Role of Hub Genes in the Occurrence and Development of Testicular Cancer Based on Bioinformatics

Chunlei Zhang1,*, Weijun Zhang1,*, Han Cui2,*, Bin Zhang1,*, Pengcheng Miao1, Qi Yang1, Mei Bai1, Hongmei Jiao1, Dehui Chang1

1Department of Urology, The 940 Hospital of Joint Logistics Support Force of Chinese PLA, Lanzhou, 730050, People’s Republic of China; 2Department of the First Clinic, The 940 Hospital of Joint Logistics Support Force of Chinese PLA, Lanzhou, 730050, People’s Republic of China

*These authors contributed equally to this work

Background: Testicular cancer severely affects male health, so finding effective diagnosis and prognostic indicators and exploring its pathogenesis are very important.

Purpose: This study aims to explore the hub genes that play important roles in the occurrence and development of testicular germ cell tumor (TGCT).

Methods: Data were obtained from Gene Expression Omnibus datasets (GSE3218 and GSE1818) and verified in The Cancer Genome Atlas database and the Genotype-Tissue Expression database and the Human Protein Atlas database. A protein–protein interaction network was constructed to obtain hub genes. GEO2R, R software and packages were used to analyze differentially expressed genes (DEGs), receiver operating characteristic curve assessment, Cox regression analysis, Kaplan–Meier survival curve assessment, Gene Ontology analysis, Kyoto Encyclopedia of Genes and Genomes analysis, the relationship with clinicopathological information, gene set enrichment analysis, the correlation with immune cells’ infiltration, and the expression in pan-cancers of the hub genes.

Results: PLK4, TRIP13, TPR, KIF18A, CDKN3, HMMR, PBK, PTTG1, CKS2, SYCP1, HSPA2, and MKI67 were selected as the hub genes. mRNA of PLK4, TRIP13, CDKN3, SYCP1, HSPA2, and MKI67 had high diagnostic values, and higher expression of CDKN3 and HSPA2 mRNA were poor prognostic factors for progression-free interval of TGCT. The hub genes involved organelle division and cell cycle, chromosome and centromeric region, heat shock protein binding, and more. Downregulated TPR and PLK4 were selected as research targets for continued study, and they may participate in multiple signaling pathways. The expression of TPR and PLK4 correlated with the infiltration of a variety of immune cells and differed in pan-cancers.

Conclusion: The mRNA levels of multiple hub genes have high diagnostic and prognostic values for TGCT. TPR and PLK4 may play a role in the occurrence and development of TGCT through cancer-related signaling pathways.

Keywords: testicular cancer, mRNA, hub gene, diagnosis, prognosis, biomarker, immune infiltration, signal pathway

Introduction

Testicular cancer accounts for 1% of newly diagnosed cancers in men worldwide. It is one of the most common cancers between the ages of 14 and 44 years for men in western countries, and its incidence has gradually increased in recent years. According to 2020 estimates by the American Cancer Society, there will be 1,806,590 new cancer cases and 606,520 cancer deaths in the United States (US), including 9,610 cases of testicular cancer and 440 deaths resulting from testicular
cancer.\(^3\) Estimated cancer statistics for US adolescents and young adults in 2020 from Leading Sites of New Cancer Cases in Adolescents and Young Adults shows that the proportion of the number of cases with testicular seminoma ranks second among the 20–29-year-old male patients.\(^4\) Because testicular cancer is relatively rare, few clinical reports exist on the topic. Currently, alpha-fetoprotein (AFP) and human chorionic gonadotropin (HCG) are molecular diagnostic markers for testicular cancer; however, because of their low specificity and the insignificant increases in serum for most patients with seminoma, challenges still exist in the screening and diagnosis of testicular cancer, especially pure seminoma.\(^5\) In addition, no good prognostic indicators exist for testicular cancer. Germ cell tumors are particularly sensitive to cisplatin-based chemotherapy, which greatly improves the patient’s long-term survival rate. However, 40–60% of the patients with germ cell tumor fail to benefit from salvage chemotherapy and have a poor prognosis. The exploration of new targeted drug therapy is indispensable for these patients.\(^6\)

At present, testicular cancer lacks ideal biomarkers with high sensitivity and specificity in diagnosis, treatment, and prognosis. Therefore, it is important to use bioinformatics methods to find clinically applicable biomarkers. Public databases, such as the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA), offer methods to study the gene spectrum and pathogenesis of tumor diseases. By extracting data, we can obtain differentially expressed genes (DEGs) in disease, predict the genes related to the disease, and analyze the possible mechanism of these genes in the process of disease. Cancer is a genomic disease that is characterized by genome instability, and a large number of point mutations and structural changes appear during tumor progression. Genomic mutation may produce tumor antigens, which may be recognized by the immune system as non-self antigens and trigger cellular immune responses. The immune system plays a vital role in immune surveillance, because immune cells of the adaptive immune system and the innate immune system can penetrate into the tumor microenvironment and help regulate tumor progression.\(^7,8\) However, current research on testicular cancer is relatively one sided and is focused on the expression of a certain gene; it lacks comprehensive, systematic, and in-depth analysis of testicular cancer sequencing results in existing databases to explore molecular markers for the diagnosis and prognosis of testicular cancer and to explore hub genes that may play a role in the development. Current efforts also fail to effectively combine and analyze the only two datasets (GSE3218 and GSE1818) for mRNA sequencing of testicular cancer tissues and normal testicular tissues. Moreover, at present, few studies have explored the correlation between gene expression and immune infiltration in testicular cancer. To make up for these shortcomings, in this study, we selected hub genes that may be involved in the occurrence and development of testicular cancer through bioinformatics analysis combined with the two datasets. As the most common testicular cancer is testicular germ cell tumor (TGCT), we used data of TGCT obtained from TCGA and Genotype-Tissue Expression (GTEx) database to explore the function of hub genes. In addition to analyzing the relationship between these genes and patient clinicopathological information, the signal pathways that may be involved, and the diagnosis and prognostic value of corresponding mRNA, we also analyzed the relationship between these genes and immune cells’ infiltration in TGCT, to explore genes as diagnostic and prognostic markers and hub molecules that may be selected as therapeutic targets.

