

Identification of genes associated with testicular germ cell tumor susceptibility through a transcriptome-wide association study

Authors

Emilio Ugalde-Morales, Rona Wilf, John Pluta, ..., Katherine L. Nathanson, Fredrik Wiklund, the Testicular Cancer Consortium

Correspondence

fredrik.wiklund@ki.se

The study integrates gene expression prediction models with a testicular germ cell tumor (TGCT) genome-wide association study, identifying 46 risk genes, including novel candidates. The findings enhance the understanding of TGCT genetics, emphasizing the need for further functional studies to inform diagnosis and treatment strategies.

Ugalde-Morales et al., 2025, *The American Journal of Human Genetics* 112, 630–643

March 6, 2025 © 2025 The Author(s). Published by Elsevier Inc. on behalf of American Society of Human Genetics.

<https://doi.org/10.1016/j.ajhg.2025.01.022>



Identification of genes associated with testicular germ cell tumor susceptibility through a transcriptome-wide association study

Emilio Ugalde-Morales,^{1,33} Rona Wilf,^{2,33} John Pluta,² Alexander Ploner,¹ Mengyao Fan,² Mohammad Damra,² Katja K. Aben,^{3,4} Lynn Anson-Cartwright,⁵ Chu Chen,^{6,7} Victoria K. Cortessis,⁸ Siamak Daneshmand,⁹ Alberto Ferlin,¹⁰ Marija Gamulin,¹¹ Jourik A. Gietema,¹² Anna Gonzalez-Niera,¹³ Tom Grotmol,¹⁴ Robert J. Hamilton,⁴ Mark Harland,⁵ Trine B. Haugen,¹⁵ Russ Hauser,¹⁶

(Author list continued on next page)

Summary

Transcriptome-wide association studies (TWASs) have the potential to identify susceptibility genes associated with testicular germ cell tumors (TGCTs). We conducted a comprehensive TGCT TWAS by integrating genome-wide association study (GWAS) summary data with predicted expression models from normal testis, TGCT tissues, and a cross-tissue panel that encompasses shared regulatory features across 22 normal tissues, including the testis. Gene associations were evaluated while accounting for variant-level effects from GWASs, followed by fine-mapping analyses in regions exhibiting multiple TWAS signals, and finally supplemented by colocalization analysis. Expression and protein patterns of identified TWAS genes were further examined in relevant tissues. Our analysis tested 19,805 gene-disease links, revealing 165 TGCT-associated genes with a false discovery rate of less than 0.01. We prioritized 46 candidate genes by considering GWAS-inflated signals, correlations between neighboring genes, and evidence of colocalization. Among these, 23 genes overlap with 22 GWAS loci, with 7 being associations not previously implicated in TGCT risk. Additionally, 23 genes located within 21 loci are at least 1 Mb away from published GWAS index variants. The 46 prioritized genes display expression levels consistent with expected expression levels in human gonadal cell types and precursor tumor cells and significant enrichment in TGCTs. Additionally, immunohistochemistry revealed protein-level accumulation of two candidate genes, *ARID3B* and *GINMI*, in both precursor and tumor cells. These findings enhance our understanding of the genetic predisposition to TGCTs and underscore the importance of further functional investigations into these candidate genes.

Introduction

Testicular cancer is the most common type of cancer among males aged 20–40 years in large parts of the world and the second most common type of cancer among males aged 15–19 years.^{1,2} Testicular germ cell tumors (TGCTs) are the main subtype, with rising global incidence over

the last two decades.³ TGCTs have a high heritability rate of 40%–50%,^{4–7} and first-degree male relatives of affected individuals have an 8- to 10-fold higher risk.⁸ TGCTs are thought to arise during fetal testis development, when a subset of primordial germ cells (PGCs) become arrested in an undifferentiated state, maintaining stem cell-like qualities that prevent them from differentiating into

¹Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; ²Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; ³Netherlands Comprehensive Cancer Organization, Radboud University Medical Center, Utrecht, the Netherlands; ⁴Radboud University Medical Center, Nijmegen, the Netherlands; ⁵Department of Surgery (Urology), University of Toronto and The Princess Margaret Cancer Centre, Toronto, ON, Canada; ⁶Epidemiology Program, Fred Hutchinson Cancer Center, Seattle, WA, USA; ⁷Department of Epidemiology, University of Washington, Seattle, WA, USA; ⁸Departments of Preventive Medicine and Obstetrics and Gynecology, Keck School of Medicine at the University of Southern California, Los Angeles, CA, USA; ⁹Departments of Urology, Keck School of Medicine at the University of Southern California, Los Angeles, CA, USA; ¹⁰Department of Medicine, University of Padova, Padua, Italy; ¹¹Department of Oncology, University Hospital Center Zagreb, University of Zagreb School of Medicine, Zagreb, Croatia; ¹²Department of Medical Oncology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands; ¹³Human Genotyping Core Unit, Spanish National Cancer Centre (CNIO), Madrid, Spain; ¹⁴Cancer Registry of Norway, Oslo Metropolitan University, Oslo, Norway; ¹⁵Department of Life Sciences and Health, Oslo Metropolitan University, Oslo, Norway; ¹⁶Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA; ¹⁷Department of Lymphoma and Myeloma, University of Texas, MD Anderson Cancer Center, Houston, TX, USA; ¹⁸Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; ¹⁹Institute of Human Genetics, University of Regensburg, Regensburg, Germany; ²⁰Institute of Clinical Human Genetics, University Hospital Regensburg, Regensburg, Germany; ²¹Department of Molecular Oncology, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway; ²²Division of Genetics & Epidemiology, The Institute of Cancer Research, London, UK; ²³William Harvey Research Institute, Queen Mary University, London, UK; ²⁴Biostatistics Research Group, Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle, UK; ²⁵Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Turin, Italy; ²⁶deCODE Genetics/Amgen, Reykjavik, Iceland; ²⁷Division of Hematology/Oncology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; ²⁸Department of Growth and Reproduction, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark; ²⁹Department of Epidemiology, Brown School of Public Health, Brown University, Providence, RI, USA; ³⁰Department of Cancer Epidemiology, Moffitt Cancer Center and Research Institute, Tampa, FL, USA; ³¹Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark; ³²Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

(Affiliations continued on next page)



Michelle A.T. Hildebrandt,¹⁷ Robert Karlsson,¹ Lambertus A. Kiemeny,⁴ Jung Kim,¹⁸ Davor Lessel,^{19,20} Ragnhild A. Lothe,²¹ Chey Loveday,^{22,23} Stephen J. Chanock,¹⁸ Katherine A. McGlynn,¹⁸ Coby Meijer,¹² Kevin T. Nead,¹⁷ Jeremie Nsengimana,²⁴ Maja Popovic,²⁵ Thorunn Rafnar,²⁶ Lorenzo Richiardi,²⁵ Maria S. Rocca,¹⁰ Stephen M. Schwartz,^{6,7} Rolf I. Skotheim,²¹ Kari Stefansson,²⁶ Douglas R. Stewart,¹⁸ Clare Turnbull,²² David J. Vaughn,²⁷ Sofia B. Winge,²⁸ Tongzhang Zheng,²⁹ Alvaro N. Monteiro,³⁰ Kristian Almstrup,^{28,31} Peter A. Kanetsky,³⁰ Katherine L. Nathanson,^{2,32} Fredrik Wiklund^{1,*} and the Testicular Cancer Consortium

spermatogonia.⁹ In this undifferentiated-arrested state, these precursor cells are known as germ cell neoplasia *in situ* (GCNIS). Transformation into TGCTs is thought to begin before puberty and testicular enlargement, when specific genomic alterations—such as the formation of isochromosome 12p—can be identified.^{10,11} Malignant transformation is then observed in early adulthood, peaking at ages 20–30 years and commonly presenting as a unilateral testicular mass.¹²

Genomic research advancements, particularly through genome-wide association studies (GWASs), have advanced our understanding of the mechanisms underlying many complex traits.¹³ A recent TGCT GWAS meta-analysis¹⁴ identified 22 additional loci, increasing the total to 78 known TGCT risk loci.^{14–20} However, the presence of multiple genes within some loci and linkage disequilibrium (LD) complicates the identification of specific causal genes. Transcriptome-wide association studies (TWASs), which utilize expression quantitative trait loci (eQTL) data to link GWAS-identified variants with changes in gene expression, offer a promising method for uncovering potential causal genes.²¹ This approach has been effectively applied in various phenotypes following GWASs.²²

To identify potential predisposition genes for TGCTs, we performed a TWAS of TGCTs by integrating genetic prediction models of gene expression with GWAS summary statistics. To validate the findings, expression patterns of the identified TWAS genes were explored across relevant tissues.

Material and methods

GWAS summary statistics

We utilized GWAS summary statistics from the most recent meta-analysis of TGCT risk, which included 10,156 men with TGCTs and 179,683 male control subjects. All men were of European descent and ascertained from 19 centers.¹⁴ To ensure a good overlap between GWAS and gene panels, we performed harmonization and imputation of SNP-level Z scores using the GWAS summary-imputation pipeline (<https://github.com/hakyimlab/summary-gwas-imputation>), resulting in 8,638,344 available SNPs and Z scores. The analytical set of SNPs consisted of 1,161,660 SNPs that were present in the 1000 Genomes LD reference panel (97.6% overlap). SNP rsIDs were used for matching.

