Association Study between Polymorphisms in DNA Methylation-Related Genes and Testicular Germ Cell Tumor Risk

Chiara Grasso[1*], Maja Popovic[1], Elena Isaevska[1], Fulvio Lazzarato[1], Valentina Fiano[1], Daniela Zugna[1], John Pluta[2], Benita Weathers[2], Kurt D'Andrea[2], Kristian Almstrup[3,4], Lynn Anson-Cartwright[5], D. Timothy Bishop[6], Stephen Chanock[7], Chu Chen[8], Victoria K. Cortessis[9], Marlene D. Dalgaard[10], Siamak Daneshmand[11], Alberto Ferlin[12], Carlo Foresta[13], Megan Frone[7], Marija Gamulin[14], Jourik A. Gietema[15], Mark H. Greene[7], Tom Grotmol[16], Robert J. Hamilton[4], Trine B. Haugen[17], Russ Hauser[18], Robert Karlsson[19], Lambertus A. Kiemeney[20], Davor Lessel[21], Patrizia Lista[22], Ragnhild A. Lothe[23,24], Chey Loveday[25], Coby Meijer[15], Kevin T. Nead[26], Jérémie Nsengimana[27], Rolf I. Skotheim[23,28], Clare Turnbull[29], David J. Vaughn[30,31], Fredrik Wiklund[19], Tongzhang Zheng[32], Andrea Zitella[33], Stephen M. Schwartz[8], Katherine A. McGlynn[7], Peter A. Kanetsky[34], Katherine L. Nathanson[2,31], Lorenzo Richiardi[1], for the Testicular Cancer Consortium

ASSOCIATION STUDY OF POLYMORPHISMS IN DNA METHYLATION-RELATED GENES WITH THE RISK OF TESTICULAR GERM-CELL TUMOR

ASSOCIATION STUDY OF POLYMORPHISMS IN DNA METHYLATION-RELATED GENES WITH THE RISK OF TESTICULAR GERM-CELL TUMOR

Chiara Grasso[1*], Maja Popovic[1], Elena Isaevska[1], Fulvio Lazzarato[1], Valentina Fiano[1], Daniela Zugna[1], John Pluta[2], Benita Weathers[2], Kurt D'Andrea[2], Kristian Almstrup[3,4], Lynn Anson-Cartwright[5], D. Timothy Bishop[6], Stephen Chanock[7], Chu Chen[8], Victoria K. Cortessis[9], Marlene D. Dalgaard[10], Siamak Daneshmand[11], Alberto Ferlin[12], Carlo Foresta[13], Megan Frone[7], Marija Gamulin[14], Jourik A. Gietema[15], Mark H. Greene[7], Tom Grotmol[16], Robert J. Hamilton[4], Trine B. Haugen[17], Russ Hauser[18], Robert Karlsson[19], Lambertus A. Kiemeney[20], Davor Lessel[21], Patrizia Lista[22], Ragnhild A. Lothe[23,24], Chey Loveday[25], Coby Meijer[15], Kevin T. Nead[26], Jérémie Nsengimana[27], Rolf I. Skotheim[23,28], Clare Turnbull[29], David J. Vaughn[30,31], Fredrik Wiklund[19], Tongzhang Zheng[32], Andrea Zitella[33], Stephen M. Schwartz[8], Katherine A. McGlynn[7], Peter A. Kanetsky[34], Katherine L. Nathanson[2,31], Lorenzo Richiardi[1], for the Testicular Cancer Consortium[#]

[1]Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO Piedmont, Turin, 10126, Italy

[2]Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.

[3]Department of Growth and Reproduction, Copenhagen University Hospital – Rigshospitalet, Copenhagen, 2300, Denmark.

[4]Department of Cellular and Molecular Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, 2200, Copenhagen

[5]Department of Surgery (Urology), University of Toronto and The Princess Margaret Cancer Centre, Toronto, ON, M5G 1Z6, Canada.

[6]Department of Haematology and Immunology, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, LS9 7TF, UK.

[7]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, 20892, USA.

[8]Program in Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109; Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA.

[9]Departments of Population and Public Health Sciences, and Obstetrics and Gynecology, Keck School of Medicine at the University of Southern California, Los Angeles, CA, 90033, USA.

[10]Department of Health Technology, Technical University of Denmark, Lyngby, 2800, Denmark.

[11]Department of Urology, Keck School of Medicine at the University of Southern California, Los Angeles, CA, 90033, USA.
[12] Unit of Andrology and Reproductive Medicine, Department of Medicine, University of Padova, Padova, Italy.
[13]Unit of Andrology and Reproductive Medicine, Department of Medicine, University of Padova, Padova, 35128, Italy.
[14]Department of Oncology, University Hospital Centre Zagreb, University of Zagreb School of Medicine, Zagreb, 10000, Croatia.
[15]Department of Medical Oncology, University Medical Center Groningen, University of Groningen, Groningen, 9713 GZ, The Netherlands.
[16]Department of Research, Cancer Registry of Norway, Oslo, N-0304, Norway.
[17]Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Oslo, NO-0130, Norway.
[18]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, 02115, USA.
[19]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE-171 6377, Sweden.
[20]Radboud University Medical Center, Nijmegen, 6525 XZ, The Netherlands.
[21]Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, 20246, Germany.
[22]Medical Oncology Division 1, AOU "Città della Salute e della Scienza di Torino", Turin, 10126, Italy.
[23]Department of Molecular Oncology, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, N-0379, Norway.
[24]Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, 0318, Norway.
[25]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK.
[26]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, TX 77030, USA.
[27]Biostatistics Research Group, Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle, NE1 7RU, UK.
[28]Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, 0371, Norway.
[29]Division of Genetics & Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK; William Harvey Research Institute, Queen Mary University, London, EC1M 6BQ, UK.

[30]Division of Hematology and Oncology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.
[31]Abramson Cancer Center, Perelman School of Medicine, Philadelphia, PA, 19104, USA
[32]Department of Epidemiology, Brown School of Public Health, Brown University, Providence, RI, 8602903, USA.
[33]Division of Urology, Department of Surgical Science, AOU "Città della Salute e della Scienza di Torino", University of Turin, Turin, 10126, Italy.
[34]Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, 33612, USA.
[#]The complete list of members and affiliations appears at the end of the paper.

Corresponding author:
Chiara Celestina Grasso, University of Turin, Department of Medical Sciences, via Santena 7, 10126, Turin, Italy. Phone: (+39) 0116336863; Fax: (+39) 0116334571
E-mail: chiaracelestina.grasso@unito.it

Abstract 250 words; current 248

Testicular germ cell tumors (TGCTs), histologically grouped as seminomas and nonseminomas, are believed to arise from primordial gonocytes with the maturation process blocked at the point when they are subjected to DNA methylation reprogramming. Single-nucleotide polymorphisms (SNPs) in DNA methylation machinery and folate-dependent one-carbon metabolism genes have been suggested to affect the proper establishment of DNA methylation.

We aimed to evaluate the association between SNPs in methylation-related genes and the risk of TGCT. We selected 273 tag SNPs from 28 DNA methylation-related genes and we carried out association analysis both at individual-SNPs level and at gene-based level by using the summary statistics from the Testicular Cancer Consortium, including 10,156 TGCT cases and 179,683 controls.

In individual-SNP analyses, seven tag SNPs, four mapping within *MTHFR*, were associated with the risk of TGCT after correction for multiple testing (q-value ≤0.05). Queries of public databases showed that three of these SNPs were associated with changes in enzymatic activity (rs1801133) and expression level of *MTHFR* in testis tissue (rs12121543 and rs1476413). Gene-based analyses revealed *MTHFR* (q-value=$8.4x10^{-4}$), *MECP2* (q-value=$2x10^{-3}$) and *ZBTB4* (q-value=0.03) as the top TGCT-associated genes. In analysis stratified by tumor histology, four *MTHFR* SNPs were associated with seminoma. In gene-based analysis *MTHFR* was associated with the risk of seminoma (q-value=$2.8x10^{-4}$), but not with non-seminomatous tumors (q-value=0.22).