## Materials and Methods

### Data Collection

The data on differentially expressed genes (DEGs) came from two datasets—GSE3218 and GSE1818—downloaded from the GEO (http://www.ncbi.nlm.nih.gov/geo/) database.\(^9\) The gene expression profile of GSE3218, including 101 testicular cancer tissues and 5 normal tissues, was assessed by the GPL96 [HG-U133A] Affymetrix Human Genome U133A Array platform. The gene expression profile of GSE1818, including 20 testicular cancer tissues and 3 normal tissues, was assessed by the GPL885 Agilent-011521 Human 1A Microarray G4110A (Feature Number version) platform. For verification of DEGs, data in the UCSC XENA database (https://xenabrowser.net/datapages/)\(^10\) and RNA sequencing (RNAseq) in TPM format from TCGA and GTEx databases,\(^11,12\) processed uniformly by the Toil process,\(^13\) were extracted; data sources included 154 cases of testicular cancer tissue samples in TCGA and 165 cases of normal testicular tissue samples in the GTEx database. The pan-cancer analysis included 10,363 cancer tissue samples and 5,413 paracancerous and normal tissue samples in TCGA and the GTEx database. The RNAseq data in the TPM format were log2 transformed and compared between samples. The pan-
cancer groups included adrenocortical carcinoma (ACC); bladder urothelial carcinoma (BLCA); breast invasive carcinoma (BRCA); cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC); cholangiocarcinoma (CHOL); colon adenocarcinoma (COAD); lymphoid neoplasm diffuse large b-cell lymphoma (DLBCL); esophageal carcinoma (ESCA); glioblastoma multiforme (GBM); head and neck squamous cell carcinoma (HNSC); kidney chromophobe (KICH); kidney renal clear cell carcinoma (KIRC); kidney renal papillary cell carcinoma (KIRP); acute myeloid leukemia (LAML); brain lower grade glioma (LGG); liver hepatocellular carcinoma (LIHC); lung adenocarcinoma (LUAD); lung squamous cell carcinoma (LUSC); mesothelioma (MESO); ovarian serous cystadenocarcinoma (OV); pancreatic adenocarcinoma (PAAD); pheochromocytoma and paraganglioma (PCPG); prostate adenocarcinoma (PRAD); rectum adenocarcinoma (READ); sarcoma (SARC); skin cutaneous melanoma (SKCM); stomach adenocarcinoma (STAD); testicular germ cell tumors (TGCT); thyroid carcinoma (THCA); thymoma (THYM); uterus corpus endometrial carcinoma (UCEC); uterine carcinosarcoma (UCS); uveal melanoma (UVM).

Identification of DEGs
GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) was used to identify DEGs between testicular cancer samples (including seminoma and nonseminoma) and normal samples using R 3.2.3, Biobase 2.30.0, GEOquery 2.40.0, and Limma 3.26.8. The cutoff criterion was an adjusted p<0.05.

Software and Packages for Analysis Between Gene Expression and Clinicopathological Information
R software (version 3.6.3, for statistical analysis and visualization) was used for differential expression analysis. The R packages were ggplot2 (version 3.3.3, for visualization) and the basic R package. The pROC package (version 1.17.0.1, for analysis) and ggplot2 package (version 3.3.3, for visualization) were used for ROC curve construction. The Survminer package version 0.4.9, for visualization) and the survival package (version 3.2–10, for statistical analysis of survival data) were used for prognostic analysis, and the survival package (version 3.2–10, for statistical analysis of survival data) was used for Cox regression analysis. Supplementary data were obtained from the study of Liu et al. The clusterProfiler package (version 3.14.3) was used for gene set enrichment analysis (GSEA); the reference gene set was c2.cp.v7.2. symbols.gmt (curated), and the gene set database was MSigDB Collections. The false discovery rate (FDR) was <0.25, and adjusted p values <0.05 were considered significant. Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were based on clusterProfiler package [version 3.14.3] (for enrichment analysis) and org.Hs.eg.db package [version 3.10.0] (for ID conversion). The GSVA package [version 1.34.0] was used to analyze the correlation between gene expression and immune cells' infiltration. The immune infiltration algorithm was ssGSEA. Immune cells included aDC (activated DC); B cells; CD8 T cells; cytotoxic cells; DC; eosinophils; iDC (immature DC); macrophages; mast cells; neutrophils; NK CD56bright cells; NK CD56dim cells; NK cells; pDC (plasmacytoid DC); T cells; T helper cells; Tcm (T central memory); Tem (T effector memory); Tfh (T follicular helper); Tgd (T gamma delta); Th1 cells; Th17 cells; Th2 cells; Treg. Immune cell markers referred to the study of Bindea et al.