Gene expression prediction models

We used pre-computed gene expression prediction models from the following source panels: (1) normal testis from the Genotype-Tissue Expression (GTEx)²³ release v.8, (2) TGCT tissue from The Cancer Genome Atlas (TCGA),^{24,25} and (3) cross-tissue prediction models derived from 22 GTEx normal tissues (Note S1).²⁶ In FUSION,²⁷ genetic prediction models were trained using *cis*-SNPs (located within a ± 500 kb window from the gene boundary) present in the 1000 Genomes LD reference panel.²⁸ Available models were computed using either single SNPs (eQTLs) or a linear combination of multiple *cis*-SNPs (*susie*, *blup*, *elastic-net*, and *lasso* methods). The best model was selected based on the largest out-of-sample R^2 computed via 5-fold cross-validation of each model. In our analysis, only models with both significant *cis*-heritability ($p < 0.05$) and predictive performance ($p < 0.05$ and cross-validation $R^2 > 0.01$) were considered. The cross-tissue panel consisted of prediction models derived from up to three main cross-tissue features based on sparse canonical correlation analysis (sCCA).²⁶ When a gene had more than one significant sCCA feature available, we selected the cross-tissue feature with the highest performance (i.e., cross-validation R^2). In total, 26,710 gene prediction models with significant *cis*-SNP heritability and predictive performance were available: 12,195 from normal testis, 1,254 from TGCTs, and 13,261 from the cross-tissue panel.

Summary-statistics-based TWAS

The summary-statistics-based TWAS was performed using FUSION software,²⁷ which implements a correlation test between genetically predicted gene expression, by using the per-SNP LD-adjusted expression weights learned for the prediction models, and common polygenic predisposition to TGCT, as captured by the per-SNP GWAS Z score, across the *cis*-SNPs of a gene of interest. SNP LD was estimated using 489 samples of European ancestry from the 1000 Genomes reference panel.²⁸ The statistical significance threshold for the summary-based correlation test was set at a false discovery rate (FDR) of 0.01 over the total number of tests. A TWAS hit was considered as overlapping with any of 78 previously reported TGCT risk loci¹⁴ if the transcription start site of the gene resided within a ± 1 Mb window around the top GWAS SNP.

TWAS conditional on GWAS effects

To control for inflation of TWAS statistics due to strong SNP-trait signals and extensive LD and to prioritize the observed associations, a conservative gene permutation test was applied on the FDR-significant TWAS hits. The test randomly shuffles expression weights (i.e., sampling with replacement) and recomputes an empirical association statistic conditional on the GWAS effects at

³³These authors contributed equally

*Correspondence: fredrik.wiklund@ki.se

<https://doi.org/10.1016/j.ajhg.2025.01.022>.

the locus.²⁷ An effective number of 100,000 permutations were performed on each test, and the significance level was set at 0.05.

TWAS joint-conditional analysis

We organized genes that passed summary-based and permutation-based correlation tests into separate genetic loci, defined as non-overlapping genomic regions containing neighboring genes within a 100 kb overlap window from their boundaries. Joint-conditional analysis was performed at each locus to identify conditionally independent genes following the FUSION post-process function (<http://gusevlab.org/projects/fusion/>) (Note S2). The significance threshold for gene selection at each locus was set at a nominal level of 0.05 for the joint-conditional *p* value and Bonferroni adjusted by the number of TWAS significant genes located on the same chromosome.

Colocalization analysis

Colocalization analysis was performed on the set of conditionally independent TWAS genes in order to identify gene-trait associations supported by shared genetic variants. For this purpose, we employed COLOC, a Bayesian method to estimate the posterior probabilities (PPs) of different variant configurations across two traits.²⁹ For each gene, we utilized *cis*-SNP summary statistics for both gene expression (eQTL weights) and TGCT risk (GWAS *Z* scores) to test the hypothesis of a shared causal variant (H4). Effect sizes were estimated based on the assumption that the standard error is inversely proportional to the square root of the sample size. For the TGCT GWAS dataset, we also incorporated the proportion of affected individuals (*n* = 10,156) as reported by Pluta et al.¹⁴ A significance threshold was set at PP4 > 0.5 to determine colocalization.

GWAS association conditional on TWAS effect

Following Mancuso et al.,³⁰ we evaluated residual association between SNPs at each TWAS risk locus and TGCT after adjusting for the predicted gene expression-TGCT association strength. Initially, the leading genes was derived as detailed in the joint-conditional TWAS analysis. Next, GWAS *Z* scores were imputed for each SNP at the locus conditioned on joint TWAS *Z* scores of selected genes. The resulting residual GWAS *Z* score was standardized to express remaining SNP-TGCT associations after conditioning on the predicted gene expression association.

Gene expression across human gonadal development, GCNIS, and TGCT

We utilized public expression datasets from human gonadal development, GCNIS, and TGCTs to study the expression patterns of TWAS-identified genes.

Human gonadal development

Preprocessed and annotated gonadal development gene expression data from human embryos (female: *n* = 33; male: *n* = 22)³¹ were downloaded from The Reproductive Cell Atlas (<https://www.reproductivecellatlas.org/gonads.html>). We used these data to evaluate whether TWAS candidate genes were found to be expressed in germ cell lineages at different developmental stages (i.e., PGCs, pre-spermatogonia, and pre-oocytes), as well as in male germ cells and somatic cells of the testis (Sertoli and fetal Leydig). In total, the expression levels of 3,764 PGCs, 621 pre-spermatogonia cells, 812 pre-oocyte cells, 3,624 germ cells, 34,927 Sertoli cells, and 2,420 fetal Leydig cells were obtained for 13,963 of

the 19,805 genes included in the TWAS analysis. As described by the authors, cells from doublets and those with high mitochondrial content were excluded from the analysis. After filtering, 3,739 PGCs, 486 pre-spermatogonia cells, 757 pre-oocyte cells, 2,883 germ cells, 15,273 Sertoli cells, and 1,769 fetal Leydig cells remained. We defined “no detectable gene expression” as cells with zero expression values. To estimate the average gene expression, values were variance scaled for each cell type, separately.

Precursor data (GCNIS)

Normalized microarray expression data from six GCNIS samples previously published by Sonne et al.³² were obtained from the publicly available ArrayExpress database (<https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-TABM-488>). The expression levels were log10 transformed, and the average values were obtained for samples characterized as cultured human embryonic stem cells (ESCs), microdissected spermatogenic tubules, and microdissected tubules containing GCNIS.

TGCT expression data

RNA sequencing (RNA-seq) data from 150 TGCT samples previously published by TCGA network in Shen et al.²⁵ were obtained from the Broad Institute FireBrowse portal (http://firebrowse.org/?cohort=TGCT&download_dialog=true). The preprocessed data were converted to transcripts per million and quantile normalized using the R package PreProcessCore (<https://bioconductor.org/packages/release/bioc/html/preprocessCore.html>), normalizing each sample to its mean expression across all genes. For each histopathology group (seminomas, *n* = 68; non-seminomas, *n* = 79), the log10-transformed gene expression levels were averaged.

Enrichment of gene expression

We used a permutation approach to test whether genes identified by TWASs were more likely to be expressed at different stages of TGCT pathogenesis (normal gonadal development, GCNIS, and TGCTs). To this end, we compared the median expression values observed for the set of leading genes (Bonferroni-adjusted joint-conditional *p* < 0.05; see above) to the median expressions for 10,000 random sets of genes of the same size, sampled from the subset of genes included in both the TWAS analysis and the respective expression dataset. For single-cell data, we also compared the median proportion of cells with non-zero expression values.

Pathway enrichment analysis

We conducted enrichment analysis using PathfindR,³³ following default settings. PathfindR is an active-subnetwork-oriented approach that leverages annotated protein-protein interaction (PPI) networks to identify clusters of disease-relevant genes that interact within active subnetworks. These subnetworks are then subjected to enrichment analysis via a one-sided hypergeometric test, using the PPI network as the background set. The analysis is performed iteratively (default of 10 iterations), with results reporting the number of times a pathway is enriched and the range of adjusted *p* values across iterations. For this analysis, we utilized Reactome^{34,35} as the pathway database and BioGRID³⁶ for the PPI network. The input data consisted of gene symbols and FDR-adjusted *p* values for the protein-coding TWAS hits with an FDR threshold of <0.01. The significance threshold for pathway enrichment was set at a Bonferroni-adjusted *p* value of <0.05.

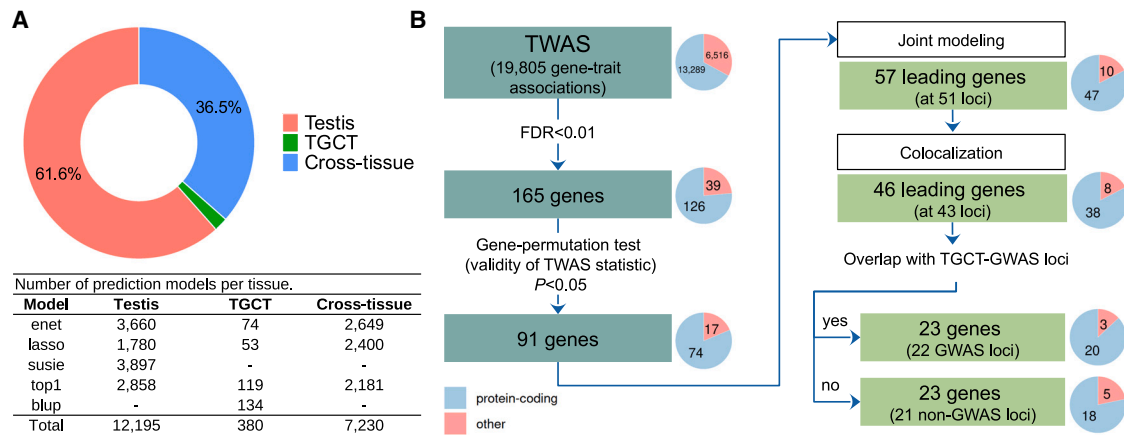


Figure 1. Overview of the study design

(A) Prediction models for heritable gene expression included in the analysis. Normal testis from the GTEx panel was selected as the primary target tissue, complemented by additional models from TGCTs from TCGA and cross-tissue features from 22 normal tissues from GTEx (mutually exclusive gene models selected in that order). Prediction models: blup, best linear unbiased predictor; enet, elastic-net regression; lasso, least absolute shrinkage and selection operator regression; susie, sum of single effects regression; top1, single best eQTL. (B) TWAS workflow to identify genes associated with predisposition to TGCTs. Significant genes were defined using a false discovery rate less than 0.01. Robust associations were identified using a gene permutation test $p < 0.05$. Genes were jointly modeled at each locus to identify conditionally independent genes (Bonferroni-adjusted joint-conditional $p < 0.05$). Colocalization analysis was subsequently conducted using a posterior probability (PP4) threshold of >0.5 to identify genes likely influenced by a shared causal variant affecting both gene expression and TGCT risk.