In conclusion, genetic variants of *MTHFR*, some with plausible functional roles, are associated with the risk of TGCT. This could support a possible involvement of aberrant epigenetic mechanisms in the testicular germ cell tumor pathogenesis.

Introduction

Testicular cancer is the most common malignancy among men aged 15-40 years of European ancestry. Since the mid-20th century, testicular cancer incidence rates have been increasing in many countries and are predicted to further increase over the next decades (1,2).

Testicular germ cell tumors (TGCTs) account for 98% of all testicular cancers and are histologically grouped as seminoma and non-seminomatous tumors. The latter include embryonal carcinomas, teratomas, choriocarcinoma, and yolk sac tumors. Mixed germ cell tumors, composed of two or more germ cell tumor types, are typically classified in the non-seminomatous group since they have similar molecular features and prognosis (3,4). Established risk factors for TGCT include age, ancestry, contralateral testicular cancer, adult height, cryptorchidism and positive family history (5).

A strong genetic component has been described in TGCT, with an estimated 37% heritability in twin studies (6). Genome-wide association studies (GWAS) have identified multiple independent common variants associated with TGCT risk, strongly suggesting that the genetic susceptibility for TGCT is not due to a few major high-penetrance genes, but rather to multiple genetic variants with modest to small effect sizes (7,8).

Both seminoma and non-seminomatous germ cell tumors are believed to arise from primordial gonocytes that have failed to differentiation normally into pre-spermatogonia in early fetal life (9). Accordingly, these immature fetal germ cells accumulate within the seminiferous tubule forming pre-invasive neoplastic lesions called germ-cell neoplasia in situ (GCNIS). The current pathogenetic model for TGCT is based on the hypothesis that the GCNIS cell could begin to proliferate at puberty and eventually acquire malignant potential (10,11).

During early embryonic development, gonocytes arrested in mitosis undergo extensive epigenetic remodelling including the genome-wide erasure of DNA methylation markers and *de novo* re-establishment of a parental imprinting pattern that is completed prior to birth (12). Studies have shown that the genome of GCNIS in the human adult testis exhibits global DNA methylation erasure (13,14), a common feature of primordial gonocytes (15,16).

Striking differences in methylation profiles in TGCT subtypes have been described: non-seminomatous tumors show aberrantly increased promoter methylation, whereas in seminomas the genome is mostly maintained in an unmethylated state (13,14,17,18). This suggest that DNA

methylation could be important for the pathogenesis of TGCTs, particularly of specific TGCT sub-types.

The proper establishment of DNA methylation patterns requires the activity of several proteins which together comprise the DNA methylation machinery. These proteins are responsible for: i) active removal of methyl groups (DNA demethylases or so-called “DNA erasers”), ii) establishment of the *de-novo* methylation and maintenance of the methylation pattern during DNA replication (DNA methyltransferases or so-called “DNA writers”); iii) reading the methylation pattern by binding the 5-methylcytosine base (methyl-CpG binding proteins or so-called “DNA readers”) (19). Methyl groups, essential for methylation reactions, are uniquely provided by the universal methyl donor S-adenosylmethionine, which is synthesized by the folate-dependent one-carbon metabolism using B-vitamins as coenzymes (20).

Studies have shown that single-nucleotide polymorphisms (SNPs) of genes coding for proteins and enzymes involved in DNA methylation machinery and in folate-dependent one-carbon metabolism are able to alter promoter activity and expression of the gene itself, thus influencing the establishment of individual methylation patterns (21-23).

In this study, we hypothesized that genetic variants in genes involved in the DNA methylation machinery and in one-carbon metabolism can influence the risk of developing TGCT. We aimed to evaluate the associations between individual SNPs in DNA methylation-related genes and the risk of TGCT, and to assess their potentially collective effect by performing gene-based analyses.

Results

Table 1 reports the number of TGCT cases and controls for the eight studies involved in the meta-analysis, as well as the number of cases stratified by histologic subtype (not available for 3% of the cases), family history of TGCT (not available for 24.7% of cases and 93.4% of controls) and history of cryptorchidism (not available for 24.7% of cases and 93.4% of controls).

Individual SNP- and gene-based analysis on all TGCT cases

The main analyses involved 10,156 cases and 179,683 controls. After correction for multiple testing, seven SNPs were associated with TGCT risk with q-values≤ 0.05, as reported in Table 2. The OR estimates ranged from 0.90 to 1.11 (Table 2). Four were located in *MTHFR* (rs1801133, rs12121543, rs1476413 and rs13306556), two in *MECP2* (rs1734791 and rs1624766), and one in *ZBTB4* (rs4796420). None of these SNPs were found associated with TGCT risk at genome-wide

scale (8). With the exception of rs4796420, the heterogeneity for *MTHFR* and *MECP2* polymorphisms among the eight studies was low. Considering the specific studies, no obvious study characteristic explaining the heterogeneity seen for the rs4796420 has been found. Complete results of all the analysed SNPs are reported in the Supplementary Appendix.

In the gene-based analysis three of the 28 analysed genes showed an association with risk of TGCT, with a q-value below 0.05: *MTHFR* (q-value=$8.4x10^{-4}$), *MECP2* (q-value=$2x10^{-3}$) and *ZBTB4* (q-value=0.03) (Table 3).

Stratified analyses

The analyses stratified by histologic subtype included 4,529 seminomas and 4,630 non-seminomatous germ cell tumors.

After adjustment for multiple testing, *MTHFR* SNPs rs1801133, rs12121543, rs6541003 and rs1476413 were associated with seminoma with a q-value ≤ 0.05 (q-values =$1.6x10^{-4}$; 0.02; 0.03; 0.05; respectively). Three of these SNPs were also among those top-ranked in non-stratified individual-SNP analysis (see above). P value for heterogeneity and I^2 index calculation revealed no substantial heterogeneity among studies (Supplementary Table S1 and Supplementary Appendix). None of the SNP were associated with the risk of non-seminomatous tumors with a q-value ≤ 0.05 (data not shown); furthermore, none of the top-ranked SNPs were included in the top positions of the main analysis (Supplementary Appendix).

As shown in Table 4, gene-based analyses stratified by histological subtype revealed an association between *MTHFR* and seminoma risk (q-value=$2.8x10^{-4}$), and no clear evidence of an association with non-seminomatous tumors for any of the 28 selected genes (Table 4).

Analyses restricted to men with a positive family history of TGCT and those with a history of cryptorchidism were carried out on 356 and 521 cases, respectively, from the Replication, NCI, UPENN and UK studies, which were compared with the 11,927 controls included in the same studies. In individual-SNP analysis restricted to history of cryptorchidism four polymorphisms, all those mapping in *MECP2*, were excluded since their summary statistics results were available for one study only (the Replication study). Then, two hundred and sixty-three SNPs were used in this analysis.

In both individual-SNP analysis, no SNP was associated with risk of TGCT after correction for multiple testing (Supplementary Appendix).

In both gene-based analyses restricted to cases with family history for TGCT and to those with history of cryptorchidism, no gene was associated with TGCT risk, though *AHCY* and *SHMT1* were two of the top-three most strongly associated genes in both analysis (Supplementary Table S2).

Functional assessment of top SNPs and and expression analysis in TGCT subtypes

Functional annotations of the tag SNPs most strongly associated with the TGCT risk in the main analysis are listed in Table 5. Six of the seven top variants were in intronic regions, whereas rs1801133 was located in the coding region of *MTHFR* gene. Rs1801133 was found to be a missense variant causing an amino acid substitution and defined as damaging by two *in silico* prediction tools, since it maps in a highly conserved sequence. The evaluation of the putative function of the seven top SNPs on regulatory motifs revealed that four of them were predicted to map to protein-binding sites, while all but rs1476413 could alter binding motifs for transcription factors.

In theSNiPA database, the rs1801133 locus was reported as associated with a range of diseases and human traits such as plasma homocysteine and folate levels.

No other common variants were reported in the same linkage disequilibrium (LD) block of rs1801133, while in that of the other six top SNPs several polymorphisms, ranging from 3 to 105, were located.

