Identification of aberrantly methylated differentially expressed genes in papillary thyroid carcinoma using integrated bioinformatic analysis

Qingxian Li
Guangdong Medical University Affiliated Longhua Central Hospital
Weiyi Cai
Guangdong Medical University Affiliated Longhua Central Hospital
Jianhong Chen
Huizhou First Maternal and Child Health Care Hospital
Jie Ning (790192539@qq.com)
Guangdong Medical University Affiliated Longhua Central Hospital

Research Article

Keywords: DLG2, papillary thyroid carcinoma, Microarray, RNA-sequencing, methylation

Posted Date: April 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-410194/v1

Abstract

Background: DNA methylation has been reported as one of the most critical epigenetic aberrations during the tumorigenesis and development of papillary thyroid carcinoma (PTC). Although PTC has been explored by gene expression and DNA methylation studies, the regulatory mechanisms of the methylation on the gene expression was poorly clarified.

Results: In this study, the comparisons between PTC and NT revealed 4995 methylated probes and 1446 differentially expressed transcripts cross-validated by The Cancer Genome Atlas (TCGA) database. The integrative analysis between DNA methylation and gene expression revealed 123 and 29 genes with hypomethylation/overexpression and hypermethylation/downexpression correlation, respectively. The DNA methylation pattern of seven selected CpGs (A: UNC80-cg04507925; B: TPO-cg09757588; C: LHX8-cg11842415; D: DLG2-cg16986720; E: FOXJ1-cg20373432; F: PALM2-cg21204870; G: IPCEF1-cg24635109, of which the candidate promoter CpG sites were preliminarily identified with the least absolute shrinkage and selection operator (LASSO) regression analysis. Then, the risk prognosis model was constructed by stepwise regression analysis. Furthermore, the receiver operating characteristic (ROC) and nomogram based on the verified independent prognostic factors was established for the prognostic prediction showed that it was able to predict 3-, 5-, and 7-year survival accurately. Kaplan-Meier survival estimate demonstrated that low DLG2 expression and DLG2-cg16986720 hypermethylation were independent biomarkers for OS. From the comprehensive meta-analysis, the combined Standardised Mean Difference (SMD) of DLG2 was 0.94 with 95% CI of (0.46,1.43), indicating that less DLG2 was expressed in the PTC tissue than in the normal tissue ($P<0.05$). Bisulfite sequencing PCR also showed that DLG2 methylation was higher in tumor group than in normal group. Components of immune microenvironment were analyzed using TIMER, and the correlation between immune cells and DLG2 was found to be distinct across cancer types. Based on Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, DLG2 was implicated in pathways involved in immunity, metabolism, cancer, and infectious diseases. PCT patients with DLG2-cg16986720 hypermethylation showed significantly short survival rates in progression- free survival concomitant with reduced infiltration of myeloid dendritic cells.

Conclusions: The current study validated that DLG2 was lowly expressed in PTC. More importantly, DLG2 hypermethylation might function as a latent tumor biomarker in the prognosis prediction for PTC. The results of bioinformatics analyses may present a new method for investigating the pathogenesis of PTC. DNA methylation loss in non-promoter, poor CGI and enhancer-enriched regions was a significant event in PTC. In addition to the promoter region, gene body and 3'UTR methylation have also the potential to influence the gene expression levels (both, repressing and inducing). The integrative analysis revealed genes potentially regulated by DNA methylation pointing out potential drivers and biomarkers related to PTC development.

Background

DNA methylation has been reported as one of the most critical epigenetic aberrations during the tumorigenesis and development of papillary thyroid carcinoma (PTC) [1]. Since the implementation of salt iodization policy, the incidence of iodine-deficient goiter has been significantly reduced, but the overall prevalence of thyroid diseases is still as high as 30% [2]. At present, thyroid cancer has become the most common malignant tumor in the department of Metabolism and Endocrinology, accounting for about 1% of systemic malignant tumors, and its global incidence is on the rise [3-4]. Which in PTC incidence is given priority to, make up about 70% of adults and children almost all of PTC, slow growth better prognosis, but prone to early lymph node metastasis and recurrence [5]. To explore the pathogenesis of PTC and search for specific tumor markers is of great practical significance for the early diagnosis, treatment and prognosis of PTC.

