Andrea for expert assistance with SNP genotyping. We thank D.R. Stewart and J. T. Loud for critical support of the NCI Clinical Genetics Branch Familial Testicular Cancer Project (NCI 02-C-0178; NCT-00039598). We also thank all previous contributors to the GWAS analyzed in this study, including N. Weinhold, D. Edgardo, H. Leffers, and A. Juul, N.E. Skakkebaek, and S. Brunak from the Danish study.

The Testicular Cancer Consortium is supported by National Institutes of Health grant U01CA164947 to K.L.N., P.A.K., and S.M.S. A portion of this work was supported by the Intramural Research Program of the National Cancer Institute and by support services contract HHSN261201300003C 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 K.L.N. and P.A.K. 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 2 (WTCCC2). C.T. is supported by the Movember Foundation. K.L. is supported by a PhD fellowship from Cancer Research UK. L.C.P. is supported by the Intramural Research Program of the National Cancer Institute (NCT-00039598). We also thank all previous contributors to the GWAS analyzed in this study.

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.B.H., M.D.D., M.H.G., R.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 and to study, data management. Z.W., K.A.M., E.R.-D.M., D.T.B., M.D.D., M.H.G., R.G., T.B.H., R.K., K.L., N.M., K.N., S.V., F.W., C.T., S.I.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.I.C., P.A.K., and K.L.N. carried out the meta-analysis and the additional bioinformatics analyses, including using GTEx and TCGA TCGA data. Z.W., P.A.K., and K.L.N. drafted the initial manuscript, and all authors reviewed and contributed to the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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, 1095–1106 (2014).
4. Broman, K. et al. Testicular, other genital, and breast cancers in first-degree relatives of testicular cancer patients and controls. Cancer Epidemiol. Biomarkers Prev. 13, 1316–1324 (2004).
5. Chen, S., Sonneveld, D.J. et al. Risk of cancer in first- and second-degree relatives of testicular germ cell tumors and controls, Int. J. Cancer 124, 952–957 (2009).
6. Heidmali, K. et al. Risk of cancer in relatives of testicular cancer patients, Br. J. Cancer 73, 970–973 (1996).
7. Schwaab, J., Mclennan, D., Beattie, B. & Meade, T. Adult and adolescent risk factors for testicular cancer. Nat. Rev. Urol. 9, 339–349 (2012).
8. Litchfield, K. et al. Quantifying the heritability of testicular germ cell tumours using both population-based and genomic approaches. Sci. Rep. 5, 13889 (2015).
9. Mcucci, L.A. et al. Familial risk and heritability of cancer among twins in Nordic countries. J. Am. Med. Assoc. 315, 68–76 (2016).
10. Crockford, G.P. et al. Genome-wide linkage screen for testicular germ cell tumour susceptibility loci. Hum. Mol. Genet. 15, 443–451 (2006).
11. Chung, C.C. et al. Meta-analysis identifies four new loci associated with testicular germ cell cancer. Nat. Genet. 45, 680–685 (2013).
12. Kanetsky, P.A. et al. Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. Nat. Genet. 41, 811–815 (2009).
13. 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 2 (WTCCC2). C.T. is supported by the Movember Foundation. K.L. is supported by a PhD fellowship from Cancer Research UK. L.C.P. is supported by the Intramural Research Program of the National Cancer Institute (NCT-00039598). We also thank all previous contributors to the GWAS analyzed in this study.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
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, 719–724 (2011).
39. Eichler, E.E. et al. Complex β-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, 5385–5394 (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, 3898–3905 (2014).
43. Mancuso, N. et al. The contribution of rare variation to prostate cancer heritability. Nat. Genet. 48, 30–35 (2016).
44. Michaelidou, K. et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat. Genet. 47, 373–380 (2015).

Testicular Cancer Consortium:

Katherine A McGlynn1, Ewa Rajpert-De Meyts2, D Timothy Bishop3, Mark H Greene1, Ramneek Gupta4, Tom Grotmol5, Trine B Haugen6, Stephen M Schwartz12, Fredrik Wiklund7, Clare Turnbull8,15, Stephen J Chanock1, Peter A Kanetsky16, Katherine L Nathanson10,17, Alberto Ferlin21, Jourik A Gietema22, Victoria Cortessis23, Russ Hauser24, Michelle Hildebrandt25, Lambertus A Kiemeney26, Christian Kubisch27, Davor Lesssel27, Thorunn Rafnar28, Lorenzo Richiardi29, Rolf Skotheim30 & Tongzhang Zheng31