In the sample of 322 normal adult testis tissues with available genotypes in the GTEx v7 database, the tag SNPs rs12121543 and rs1476413 were associated with *MTHFR* expression quantitative trait loci (eQTLs) in human adult testis tissue (Table 5 and Supplementary Figure S1, upper panel). Each SNP was in strong LD with another locus which was associated with *MTHFR* eQTL in testis tissue: rs3818762 was tagged by rs12121543 (pairwise r^2=0.81), while rs1023252 was a proxy for rs1476413 (pairwise r^2=0.84) (Table 5 and Supplementary Figure S1, lower panel). The C allele (major) of rs12121543 and the C allele (major) of rs1476413, both associated with decreased expression of *MTHFR* (Supplementary Figure S1, upper panel), were associated with an increased TGCT risk in the individual-SNP analysis (Table 2).

Expression analysis by histologic subtypes were limited to *MTHFR*, which was associated with the risk of seminoma, but not with the risk of non-seminomatous germ cell tumors.

Since expression data on the adjacent non neoplastic tissue were not available in the publicly available TGCT dataset, we carried out this analysis on the expression data obtained in the tumor

tissue. From the TGCT dataset we retrieved data on *MTHFR* expression evaluated on 43 seminoma and 68 non-seminomatous tumor tissues.

The p-value for comparison of *MTHFR* expression level between the two histologic subtypes was 0.098. Means of z-scores for seminoma and non-seminomatous tumors were -0.29 and -0.12, respectively.

Discussion

It has been suggested that epigenetic mechanisms may be important driving factors in the pathogenesis of testicular germ cell tumors. A recent large meta-analysis of GWAS on TGCT carried out by the Testicular Cancer Consortium has identified, as associated with TGCT susceptibility, genes critically involved in epigenetic reprogramming through chromatin remodelling and histone modifications (8). We used the genome wide association dataset from the Testicular Cancer Consortium (TECAC; www.tecac.org) to conduct a pathway-focused study on polymorphisms within selected genes involved in DNA methylation, and we found a robust association between variants in *MTHFR* and TGCT risk, some having a possible functional role. We found associations, although weaker, also for variants in *MECP2* and *ZBTB4*.

MTHFR encodes the one-carbon metabolism enzyme 5,10-methylenetetrahydrofolate reductase, necessary for the synthesis of methyl donor S-adenosylmethionine, the primary substrate for DNA methyltransferases. *MTHFR* is a well-studied gene, expressed in several human tissues: according to the Human Protein Atlas database, the highest levels have been reported in glandular cells of the epididymis (24). Moreover, mouse studies have revealed that *MTHFR* is expressed in fetal germ cells, from which the precursor GCNIS is thought to arise, and most highly during the phase of late *de novo* DNA methylation (25,26). However, no eQTL studies on human fetal germ cells are yet available, hence if expression of MTHFR is particularly high also in the embryonic gonad of human males during the times when DNA methylation is acquired remains to be elucidated.

Common genetic variants of *MTHFR* have been studied in relation to several multifactorial disorders as cardiovascular diseases, pregnancy complications, congenital anomalies including neural tube defects, neuropsychiatric diseases and some types of cancer. Results of these studies have been conflicting, making the biological and clinical significance of these polymorphisms still uncertain (27). Moreover, no *MTHFR* polymorphism has been associated neither with congenital anomalies of the genitourinary system, that include both well-established (cryptorchidism) and suggested (hypospadias, inguinal hernia) risk factors for TGCT (28), nor with the risk of testicular cancer itself.

Rs1801133, one of the most well-studied *MTHFR* polymorphisms, is a coding non-synonymous variant, causing an amino acid substitution in the catalytic domain that leads to the synthesis of a thermolabile isoform with reduced activity. Compared with the wild-type GG, the AA and GA genotypes are associated with only ~10-20% and ~65% efficiency of the enzyme, respectively, in converting folic acid into 5-methyltetrahydrofolate that is the biologically active and usable form of folate. This mild *MTHFR* deficiency affects 5–20% of North Americans and Europeans (21). Our individual-SNP analysis showed that the major allele G, encoding the isoform of the enzyme with normal level of activity, is inversely associated with risk of TGCT (per-allele OR=0.90; 95% IC 0.87-0.94).

The association betweeen rs1801133 with folate deficiency and high levels of homocysteine, a folate derivative, has been reported in many studies (29). Both conditions might induce epigenetic changes, leading to global DNA hypomethylation, DNA repair defects, and chromosomal instability, and have been also related to an increased risk of cancer (all types combined) (30).We could hypothesize that the thermolabile isoform of *MTHFR*, coded by the rs1801133 minor allele A, might contribute to a hypomethylated environment by perturbating the folate cycle, and the fact that rs1801133 has been also related toDNA hypomethylation in lymphocytes of healthy adult people (31) might be consistent with this hypothesis.

Two other *MTHFR* polymorphisms, the intronic variants rs12121543 and rs1476413, were associated with TGCT in the individual-SNP analysis. Their major alleles, associated with an increased risk of TGCT, were also associated with a decreased expression of *MTHFR* in testis tissue according to our functional assessment. Albeit these SNPs apparently do not have the same deleterious effect on protein structure as that for rs1801133, they might exert modulating effect on *MTHFR* expression in testis tissue, with possible implications for the establishment of the DNA methylation patterns.

TGCT subtypes originate from the same preneoplastic cell; however, seminoma and non-seminomatous tumors exhibit different global DNA methylation patterns, being seminomas mostly hypomethylated and non-seminomatous tumors retaining high levels of DNA methylation (13). In the stratified analysis we found that rs1801133 was specifically associated with seminomas, and not with non-seminomatous tumors. Similarly, in gene-based analysis stratified by histologic subtype *MTHFR* was found associated only with seminomas.We hence could speculate that common *MTHFR* variants, by causing a decreased *MTHFR* expression or activity and thus theoretically a

lower amount of methyl groups produced, might be involved in the sub-type-specific pathogenesis of hypomethylated seminomas. We could not prove that *MTHFR* is downregulated in seminoma compared to non-seminomatous tumors with the few expression data available from public data repositories, hence these analyses need replication in a larger sample and a dedicated study design. Moreover, to demonstrate if *MTHFR* is differentially regulated in the tissue from which seminoma and non-seminomatous tumors originate, expression data obtained on the tissue adjacent to tumor of the two histologic subtypes seem to be more suitable.

It is known that TGCT and infertility problems are associated (32). Folate-mediated one carbon metabolism participates in the complex mechanism of spermatogenesis, and frequent inactivation of *MTHFR* by epigenetic silencing in sperm DNA of idiopathic infertile men has been reported (33). Moreover the minor allele A of rs1801133, that we found associated with an increased risk of TGCT in our study, was strongly associated with male infertility risk in populations of Asian and European ancestry, and described as more frequent in men with decreased sperm count (34,35). Further experimental studies are necessary to evaluate if common *MTHFR* variants might have a role in the failure of gonocyte maturation during fetal life and, in general, in TGCT development.

The other genes that showed an association with TGCT risk in gene-based analysis are less well studied, and little is known about their involvement in cancer predisposition. Alterations in *MECP2* (methyl-CpG-binding-protein 2) sequence have been related to congenital diseases and cancer (36). According to functional assessment, the top *MECP2* SNPs associated with TGCT were predicted to alter regulatory motifs, suggesting they could influence *MECP2* expression.

The main strength of this study is its very large sample size (for TGCT, a relatively rare malignancy), combined with a pre-selected panel of genes and a gene-based analysis with a specific focus on the methylation machinery. TECAC, by pooling the efforts and resources of all its members, made it possible to analyze genome-wide data on more than 10,000 cases, which represents a crucial advantage, since TGCT has a significant heritable basis due to multiple minor genetic factors. Another strength is the simultaneous modelling of the collective effect of multiple genetic variants within the same gene, as individual effects could be too weak effects to be detected.