Protein Interaction Network Construction and Screening of Hub Genes
The construction of a PPI network relied on the STRING database (version 11.5 https://www.string-db.org/). The minimum required interaction score was 0.4 and maximum number of interactors was 5. Hubba plug-in in Cytoscape software was used to select top 20 hub nodes ranked by degree. Functional gene clusters were constructed via MCODE software. Network scoring degree cutoff was 2, node score cutoff was 0.2, K-core was 2, and Max. depth was 100.

Protein Immunohistochemical Expression Analysis
The extraction of immunohistochemical results relied on the Human Protein Atlas database (https://www.proteinatlas.org/). The scoring method of immunohistochemistry results was intensity plus quantity. For intensity, negative, weak, medium, and strong corresponded to 0, 1, 2, and 3 points, respectively; for quantity, negative, <25%, 25–75%, and >75% corresponded to 0, 1, 2, and 3 points, respectively.

Statistical Analysis
The Mann–Whitney U-test or t test was used to analyze the difference of gene expression in samples according to...
whether the data were normally distributed based on the normality test. The chi-squared test and Fisher’s exact test were used to analyze the difference between clinicopathological information groups according to whether the data met the conditions of theoretical frequency >5 or total sample size >40. The Log rank test was used in the Kaplan-Meier prognostic analysis, and the prognostic factors of TGCT were analyzed by Cox regression analysis. The Spearman correlation coefficient or the Pearson correlation coefficient was calculated in correlation analysis according to whether the result of the Shapiro–Wilk normality test satisfied normal distribution. Nonsignificance (ns) was set as p≥0.05; p<0.05 was considered statistically significant.

Results
Screening for Differentially Expressed Genes
The two datasets, GSE3218 and GSE1818, in the GEO database were selected to analyze the DEGs between testicular cancer and normal tissues. Specimens represented 101 adult male germ cell cancers (including choriocarcinoma, embryonal carcinoma, seminoma, teratoma, and yolk sac tumor), and 5 normal testis specimens were obtained from the GSE3218 dataset; there was a relatively high degree of discrimination between the two groups of data (Figure 1A). A total of 15,518 genes were identified, and 6,765 genes were differentially expressed (adjusted p<0.05, Figure 1B), of which 3,027 genes were upregulated and 3,738 were downregulated in cancer tissues compared with normal tissues (Figure 1C). Twenty male germ cell cancer specimens (including choriocarcinoma, embryonal carcinoma, intratubular germ cell tumor, seminoma, teratoma, and yolk sac tumor) and 3 normal testis specimens were in the GSE1818 dataset. The distinction between the two groups of data was relatively high (Figure 1D), and 14,379 genes were identified; 944 genes were differentially expressed (adjusted p<0.05, Figure 1E), of which 81 were upregulated and 863 were downregulated in cancer tissues compared with normal tissues (Figure 1F).

Screening for Hub Genes
The differentially expressed and co-expressed genes with log|fold change|>1 and adjusted p<0.05 were selected in

![Figure 1](https://doi.org/10.2147/IJGM.S342611)
the two datasets. Of the results, 32 genes were upregulated and 328 genes were downregulated in cancer tissues compared with normal tissues (Figure 2A and B); all DEGs are shown in Table 1. To screen for hub genes, we constructed a PPI network of these genes and selected the top 20 hub nodes ranked by degree with Hubba software (Figure 3A) and the top 2 functional gene clusters ranked by MCODE score with MCODE software (Figure 3B and C). Twelve hub genes were used as the target genes for the next analysis by taking the intersection; these genes were