21Department of Medicine, Unit of Andrology and Reproductive Medicine, University of Padova, Padova, Italy. 22Medical Oncology, University Medical Center Groningen, Groningen, the Netherlands. 23Department of Clinical Preventive Medicine, Keck School of Medicine at the University of Southern California, University of Southern California, Los Angeles, California, USA. 24Department of Environmental Health, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA. 25Department of Epidemiology, Division of OVP, Cancer Prevention and Population Sciences, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. 26Radboud University Medical Centre, Radboud Institute for Health Sciences, Nijmegen, Nijmegen, the Netherlands. 27Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. 28deCODE Genetics, Reykjavik, Iceland. 29Department of Biostatistics and Epidemiology, University of Turin, Turin, Italy. 30Department of Molecular Oncology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. 31Department of Epidemiology, Brown University, Providence, Rhode Island, USA.
CONTRIBUTIONS TO THE 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\textsuperscript{12,13,14,22}. 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 of <95%, a Hardy–Weinberg proportion test value of $P < 0.000001$ or a MAF of <1% were removed prior to imputation. Imputation was conducted using IMPUTE2 software v2.2.2 and v3 of the 1000 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 liftOver tool. Second, the strand of the inference data was aligned with the 1000 Genomes Project data as informed by allele state comparison or allele frequency matching for A/T and G/C SNPs. A prephasing strategy with SHAPEIT software version 1 was adopted to improve imputation performance. The phased haplotypes from SHAPEIT were imported directly into the IMPUTE2 program. We applied sliding windows of 4 Mb with 250 kb as an overlapping buffer and generated 744 segments for the imputation of autosomes. For the X chromosome, pseudautosomal region 1 (PAR1), PAR2, and the remaining region, which was split into 37 segments, were imputed separately. We excluded imputed loci with an INFO score of <0.3 or a MAF of <0.01 from further association analysis. Further, we acknowledge the limitations of imputation, including that the accuracy of imputation depends on the LD between markers in the reference panel and the 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.

Statistical analysis. Within each data set, a test for trend was performed for each SNP using SNIPTEST software v2.2 or v2.5. Fixed-effects meta-analysis was used to combine individual within-study association estimates from five imputed GWAS scans. Genetic effect heterogeneity across studies was assessed using the $I^2$ statistic and the $P$ value calculated from Cochran’s $Q$ statistic. To refine the association signals of each risk-associated region, we first performed LD pruning using pairwise $r^2 > 0.3$ and then conducted conditional association analyses to estimate the independent effect of each SNP by simultaneously including, in the same logistic regression model, specified SNPs from the same region that surpassed an unconditional $P$ value of $5 \times 10^{-8}$.

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

In silico bioinformatics analysis. We used HaploReg v4.1 and RegulomeDB v1.1 to explore potential noncoding functional annotation within the ENCODE database in the genomic regions surrounding our SNPs of interest, with particular attention to annotations in induced pluripotent stem cells (iPSCs) and embryonic stem cells (ESCs), as we considered these tissue types to be the best proxies for TGCT. Specifically, we interrogated the LD block of neighboring SNPs in a haplotype block defined by pairwise $r^2 > 0.4$ with the index SNP (Supplementary Table 4). We also searched the GTEx v6 database to determine whether the haplotype block SNPs were implicated as eQTLs in their sample of 157 normal adult testis tissues with available genotype. Of note, normal testis contains an abundance of stromal cells (i.e., Sertoli and Leydig cells), so it 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 new susceptibility loci for eQTLs among the 128 cases of TGCT with linked genotype data available in TCGA.

For correlation between genotype and expression data in the TCGA TGCT data set, the genotype data were downloaded from NCI’s Genomic Data Commons (https://gdc.ncri.nih.gov/). Data were converted to PLINK (v1.07)\textsuperscript{47} format. Subjects were screened for discordant sex, insufficient genotype call rate (>0.05), and excessive heterozygosity (>±3 s.d. 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 5 subjects were removed (all for heterozygosity violations), leaving 145 valid for analysis. The 1000 Genomes Project Phase 3 panel was used as the reference set\textsuperscript{41}. Alignment to the reference set and haplotype estimation were performed using SHAPEIT (v2)\textsuperscript{48}, and additional SNPs were imputed using IMPUTE2 (ref. 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 retrieved from the TCGA Data Coordinating Center and processed through the TCGA pipeline at the TCGA Genoma 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\textsuperscript{50}. Imputed genotypes for all new SNPs reported in this paper were related to gene expression for the 128 cases with both genotype and gene expression data available. Associations 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 according to the standard pipeline at the Cancer Genomics Research Laboratory at the National Cancer Institute (Supplementary Table 7). For six loci that failed initial TaqMan assay design, surrogate SNPs in LD with the original SNPs were used. We randomly selected about 1,000 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 were 0.98, 0.97, and 0.93 for NCI, Penn, and Norway/Sweden, respectively (Supplementary Table 7).

Data availability. Individual-level data from the UK GWAS have been deposited in the European Genome-phenome Archive (EGA) under accession EGAS000001001302. Individual-level data have been deposited in the database of Genotypes and Phenotypes (dbGaP) for the University of Pennsylvania (phs001307.v1.p1) and NCI (phs001303.v1.p1). The summary data from the TECAC meta-analysis also have been deposited in dbGaP (phs001349.v1.p1).

45. Park, J.H. et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat. Genet. 42, 570–575 (2010).
46. Hemminki, K. & Li, X. Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. Br. J. Cancer 90, 1765–1770 (2004).
47. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
48. Delanoue, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods 10, 5–6 (2013).
49. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
50. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323 (2011).