A limitation of our approach is that the analyses were restricted only to genes known to be implicated in DNA methylation processes. It is known that epigenetic reprogramming is a very complex process involving other genes, such as those implicated in DNA repair, histone modifications and chromatin remodeling, or in microRNA biosynthesis and regulation. Additional

studies are required, since the comprehensive examination of the association between genetic variants of the whole epigenetic machinery and TGCT risk is of interest, but outside the scope of this study.

In conclusion, in a large pathway-focused meta-analysis we found that common polymorphisms in *MTHFR*, some of which appear to be functional, are associated with TGCT risk. This finding may contribute to support a potential involvement of epigenetic mechanisms in the pathogenesis of TGCT.

Material and Methods

Study population

The TECAC Consortium assembled multiple TGCT case-control studies conducted by more than 20 institutions from Europe and North America (8). All studies involved in the Consortium have collected blood or saliva samples, from which DNA has been extracted, and a selection of phenotype and questionnaire data on potential TGCT risk factors.

Data from eight sources were obtained by TECAC: (i-v) summary statistics from 5 independently conducted GWASes on TGCT (37-41) and previously published as a meta-analysis (42); (vi) individual level genotype data from the Replication study involving 14 case-control studies conducted by the TECAC institution members in Europe and the United States, with genotyping centrally conducted at the Center for Applied Genomics at the Children's Hospital of Philadelphia (13 studies) or MD Anderson Cancer Center (one study) using the Illumina Human Core array technology (8); (vii) the deCODE genetics company (https://www.decode.com/; 43) study in Iceland; and (viii) the UK biobank study (https://www.ukbiobank.ac.uk/; 44). These studies were described in detail elsewhere (8).
In total, the Consortium assembled 10,156 cases and 179,683 controls (Figure 1 and Table 1).

For most of these studies, information was available on the histological subtype classified as pure seminoma and non-seminomatous tumors (the latter including TGCTs with mixed histology), family history of TGCT, history of cryptorchidism and other selected key characteristics.

The current study was carried out on summary statistics data from the meta-analysis of the eight sources performed by the TECAC Consortium (8). Data from participants in each contributing study were collected and analyzed in accordance with the local ethical permissions and informed consent.

Selection of genes and SNPs

To obtain a list of DNA methylation machinery and one-carbon metabolism genes, we conducted a search in public pathway catalogues in 2014, including BioCarta (http://www.biocarta.com/genes/index.asp), Reactome (http://www.reactome.org), KEGG pathway (http://www.genome.jp/kegg/kegg2.html) and NCI-PID (http://pid.nci.nih.gov/index.shtml) using the following queries: “DNA methylation”, “DNA methylation pathway”, “mechanisms of transcriptional repression by DNA methylation”, “epigenetic regulation of gene expression”, “folate cycle”, “one-carbon metabolism”, and “one carbon pool by folate”. We identified a preliminary list of protein-coding genes from these pathways and checked the function of each gene manually using the public databases GENEcards (https://www.genecards.org/) and UniProtKB (http://www.uniprot.org/help/uniprotkb), keeping in the final list only genes strictly involved in the DNA methylation process.

We identified 28 DNA methylation pathway genes (Supplementary Table S3), classified into two groups based on the molecular mechanism in which they are involved: one-carbon metabolism (N=11 genes) and DNA methylation machinery (N=17 genes), the latter further classified in the following three subgroups: i) “writers” (N=4 genes), ii) “erasers” (N=4 genes), and iii) “readers” (N=9 genes). No significant changes to this selection were identified in a quick search carried out in 2021 in the Biocarta and the Reactome pathway databases.

For each gene, we selected a list of tag SNPs using Haploview 4.2 software, implemented with the Tagger pairwise method (Broad Institute, Cambridge, MA) applied to genotype data of the public database of the International HapMap Project (45). We used the phased genotype data (Human Genome Build 37p13) from the CEU (Utah Residents with Northern and Western European Ancestry) population, the population sample that most closely resembles the subjects used in this study. We selected tag SNPs with the following characteristics: minor allele frequency (MAF) of ≥5% to select only common variants inpersons of European ancestry, and an r^2 of 0.8 as the LD threshold. To include the 5′- and 3′-untranslated regulatory regions, tag SNP search was expanded by10 kb-upstream and 10 kb-downstream of each gene sequence, as predicted clusters of transcription factor binding sites are most enriched in these sequences (46). Moreover, potential functional SNPs were included by searching in public databases, including Ensembl (https://www.ensembl.org), SNPedia (https://www.snpedia.com/index.php/SNPedia), and PubMed (https://www.ncbi.nlm.nih.gov/pubmed/).

In total, 273 polymorphisms were selected for the current study. The SNPs were included as part of the custom content on the Illumina Human Bead Core array. The complete list of the candidate genes, together with their annotations, and the number of tag SNPs selected for each gene, is provided in Supplementary Table S3.

Individual-SNP analysis

Summary statistics of the association analysis of the selected tag SNPs and the risk of TGCT were provided by the TECAC Consortium. The estimates of the fixed-effect meta-analysis (overall summary p-values, odds ratios (ORs) and corresponding 95% confidence intervals (95% CI)) were obtained as previously described (8).

Four out of the 273 selected tag SNPs were neither genotyped nor imputed in any of the individual studies. We included only polymorphisms with available summary statistics from at least two of the eight studies, leading to the exclusion from the analysis of two other polymorphisms, one in *MBD4* and one in *DNMT3L*, leaving a total of 267 tag SNPs in 28 genes for the final analytic data set (Supplementary Table S3).

We conducted stratified analyses for seminoma and non-seminomatous tumors in all studies except deCODE (which includes 3% of the total number of cases and 84.6% of the total number of controls); analyses restricted to the subgroup of cases with positive TGCT family history, or positive history of cryptorchidism were carried out on cases and controls of the NCI, UPENN and UK studies, and on a sub-set of the Replication study for which this information were available.

Association p-values were adjusted for multiple comparisons using the Benjamini-Hochberg FDR method (47).

Gene-based analysis

Gene-based analysis was carried out using MAGMA (Multi-marker Analysis of GenoMic Annotation) v1.07b, which combines the individual-SNP p-values to test the collective effect of multiple markers from a gene by properly incorporating LD between markers (48). In MAGMA, two types of gene test statistics are implemented. The SNP-wise Mean model is more attuned to the mean SNP association, though it is biased towards association in areas of higher gene LD. The SNP-wise Top model is more sensitive when only a small proportion of the analyzed SNPs in a gene show an association (48). We preferred this second approach and calculated a permutation-based p-value for each gene.

Analyses on MAGMA were conducted using the summary p-values for the associations between the tag SNPs and TGCT, and 100,000 permutations were computed for each gene. The European ancestry population from the 1000 Genomes Project Phase 3 (Build 37/European data only) was taken as the reference for LD patterns.

MAGMA analyses were stratified by histological subtypes as in the individual-SNP analyses, and further restricted to cases with a TGCT family history or a history of cryptorchidism. The Benjamini-Hochberg FDR method was used to adjust for multiple comparisons (47).

Functional assessment of SNPs and gene expression analysis in TGCT subtypes

The dbSNP database (https://www.ncbi.nlm.nih.gov/snp/) was interrogated to explore the potential functional consequences of the selected SNPs on gene expression and regulation, and on amino acid change (49). HaploReg v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) was used to evaluate their possible effects on protein binding sites and regulatory motifs (50). SNPnexus web server (https://www.snp-nexus.org/) was interrogated to predict the possible functional impact of each SNP at transcriptome and proteome levels and on regulatory elements (51). From SNiPA (https://snipa.helmholtz-muenchen.de/snipa3/) (52) we retrieved information on possible clinical significance and previously reported associations with other traits and human diseases. SNiPA also was applied, drawing on 1000 Genomes Project Phase 1 v.3 and Phase 3 v.5 data, to define the size of LD block spanning each SNP and to identify any proxy variants in high LD (r^2>0.8). GTEx v7 (http://www.gtexportal.org/) was explored to predict the possible association with expression quantitative trait loci of each tag SNP and of each SNP in high LD with the tags in a sample of 322 normal adult testis tissues with donor genotypes available (53).