At present, the common PTC-related genes studied at home and abroad include CK19, 34bE12, HBME-1, VEGF, PTEN, Survivin, COX-2, Galectin-3, TPO, RET, etc. [6-8]. Many studies have shown that these tumor-related genes have certain value in the diagnosis and differential diagnosis of PTC. The first studies in this field evaluated the most common mutations and rearrangements that could identify approximately 60% of the thyroid carcinomas with indeterminate results based on cytological diagnosis [9,10]. The addition of several other mutations and gene fusions using next-generation sequencing panels brought an increase in the sensitivity up to 91% [11]. DNA methylation is one of the major epigenetic modifications, which is closely related to gene expression regulation, embryonic development regulation, genomic imprinting, female X chromosome inactivation, and cell differentiation and proliferation. Existing studies have confirmed that epigenetic silencing caused by DNA methylation plays an important role in the occurrence and development of tumors [12]. DNA methylation of the gene promoter region typically inhibits gene expression. Most methylated CpG islands are located within genes or intergenic regions, while less than 3% in the gene promoter region. Intra- or intergene DNA methylation may regulate gene expression [13]. Many studies in recent years have shown that DNA methylation is closely related to tumor progression.

In this study, we used The Cancer Genome Atlas (TCGA) data as training set aimed to explore novel survival-associated promoter CpG dinucleotide sites for prognostic prediction in PTC patients. The integrative analysis between gene expression and DNA methylation revealed 123 and 29 genes with hypomethylation/overexpression and hypermethylation/downexpression correlation, respectively. First, the differential methylation CpG (dmCpG) sites were screened in TCGA dataset, of which the candidate promoter CpG sites were preliminarily identified with the least absolute shrinkage and selection operator (LASSO) regression analysis. Second, the promoter CpG-based signature was constructed with stepwise regression analysis. Receiver operating characteristic (ROC) and nomogram showed that it was able to predict 3-, 5-, and 7-year survival accurately. Meanwhile, a Kaplan-Meier analysis with a log-rank test was used to assess the survival difference between hypermethylated group hypomethylated group. Furthermore, we investigated the correlations between DLG2-cg16986720 hypermethylation and clinical characteristics through data collected from Gene Expression Omnibus (GEO) microarrays, the relevant literature. Bisulfite sequencing PCR analyses also used to determine the clinical role of DLG2 in PTC. At

last, we investigated the biological and immune function of the CpG-sites target gene by Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses and TIMER.

Materials And Methods

Data acquisition and processing

We downloaded the genome methylation and expression data of 611 samples from the TCGA database (https://portal.gdc.cancer.gov/), including 59 tumor samples and 59 matched normal samples. Further methylation groups and expression profiles were visualized with the principal component analysis (PCA) respectively.

Data processing and identification of DEGS and DMGS

To explore the differentially expressed genes (DEGs) and differentially methylated genes (DMGs), we applied the “limma” packages for processing RNA-seq datasets and ChAMP R package for methylated datasets. The DEGs were screened with criteria |logFC| > 0.2 and adj-P-value < 0.05, while DMGs were identified with FDR < 0.05 and |logFC| > 0.2. Further screened and visualized with the pheatmap R package. To illustrate the intersection among DEGs, DMGs, and oncogenes, an Venn diagram program was employed. As a result, upregulated hypomethylated oncogenes as well as downregulated expression genes were filtered out.

Construction of PTC-specific dmCpG sites signature

LASSO regression analysis was further performed to explore the key dmCpG sites with glmnet R package. Finally, 7 dmCpG sites were identified and stepwise regression analysis was applied to optimize the model. Under the optimal situations, the dmCpG sites and corresponding coefficients were presented, and the formula for calculating the prognostic index (designated as risk score) based on methylation levels of dmCpG sites was obtained.

Independent Prognostic Prediction Analysis

We utilized the time ROC package to conduct the receiver operating characteristic (ROC) curve to show the 3-, 5-, and 7-year OS prediction. Area under the curve (AUC) of the ROC curve was also provided. The nomogram of risk score, age, and TNM (tumor, node, metastasis), was also constructed. Meanwhile, a Kaplan-Meier analysis with a log-rank test was used to assess the survival difference between hypermethylated group hypomethylated group.

Meta-analysis

A microarray search of DLG2 in PTC was conducted in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) with the following keywords: (Methylation OR “DNA Methylation” OR “Methylation, DNA” OR "Methylations, DNA") AND (Thyroid OR "Thyroid Gland" OR "Papillary Thyroid")