Immunohistochemistry

As a proof of concept, we investigated the presence of two specific TWAS-identified genes at the protein level. We selected one gene from regions not implicated by GWASs, along with another gene from a GWAS region that had not previously been associated with TGCT risk. Following an assessment of commercially available antibodies, ARID3B (from a non-GWAS region) and GINM1 (from a GWAS region) were chosen for analysis. Immunohistochemistry was conducted as previously described.³⁷ Tissue sections were obtained from the tissue biobank at the Department of Growth and Reproduction (Rigshospitalet, Copenhagen, Denmark) containing orchiectomy specimens from individuals with testicular cancer (The Danish Data Protection Agency, permit number J.nr. 2001-54-0906). Informed consent was provided by all tissue donors. The use of human samples was approved by the regional medical research ethics committee of the capital region of Copenhagen (H-16019637). Fresh testicular tissue was fixed in our in-house fixative (7.4% formaldehyde, 4% acetic acid, 2% methanol, 0.57% sodium phosphate [dibasic], and 0.11% potassium phosphate [monobasic]) overnight (for at least 16 h) at 4°C, dehydrated, and embedded in paraffin. Immunohistochemistry was conducted on 4 μ m sections as previously described.⁷ Briefly, tissue sections were subjected to heat-induced antigen retrieval in a pressure cooker (medical decloaking chamber, Biocare, Concord, CA, USA) in citrate buffer (ARID3B) or in a Tris-EGTA buffer (GINM1, D2-40, and SOX2; 10 mM Tris, 1 mM EDTA, and 0.05% Tween 20 [pH 8.5]) at 110°C for 30 min. Endogenous peroxidase was blocked with 1% (v/v) H₂O₂ in methanol for 30 min. Unspecific staining was blocked using 0.5% skimmed milk in Tris-buffered saline (TBS) for 30 min. Sections were incubated overnight at 4°C with primary antibodies against ARID3B (NBP2-33596, Novus Biologicals, San Diego, CO, USA) diluted 1:75, GINM1 (PA5-59671, Thermo Fisher Scientific, Waltham, MA, USA) diluted 1:400, D2-40 (M3619, Dako, Denmark) diluted 1:2,500, SOX2 (AF2018, R&D Systems, Minneapolis, MN, USA) diluted 1:150, or the negative control normal rabbit immunoglobulin (Ig)G (PP501P, Origene,

Rockville, MD, USA) diluted 1:75 (for ARID3B) or 1:400 (for GINM1) in a humidified chamber and then incubated for 30 min with the species-specific ImmPRESS HRP (peroxidase) secondary antibodies (Vector Laboratories, Newark, CA, USA). Between all steps (except after the blockage of unspecific staining), the sections were washed in TBS. Visualization was performed using ImmPACT AEC peroxidase (HRP) substrate (Vector Laboratories). The sections were subsequently counterstained with Mayer's hematoxylin and mounted with Aquatex mounting medium (Merck KGaA, Germany). The slides were scanned using a Hamamatsu NanoZoomer v.2.0 (Hamamatsu Photonics, Japan). Protein accumulation of GINM1 in precursor and tumor cells is illustrated in Figure S2. Negative controls and markers specific to tumor components are shown in Figure S3.

Results

TWAS of TGCT

To identify TGCT-associated genes, we conducted TWAS analyses using GTEx and TCGA data against TGCT risk variants (Figure 1A). Our analysis workflow prioritized available prediction models from tissues most biologically related to TGCTs to minimize potential false positive findings and increase statistical power. We selected 12,195 models from normal testis (main target tissue), complemented by 380 models from TGCTs and 7,230 from cross-tissue models capturing common regulatory features of gene expression among 22 normal tissues (testis included). Among these genes, the average *cis*-SNP heritability of expression was 24.5%, and the average predictive performance standardized by SNP heritability was 66.5% (Figure S1).

We tested 19,805 gene-disease associations for TGCT risk (Table S1) and visualized the results in Figure 2. Of these,

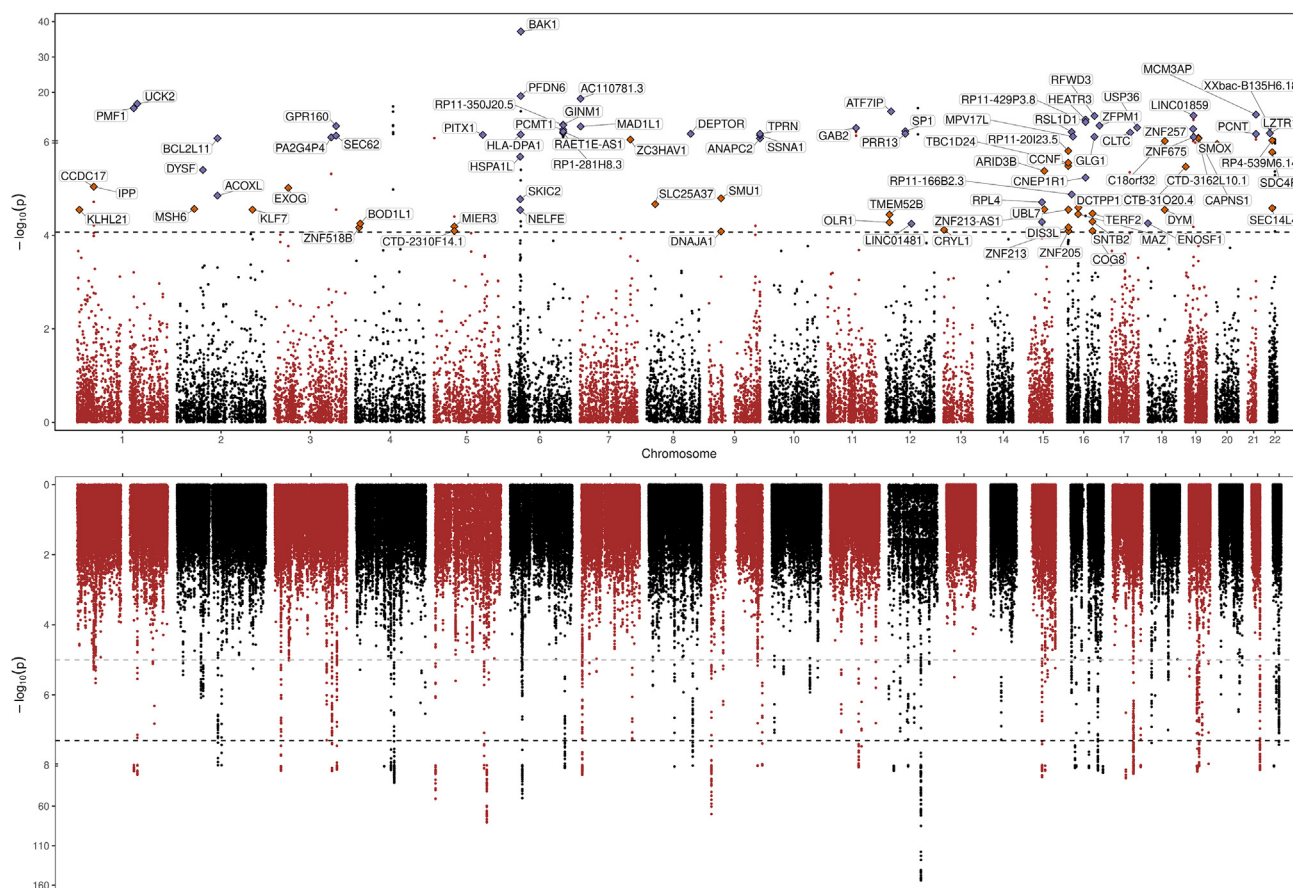


Figure 2. Manhattan/Miami plot summarizing the TWAS gene-TGCT associations in relation to GWAS signals

Top: summary of TWAS results showing the strength of association of TGCT genetic risk for 19,805 genes. Named TWAS hits indicated by diamonds were independent of GWAS signals (i.e., permutation $p < 0.05$); color indicates whether the TWAS hit is located at a GWAS locus (purple diamond) or at a non-GWAS region (red diamond). The dotted horizontal line represents the significance threshold at false discovery rate < 0.01 . Bottom: strength of SNP-level associations for TGCT risk ($-\log_{10} p$ value) as reported in the TGCT GWAS meta-analysis results.¹⁴ Shown are 1,161,660 SNPs that were included in the FUSION prediction models (i.e., SNPs that are present in the 1000 Genomes reference panel). Dotted horizontal lines indicate the GWAS-suggestive threshold ($p < 1 \times 10^{-5}$) in light gray and the GWAS-significance threshold ($p < 5 \times 10^{-8}$) in dark gray.