We analyzed publicly available gene expression datasets for genes showing different association patterns between seminoma and non-seminomatous tumors. Expression data from 43 seminoma and 68 non-seminomatous tumors were downloaded from the cBioPortal for Cancer Genomics (https://www.cbioportal.org/) (54,55). We used the mRNA expression z-scores relative to diploid samples (RNA Seq V2 RSEM) from the TGCA PanCancer Atlas dataset. Gene expression between the two histologic groups was compared using Wilcoxon-Mann-Whitney tests. Samples with z-scores above 2 and below -2 were excluded from the analysis.

Acknowledgements

We thank the participants in the testicular cancer germ cell studies around the world that contributed to this study.

A.F. would like to thank Dr. Maria Santa Rocca for technical assistance. J.A.G. would like to thank Nynke Zwart and Gerrie Steursma for their contributions to the study. J.N. would like to thank Louise Parkinson for coordinating the recruitment. R.A.L. and R.I.S. are grateful to Anne C. Bakken, Sigrid M. Kraggerud and Sophie D. Fosså for their contributions. D.J.V. would like to thank Dr. Linda Jacobs and Ms. Donna Pucci for their contributions to the local Penn TGCT study. K.A. would like to thank Drs Ewa Rajpert-De Meyts and Ramneek Gupta for their contribution to the local study. T.G., T.B.H., R.K. and F.W. would like to thank Wenche Kristiansen for her contribution to the Norwegian/Swedish study.

The Testicular Cancer Consortium is comprised of investigator teams from around the world with the common aim to pool resources of all GWAS to identify new genetic markers of risk, further characterize the existing genomic regions with risk markers, examine potential maternal or parent-of-origin effects though case-parent triads and dyads and finally create a Consortium that formalizes the existing research alliances.

Conflicts of Interest Statement

The authors declare no competing interests.

Funding

This work was supported by: the National Institutes of Health [U01CA164947 to K.L.N., P.A.K. and S.M.S., CA114478 to K.L.N. and P.A.K., R01 CA104786 to T.Z.]; the Abramson Cancer Center at the University of Pennsylvania [P30 CA016520 to K.L.N. and P.A.K.]; the Villum Kann Rasmussen Foundation, the Danish Strategic Research Council [NABIIT grant], the Novo Nordisk Foundation, the Danish Cancer Society, the Danish Childhood Cancer Foundation [to K.A.]; the Cancer Research UK Programme [Award C588/A19167]; the Institute of Cancer Research, the Cancer Research UK and the Wellcome Trust Case Control Consortium (WTCCC) [to D.T.B., J.N., C.L., and C.T.]; the National Cancer Institute [R01CA085914, contracts CN-67009 and PC-35142, to S.M.S., R01 CA102042 to V.K.C.]; the California Cancer Research Program [03-00174-30021, 53299-0050-V-10260]; the Robert E. and May R. Wright Foundation and the Whittier Foundation [to V.K.C.]; the Intramural Research Program of the National Cancer Institute [ZO1-CP-10144: "Clinical/Genetic Studies of Familial and Hereditary Cancer Syndromes" to M.H.G. and K.A.M.]; the IMS, Inc support services contract [HHSN26120130003C to M.H.G. and K.A.M.]; the Norwegian Cancer Society [418975 – 71081 – PR-2006-0387, PK01-2007-0375]; the Nordic Cancer Union [S-53712/07 to T.G., T.B.H. and F.W]; the Swedish Cancer Society [2008/708, 2010/808, CAN2011/484, CAN2012/823 to T.G., T.B.H. and F.W]; the Piedmont Region [to L.R.]; the Italian Ministry for Education, University and Research under the programme "Dipartimenti di Eccellenza 2018–2022"

[D15D18000410001 to L.R.]; Dell'Elce Family Fund and Princess Margaret Cancer Foundation [to R.J.H.]; the Deutsche Krebshilfe grant [70113348 to D.L.]; the Norwegian Cancer Society and the South-Eastern Norway Regional Health Authority [to R.A.L. and R.I.S.]; the Fred Hutchinson Cancer Research Center institutional funds [to S.M.S.]; the Movember foundation [to C.T.]; the Fondazione Umberto Veronesi [for a postdoctoral fellowship to C.G.].

Testicular Cancer Consortium

Kristian Almstrup[3,4], Victoria K. Cortessis[9], Alberto Ferlin[12]; Jourik A. Gietema[15], Anna Gonzalez-Niera[35], Mark H. Greene[7], Tom Grotmol[16], Robert J. Hamilton[4],Peter A. Kanetsky[34], Lambertus A. Kiemeney[20], Davor Lessel[21], Katherine A. McGlynn[7], Katherine L. Nathanson[2,31], Kevin T. Nead[26], Jérémie Nsengimana[27], Thorunn Rafnar[36], Lorenzo Richiardi[1], Stephen M. Schwartz[8], Rolf I. Skotheim[23,28], Clare Turnbull[29], Fredrik Wiklund[19],Tongzhang Zheng[32].

[35]Human Genotyping Core Unit, Spanish National Cancer Centre (CNIO), Madrid, 28029, Spain.
[36]deCODE Genetics, Reykjavik, Iceland, Reykjavik, 102, Iceland.

References

1. Gurney, J.K., Florio, A.A., Znaor, A., Ferlay, J., Laversanne, M., Sarfati, D., Bray, F. and McGlynn, K.A. (2019) International trends in the incidence of testicular cancer: lessons from 35 years and 41 countries. *Eur. Urol.*, **76**, 615-623.

2. Znaor, A., Skakkebaek, N.E., Rajpert-De Meyts, E., Laversanne, M., Kuliš, T., Gurney, J., Sarfati, D., McGlynn, K.A. and Bray, F. (2020) Testicular cancer incidence predictions in Europe 2010-2035: A rising burden despite population ageing. *Int. J. Cancer*, **147**, 820-828.

3. Shen, H., Shih, J., Hollern, D.P., Wang, L., Bowlby, R., Tickoo, S.K., Thorsson, V., Mungall, A.J., Newton, Y., Hegde, A.M. *et al.* (2018) Integrated molecular characterization of testicular germ cell tumors. *Cell. Rep.*, **23**, 3392-3406.

4. Akan, S., Ediz, C., Tavukcu, H.H., Ozgun, A. and Yilmaz O. (2020) The clinical significance of seminoma component in testicular mixed germ cell tumour. *Urol. Int.*, **104**, 489-496.

5. Baird, D.C., Meyers, G.J. and Hu, J.S. (2018) Testicular cancer: diagnosis and treatment. *Am. Fam. Physician*, **97**, 261-268.

6. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Möller, S., Unger, R.H. *et al.* (2016) Familial risk and heritability of cancer among twins in nordic countries. *JAMA*, **315**, 68-76.

7. Litchfield, K., Loveday, C., Levy, M., Dudakia, D., Rapley, E., Nsengimana, J., Bishop, D.T., Reid, A., Huddart, R., Broderick, P. *et al.* (2018) Large-scale sequencing of testicular germ cell tumour (TGCT) cases excludes major TGCT predisposition gene. *Eur. Urol.*, **73**, 828-831.

8. Pluta, J., Pyle, L.C., Nead, K.T., Wilf, R., Li, M., Mitra, N., Weathers, B., D'Andrea, K., Almstrup, K., Anson-Cartwright, L. *et al.* (2021) Identification of 22 novel susceptibility loci associated with testicular germ cell tumors. *Nat. Commun.*, **12**, 4487.

9. Almstrup, K., Hoei-Hansen, C.E., Wirkner, U., Blake, J., Schwager, C., Ansorge, W., Nielsen, J.E., Skakkebaek, N.E., Rajpert-De Meyts, E. and Leffers, H. (2004) Embryonic stem cell-like features of testicular carcinoma in situ revealed by genome-wide gene expression profiling. *Cancer Res.*, **64**, 4736-4743.

10. Rajpert-De Meyts, E., McGlynn, K.A., Okamoto, K., Jewett, M.A. and Bokemeyer, C. (2016) Testicular germ cell tumours. *Lancet*, **387**, 1762-1774.