Table 1 Differentially Expressed Genes

| Differential Expression | Gene Name |
|-------------------------|-----------|
| Upregulated (32)        | GSTP1, SEMA4C, ZFP36L2, CCND2, IER2, DERA, LY6E, CDC42SE1, ZMIZ1, CRIP1, AKR1A1, MSN, MDK, TSPAN9, TMEM51, LOXL2, RGS10, AFF1, CD53, TRIM2, GOLT1B, LYN, EMG1, LILRB2, MKI67, KRAS, HLA-G, CDH3, MRPL15, L2, ANX3A, NF2L2 |
| Downregulated (328)     | TSSK2, MSH4, CREM, RNF32, ADAM2, GK2, ADCY10, ACSL6, ACTL7A, PRM2, CCT6B, CRISP2, MS4A5, ZRPP, TCF11, ACSBG2, HCFC2, YBX2, RIBC2, CCIN, ACRV1, ODF1, LDHC, SYCP1, ART3, SSOX, ANKRD7, ODF2, PRKAA1, C10orf11, APOB, APH1B, TNP1, SLCA6A16, INI56, YFEL1, GGA1, PSORS1C1, TEX14, CYB5R2, TTC12, ADAM20, ZCHC14C, Celorfo, C2orf62, FEM1B, KIF2A, DPKX, ZBX8, AURKC, DZP1, GRAMD1C, CLGN, DZP3, SQX5, CDC4A, STAG3, SRPK2, RAD17, CASCI, ZCWPW1, NME5, PRG5, PIKR3, TBL1, MNS1, CDDC88A, LGALS8, HIRIP3, CEP63, LRPL8, MED7, FGFR10/RPA2, SFMBT1, LINTA, SLCA25, ASB9, PPAPC2, TNP2, SPAT7A, PCT2, ADAM20, ZAP8L3, RANBP9, TCF5L, TBL2, MYBL1, PHKG2, BCA2P9, SNS3, SHGL3, RMNDSB, NLRP3, CCHR1, NLRF141, ADO, ARL4A, RORA, AQP5, AFF4, C4R, PIBFI, SCCDH, KIF18A, AM22, ZNF821, MAP7, C1orf88, TRIM13, PLK4, B3GALT4, BAG1, TP2, EIF4G3, CST8, TBP, RNF141, NGLY1, C6orf106, CDDC53, CDKN3, VCK2, NEK4, DNAH7, SCAPER, NDUFAF1, TRIP12, STAM2, HSPA5, OPALH, CETN3, KHL1L, SLP49, SAP2, SKAP2, HAGL2, CHOL, USP32, PAFAH1B1, DYRK3, RANGSF, TEX1, YBX1, KCNH2, CLRX2, MEAI, MCF2, INPP1, ITCH, IFIB1, C1orf93, ST6GALNAC2, BA21A, DYNC2LI1, SCML1, BRD1, EFHC1, CBL1A, KATNA1, PAM2, AP4S1, SH3GL1, NNX3-1, GIN1, EBP41L3, HMOX2, WASF1, GMT3, COX7A2, SERPINAS5, PPP1R2, PPP1B, NUP155, FIG54, GTF2A2, INP3, PBK, TSP2, USF3A, PRM1, GOLGA1, CDH18, RBL2, ANKRD12, WDR60, SECISBP2, PDZD3, RPL9, USF2, PPR2R3, ASRGL1, BTBD1, KIF5B, ACAT1, SAP130, PRP4, ZDHHC13, NUP88, VAMP4, LIN37, CMAPS, NFX3, GPN3, PPM1G, DBF4, ACY1, ATPFI, KLHL12, MKRN2, PAIF1, CDDN2AIP, SF3A1, CTSF, MDC1, PMSF1, TORD7, RAB8B, FBX7O7, PSMG1, NUC28, MAK, N5F1, SEN2, CDC27, TPR, CAMLG, TMEM5, RNF139, RPL26L1, GMC1, ALG2, ATP7, ING2, ARCP5L, HIPK1, FERM1C, COP3, MAGEC2, ISYNA1, MKKS, PTG1, TCEA2, TFB2, PDZ2, ZWILCH, GPR18, RPL2, SKIV2L, USP1, LZTFL1, RNF6, MALT1, TAF10, MEF1E, HPRT1, TMBC14, MTX2, FIP1L1, PIP1R3B, USP25, PRPF18, EPS15, EXOSC4, KTN1, ATAD2, SMG6, UCHL3, GPR18, RAD1, FXR1, PROX1, PLEKHO1, TBC1D15, ARL3, ZNF688, OPTN, KCMF1, NCOA6, UBA1, HMMR, TEF2ZIP, C1orf166, HSPA14, STS19, ZNF516, TSN, CKLF, PNN1A1, KATNB1, Y11AP1, COP5, C1orf12, MRPL42, KNYU, PPR2RSC, GOLGA4, EZH2, NDUFA4, ACT16A, BCL2L14, FND3C3A, DIB1, SLCA24A, TIPARP, C1orf59, TIPIN, KPN4A, PKRF, UBA2, EIF1B, CTLL1, EXO2, CKS2, CDV3, TRIP13, CEPT0, AKAP8L, C10orf1, MAGEA4 |
PLK4, TRIP13, TPR, KIF18A, CDKN3, HMMR, PBK, PTTG1, CKS2, SYCP1, HSPA2, and MKI67. Only the mRNA of MKI67 was upregulated in testicular cancer tissues. To verify the accuracy of the differential expression of these genes, we used data from 154 TGCT tissue samples in TCGA and from 165 normal testicular tissue samples in the GTEx database to analyze the mRNA expression level. The results of gene differential expression trends were consistent with the results from the GEO database (Figure 4). In addition, we analyzed the protein expression of these genes in normal testicular tissues and in cancer tissues using the Human Protein Atlas database, and immunohistochemical scores of these proteins in the two groups were compared. The results showed that PLK4, TRIP13, TPR, HMMR, PBK, PTTG1, SYCP1, and HSPA2 were downregulated in cancer tissues (all p<0.05), and there was no significant difference in the expression of KIF18A, CKS2 and MKI67. CDKN3 data could not be found in the database (Figure 5).