165 candidate genes were associated at $FDR < 0.01$ (Table S2). The majority (87, 52.7%) were from predicted expression in normal testis, 68 (41.2%) from the cross-tissue panel, and 10 (6.1%) from TGCT tissue. Pathway enrichment analysis, performed on the 123 protein-coding genes identified from the 165 TWAS hits that mapped to established protein interaction networks (as detailed in the material and methods), identified significant over-representation in 154 Reactome pathways (Table S3). Notably, pathways involved in cell cycle regulation, particularly those related to chromosomal segregation, were prominently enriched. Specifically, the mitotic spindle checkpoint and separation of sister chromatids pathways showed substantial fold enrichments of 9.82 and 6.35, respectively, with strong statistical significance ($p = 8.80 \times 10^{-7}$ and 5.1×10^{-6}). In addition to cell-cycle-related pathways, the analysis underscored the role of DNA repair mechanisms, evidenced by the enrichment of the SUMOylation of DNA replication proteins pathway (fold enrichment of 8.04, $p = 4.7 \times 10^{-5}$).

To mitigate potential TWAS statistic inflation from GWAS signals, we performed gene-wise permutation tests, identifying 91 out of 165 TWAS candidate genes as robustly associated with TGCTs (Table S2; Figure 1B). These genes coalesce into 51 genomic loci, with 25 loci harboring a single gene and 26 accommodating multiple genes. Subsequently, joint-conditional TWASs within loci with multiple genes identified 57 leading genes showing robust and independent associations with TGCTs (Tables 1, S4, and S5; Figure 1B). Following this analysis, only five loci contained multiple genes. Colocalization analysis was conducted on the 57 genes remaining after joint-conditional analysis. Using a PP (PP4) threshold greater than 0.5, evidence was found that 46 genes harbor a shared causal variant influencing both gene expression and TGCT risk. Among these 46 leading genes, the mean PP4 was 0.87 (interquartile range: 0.82–0.98). Of these, 23 genes were situated within 22 previously identified TGCT GWAS risk loci, while the other 23 genes were mapped to 21 regions not previously associated with TGCT risk in GWASs (Table 1).

Table 1. TWAS joint conditionally independent genes associated with TGC

Cytoband	Gene symbol	Position (hg38)	Gene type ^a	Tissue	Joint Z score	Joint <i>p</i> value	PP4	GWAS locus ^b
1p36.31	<i>KLHL21</i>	chr1:6,590,723–6,614,607	PC	TGCT	4.19	2.83×10^{-5}	0.608	N
1p34.1	<i>CCDC17</i>	chr1:45,620,043–45,624,057	PC	cross-tissue	−4.44	9.08×10^{-6}	0.905	N
1q22	<i>PMF1</i>	chr1:156,212,992–156,240,042	PC	TGCT	−8.17	3.02×10^{-16}	0.927	Y ^c
1q24.1	<i>UCK2</i>	chr1:165,827,755–165,908,109	PC	testis	8.52	1.62×10^{-17}	0.998	Y ^c
2p16.3	<i>MSH6</i>	chr2:47,695,529–47,810,101	PC	testis	−4.20	2.67×10^{-5}	0.916	N
2p13.2	<i>DYSF</i>	chr2:71,453,721–71,686,768	PC	testis	−4.61	4.03×10^{-6}	0.948	Y
2q13	<i>BCL2L11</i>	chr2:111,119,377–111,168,447	PC	cross-tissue	5.33	9.95×10^{-8}	0.990	Y ^c
2q33.3	<i>KLF7</i>	chr2:207,074,136–207,167,267	PC	testis	4.19	2.79×10^{-5}	0.933	N
3p22.2	<i>EXOG</i>	chr3:38,496,126–38,525,951	PC	testis	−4.43	9.64×10^{-6}	0.921	N
3q25.31	<i>PA2G4P4</i>	chr3:156,809,550–156,810,732	Pps	cross-tissue	5.44	5.32×10^{-8}	0.545	Y
3q26.2	<i>GPR160</i>	chr3:170,037,928–170,085,403	PC	testis	6.63	3.31×10^{-11}	0.863	Y ^c
4p16.1	<i>ZNF518B</i>	chr4:10,439,873–10,457,408	PC	cross-tissue	−3.98	6.81×10^{-5}	0.932	N
4p15.33	<i>BOD1L1</i>	chr4:13,568,737–13,627,723	PC	cross-tissue	4.03	5.60×10^{-5}	0.654	N
5q11.2	<i>MIER3</i>	chr5:56,919,601–56,971,675	PC	testis	3.99	6.47×10^{-5}	0.540	N
6p21.33	<i>HSPA1L</i>	chr6:31,809,618–31,815,065	PC	cross-tissue	4.75	2.08×10^{-6}	0.850	Y
6p21.31	<i>BAK1</i>	chr6:33,572,546–33,580,293	PC	TGCT	−9.02	1.91×10^{-19}	0.683	Y ^c
6q25.1	<i>GINM1</i>	chr6:149,566,293–149,591,748	PC	testis	6.76	1.40×10^{-11}	0.863	Y
7p22.3	<i>AC110781.3</i>	chr7:1,838,585–1,849,931	Ant	testis	−8.89	5.93×10^{-19}	0.977	Y
7q34	<i>ZC3HAV1</i>	chr7:139,043,519–139,109,719	PC	cross-tissue	5.17	2.29×10^{-7}	0.979	N
8q24.12	<i>DEPTOR</i>	chr8:119,928,399–120,050,913	PC	TGCT	−5.84	5.21×10^{-9}	0.913	Y ^c
9p21.1	<i>SMU1</i>	chr9:33,041,763–33,076,659	PC	testis	4.31	1.61×10^{-5}	0.874	N
9q34.3	<i>TPRN</i>	chr9:137,191,616–137,200,798	PC	testis	−5.84	5.34×10^{-9}	0.994	Y ^c
11q14.1	<i>GAB2</i>	chr11:78,215,296–78,418,348	PC	cross-tissue	6.44	1.18×10^{-10}	0.994	Y ^c
12p13.2	<i>TMEM52B</i>	chr12:10,170,541–10,191,801	PC	testis	−4.13	3.59×10^{-5}	0.701	N
12p13.1	<i>ATF7IP</i>	chr12:14,365,631–14,502,935	PC	cross-tissue	−7.92	2.43×10^{-15}	0.949	Y ^c
12q13.13	<i>SP1</i>	chr12:53,380,175–53,416,446	PC	testis	6.12	9.58×10^{-10}	0.952	Y ^c
13q12.11	<i>CRYL1</i>	chr13:20,403,666–20,525,857	PC	testis	−3.96	7.64×10^{-5}	0.543	N
15q24.1	<i>ARID3B</i>	chr15:74,541,176–74,598,131	PC	testis	−4.60	4.18×10^{-6}	0.973	N
16p13.3	<i>RP11-20I23.5</i>	chr16:2,575,627–2,577,373	Tec	cross-tissue	4.48	7.48×10^{-6}	0.956	N
16p13.3	<i>ZNF213-AS1</i>	chr16:3,110,459–3,134,882	Ant	testis	3.91	9.36×10^{-5}	0.853	N
16p13.11	<i>MPV17L</i>	chr16:15,395,753–15,413,268	PC	cross-tissue	5.53	3.12×10^{-8}	0.964	Y ^c
16p11.2	<i>DCTPP1</i>	chr16:30,423,618–30,424,533	PC	cross-tissue	−4.21	2.55×10^{-5}	0.870	N
16q12.1	<i>HEATR3</i>	chr16:50,065,940–50,106,387	PC	testis	7.23	4.80×10^{-13}	0.996	Y ^c
16q22.1	<i>TERF2</i>	chr16:69,355,560–69,408,571	PC	cross-tissue	−4.14	3.41×10^{-5}	0.782	N
16q24.2	<i>ZFFM1</i>	chr16:88,453,316–88,537,016	PC	testis	6.68	2.47×10^{-11}	0.812	Y ^c
17q25.3	<i>USP36</i>	chr17:78,787,380–78,841,441	PC	testis	6.48	9.12×10^{-11}	0.914	Y ^c
18q21.1	<i>C18orf32</i>	chr18:49,477,249–49,487,252	PC	testis	3.79	1.48×10^{-4}	0.545	N
19p13.3	<i>CTB-31020.4</i>	chr19:1,822,088–1,824,542	Ptr	testis	4.65	3.40×10^{-6}	0.786	N
19p12	<i>ZNF257</i>	chr19:22,052,451–22,091,480	PC	testis	5.92	3.28×10^{-9}	0.980	Y ^c
19p12	<i>ZNF675</i>	chr19:23,525,630–23,687,220	PC	cross-tissue	−5.11	3.18×10^{-7}	0.983	Y

(Continued on next page)

Table 1. Continued

Cytoband	Gene symbol	Position (hg38)	Gene type ^a	Tissue	Joint Z score	Joint p value	PP4	GWAS locus ^b
19q13.12	<i>CAPNS1</i>	chr19:36,139,574–36,150,353	PC	testis	–3.05	2.28×10^{-3}	0.992	N
19q13.12	<i>CTD-3162L10.1</i>	chr19:36,304,579–36,312,668	Lnc	testis	–2.93	3.39×10^{-3}	0.993	N
20p13	<i>SMOX</i>	chr20:4,120,979–4,187,747	PC	TGCT	–4.87	1.11×10^{-6}	0.551	N
21q22.3	<i>MCM3AP</i>	chr21:46,236,828–46,286,297	PC	testis	7.65	1.97×10^{-14}	0.984	Y ^c
22q11.21	<i>XXbac-B135H6.18</i>	chr22:20,981,360–20,981,755	Lnc	testis	–6.59	4.47×10^{-11}	0.999	Y
22q12.2	<i>RP4-539M6.14</i>	chr22:30,475,363–30,492,804	Ant	testis	5.07	4.03×10^{-7}	0.989	N

Joint-conditional results for 46 out of 57 independent genes surpassing colocalization analysis at the posterior probability of a common genetic variant ($PP4 > 0.5$), located at 43 genomic regions. Joint analysis was based on 91 genes with robust association ($FDR_{TWAS} < 0.01$ and gene permutation test $p < 0.05$) mapped onto 51 non-overlapping regions. PC, protein-coding; pps, processed pseudogene; ant, antisense; lnc, long intergenic non-coding RNA; tec, to be experimentally confirmed; ptr, processed transcript; Y, yes; N, no.