11. Batool, A., Karimi, N., Wu, X.N., Chen, S.R. and Liu, Y.X. (2019) Testicular germ cell tumor: a comprehensive review. *Cell. Mol. Life Sci.*, **76**, 1713-1727.

12. Stewart, K.R., Veselovska, L. and Kelsey, G. (2016) Establishment and functions of DNA methylation in the germline. *Epigenomics*, **8**, 1399-1413.

13. Netto, G.J., Nakai, Y., Nakayama, M., Jadallah, S., Toubaji, A., Nonomura, N., Albadine, R., Hicks, J.L., Epstein, J.I., Yegnasubramanian, S. *et al.* (2008) Global DNA hypomethylation in intratubular germ cell neoplasia and seminoma, but not in nonseminomatous male germ cell tumors. *Mod. Pathol.*, **21**, 1337-1344.

14. Kristensen, D.G., Nielsen, J.E., Jørgensen, A., Skakkebæk, N.E., Rajpert-De Meyts, E. and Almstrup, K. (2014) Evidence that active demethylation mechanisms maintain the genome of carcinoma in situ cells hypomethylated in the adult testis. *Br. J. Cancer*, **110**, 668-678.

15. Almstrup, K., Nielsen, J.E., Mlynarska, O., Jansen, M.T., Jørgensen, A., Skakkebæk, N.E. and Rajpert-De Meyts, E. (2010) Carcinoma in situ testis displays permissive chromatin modifications similar to immature foetal germ cells. *Br. J. Cancer*, **103**, 1269-1276.

16. Kristensen, D.G., Skakkebæk, N.E., Rajpert-De Meyts, E. and Almstrup, K. (2013) Epigenetic features of testicular germ cell tumours in relation to epigenetic characteristics of foetal germ cells. *Int. J. Dev. Biol.*, **57**, 309-317.

17. Mirabello, L., Kratz, C.P., Savage, S.A. and Greene, M.H. (2012) Promoter methylation of candidate genes associated with familial testicular cancer. *Int. J. Mol. Epidemiol. Genet*., **3**, 213-227.

18. Pedersen, L.H., Nielsen, J.E., Daugaard, G., Hansen, T.V., Rajpert-De Meyts, E. and Almstrup, K. (2016) Differences in global DNA methylation of testicular seminoma are not associated with changes in histone modifications, clinical prognosis, BRAF mutations or gene expression. *Cancer Genet.*, **209**, 506-514.

19. Weissman, J., Naidu, S. and Bjornsson, H.T. (2014) Abnormalities of the DNA methylation mark and its machinery: an emerging cause of neurologic dysfunction. *Semin. Neurol.*, **34**, 249-257.

20. Friso, S., Udali, S., De Santis, D. and Choi, S.W. (2017) One-carbon metabolism and epigenetics. *Mol. Aspects Med.*, **54**, 28-36.

21. Stover, P.J. (2011) Polymorphisms in 1-carbon metabolism, epigenetics and folate-related pathologies. *J. Nutrigenet. Nutrigenomics*, **4**, 293-305.

22. Murgatroyd, C. and Spengler, D. (2012) Genetic variation in the epigenetic machinery and mental health. *Curr. Psychiatry Rep.*, **14**, 138-149.

23. Inoue-Choi, M., Nelson, H.H., Robien, K., Arning, E., Bottiglieri, T., Koh, W.P. and Yuan, J.M. (2013) Plasma S-adenosylmethionine, DNMT polymorphisms, and peripheral blood LINE-1 methylation among healthy Chinese adults in Singapore. *BMC Cancer*, **13**, 389.

24. Pontén, F., Jirström, K. and Uhlen, M. (2008) The Human Protein Atlas-a tool for pathology. *J. Pathol.*, **216**, 387-393.

25. Garner, J.L., Niles, K.M., McGraw, S., Yeh, J.R., Cushnie, D.W., Hermo, L., Nagano, M.C. and Trasler, J.M. (2013) Stability of DNA methylation patterns in mouse spermatogonia under conditions of MTHFR deficiency and methionine supplementation. *Biol. Reprod.*, **89**, 125.

26. Li, L., Dong, J., Yan, L., Yong, J., Liu, X., Hu, Y., Fan, X., Wu, X., Guo, H., Wang, X. *et al.* (2017) Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell*, **20**, 858-873.

27. Levin, B.L. and Varga, E. (2016) MTHFR: addressing genetic counseling dilemmas using evidence-based literature. *J. Genet. Couns.*, **25**, 901-911.

28. Trabert, B., Zugna, D., Richiardi, L., McGlynn, K.A. and Akre, O. (2013) Congenital malformations and testicular germ cell tumors. *Int. J. Cancer*, **133**, 1900-1904.

29. Colson, N.J., Naug, H.L., Nikbakht, E., Zhang, P. and McCormack, J. (2017) The impact of MTHFR 677 C/T genotypes on folate status markers: a meta-analysis of folic acid intervention studies. *Eur. J. Nutr.*, **56**, 247-260.

30. Zhang, D., Wen, X., Wu, W., Guo, Y. and Cui, W. (2015) Elevated homocysteine level and folate deficiency associated with increased overall risk of carcinogenesis: meta-analysis of 83 case-control studies involving 35,758 individuals. *PLoS One*, **10**, e0123423.

31. Castro, R., Rivera, I., Ravasco, P., Camilo, M.E., Jakobs, C., Blom, H.J. and de Almeida, I.T. (2004) 5,10-methylenetetrahydrofolate reductase (MTHFR) 677C-->T and 1298A-->C mutations are associated with DNA hypomethylation. *J. Med. Genet.*, **41**, 454-458.

32. Del Giudice, F., Kasman, A.M., De Berardinis, E., Busetto, G.M., Belladelli, F. and Eisenberg, M.L. (2020) Association between male infertility and male-specific malignancies: systematic

review and meta-analysis of population-based retrospective cohort studies. *Fertil. Steril.*, **114**, 984-996.

33. Wu, W., Shen, O., Qin, Y., Niu, X., Lu, C., Xia, Y., Song, L., Wang, S. and Wang, X. (2010) Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolatereductase (MTHFR). *PLoS One*, **5**, e13884.

34. Gong, M., Dong, W., He, T., Shi, Z., Huang, G., Ren, R., Huang, S., Qiu, S. and Yuan, R. (2015) MTHFR 677C>T polymorphism increases the male infertility risk: a meta-analysis involving 26 studies. *PLoS One*, **10**, e0121147.

35. Liu, K., Zhao, R., Shen, M., Ye, J., Li, X., Huang, Y., Hua, L., Wang, Z., Li, J. (2015) Role of genetic mutations in folate-related enzyme genes on male infertility. *Sci. Rep*., **5**, 15548.

36. Ezeonwuka, C.D. and Rastegar, M. (2014) MeCP2-related diseases and animal models. *Diseases*, **2**, 45-70.

37. Kanetsky, P.A., Mitra, N., Vardhanabhuti, S., Vaughn, D.J., Li, M., Ciosek, S.L., Letrero, R., D'Andrea, K., Vaddi, M., Doody, D.R. *et al.* (2011) A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. *Hum. Mol. Genet.*, **20**, 3109-3117.

38. Dalgaard, M.D., Weinhold, N., Edsgärd, D., Silver, J.D., Pers, T.H., Nielsen, J.E., Jørgensen, N., Juul, A., Gerds, T.A., Giwercman, A. *et al.* (2012) A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation. *J. Med. Genet.*, **49**, 58-65.

39. Schumacher, F.R., Wang, Z., Skotheim, R.I., Koster, R., Chung, C.C., Hildebrandt, M.A., Kratz, C.P., Bakken, A.C., Bishop, D.T., Cook, M.B. *et al.* (2013) Testicular germ cell tumor susceptibility associated with the UCK2 locus on chromosome 1q23. *Hum. Mol. Genet.*, **22**, 2748-2753.

40. Litchfield, K., Holroyd, A., Lloyd, A., Broderick, P., Nsengimana, J., Eeles, R., Easton, D.F., Dudakia, D., Bishop, D.T., Reid, A. *et al.* (2015) Identification of four new susceptibility loci for testicular germ cell tumour. *Nat. Commun.*, **6**, 8690.