AND ("Cancer, Papillary Thyroid" OR "Cancers, Papillary Thyroid" OR "Papillary Thyroid Cancer" OR "Papillary Thyroid Cancers" OR "Thyroid Cancers, Papillary" OR "Thyroid Carcinoma, Papillary" OR "Carcinoma, Papillary Thyroid" OR "Carcinomas, Papillary Thyroid" OR "Papillary Thyroid Carcinomas" OR "Thyroid Carcinomas, Papillary" OR "Papillary Carcinoma Of Thyroid" OR "Thyroid Carcinomas, Papillary" OR "Papillary Carcinoma Of Thyroid" OR "Papillary Thyroid Carcinoma" OR "Familial Nonmedullary Thyroid Cancer" OR "Nonmedullary Thyroid Carcinoma" OR "Carcinoma, Nonmedullary Thyroid" OR "Carcinoma, Nonmedullary Thyroid" OR "Carcinomas, Nonmedullary Thyroid" OR "Nonmedullary Thyroid Carcinomas" OR "Thyroid Carcinoma, Nonmedullary" OR "Thyroid Carcinomas, Nonmedullary" OR PTC) AND "Homo sapiens". As this is the first study to investigate the prognostic role of DLG2-cg16986720 methylation in PTC, no previous studies were obtained from the databases. Therefore, we could utilize the meta- analysis to assess the overall prognostic significance of DLG2-cg16986720 methylation in patients with PTC from 5 datasets. Combined HR and 95% CI were calculated to evaluate the correlation of DLG2-cg16986720 hypermethylation with prognosis of PTC patients. The heterogeneity across four datasets was assessed by the Q test (I^2 statistics). A fixed-effects model would be selected for combination if no obvious heterogeneity (I^2< 50%). Otherwise, a random-effects model would be applied.

The criteria for inclusion were as follows: (1) patients diagnosed with PTC and its subtypes were investigated; (2) cancerous and noncancerous samples were involved; (3) the healthy and malignant groups included at least three samples in the form of tissue, blood, or plasma; (4) and the expression profiling data for DLG2 were available. Related studies were retrieved from the PubMed, Google Scholar, China National Knowledge Infrastructure (CNKI), Chongqing VIP electronic (VIP) and Chinese Wanfang databases.

Bisulfite sequencing PCR GSE86961

Bisulfite sequencing PCR for the current study, 100 matched samples were provided by the Department of Endocrinology of the Shenzhen Longhua District Central Hospital. Formalin fixation and paraffin embedding (FFPE) were performed to preserve the specimens. This aspect of the research was authorized by the hospital’s ethics committee. Next, the cg16986720 expression in 125 paired clinical samples was detected by bisulfite sequencing PCR with the Applied Biosystems. The DLG2-cg16986720 sequence was as follows: Left primer ATTTTTAGTAAAAGGAGGGTGGAAA Right primer ATCCTTCATAAACCCCAAAATAATC. The formula for 2-Δcq was used for the calculation of the DLG2 expression value.

Functional analysis for promising target CpG sites

The GO vocabularies were enriched by Metascape (http://metascape.org/gp/). The functional annotation of the underlying target genes was then elucidated by KEGG pathway analysis with Metascape tool.

Explorations of associations between promising target CpG sites and immune infiltrates

TIMER is a comprehensive resource for systematic analysis of immune infiltrates across multiple malignancies. The abundances of six immune infiltrates (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells) are estimated by the statistical method, which is validated by pathological estimations. This web server allows investigators to input function-specific parameters, with resulting figures dynamically displayed to conveniently access the tumour's immunological, clinical, and genomic features. Based on the public resources, we could discover the identified immune signature with tumour-infiltrating immune cells. Pearson's correlation coefficient and the estimated P value were calculated to evaluate the associations between tumor group and normal group.

Results

DNA methylation and gene expression profiles in PTC

DEGs and DMGs in endometrial cancers

The PCA of genome methylation and expression data of 611 samples were presented in Figure 1 A, B. The DMGs and DEGs were presented in Figure 2 A, B. We obtained 1718 DEGs, including 824 upregulated and 894 downregulated. A total of 2621 DMGs were also obtained, including 510 hypermethylated genes and 2111 hypomethylated genes were also screened out. The variation tendency of the methylation level of dmCpG sites in heatmap was consistent with their coefficients in prognostic signature (Figure 2C, D).

Through Venn analysis, we identified the genes with overlapped expression between the upregulated genes and hypomethylated genes. we searched 123 downregulated hypermethylated genes and 29 upregulated hypomethylated genes. Subsequently, we also identified the genes with overlapped expression between downregulated genes and hypermethylated genes. Correspondingly, we retrieved seven upregulated hypomethylated oncogenes (Figure 3A).

The Construction of dmCpG-Based

LOSS regression analysis was further applied to screen the candidate dmCpG sites for constructing the prognostic signature (Figure 3B, C). Seven candidate dmCpGs sites were screened in the stepwise analysis. These prognostic dmCpG sites were finally included in the model for calculating the risk score. In the optimal model, the AIC was 901.8, and concordance index was 0.77. The risk score was calculated with the following formula: risk score = cg04507925× -1.3175 + cg09757588× 0.2571 + cg11842415× -3.1059 +cg16986720× 9.3670 + cg20373432× -6.5045 + cg21204870× 4.2238 + cg24635109× -1.8893.