**Diagnostic and Prognostic Value of Hub Gene mRNA**

The data of 139 TGCT samples with relatively complete clinicopathological information in TCGA were analyzed for the clinically related indicators (and clinicopathological information is shown in Table 2). The ROC curve was constructed for mRNA expression of these hub genes in TGCT and normal tissues. The results showed that the

![Figure 3](image-url)  
**Figure 3** Screen for hub genes. (A) Top 20 hub nodes ranked by degree with Hubba software; (B and C) top 2 functional gene clusters ranked by MCODE score with MCODE software.

![Figure 4](image-url)  
**Figure 4** mRNA expression of PLK4, TRIP13, TPR, KIF18A, CDKN3, HMMR, PBK, PTTG1, CKS2, SYCP1, HSPA2, and MKI67 in 154 TGCT cancer tissue samples in TCGA database and 165 normal testicular tissue samples in the GTEx database. ***p<0.001.
mRNA of PLK4, TRIP13, CDKN3, SYCP1, HSPA2, and MKI67 had high diagnostic value, with AUCs of 0.906, 0.926, 0.924, 0.986, 0.972, and 0.989, respectively (Figure 6). In the prognostic analysis, the progression-free interval was used as an endpoint because of the high censorship rate in patients with overall survival and disease-specific survival. Unfortunately, only high expression of CDKN3 (p=0.034, Figure 7) and HSPA2 (p=0.033, Figure 7) mRNA were factors for the poor prognosis of TGCT. We used univariate and multivariate Cox regression analyses to determine whether these hub gene mRNAs were independent risk factors for poor prognosis in terms of progression-free interval. The results showed that pathologic N stage and serum tumor markers were independent risk factors (all p<0.05). Although CDKN3 and HSPA2 mRNA also affected the prognosis of patients in univariate analysis (all p<0.05), the results of multivariate analysis did not support this conclusion (all p>0.05, Table 3).

Mechanism of Hub Genes in the Occurrence and Development of TGCT

Because the protein expression of KIF18A, CKS2 and MKI67 was not significantly different between cancer tissues and normal tissues, the remaining 9 genes were selected for functional analysis. Gene Ontology (GO) analysis results indicated that biological processes included meiotic nuclear division, nuclear division, meiotic cell cycle process, and organelle fission; the cellular components included male germ cell nucleus, germ cell nucleus, chromosome, centromeric region, and synaptonemal complex; and the molecular functions included heat shock protein binding, microtubule plus-end binding, ATPase activity, coupled, and hyaluronic acid binding (Figure 9). Unfortunately, no Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling was matched to these genes. Among all downregulated DEGs in cancer tissues, only TPR was downregulated in the higher T-stage group and the lymphatic metastasis group, and PLK4 was downregulated in the lymphatic metastasis group. The mRNA expression trend was consistent with the protein expression. Therefore, TPR and PLK4 were chosen for additional functional mechanism analysis. GSEA

Relationship Between Hub Genes and Clinicopathological Characteristics in Patients with TGCT

The results showed that, in samples with higher T stage, the mRNA of HMMR and PTTG1 was overexpressed, and the mRNA of TPR was downregulated (all p<0.05). In samples with lymphatic metastasis, the mRNA of TPR and PLK4 was downregulated (all p<0.05), and no differential gene mRNA was found between M0 and M1 samples (all p>0.05). In the group with higher serum tumor markers, the mRNA of CDKN3, HMMR, and PTTG1 was overexpressed, and no mRNA was downregulated (all p<0.05, Figure 8).

Figure 5 Protein immunohistochemical analysis of PLK4, TRIP13, TPR, KIF18A, HMMR, PBK, PTTG1, CKS2, SYCP1, HSPA2, and MKI67 in TGCT and normal tissues from the Human Protein Atlas database. Nonsignificant (ns): p≥0.05; *p<0.05; **p<0.01.
Table 2 Clinicopathological Information of TGCT Patients

| Characteristics                           | Levels | Overall |
|------------------------------------------|--------|---------|
| n                                        | 139    |         |
| Pathologic T stage, n (%)                |        |         |
| T1                                       | 80 (58)|         |
| T2                                       | 52 (37.7)|       |
| T3                                       | 6 (4.3)|         |
| Pathologic N stage, n (%)                |        |         |
| N0                                       | 51 (79.7)|       |
| N1                                       | 11 (17.2)|       |
| N2                                       | 2 (3.1)|         |
| Pathologic M stage, n (%)                |        |         |
| M0                                       | 120 (96.8)|       |
| M1                                       | 4 (3.2)|         |
| Radiation therapy, n (%)                 |        |         |
| No                                       | 113 (82.5)|       |
| Yes                                      | 24 (17.5)|       |
| Race, n (%)                              |        |         |
| Asian                                    | 6 (4.5)|         |
| Black or African American                |        |         |
| White                                    | 124 (92.5)|       |
| Age, n (%)                               |        |         |
| ≤30                                      | 67 (48.2)|       |
| >30                                      | 72 (51.8)|       |
| Serum tumor markers(5), n (%)            |        |         |
| S0                                       | 43 (34.4)|       |
| S1                                       | 41 (32.8)|       |
| S2                                       | 36 (28.8)|       |
| S3                                       | 5 (4)|         |
| Lymphovascular invasion, n (%)           |        |         |
| No                                       | 79 (58.5)|       |
| Yes                                      | 56 (41.5)|       |
| Family history of testicular cancer, n (%)|        |         |
| No                                       | 107 (87.7)|       |
| Yes                                      | 15 (12.3)|       |
| Laterality, n (%)                        |        |         |
| Left                                     | 74 (55.2)|       |
| Right                                    | 60 (44.8)|       |
| PFI event, n (%)                         |        |         |
| Alive                                    | 99 (71.2)|       |
| Dead                                     | 40 (28.8)|       |
| Age, median (IQR)                        |        |         |
|                                           | 31 (26, 37)|       |