^aENSEMBL biotype annotation.

^bOverlap with a GWAS loci.

^cGene previously reported as implicated in TGCT risk.

GWAS association conditional on TWAS effect

To assess the proportion of the GWAS signal accounted for by the identified TWAS genes, we investigated the residual SNP association with TGCT after adjusting for the local TWAS signal at each of the 22 GWAS loci (Table S7). The median proportion of GWAS signals explained across the 22 GWAS loci was 86%, with an interquartile range of 75%–92%. In Figure 3, the effect of conditioning on TWAS effects on GWAS signals is illustrated for the two loci containing *UCK2* and *GINM1*, capturing 82% and 100% of the GWAS signal, respectively.

TWAS genes are expressed at various stages of TGCT development

We examined expression patterns of the 46 conditional independent TWAS genes showing evidence of colocalization compared to non-TWAS genes in datasets covering human gonadal development, GCNIS, and TGCT samples using a permutation method (see material and methods; Table S8). The expression levels and proportions of cells expressing the 46 leading TWAS genes were consistent with the anticipated expression levels and cell proportions within specific human gonadal cell types, including germ cell differentiation, germ cells, and supporting cells. Similarly, comparable expression levels were detected in microdissected GCNIS samples. Furthermore, a significant enrichment was observed in both seminoma ($p = 0.004$) and non-seminoma ($p = 0.01$) tumor tissue samples. Expression patterns of the ten most significant TWAS genes within GWAS and non-GWAS loci related to specific cell type marker genes are illustrated in Figures 4A–4C.

Protein-level abundance

Immunohistochemistry was performed to investigate protein-level accumulation of two candidate genes, *ARID3B* (Figure 4D) and *GINM1* (Figure S2), in relevant tissues. The proteins encoded by these two genes exhibited elevated accumulation in precursor (GCNIS) cells adjacent

to both seminoma and non-seminoma, as well as within seminoma. *ARID3B* protein levels also were detected in non-seminoma (embryonal carcinoma), whereas *GINM1* protein levels were not detected. Negative controls and markers specific to tumor components are shown in Figure S3.

Discussion

We conducted a TGCT-focused TWAS using GWAS summary data and genetically predicted expression models across multiple tissues. Over 19,000 prediction models were explored, identifying 165 TGCT-associated genes with $FDR < 0.01$. Following a conservative permutation test, joint-conditional analyses, and colocalization assessment, 46 leading genes remained associated with TGCTs. Among them, 23 genes were in known GWAS loci, 7 of which have not been previously implicated in the risk of TGCTs, and 23 genes were in 21 loci not previously identified by GWASs.

Gene expression prediction models were trained in normal and tumor testicular tissue and a set of 22 normal tissues. By employing this multi-tissue methodology, we identified an increased number of genes compared to using testis alone. For example, the multi-tissue approach identified *ATF7IP* as implicated in TGCT etiology.³⁸ *ATF7IP* acts as a transcription factor, facilitating the expression of *TERT*, also implicated as being involved in TGCT susceptibility through GWASs, and its corresponding RNA component, *TERC*.³⁹ *ATF7IP* also interacts with *SP1*,³⁹ which was also identified in this study, linking the pathway of chromatin organization to TGCT susceptibility.

Our study identified 23 genes across 21 loci previously unreported in GWASs. Notably, *MSH6* and *ARID3B*, identified through TWASs, are critical in DNA damage repair pathways linked to TGCT pathogenesis. *MSH6*, a key DNA mismatch repair gene,⁴⁰ is highly expressed in PGCs,^{31,41} and its mutations are linked to Lynch syndrome.^{40–42} It also

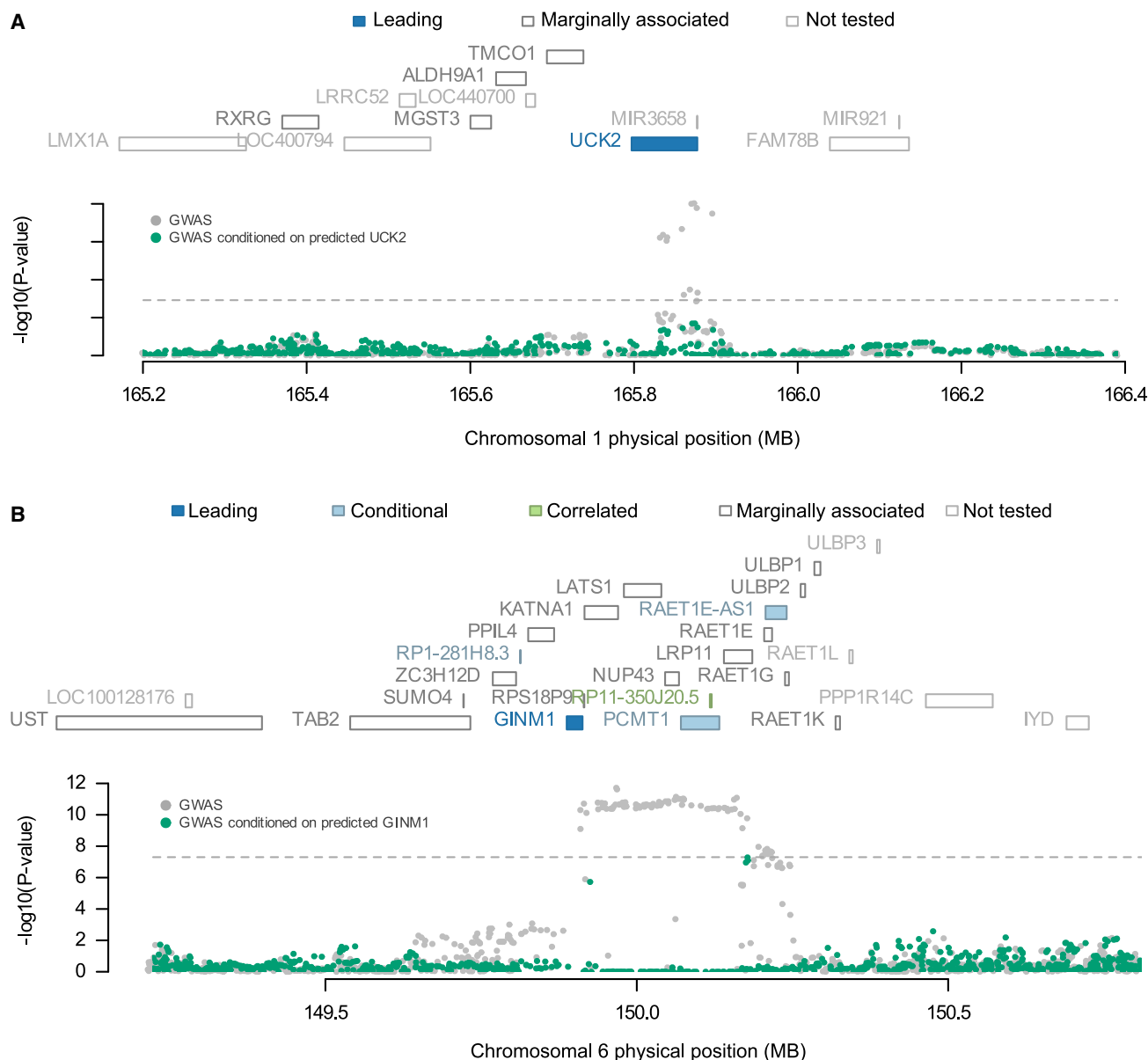


Figure 3. GWAS signals explained by top leading TWAS genes: *UCK2* and *GINM1*

UCK2 (A) and *GINM1* (B) have the lowest permutation p values among the 46 leading genes in Table 1. Each image consists of a top and bottom image. The top image shows all genes located at the risk locus according to the genomic position (hg19). Based on the joint modeling, genes are labeled as leading in dark blue (joint-conditionally independent), conditional on leading genes(s) in light blue, or correlated with leading genes in green (predicted expression $R^2 > 0.9$); marginally associated genes ($p_{\text{TWAS}} \geq 0.01$ or $p_{\text{permutation}} \geq 0.05$) and genes not tested (e.g., due to lack of a significant prediction model) are indicated with empty boxes with dark and light outlines, respectively. The bottom image shows results from the SNP-level GWAS-conditional analysis. Each point illustrates the association between a single SNP at the locus and TGCT status: gray points indicate the marginal association of a SNP with TGCT status (GWAS association) and green points indicate the association of the same SNP with TGCTs after conditioning of predicted expression of the leading gene at each locus (here, *UCK2* and *GINM1*). The horizontal dashed line indicates the genome-wide significance threshold ($p = 5 \times 10^{-8}$).