Prognostic Prediction Analysis

Kaplan–Meier survival analysis of the risk score indicated the survival probability of patients in tumor and normal groups (P<0.001, Figure 3D). The AUC value indicated good prognostic prediction efficacy. The nomogram was plotted for the TCGA cohort to calculate the risk score and predict 3-, 5-, and 7-year OS of PTC patients (Figure 4A). Further, 3-, 5-, and 7-year ROC curves of risk score were plotted, with the

AUCs of 0.758, 0.784, and 0.854, respectively (Figure 4B, C, D). The optimal model was released online at the following: https://caiweiyi123.shinyapps.io/DyNom/.

Results of meta-analysis of gene expression omnibus datasets

A total of 6 microarrays from the GEO database met the entry criteria. The features of the included GEO datasets are depicted in Table 1. Of the microarrays, 5 were obtained from tissue (GSE53051, GSE51090, GSE86961, GSE89093 and GSE97466), and 1 were derived from blood (GSE121377). In addition, the expression data from the PTC and control groups were collected on the basis of the GEO database. With respect to the data from the tissue samples, the PTC groups had a significantly lower level of cg16986720 expression than the control groups in GSE53051, GSE51090, GSE86961, GSE97466 and GSE121377 ($p < 0.0001$, $p < 0.0001$, $p < 0.0001$, $p < 0.0001$, $p < 0.0001$, $p < 0.0001$, and $p < 0.0001$, respectively (Fig. 2). In contrast, regarding the data from the blood samples, no notable distinction in cg16986720 expression was detected between the PTC and the control groups in the GSE89093($p = 0.2098$).

A meta-analysis was conducted on the basis of the 19 included microarrays from the GEO database. The results are demonstrated in Fig. 5. Given the apparent heterogeneity ($p <0.01$, $I^2 = 96\%$), a random effects model was applied, and remarkable down-regulation (SMD =0.94; 95% CI 0.46,1.43; $p = 0.000$) of cg16986720 was found in the PTC groups. A sensitivity analysis was later conducted to explore whether a particular microarray played a vital role in significant heterogeneity (Figure 5B). To further clarify the heterogeneity source, a subgroup analysis was performed. It was based on multiple characteristics: sample source (tissue vs. blood). As is illustrated in Figure 5C, significant heterogeneity was observed in the tissue subgroup ($I^2= 44\%$, $p= 0.1$).

Bisulfite sequencing PCR

The DLG2 methylation and its significance in PTC prognoses using bisulfite sequencing PCR, the clinical methylation value of DLG2 in 125 matched tissues was evaluated. The PTC samples exhibited a significantly hypermethylation level of DLG2 than the non-cancerous samples (0.65±0.1539 vs. 0.48±0.133, $p < 0.001$).

Relationships of DLG2 with immune cells

Given the important roles of infiltrating immune cells in the tumor microenvironment, we integrated the comprehensive analysis of immune signatures combined with immune infiltrates. Based on the TIMER database, we excluded the samples with a calculated P value <0.05 to guarantee the accuracy of the analysis, and we illustrated the results in a box plot, where the differential immune cells were annotated with various colours and the sum of immune fractions in each sample was equal to one (Figure 6). DLG2-cg16986720 hypermethylation showed significantly short survival rates in progression- free survival concomitant with reduced infiltration of myeloid dendritic cells.

Functional analysis for promising target CpG sites

The characteristics of involved dmCpG sites were extracted, including located gene and regions were explored to further understand the function mechanism of methylation. A total of 32 screened DMRs were annotated, and the KEGG pathway enrichment analysis was performed. The enriched pathways were provided in Fig. 6D. including Oncocytoma, renal, stuttering, Thyrotoxicosis, Hypersomnia, Thyroid Diseases, Sleep Apnea, Obstructive, Abdomen distended, Autoimmune thyroid disease (AITD), Hypothyroidism, Dysautonomia, Thyroid Nodule, etc., pathway showed most involved DMRs.

Discussion

Methylation of tumor cells is characterized by the overall methylation level of genomic DNA, which is usually lower than that of normal cells. At present, there are still few studies on the gene DNA methylation level of patients with PTC cancer, and the detection of PTC-related genes based on DNA methylation has potential research significance. The purpose of this study was to characterize the DNA methylation pattern of PTC and to elucidate its effect on gene expression deregulation. The hierarchical clustering analysis showed a substantial overlapping between transcripts and methylation profiles.