Analysis of Hub Genes Involved in Immune Cell Infiltration in TGCT

After analyzing the relationship between hub genes and immune cell infiltration in TGCT, we found that immune cells with infiltration that positively correlated with TPR expression were Tcm, Th helper cells, Th2 cells, and Tem (all p<0.05, Figure 10C); those that negatively correlated with TPR expression were aDC, cytotoxic cells, NK CD56dim cells, TReg, Tgd, CD8 T cells, mast cells, NK cells, neutrophils, macrophages, Th1 cells, pDC, DC, iDC, and NK CD56bright cells (all p<0.05, Figure 10C). Immune cells with infiltration degrees that positively correlated with the expression of PLK4 were Tcm, Th2 cells, and T helper cells (all p<0.05, Figure 10D), and those with infiltration degrees that negatively correlated with the expression of PLK4 were aDC, TFH, Th17 cells, B cells, NK CD56dim cells, TReg, Tgd, cytotoxic cells, eosinophils, CD8 T cells, NK cells, NK CD56bright cells, neutrophils, mast cells, Th1 cells, macrophages, DC, pDC, and iDC (all p<0.05, Figure 10D).
Overall, 33 related samples of cancer tissues were included in the gene expression in pan-cancer analysis. The results showed that, compared with normal tissues, TPR was differentially expressed in 24 cancer tissues, of which it was upregulated in all cancers except ACC, KICH, OV, TGCT, and THCA (all p<0.05, Figure 11A). Compared with normal tissues, PLK4 was downregulated in TGCT and LAML, and was upregulated in almost all the other cancers (all p<0.05, Figure 11B).

Discussion

Although the incidence of testicular cancer is increasing year by year, its proportion is not high among all tumors; however, it seriously affects male health, including fertility, especially that of young men.1–4 Therefore, finding effective diagnosis and prognostic indicators and exploring its pathogenesis to seek treatment targets are important research goals. However, the current basic and translational research on testicular cancer is relatively scarce. Using bioinformatics analysis technology to assess the sequencing results of tissue samples in public databases, such as the GEO database, TCGA, and the GTEx database, provides a convenient way to obtain gene expression data and clinical information about patients. In this study, we used the tissue sequencing information of testicular cancer in the two datasets GSE3218 and GSE1818 from the GEO database to select 32 upregulated genes and 328 downregulated genes in testicular cancer tissues to compare with normal tissues through bioinformatics analysis. For additional study, we constructed a PPI network and screened out 12 hub genes as research targets: PLK4, TRIP13, TPR, KIF18A, CDKN3, HMMR, PBK, PTTG1, CKS2, SYCP1, HSPA2, and MKI67. The mRNA expression of these DEGs was verified in TGCT tissues from TCGA and in normal testis tissues from the GTEx database. We found via immunohistochemical analysis that mRNA expression was not completely consistent with the
protein expression. For 3 genes, protein expression was not differential between cancer and normal tissues. One possible reason is that the process of mRNA translation into protein may be regulated by noncoding RNA, including circRNA, lncRNA, and miRNA,22–24 or may be subject to epigenetic modifications, such as m6A modification.25 Another reason may be the complex composition of normal testicular tissue, including spermatogonia, spermatocytes, and more. There was cellular heterogeneity in the expression of these genes in testicular tissues, but tumor cells are usually derived from the same type of cell. For example, CKS2 expresses in round or early spermatids in normal testicular tissues and not in other cell types, but it is uniformly expressed in cancer tissues.

mRNA carries genetic information and is a direct template for instructing protein biosynthesis. It links the genetic information in DNA with the translation and expression of proteins, and it plays a vital role in all cells, including tumor cells. The mRNA expression level can reflect the expression of hub genes to a certain extent and often can distinguish diseases and disease subgroups as a gene transcription manuscript.26–28 Therefore, mRNA has the potential to be used as a diagnostic and prognostic molecular of diseases, including cancer. The AUCs of mRNA of PLK4, TRIP13, CDKN3, SYCP1, HSPA2, and