contributes to genomic imprinting,⁴³ a critical process in germ cell development connected to TGCTs by promoting cancer stemness.^{44–46} ARID3B, known for its role in chromatin remodeling and gene expression regulation,⁴⁷ promotes cancer stemness and tumor growth in ovarian cancer xenografts.^{48,49} It regulates stem cell markers such as *POU5F1* (*OCT4*) and *SOX2*,^{50,51} associated with GCNIS and embryonal carcinoma.^{32,44,45} Notably, in mouse embry-

onic development, *Arid3b* is highly expressed specifically in developing testes, suggesting a similar role for its highly conserved human counterpart.⁵²

Our pathway enrichment analysis provided additional evidence supporting the involvement of genes associated with chromosomal segregation and the DNA damage response in TGCTs. After carefully reviewing the literature, we annotated TWAS- and GWAS-identified genes in the

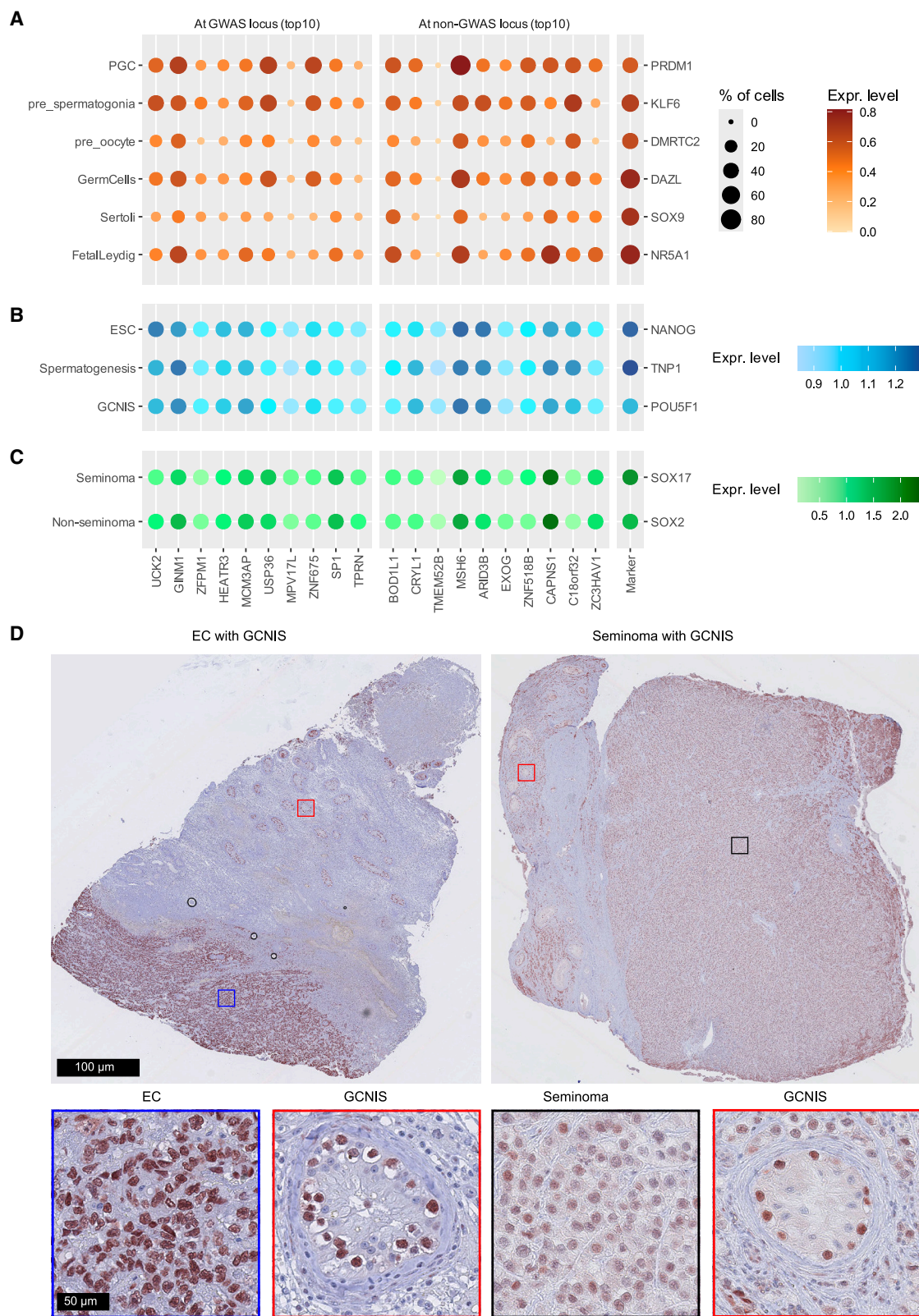


Figure 4. Expression patterns of top leading genes at various tissues

(A–C) Expression levels of ten most significant genes among genes overlapping a GWAS locus (left) or not (right). Genes are ordered by strength of evidence ($p_{\text{permutation}}$) with decreasing significance from left to right; expression levels of tissue-specific marker genes are shown for reference on the right.

(legend continued on next page)

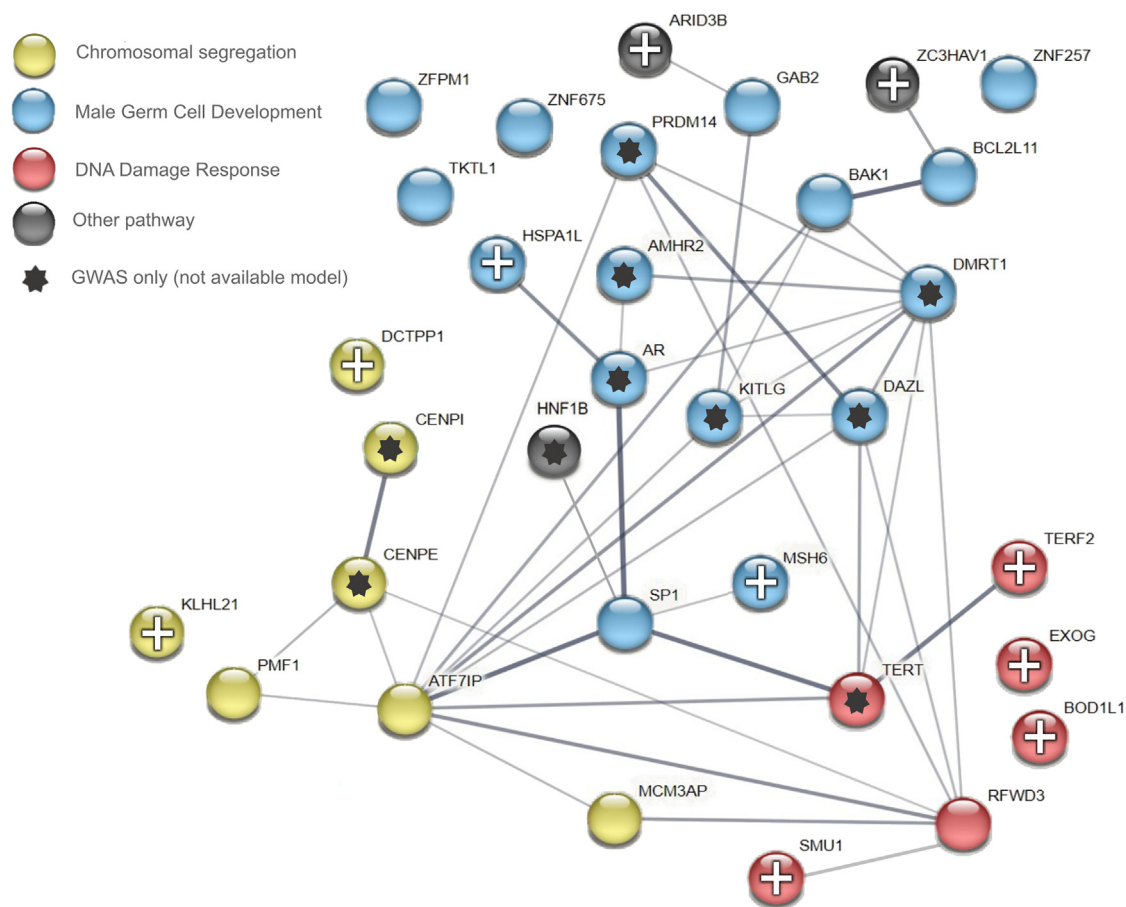


Figure 5. Network interactions and TGCT-relevant annotations for TWAS leading genes

A protein-protein interaction network in STRING (string-db.org) for TWAS leading genes along with genes implicated by GWASs¹⁴ that were not included in the current study due to lack of a prediction model. Disconnected gene nodes without relevant annotations were removed. Line weights indicate the degree of confidence of an interaction between any two proteins based on STRING presets. Pathway annotation was based on manual curation as described in the [discussion](#) and [Table S9](#).

context of three main pathways relevant to TGCT etiology—male germ cell development, chromosomal segregation, and the DNA damage response ([Figure 5](#); [Table S9](#)). [Figure 5](#) includes genes identified through GWASs that could not be evaluated in TWASs because a prediction model was not available. Integration of the findings from both GWASs and TWASs broadens the biological inference and expands the genes in these pathways, in which variation is associated with susceptibility to TGCTs.