Among the 29 differentially methylated genes, seven (UNC80-cg04507925, TPO-cg09757588, LHX8-cg11842415, DLG2-cg16986720, FOXJ1-cg20373432, PALM2-cg21204870, IPCEF1-cg24635109) were selected for optimize the model. From these genes, only DLG2 was found hypomethylated (promoter region) and OS of our gene list. DLG2 has been reported as an important candidate involved in the progression of thyroid cancer [14]. Additionally, one of the isoforms may be a good candidate for molecular differential diagnosis of renal oncocytoma [15]. Much of what is known about DLG cellular function has come from a series of elegant studies of DLG mutations in Drosophila melanogaster and in Caenorhabditis elegans. One of the key findings from these studies was that DLG had tumor suppressor properties in Drosophilalarvae: loss of DLG function was associated with excessive over proliferation and neoplastic transformation of epithelial cells within the imaginal discs [16]. Disruption of Drosophila DLG also resulted in acute disorganization of epithelial structure with disruption of intercellular junction formation and a loss of apico-basal cell polarity [17], whereas in C.elegans the single homologue of Drosophila DLG, homologue of Drosophila DLGs required for proper adherens junction formation.

From these genes, only DLG2-cg16986720 was in promoter regions, the other were in gene bodies. Gene bodies are reported as having a limited number of CpG islands and broadly methylated. Interestingly, this region harbors multiple repetitive and transposable elements [18]. However, the biological significance of DNA methylation in this region is poorly clarified. Gene body methylation does not necessarily block the transcription as observed in promoter regions, but might increase the transcriptional activity, stimulate the transcription elongation, and impact in the splicing [19-21]. Accordingly, 80% of the probes presenting positive correlation between methylation/expression were annotated in gene body or 3’UTR. Interestingly, differentially methylated probes exclusively mapped in body or in promoter regions presented a similar proportion of differentially expressed genes (14 and 18%, respectively) [22, 23]. Furthermore, an opposite

relation between methylation and expression (hypomethylation/overexpression or hypermethylation/downexpression) of probes mapped exclusively in the promoter or in body gene regions were also detected (73 and 93%, respectively). These findings suggest that body gene methylation is a process involved in gene expression regulation, similar to those described in the promoter regions. DNA methylation loss in non-promoter, poor CGI and enhancer-enriched regions was a significant event in PTC. In addition to the promoter region, gene body and 3'UTR methylation have also the potential to influence the gene expression levels (both, repressing and inducing). The integrative analysis revealed genes potentially regulated by DNA methylation pointing out potential drivers and biomarkers related to PTC development.

Our meta-analysis containing 522 PCT patients from 6 different databases further demonstrated that DLG2-cg16986720 hypermethylation was an independent prognostic variable of OS in PCT patients. Our analyses also revealed that the levels of immune infiltration and a list of immune markers were significantly correlated with DLG2 expression in PCT. To our knowledge, our analyses provide novel insights into the prognostic role of DLG2 and potential role of DLG2 in the tumor immunology of PCT. The DNA methylation located at gene promoter would generally suppress gene transcription, thus down-regulating the expression level of target genes. Considering the benefit of early and non-invasive diagnosis, blood sample–based DNA methylation was developed. One comprehensive study identified DNA methylation markers in blood for diagnosis or risk evaluation of PTC. Variations in DNA methylation profiles, both at overall genomic level and specific loci, have been associated with PTC risk [24].

Conclusion

In recent years, important progress has been made in the study of thyroid cancer molecular genetics, which provides a lot of theoretical basis for the accurate diagnosis and prognosis of thyroid cancer. The application of some molecular markers significantly improves the diagnostic accuracy of thyroid nodule nature, avoids the occurrence of diagnostic thyroid surgery, and has a certain significance for the judgment of prognosis. Compared with single gene detection, combined detection of multiple genes has higher sensitivity and specificity, so it has higher diagnostic value, and is becoming a research hotspot.

Abbreviations

PTC: papillary thyroid carcinoma; dmCpG: differential methylation CpG; TCGA: The Cancer Genome Atlas; LASSO: Least absolute shrinkage and selection operator; ROC: receiver operating characteristic; SMD: Standardised Mean Difference; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GEO: Gene Expression Omnibus; PCA: principal component analysis; DEGs: differentially expressed genes; DMGs: differentially methylated genes; AUC: Area under the curve

Declarations

Ethics Approval

This study was approved by the Ethics Committee of Shenzhen Longhua District Central Hospital, and written informed consent was obtained from the patient.

Consent for publication

The authors declare that they have no competing interests and the patient in this case report had provided her consent for publication.

Availability of data and materials

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request, The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus database repository, https://www.cancer.gov/, https://www.ncbi.nlm.nih.gov/geo/

Competing Interests

The authors declare no conflicts of interest.

Funding

None

Authors' contributions

All authors have materially participated in the study and manuscript preparation. Weiyi Cai carried out all the molecular genetic analysis, and paticipated in the design of the work; Qingxian Li collected all clininal data and participated in conceiving the work; Jie Ning and Jianhong Chen designed the work, drafted and revised the manuscript. All authors have approved the final article.