### Table 3 Associations Between Clinicopathological Characteristics and PFI in TGCT Patients Using Cox Regression Analysis

| Characteristics                        | Total(n) | Univariate Analysis | Multivariate Analysis |
|----------------------------------------|----------|---------------------|----------------------|
|                                        |          | Hazard Ratio (95% CI) | p value  | Hazard Ratio (95% CI) | p value  |
| Pathologic T stage (T2 & T3 vs T1)    | 138      | 1.158 (0.621–2.160) | 0.645    |                       |          |
| Pathologic N stage (N1 & N2 vs N0)    | 64       | 9.292 (1.261–68.450) | 0.029*   | 12.115 (1.595–92.045) | 0.016*   |
| Pathologic M stage (M1 vs M0)         | 124      | 0.000 (0.000–Inf)   | 0.996    |                       |          |
| Race (White vs Asian & Black or African American) | 134 | 0.477 (0.168–1.351) | 0.163    |                       |          |
| Age (>30 vs ≤30)                      | 139      | 0.697 (0.373–1.301) | 0.257    |                       |          |
| Serum tumor markers (S) (S2 & S3 vs S0 & S1) | 125 | 2.023 (1.075–3.804) | 0.029*   | 2.643 (1.077–6.486) | 0.034*   |
| Radiation therapy (Yes vs No)         | 137      | 0.873 (0.386–1.978) | 0.745    |                       |          |
| Lymphovascular invasion (Yes vs No)   | 135      | 1.364 (0.733–2.538) | 0.327    |                       |          |
| Laterality (Right vs Left)            | 134      | 1.231 (0.639–2.371) | 0.334    |                       |          |
| History of undescended testis (Yes vs No) | 132 | 0.560 (0.219–1.434) | 0.227    |                       |          |
| PLK4 (High vs Low)                    | 139      | 1.214 (0.651–2.265) | 0.542    |                       |          |
| TRIP13 (High vs Low)                  | 139      | 1.280 (0.685–2.391) | 0.440    |                       |          |
| TPR (High vs Low)                     | 139      | 1.005 (0.537–1.881) | 0.988    |                       |          |
| KIF1BA (High vs Low)                  | 139      | 1.242 (0.664–2.323) | 0.497    |                       |          |
| CDKN3 (High vs Low)                   | 139      | 2.044 (1.054–3.965) | 0.034*   | 0.725 (0.277–1.898) | 0.512    |
| HMMR (High vs Low)                    | 139      | 1.164 (0.618–2.193) | 0.638    |                       |          |
| PBK (High vs Low)                     | 139      | 1.044 (0.561–1.943) | 0.893    |                       |          |
| PTTG1 (High vs Low)                   | 139      | 1.322 (0.704–2.482) | 0.386    |                       |          |
| CKS2 (High vs Low)                    | 139      | 1.451 (0.774–2.720) | 0.246    |                       |          |
| SYCP1 (High vs Low)                   | 139      | 0.785 (0.420–1.465) | 0.446    |                       |          |
| HSPA2 (High vs Low)                   | 139      | 2.145 (1.062–4.332) | 0.033*   | 1.262 (0.479–3.327) | 0.637    |
| MKI67 (High vs Low)                   | 139      | 1.204 (0.645–2.247) | 0.559    |                       |          |

*Note: p values with “*” were considered statistically significant.*
MKI67 in this study were all above 0.9, and this measure can distinguish cancer tissues from normal tissues well. Therefore, these genes have good potential to be diagnostic molecular markers. In the prognostic analysis, mRNA for CDKN3 and HSPA2 was a poor prognostic factor for the progression-free interval in TGCT; however,
the Cox regression analysis did not indicate CDKN3 and HSPA2 mRNA expression but suggested that pathologic N stage and serum tumor markers were independent risk factors for the poor prognosis of TGCT. These results were consistent with the conclusions that lymphadenectomy can bring benefits to patients and that serum tumor markers can predict the prognosis of patients. 29,30

In this study, we found that the mRNA expression of almost all hub genes, except MKI67, in cancer tissues was lower than that in normal tissues; but many of the genes were upregulated in higher tumor grades. For example, the mRNA of HMMR and PTTG1 was overexpressed in samples with higher T stages, and the mRNA of CDKN3, HMMR, and PTTG1 was overexpressed in the group with higher serum tumor markers. A possible explanation is that downregulation of genes is the initiating factor of TGCT, but the expression level gradually increases during the disease progression or as a result of other regulation of
oncogene expression. Another possible explanation is that these genes may be related to the infiltration of immune cells, whereas the result of tissue sequencing referred to all the components, and the high expression of these genes in immune cells leads to genes higher expression. Immune cells also play an important role in the disease process of cancer. Innate immune cells are composed of natural killer (NK) cells, eosinophils, basophils, and phagocytes (including mast cells, neutrophils, monocytes, macrophages, and dendritic cells), and they participate in cancer suppression by directly killing tumor cells or triggering an adaptive immune response. The adaptive immune system works through lymphocytes (including B cells and T cells), among which B cells play a major role in humoral immune responses, whereas T cells participate in cell-mediated immune responses. The expression of genes may affect the immune cells’ infiltration in cancer. In this study, we found that the expression of the selected genes PLK4 and TPR was related to the content of specific immune cells, which may be used as interference targets for immunotherapy.