41. Kristiansen, W., Karlsson, R., Rounge, T.B., Whitington, T., Andreassen, B.K., Magnusson, P.K., Fosså, S.D., Adami, H.O., Turnbull, C., Haugen, T.B. *et al.* (2015) 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**, 4138-4146.

42. Wang, Z., McGlynn, K.A., Rajpert-De Meyts, E., Bishop, D.T., Chung, C.C., Dalgaard, M.D., Greene, M.H., Gupta, R., Grotmol, T., Haugen, T.B. *et al.* (2017) Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor. *Nat. Genet.*, **49**, 1141-1147.

43. Gudbjartsson, D.F., Helgason, H., Gudjonsson, S.A., Zink, F., Oddson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B.V., Hjartarson, E. *et al.* (2015) Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.*, **47**, 435-444.

44. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203-209.

45. Barrett, J.C. (2009) Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harb. Protoc.*, **4**.

46. Ouyang, C., Smith, D.D. and Krontiris, T.G. (2008) Evolutionary signatures of common human cis-regulatory haplotypes. *PLoS One*, **3**, e3362.

47. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, **57**, 289-300.

48. de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PloS Comput. Biol.*, **11**, e1004219

49. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308-311.

50. Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930-D934.

51. Dayem Ullah, A.Z., Oscanoa, J., Wang, J., Nagano, A., Lemoine, N.R. and Chelala, C. (2018) SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res.*, **46**, W109-W113.

52. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. and Kastenmüller, G. (2015) SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics*, **31**, 1334-1336.

53. GTEx Consortium. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580-585.

54. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401-404.

55. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal*., **6**, pl1.

Legends to Figures

Figure 1. Flow chart of cases and controls from the eight studies involved in the main analysis

716**Tables**

717**Table 1.** Number of TGCT cases and controls included for testing associations with SNPs in 718methylation-related genes, by originating study, and for cases by histologic type, family history of 719TGCT, and history of cryptorchidism

TECAC Study	TGCT Cases					Controls (N)
	ALL (N)	Seminoma histology (N)	Non-seminomatous histology (N)	Family history (N)	Cryptorchidism (N)	
GWAS-DENMARK	183	88	55	na	na	363
GWAS-NCI	581	243	334	76	131	1,056
GWAS-UPENN	481	171	299	49	39	919
GWAS-NORWAY/ SWEDEN	1,326	766	549	na	na	6,687
GWAS-UK	986	410	410	136	56	4,945
REPLICATION STUDY	5,602	2,456	2,760	95	295	5,006
deCODE ICELAND	300	na	na	na	na	151,991
UK BIOBANK	697	395	223	na	na	8,716
Total	**10,156**	**4,529**	**4,630**	**356**	**521**	**179,683**

720
721GWAS: genome-wide association study
722na: information not available by TECAC study: cases not included in stratified/restricted analysis

Table 2. Individual SNP association results for the whole dataset

SNP ID	GENE; location	Allele1/ Allele2[$]	Allele2 frequency	q-value	I^2	p-het[#]	Direction*	OR (95% CI)[§]
rs1801133	MTHFR; Exon #4	A/G	0.66	$3.6x10^{-4}$	8.6	0.36	------+-	0.90 (0.87-0.94)
rs1734791	MECP2; Intronic	A/T	0.15	$7.8x10^{-3}$	0	0.99	+++++?++	1.09 (1.05-1.14)
rs12121543	MTHFR; Intronic	A/C	0.75	0.02	0	0.94	+?++++++	1.09 (1.04-1.14)
rs1476413	MTHFR; Intronic	T/C	0.73	0.02	0	0.99	++++++++	1.08 (1.03-1.13)
rs4796420	ZBTB4; Intronic	A/T	0.79	0.02	71.7	8×10^{-4}	+-+0++-+	1.09 (1.04-1.14)
rs1624766	MECP2; Intronic	T/C	0.20	0.02	0	0.84	+++++?++	1.07 (1.03-1.12)
rs13306556	MTHFR; Intronic	T/C	0.66	0.05	0	0.98	++++++++	1.11 (1.04-1.19)

[$]Allele1: Reference allele; Allele2: Effect allele

[#]P for heterogeneity test

*Summary of effect directions of the single studies of the meta-analysis. "+" indicates a positive (increased) effect of the alternative allele on risk of TGCT, while "-" indicates a negative (decreased) effect of the alternative allele on risk of TGCT. "0" indicate null effect and "?" indicates missing effect. Study order: Replication, deCODE, UK, NCI, Denmark, Norway/Sweden, UPENN, UK biobank

[§]OR: Odds Ratio; CI: Confidence Interval

735**Table 3.** Genes associated with TGCT risk based on analysis of all SNPs in each gene
736

GENE	N SNPs*	perm-p#	q-value§
MTHFR	13	3×10^{-5}	8.4×10^{-4}
MECP2	4	1.4×10^{-4}	2×10^{-3}
ZBTB4	7	3.2×10^{-3}	0.03
AHCY	4	0.05	0.35
MBD3L1	5	0.07	0.35
SHMT1	8	0.1	0.35
MAT1A	14	0.1	0.35
DNMT3L	16	0.1	0.35
DNMT1	8	0.17	0.42
MAT2B	9	0.18	0.42
DNMT3B	8	0.19	0.42
ZBTB38	4	0.19	0.42
UHRF1	13	0.20	0.42
MTRR	25	0.25	0.46
TET2	9	0.26	0.46
MBD2	10	0.27	0.46
CBS	17	0.30	0.50
TET3	11	0.44	0.69
MBD3	4	0.53	0.75
BHMT	13	0.53	0.75
DNMT3A	18	0.56	0.75
MAT2A	4	0.59	0.75
TET1	7	0.65	0.76
MBD2	7	0.65	0.76
CTCF	3	0.76	0.85
MTR	8	0.81	0.87
MBD4	3	0.94	0.96
GADD45b	8	0.96	0.96

737
738*Number of SNPs tested within a gene
739#Gene level p-value computed by MAGMA after 100,000 permutations
740§Gene level q-value calculated on permutation p-value

741**Table 4.** Gene-based analysis stratified by histologic subtype
742

Seminoma cases (N=4,529) vs. controls (N=27,693)			**Non-seminomatous cases (N=4,630) vs. controls (N=27,693)**		
GENE	**perm-p[#]**	**q-value[§]**	**GENE**	**perm-p[#]**	**q-value[§]**
MTHFR	1×10^{-5}	2.8×10^{-4}	**DNMT1**	7.9×10^{-3}	0.22
AHCY	0.02	0.2	**MTRR**	0.02	0.35
DNMT3L	0.02	0.2	**MBD3**	0.05	0.48
SHMT1	0.05	0.34	**MBD3L1**	0.09	0.64
ZBTB38	0.06	0.36	**MTHFR**	0.2	0.70
ZBTB4	0.07	0.40	**ZBTB4**	0.26	0.70
MAT1A	0.10	0.40	**DNMT3B**	0.26	0.70
MBD3L1	0.12	0.41	**MAT1A**	0.27	0.70
MECP2	0.13	0.41	**TET1**	0.28	0.70
CBS	0.20	0.55	**CBS**	0.30	0.70
MAT2A	0.27	0.61	**BHMT**	0.32	0.70
MTRR	0.28	0.61	**MECP2**	0.37	0.70
TET1	0.28	0.61	**UHRF1**	0.38	0.70
MBD1	0.38	0.76	**MBD1**	0.39	0.70
DNMT3B	0.41	0.77	**MAT2B**	0.39	0.70
UHRF1	0.46	0.77	**AHCY**	0.40	0.70
MBD3	0.47	0.77	**SHMT1**	0.48	0.75
MAT2B	0.51	0.79	**DNMT3L**	0.48	0.75
TET3	0.56	0.83	**TET3**	0.57	0.84
MBD2	0.61	0.84	**MTR**	0.61	0.84
DNMT1	0.67	0.84	**TET2**	0.63	0.84
BHMT	0.68	0.84	**DNMT3A**	0.68	0.87
TET2	0.69	0.84	**ZBTB38**	0.72	0.88
GADD45b	0.76	0.86	**CTCF**	0.78	0.91
DNMT3A	0.77	0.86	**MAT2A**	0.85	0.95
MBD4	0.88	0.94	**MBD2**	0.95	0.99
MTR	0.90	0.94	**MBD4**	0.95	0.99
CTCF	0.96	0.96	**GADD45b**	0.99	0.99