Acknowledgement

None

References

1. REGO A, PEREZ LF, Mantinan B, et al. Time Trends for Thyroid Cancer in Northwestern Spain: True Rise in the Incidence of Micro and Larger Forms of Papillary Thyroid Carcinoma [J]. Thyroid, 2009;19(4):333-340.
2. BERNIER MO, WITHROW DR, BERRINGTON DE, et al. Trends in pediatric thyroid cancer incidence in the United States, 19982013 [J]. Cancer, 2019;125(14):24972505.
3. CHO BY, CHOI HS, PARK YJ, et al. Changes in the clinicopathological characteristics and outco-mes of thyroid cancer in Korea over the past four decades [J]. Thyroid. 2013;23(7):797-804.
4. ALTIMARI A, GRIGIONI AD, BENEDETTINI E, et al. Diagnostic role of circulating free plasma DNA detection in patients with localized prostate cancer[J]. Am J Clin Pathol. 2008;129 (5), 756-762.

5. Li, WB, Zhou J; Xu L, et al. Identification of Genes Associated with Papillary Thyroid Carcinoma (PTC) for Diagnosis by Integrated Analysis[J]. HORMONE AND METABOLIC RESEARCH.2016;48(4):226-231.
6. LIANG G, WEISENBERGER DJ. DNA methylation aberrancies as a guide for surveillance and treatment of human cancers [J]. Epigenetics. 2017; 12(6):416-432.
7. BELTRAMI CM, DOS MB, BARROS MC, et al. Integrated data analysis reveals potential drivers and pathways disrupted by DNA methylation in papillary thyroid carcinomas[J]. Clin Epigenetics. 2017; 9: 45.
8. ORITIZ I, BARROS MC, DOS MB, et al. Loss of DNA methylation is related to increased expression of miR-21 and miR-146b in papillary thyroid carcinoma[J]. Clin Epigenetics. 2018; 10: 144.
9. White MG, Nagar S, Aschebrook-Kilfoy B, et al. Epigenetic Alterations and Canonical Pathway Disruption in Papillary Thyroid Cancer: A Genome-wide Methylation Analysis. Ann Surg Oncol. 2016;23(7):2302-9.
10. Sandoval J, Heyn H, Moran S, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692-702.
11. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 29;13(7):484-92.
12. Li WB, Zhou J, Xu L, et al. Identification of Genes Associated with Papillary Thyroid Carcinoma (PTC) for Diagnosis by Integrated Analysis. Horm Metab Res. 2016; 48(4):226-31.
13. Gntara S, Capezzone M, Marchisotta S, et al. Impact of proto-oncogene mutation detection in cytological specimens from thyroid nodules improves the diagnostic accuracy of cytology. J Clin Endocrinol Metab 95:1365-1369.
14. Xu Y, Deng Y, Ji Z, et al. Identification of thyroid carcinoma related genes with mRMR and shortest path approaches. PLoS One. 2014; 9;9(4):e94022.
15. Roberts S, Delury C, Marsh E. The PDZ protein discs-large (DLG): the 'Jekyll and Hyde' of the epithelial polarity proteins. FEBS J. 2012; 79(19):3549-3558.
16. Elsum I, Yates L, Humbert PO, et al. The Scribble-Dlg-Lgl polarity module in development and cancer: from flies to man. Essays Biochem. 2012; 53:141-68.
17. Keane S, Améen S, Lindlöf A, et al. Low DLG2 gene expression, a link between 11q-deleted and MYCN-amplified neuroblastoma, causes forced cell cycle progression, and predicts poor patient survival. Cell Commun Signal. 2020; 20;18(1):65.
18. Beltrami CM, Dos MB, Barros MC, et al. Integrated data analysis reveals potential drivers and pathways disrupted by DNA methylation in papillary thyroid carcinomas. Clin Epigenetics. 2017; 2; 9:45.
19. Park JL, Jeon S, Seo EH, et al. Comprehensive DNA Methylation Profiling Identifies Novel Diagnostic Biomarkers for Thyroid Cancer. Thyroid. 2020 Feb;30(2):192-203.

20. Bisarro M, Barros MC, Marchi FA, et al. Prognostic Classifier Based on Genome-Wide DNA Methylation Profiling in Well-Differentiated Thyroid Tumors. J Clin Endocrinol Metab. 2017 Nov 1;102(11):4089-4099.
21. Roos L, van Dongen J, Bell CG, et al. Integrative DNA methylome analysis of pan-cancer biomarkers in cancer discordant monozygotic twin-pairs. Clin Epigenetics. 2016; 20; 8:7.
22. Mancikova V, Buj R, Castelblanco E, et al. DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. Int J Cancer. 2014; 1;135(3):598-610.
23. Barros MC, Dos MB, Beltrami CM, et al. DNA Methylation-Based Method to Differentiate Malignant from Benign Thyroid Lesions. Thyroid. 2019; 29(9):1244-1254.