To analyze the role of hub genes in TGCT, we performed GO analysis and KEGG analysis on these genes. Although no signal pathway was matched to KEGG analysis, we obtained some important GO analysis results. For example, we found that biological processes involved organelle division and cell cycle, and molecular functions involved chromosome and centromeric region. The proper coordination of centromeric chromatin dynamics is essential in the cell cycle, and previous studies have shown that the level of centromeric transcripts is elevated in some tumors. Biological processes involve heat shock protein (HSP) binding. HSP is a large class of proteins involved in protein folding and maturation. Its expression is induced by heat shock or other stress factors, and it plays an important role in cell proliferation, differentiation, and carcinogenesis. HSP is the research target for cancer development and treatment. Therefore, these biological functions may be the mechanism of action of these hub genes in testicular cancer. To avoid ambiguity, we selected PLK4 and TPR as the target genes that may be involved in the mechanism of TGCT for additional research; the mRNA and protein of each gene were downregulated in cancer tissues and in higher-grade cancer tissues. GSEA was used to analyze the signal pathways in which a single gene may participate. The results showed that PLK4 may be involved in the signaling pathways NABA_MATRISOME and REACTOME_DEVELOP...
MENTAL BIOLOGY and that TPR may be involved in the signaling pathways NABA MATRISOME and REACTOME METABOLISM OF LIPIDS to regulate tumor progression. Lipid metabolism reprogramming not only can be used as a molecular marker of tumor but also can affect the proliferation of tumor cells. Whether TPR functions through lipid metabolism is worthy of further study.

Polo-like kinase 4 (PLK4) is a serine threonine kinase that is located in the centrioles throughout the cell cycle and is essential for centriole replication. It promotes cell proliferation, invasion, and migration in cancers through its important role in the centrioles. Interestingly, PLK4 was specifically downregulated in TGCT and LAML, but its expression was relatively high in normal testicular tissue compared to other organ tissues, even higher than other malignancies. This phenomenon shows that PLK4 mRNA can be effectively used as a diagnostic indicator of TGCT but also indicates that PLK4 may function in TGCT through a unique mechanism. High PLK4 expression in testis tissue may result from the spermatogonia that can continue to duplicate as a kind of cell with active spermatogenesis function. If the balance of PLK4 expression is disrupted (eg, by a gene mutation), tumor development may occur. TPR encodes a large coiled-coil protein that forms inner nuclear filaments that attach to the inner surface of the nuclear pore complex (NPC). The TPR protein directly interacts with several components of NPC and is necessary for the nuclear export of mRNA and some proteins. There is no research on this protein in tumor diseases, so it is a valuable topic for study in TGCT. TPR is downregulated in several tumors with typical endocrine function, such as ACC, OV, TGCT, and THCA, so the relationship between TPR expression and hormone level in tumors is an interesting topic that is worthy of in-depth study.

Conclusion

PLK4, TRIP13, TPR, KIF18A, CDKN3, HMMR, PBK, PTTG1, CKS2, SYCP1, HSPA2, and MKI67 may be important genes for the development of TGCT. PLK4 and TPR in particular may affect the disease process by participating in a variety of signal pathways and regulating tumor-related genes. Influencing immune cell infiltration may be one way to regulate the disease. The hub genes in this article were screened out by bioinformatics technology; whether they can regulate the biological function of TGCT cells must be determined with additional experimental studies. The specific mechanism also must be verified and explored in more detail.

Abbreviations

TGCT, testicular germ cell tumor; DEGs, differentially expressed genes; GSEA, Gene Set Enrichment Analysis; FDR, false discovery rate; PPI, protein–protein interaction; AUC, area under the curve; PFI, progression-free interval; US, the United States; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; HSP, heat shock protein.

Acknowledgments

This study benefited from the Cancer Genome Atlas (TCGA) database, the Genotype-Tissue Expression (GTEX) database, and the Gene Expression Omnibus (GEO) database.

Funding

This work was supported by the Natural Science Foundation of Gansu Province (grant number 18JR5RA404); Natural Science Foundation of Gansu Province (grant number 20JR5RA601); and in-hospital projects of The 940 Hospital of Joint Logistics Support Force of Chinese PLA (grant numbers 2021yxky057 and 2021yxky034).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–386. doi:10.1002/ijc.29210
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424. doi:10.3322/caac.21492
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7–30. doi:10.3322/caac.21590
4. Miller KD, Fidler-Benaoudia M, Keegan TH, et al. Cancer statistics for adolescents and young adults, 2020. CA Cancer J Clin. 2020;70(6):443–459. doi:10.3322/caac.21637
5. Leão R, Ahmad AE, Hamilton RJ. Testicular cancer biomarkers: a role for precision medicine in testicular cancer. Clin Genitourin Cancer. 2019;17(1):e176–e183. doi:10.1016/j.clgc.2018.10.007
6. kondagunta GV, Bacik J, Donadio A, et al. Combination of paclitaxel, ifosfamide, and cisplatin is an effective second-line therapy for patients with relapsed testicular germ cell tumors. J Clin Oncol. 2005;23(27):6549–6555. doi:10.1200/JCO.2005.19.638
7. Pulendran B, Davis MM. The science and medicine of human immunology. Science. 2020;369(6511). doi:10.1126/science.aaq4014
8. Chovanec M, Mardiak J, Mego M. Immune mechanisms and possible immune therapy in testicular germ cell tumours. Andrology. 2019;7(4):479–486. doi:10.1111/andr.12656