743
744[#]Gene level p-value computed by MAGMA after 100,000 permutations
745[§]Gene level q-value on permutation p-value

746**Table 5**. Functional annotation of tag SNPs in MTHFR, MECP2 and ZBTB4 associated with risk of TGCT identified in the individual SNP analyis

	SNP ID (GENE)						
Characteristic	**rs1801133 (MTHFR)**	**rs12121543 (MTHFR)**	**rs1476413 (MTHFR)**	**rs13306556 (MTHFR)**	**rs1734791 (MECP2)**	**rs1624766 (MECP2)**	**rs4796420 (ZBTB4)**
Consequence	Coding, missense	Intron variant	Intron variant	Intron variant	Intron variant	Intron variant	Intron variant
Amino acid change	Ala222Val	None	None	None	None	None	None
Proteins bound	CEBPB, HDAC8, POL2	na	CEBPB	na	na	ZNF263	GATA2, POL24H8, TAL1, POL2
Motifs changed	Cphx	STAT	na	PLAG1	DMRT1, GATA, HDAC2	Arid5a, Foxj2	HNF4, Pax-4
Sift Prediction	damaging, high confidence	na	na	na	na	na	na
PolyPhen Prediction	probably damaging	na	na	na	na	na	na
Variant annotation	Methotrexate response - dosage, efficacy, toxicity/adverse drug reactions (adr); Carboplatin response – efficacy; Cyclophosphamide response - toxicity/adr; Gastrointestinal stroma tumor; MTHFR deficiency, thermolabile type	na	na	na	na	na	na
Variant association (trait/p-value)	Homocysteine levels[#] p-value<4×10^{-104}; <8×10^{-35}; <1×10^{-19} Red cell distribution width p-value<1×10^{-23} Serum folate level[§] p-value<4×10^{-19}; <3×10^{-11} High altitude adaptation p-value<6×10^{-9}	na*	Coronary artery disease p-value=2.28×10^{-5}	Diastolic blood pressure via alcohol consumption interaction p-value<3×10^{-9}	na	na	Educational attainment p-value<2×10^{-8} Lung function p-value <4×10^{-16}
LD block size	1 bp	3,669 bp*	57,089 bp	67,385 bp	160,011 bp	168,512 bp	85,798 bp
Proxy SNPs in high LD (r^2>0.8)	1 variant	3 variants*	3 variants	105 variants	24 variants	52 variants	60 variants
Association with eQTLs in testis tissue	na	p-value=$2.3\text{x}10^{-8}$	p-value=$5.9\text{x}10^{-11}$	na	na	na	na
Association of high LD-SNPs (r^2>0.8) with eQTLs in Testis	na	rs3818762 (r^2=0.81) p-value=$4.4\text{x}10^{-11}$	rs1023252 (r^2=0.84) p-value=$2.8\text{x}10^{-11}$	na	na	na	na

747na: not available; LD: linkage disequilibrium; eQTLs: expression quantitative trait loci

748*For functional analysis of rs12121543 on SNiPA tool, 1000 Genome Project Phase 1 v.3 data were used

749[#]Three independent studies; [§]two independent studies

750

Abbreviations

TGCT: testicular germ cell tumor

SNP: single nucleotide polymorphism

GWAS: genome-wide association study

GCNIS: germ-cell neoplasia in situ

TECAC: testicular cancer consortium

LD: linkage disequilibrium

MAGMA: multi-marker analysis of genomic annotation

FDR: false discovery rate

eQTL: expression quantitative trait locus

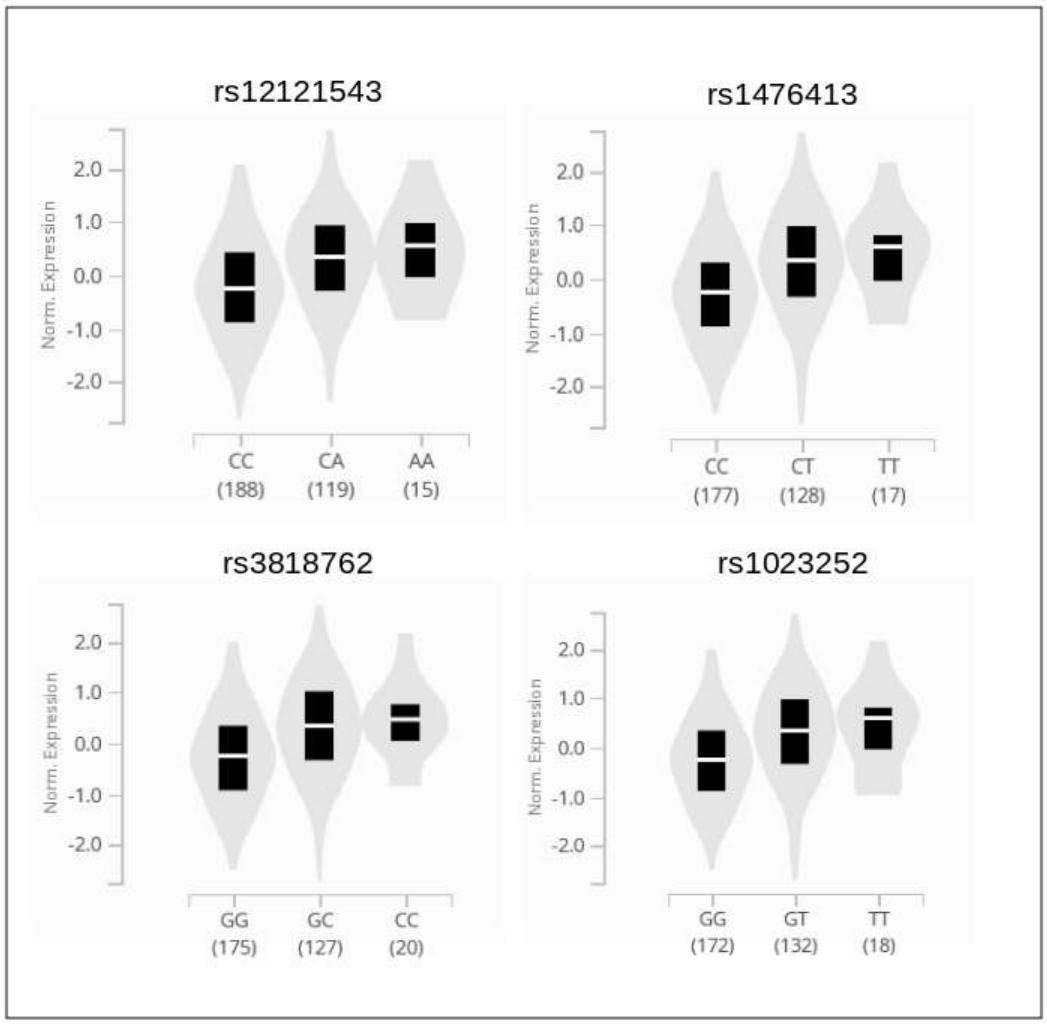

Figure S1. E-QTL violin-plots of *MTHFR* tag SNPs associated with risk of TGCT in meta-analysis and of other *MTHFR* SNPs in high LD. Upper panel: e-QTL violin-plots of SNPs rs12121543 (left) and rs1476413 (right) showing *MTHFR* expression level in human adult testis tissue according to the distributions of the three genotypes: homozygous reference, heterozygous, and homozygous alternative allele. Lower panel: e-QTL violin-plots of SNPs in high LD with rs12121543 and rs1476413, respectively: rs3818762 (left) and rs1023252 (right). Data Source: GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2)