Tables

Table 1. Association between methylation and clinicopathological parameters in PTC, based on The Cancer Genome Atlas data

Summary descriptives table by groups of `group'

	normal	tumor	p.overall
	N=503	*N=108*	
Age:			0.309
<=65Y	71 (14.1%)	20 (18.5%)	
>65Y	432 (85.9%)	88 (81.5%)	
TNM_stage:			0.252
stage i	283 (56.3%)	65 (60.2%)	
stage ii	52 (10.3%)	12 (11.1%)	
stage iii	113 (22.5%)	22 (20.4%)	
stage iv	2 (0.40%)	2 (1.85%)	
stage iva	47 (9.34%)	5 (4.63%)	
stage ivc	6 (1.19%)	2 (1.85%)	
Gender:			0.936
male	135 (26.8%)	30 (27.8%)	
female	368 (73.2%)	78 (72.2%)	
T_stage:			.
T1	42 (8.35%)	12 (11.1%)	
T1a	20 (3.98%)	4 (3.70%)	
T1b	80 (15.9%)	8 (7.41%)	
T2	167 (33.2%)	39 (36.1%)	
T3	171 (34.0%)	38 (35.2%)	
T4	9 (1.79%)	6 (5.56%)	
T4a	14 (2.78%)	1 (0.93%)	
N_stage:			0.002
N0	230 (45.7%)	49 (45.4%)	
N1	58 (11.5%)	28 (25.9%)	
N1a	93 (18.5%)	14 (13.0%)	
N1b	74 (14.7%)	11 (10.2%)	
NX	48 (9.54%)	6 (5.56%)	
M_stage:			0.012
M0	280 (55.7%)	76 (70.4%)	
M1	9 (1.79%)	2 (1.85%)	

	normal	tumor	p.overall
	N=503	*N=108*	
MX	214 (42.5%)	30 (27.8%)	

Table 2. Features of the enrolled Gene Expression Omnibus datasets

Study	tumorN	tumorMEAN	tumorSD	normalN	normalMEAN	normalSD	Source
GSE121377(2018)	27	0.6813709	0.126399	7	0.5115317	0.10654	tissue
GSE51090(2014)	83	0.5704313	0.180314	8	0.4692902	0.072152	tissue
GSE53051(2013)	70	0.5565577	0.158457	12	0.4674677	0.082452	tissue
GSE86961(2017)	41	0.6696086	0.125877	41	0.4828395	0.099128	tissue
GSE89093(2016)	46	0.9662353	0.008388	46	0.9636509	0.011047	blood
GSE97466(2017)	74	0.6597332	0.145883	67	0.4962919	0.100348	tissue
TCGA	515	0.6815435	0.174647	56	0.4797196	0.133257	tissue
Bisulfite sequencing PCR	50	0.6512	0.1539	50	0.5318	0.08738	tissue

Figures

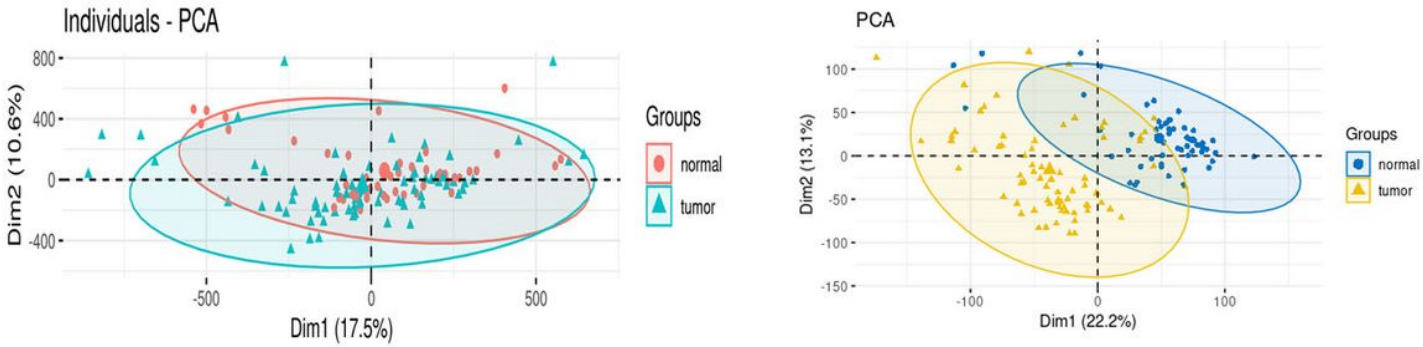

Figure 1

Aberrantly methylation and expression profiles in the PCT of TCGA with the principal component analysis (PCA). A, PCA visualizing all the methylation. B, PCA visualizing all the expression profiles. Circles represent normal tissue samples and triangles represents the tumor tissue samples.