We assessed the expression patterns of the identified TWAS genes across human gonadal development, GCNIS, and TGCT samples. Overall, the expression levels

of these TWAS genes aligned with the anticipated expression levels in gonadal development and GCNIS samples. Additionally, significant enrichment was observed in both seminoma and non-seminoma tumor tissue samples, supporting a potential causal association between the identified genes and TGCTs. Additionally, through immunohistochemistry, we demonstrated the protein-level accumulation of two candidate genes, *ARID3B* and *GINM1*, in both precursor and tumor cells. *ARID3B* is frequently co-expressed with *MYCN* in germ cell tumors, ESCs, and testis—cell types associated with pluripotency.⁵³ This co-expression is known to repress apoptosis and

(A) Human embryonal gonads. Single-cell average gene expression (color panel) and proportion of cells with non-zero expression (dot size) at different stages of germ cell differentiation (primordial germ cell [PGC], pre-spermatogonia, and pre-oocyte cell types) as well as in male germ cells and supporting cells (Sertoli and fetal Leydig).

(B) Pre-malignant tissue. Microarray expression from microdissected samples characterized as embryonal stem cells (ESCs), spermatogenesis, and germ cell neoplasia *in situ* (GCNIS). The color panel indicates the log₁₀-average expression levels at each condition.

(C) TGCTs. The color panel indicates the log₁₀-average expression levels.

(D) Protein localization of *ARID3B*, in GCNIS adjacent to a non-seminoma (embryonal carcinoma [EC]), and GCNIS adjacent to a seminoma. For each TGCT subtype, both the tumor and GCNIS component are shown in greater magnification at the bottom. Scale bars indicate 100 and 50 μ m, respectively.

promote cell cycle progression, both critical in cancer development.⁵⁴ The N-glycosylation pattern of GINM1, linked to disease progression in bladder cancer,⁵⁵ has not been previously associated with TGCTs to our knowledge.

Our study has several limitations. Firstly, our expression models, trained on adult tissues, might not capture genes specifically expressed during gonadal development. Secondly, even with a multi-tissue approach, we faced challenges in developing prediction models for some key genes associated with TGCT, such as *KITLG*, which plays a critical role in the disease's pathogenesis.⁵⁶ Thirdly, the absence of rare variants during model construction may have led to overlooking important TGCT genes identified in sequencing studies.⁵⁷ Additionally, we limited our analysis to autosomal chromosomes, ignoring potentially critical sex-linked genes like *AR*, hypothesized to contribute to TGCT pathogenesis.¹⁴ Lastly, our study solely focused on overall gene expression, neglecting the influence of RNA splicing on disease etiology, which has been shown to represent a critical link between genetic variation and complex diseases.⁵⁸

Conclusions

In summary, utilizing TWAS methodology, we have successfully identified numerous genes associated with TGCT susceptibility, warranting further investigation into their functional roles. Our approach has revealed previously unreported susceptibility regions and identified candidate genes within several well-established susceptibility loci. Future investigations employing enhanced TWAS methodologies, encompassing rare coding variants and gene expression levels specific to splicing events, hold the potential to uncover additional genes associated with susceptibility to TGCTs.

Data and code availability

The code used in this analysis is available at https://github.com/emiuga/TGCT_TWAS. Fusion software, prediction models, and reference LD are available at <http://gusevlab.org/projects/fusion/>. TGCT GWAS summary data used in this study are available at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001349.v2.p1.

Acknowledgments

Please see the [supplemental information](#) for acknowledgments.

Author contributions

F.W., K.L.N., and P.A.K. conceived and supervised the study. E.U.-M. and R.W. performed primary data analysis. E.U.-M., R.W., A.P., R.K., S.B.W., J.P., A.N.M., K.A., P.A.K., K.L.N., and F.W. contributed to the analysis and interpretation of data. M.F., M.D., K.K.A., L.A.-C., C.C., V.K.C., A.F., M.G., J.A.G., A.G.-N., T.G., R.J.H., M.H., T.B.H., R.H., M.A.T.H., L.A.K., J.K., D.L., R.A.L., C.L., S.J.C., K.A.M., C.M., K.T.N., J.N., M.P., T.R., L.R., M.S.R., S.M.S., R.I.S., K.S., D.R.S., C.T., D.J.V., T.Z., K.A., P.A.K.,

K.L.N., and F.W. contributed phenotypic data and genotypic data from men with and without TGCT. E.U.-M., R.W., F.W., and K.L.N. wrote the manuscript with critical input from all authors. All authors approved the final version of the manuscript.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2025.01.022>.

Web resources

EMBL's European Bioinformatics Institute, <https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-TABM-488>
FireBrowse, http://firebrowse.org/?cohort=TGCT&download_dialog=true
GitHub, <https://github.com/hakymilab/summary-gwas-imputation>
Gusev Lab, <http://gusevlab.org/projects/fusion/>
NIH, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001349.v2.p1
The Reproductive Cell Atlas, <https://www.reproductivecellatlas.org/gonads.html>

Received: August 5, 2024

Accepted: January 28, 2025

Published: February 24, 2025

References

1. Moch, H., Cubilla, A.L., Humphrey, P.A., Reuter, V.E., and Ulbright, T.M. (2016). The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur. Urol.* 70, 93–105. <https://doi.org/10.1016/j.eururo.2016.02.029>.
2. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* 71, 209–249. <https://doi.org/10.3322/caac.21660>.
3. 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. <https://doi.org/10.1016/j.eururo.2019.07.002>.
4. Hemminki, K., and Chen, B. (2005). Are twins at risk of cancer: results from the Swedish family-cancer database. *Twin Res. Hum. Genet.* 8, 509–514. <https://doi.org/10.1375/183242705774310150>.
5. Hemminki, K., and Li, X. (2002). Cancer risks in twins: results from the Swedish family-cancer database. *Int. J. Cancer* 99, 873–878. <https://doi.org/10.1002/ijc.10441>.
6. Hemminki, K., and Li, X. (2004). Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. *Br. J. Cancer* 90, 1765–1770. <https://doi.org/10.1038/sj.bjc.6601714>.

7. Neale, R.E., Carrière, P., Murphy, M.F.G., and Baade, P.D. (2008). Testicular cancer in twins: a meta-analysis. *Br. J. Cancer* 98, 171–173. <https://doi.org/10.1038/sj.bjc.6604136>.
8. Hemminki, K., Vaittinen, P., Dong, C., and Easton, D. (2001). Sibling risks in cancer: clues to recessive or X-linked genes? *Br. J. Cancer* 84, 388–391. <https://doi.org/10.1054/bjoc.2000.1585>.
9. Sonne, S.B., Almstrup, K., Dalgaard, M., Juncker, A.S., Edsgard, D., Ruban, L., Harrison, N.J., Schwager, C., Abdollahi, A., Huber, P.E., et al. (2009). Analysis of gene expression profiles of microdissected cell populations indicates that testicular carcinoma in situ is an arrested gonocyte. *Cancer Res.* 69, 5241–5250. <https://doi.org/10.1158/0008-5472.CAN-08-4554>.
10. 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. <https://doi.org/10.1007/s00018-019-03022-7>.
11. Litchfield, K., Levy, M., Huddart, R.A., Shipley, J., and Turnbull, C. (2016). The genomic landscape of testicular germ cell tumours: from susceptibility to treatment. *Nat. Rev. Urol.* 13, 409–419. <https://doi.org/10.1038/nrurol.2016.107>.
12. Porten, S.P., and Presti, J.J.C. (2020). Genital Tumors. In *Smith & Tanagho's General Urology*, 19e, J.W. McAninch and T.F. Lue, eds. (McGraw Hill).
13. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* 101, 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>.
14. 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 susceptibility loci associated with testicular germ cell tumors. *Nat. Commun.* 12, 4487. <https://doi.org/10.1038/s41467-021-24334-y>.
15. Chung, C.C., Kanetsky, P.A., Wang, Z., Hildebrandt, M.A.T., Koster, R., Skotheim, R.I., Kratz, C.P., Turnbull, C., Cortessis, V.K., Bakken, A.C., et al. (2013). Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat. Genet.* 45, 680–685. <https://doi.org/10.1038/ng.2634>.
16. 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. <https://doi.org/10.1093/hmg/ddr207>.
17. Kristiansen, W., Karlsson, R., Rounge, T.B., Whittington, 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. <https://doi.org/10.1093/hmg/ddv129>.
18. Ruark, E., Seal, S., McDonald, H., Zhang, F., Elliot, A., Lau, K., Perdeaux, E., Rapley, E., Eeles, R., Peto, J., et al. (2013). Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nat. Genet.* 45, 686–689. <https://doi.org/10.1038/ng.2635>.
19. Schumacher, F.R., Wang, Z., Skotheim, R.I., Koster, R., Chung, C.C., Hildebrandt, M.A.T., 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. <https://doi.org/10.1093/hmg/ddt109>.
20. 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. <https://doi.org/10.1038/ng.3879>.
21. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., et al.; GTEx Consortium (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* 47, 1091–1098. <https://doi.org/10.1038/ng.3367>.
22. Mai, J., Lu, M., Gao, Q., Zeng, J., and Xiao, J. (2023). Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun. Biol.* 6, 899. <https://doi.org/10.1038/s42003-023-05279-y>.
23. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330. <https://doi.org/10.1126/science.aaz1776>.
24. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120. <https://doi.org/10.1038/ng.2764>.
25. 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. <https://doi.org/10.1016/j.celrep.2018.05.039>.
26. Feng, H., Mancuso, N., Gusev, A., Majumdar, A., Major, M., Pasaniuc, B., and Kraft, P. (2021). Leveraging expression from multiple tissues using sparse canonical correlation analysis and aggregate tests improves the power of transcriptome-wide association studies. *PLoS Genet.* 17, e1008973. <https://doi.org/10.1371/journal.pgen.1008973>.
27. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48, 245–252. <https://doi.org/10.1038/ng.3506>.
28. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>.
29. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* 10, e1004383. <https://doi.org/10.1371/journal.pgen.1004383>.
30. Mancuso, N., Gayther, S., Gusev, A., Zheng, W., Penney, K.L., Kote-Jarai, Z., Eeles, R., Freedman, M., Haiman, C., Pasaniuc, B.; and PRACTICAL consortium (2018). Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat. Commun.* 9, 4079. <https://doi.org/10.1038/s41467-018-06302-1>.
31. Garcia-Alonso, L., Lorenzi, V., Mazzeo, C.I., Alves-Lopes, J.P., Roberts, K., Sancho-Serra, C., Engelbert, J., Marečková, M., Gruhn, W.H., Botting, R.A., et al. (2022). Single-cell roadmap of human gonadal development. *Nature* 607, 540–547. <https://doi.org/10.1038/s41586-022-04918-4>.

32. Sonne, S.B., Perrett, R.M., Nielsen, J.E., Baxter, M.A., Kristensen, D.M., Leffers, H., Hanley, N.A., and Rajpert-De-Meyts, E. (2010). Analysis of SOX2 expression in developing human testis and germ cell neoplasia. *Int. J. Dev. Biol.* 54, 755–760. <https://doi.org/10.1387/ijdb.082668ss>.
33. Ulgen, E., Ozisik, O., and Sezerman, O.U. (2019). pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks. *Front. Genet.* 10, 858. <https://doi.org/10.3389/fgene.2019.00858>.
34. Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P., Wu, G., Stein, L., D'Eustachio, P., and Hermjakob, H. (2018). Reactome graph database: Efficient access to complex pathway data. *PLoS Comput. Biol.* 14, e1005968. <https://doi.org/10.1371/journal.pcbi.1005968>.
35. Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., et al. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 50, D687–D692. <https://doi.org/10.1093/nar/gkab1028>.
36. Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., et al. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 30, 187–200. <https://doi.org/10.1002/pro.3978>.
37. Nielsen, J.E., Rolland, A.D., Rajpert-De Meyts, E., Janfelt, C., Jørgensen, A., Winge, S.B., Kristensen, D.M., Juul, A., Chalmel, F., Jégou, B., and Skakkebaek, N.E. (2019). Characterisation and localisation of the endocannabinoid system components in the adult human testis. *Sci. Rep.* 9, 12866. <https://doi.org/10.1038/s41598-019-49177-y>.
38. Turnbull, C., Rapley, E.A., Seal, S., Pernet, D., Renwick, A., Hughes, D., Ricketts, M., Linger, R., Nsengimana, J., Deloukas, P., et al. (2010). Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat. Genet.* 42, 604–607. <https://doi.org/10.1038/ng.607>.
39. Liu, L., Ishihara, K., Ichimura, T., Fujita, N., Hino, S., Tomita, S., Watanabe, S., Saitoh, N., Ito, T., and Nakao, M. (2009). MCAF1/AM is involved in Sp1-mediated maintenance of cancer-associated telomerase activity. *J. Biol. Chem.* 284, 5165–5174. <https://doi.org/10.1074/jbc.M807098200>.
40. Zhan, H., Mo, F., Xu, Q., Wang, S., Zhang, B., Liu, X., Dai, M., and Liu, H. (2021). An integrative pan-cancer analysis reveals the oncogenic role of mutS homolog 6 (MSH6) in human tumors. *Aging (Albany NY)* 13, 25271–25290. <https://doi.org/10.18632/aging.203745>.
41. Menezes, Y., Jr., Russo, G., Tosti, E., El Mouatassim, S., and Benkhalifa, M. (2007). Expression profile of genes coding for DNA repair in human oocytes using pangenomic microarrays, with a special focus on ROS linked decays. *J. Assist. Reprod. Genet.* 24, 513–520. <https://doi.org/10.1007/s10815-007-9167-0>.
42. Haraldsdottir, S., Rafnar, T., Frankel, W.L., Einarisdottir, S., Sigurdsson, A., Hampel, H., Snaebjornsson, P., Masson, G., Weng, D., Arngrimsson, R., et al. (2017). Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nat. Commun.* 8, 14755. <https://doi.org/10.1038/ncomms14755>.
43. Acurzio, B., Cecere, F., Giaccari, C., Verma, A., Russo, R., Valletta, M., Hay Mele, B., Angelini, C., Chambery, A., and Riccio, A. (2022). The mismatch-repair proteins MSH2 and MSH6 interact with the imprinting control regions through the ZFP57-KAP1 complex. *Epigenet. Chromatin* 15, 27. <https://doi.org/10.1186/s13072-022-00462-7>.
44. Cheng, L., Albers, P., Berney, D.M., Feldman, D.R., Dugaard, G., Gilligan, T., and Looijenga, L.H.J. (2018). Testicular cancer. *Nat. Rev. Dis. Primers* 4, 29. <https://doi.org/10.1038/s41572-018-0029-0>.
45. Oosterhuis, J.W., and Looijenga, L.H.J. (2005). Testicular germ-cell tumours in a broader perspective. *Nat. Rev. Cancer* 5, 210–222. <https://doi.org/10.1038/nrc1568>.
46. Oosterhuis, J.W., and Looijenga, L.H.J. (2019). Human germ cell tumours from a developmental perspective. *Nat. Rev. Cancer* 19, 522–537. <https://doi.org/10.1038/s41568-019-0178-9>.
47. Dausinas, P., Pulakanti, K., Rao, S., Cole, J.M., Dahl, R., and Cowden Dahl, K.D. (2020). ARID3A and ARID3B induce stem promoting pathways in ovarian cancer cells. *Gene* 738, 144458. <https://doi.org/10.1016/j.gene.2020.144458>.
48. Bobbs, A., Gellerman, K., Hallas, W.M., Joseph, S., Yang, C., Kurkewich, J., and Cowden Dahl, K.D. (2015). ARID3B Directly Regulates Ovarian Cancer Promoting Genes. *PLoS One* 10, e0131961. <https://doi.org/10.1371/journal.pone.0131961>.
49. Roy, L., Samyesudhas, S.J., Carrasco, M., Li, J., Joseph, S., Dahl, R., and Cowden Dahl, K.D. (2014). ARID3B increases ovarian tumor burden and is associated with a cancer stem cell gene signature. *Oncotarget* 5, 8355–8366. <https://doi.org/10.18632/oncotarget.2247>.
50. Chien, C.S., Wang, M.L., Chu, P.Y., Chang, Y.L., Liu, W.H., Yu, C.C., Lan, Y.T., Huang, P.I., Lee, Y.Y., Chen, Y.W., et al. (2015). Lin28B/Let-7 Regulates Expression of Oct4 and Sox2 and Reprograms Oral Squamous Cell Carcinoma Cells to a Stem-like State. *Cancer Res.* 75, 2553–2565. <https://doi.org/10.1158/0008-5472.CAN-14-2215>.
51. Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444, 364–368. <https://doi.org/10.1038/nature05284>.
52. Takebe, A., Era, T., Okada, M., Martin Jakt, L., Kuroda, Y., and Nishikawa, S.I. (2006). Microarray analysis of PDGFR alpha+ populations in ES cell differentiation culture identifies genes involved in differentiation of mesoderm and mesenchyme including ARID3b that is essential for development of embryonic mesenchymal cells. *Dev. Biol.* 293, 25–37. <https://doi.org/10.1016/j.ydbio.2005.12.016>.
53. Kobayashi, K., Jakt, L.M., and Nishikawa, S.I. (2013). Epigenetic regulation of the neuroblastoma genes, Arid3b and Mycn. *Oncogene* 32, 2640–2648. <https://doi.org/10.1038/onc.2012.285>.
54. Gerl, R., and Vaux, D.L. (2005). Apoptosis in the development and treatment of cancer. *Carcinogenesis* 26, 263–270. <https://doi.org/10.1093/carcin/bgh283>.
55. Sathe, G., George, I.A., Deb, B., Jain, A.P., Patel, K., Nayak, B., Karmakar, S., Seth, A., Pandey, A., and Kumar, P. (2020). Urinary glycoproteomic profiling of non-muscle invasive and muscle invasive bladder carcinoma patients reveals distinct N-glycosylation pattern of CD44, MGAM, and GINM1. *Oncotarget* 11, 3244–3255. <https://doi.org/10.18632/oncotarget.27696>.
56. Kanetsky, P.A., Mitra, N., Vardhanabhuti, S., Li, M., Vaughn, D.J., Letrero, R., Ciosek, S.L., Doody, D.R., Smith, L.M., Weaver, J., et al. (2009). Common variation in KITLG and at

- 5q31.3 predisposes to testicular germ cell cancer. *Nat. Genet.* 41, 811–815. <https://doi.org/10.1038/ng.393>.
57. Pyle, L.C., Kim, J., Bradfield, J., Damrauer, S.M., D'Andrea, K., Einhorn, L.H., Godse, R., Hakonarson, H., Kanetsky, P.A., Kember, R.L., et al. (2024). Germline Exome Sequencing for Men with Testicular Germ Cell Tumor Reveals Coding Defects in Chromosomal Segregation and Protein-targeting Genes. *Eur. Urol.* 85, 337–345. <https://doi.org/10.1016/j.eururo.2023.05.008>.
58. Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. *Science* 352, 600–604. <https://doi.org/10.1126/science.aad9417>.