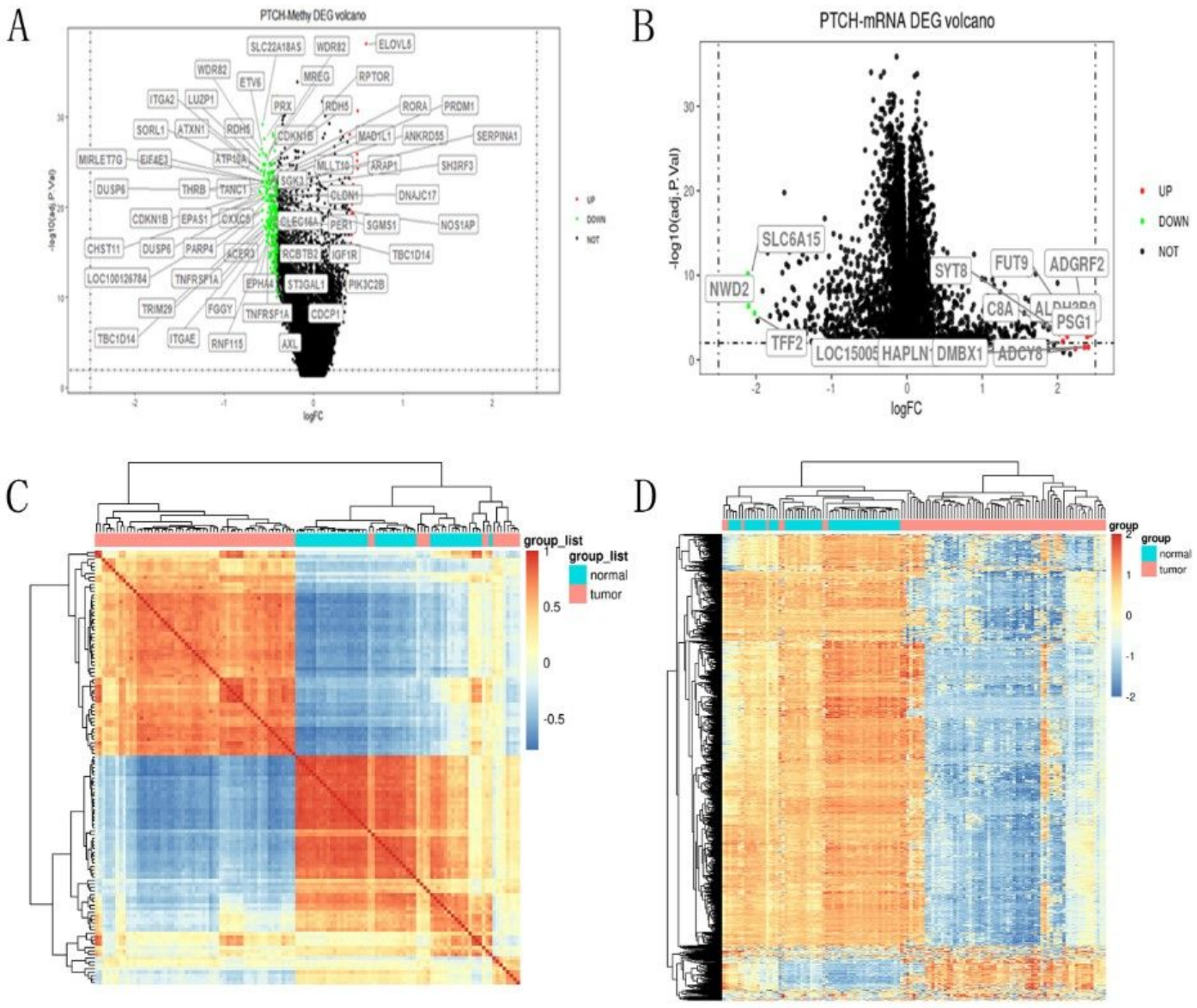

Figure 2

Aberrantly methylated DMGs and DEGs in the PCT of TCGA. (A) Volcano Plot visualizing all the DMGs. B, Volcano Plot visualizing all the DEGs. Red dots represent upregulated genes and dmCpG, green dots represent downregulated genes and dmCpG, and black dots represent genes without differential expression and methylation. (C) Heatmap of the top 500 DMGs. (D) Heatmap of the top 1000 DEGs.

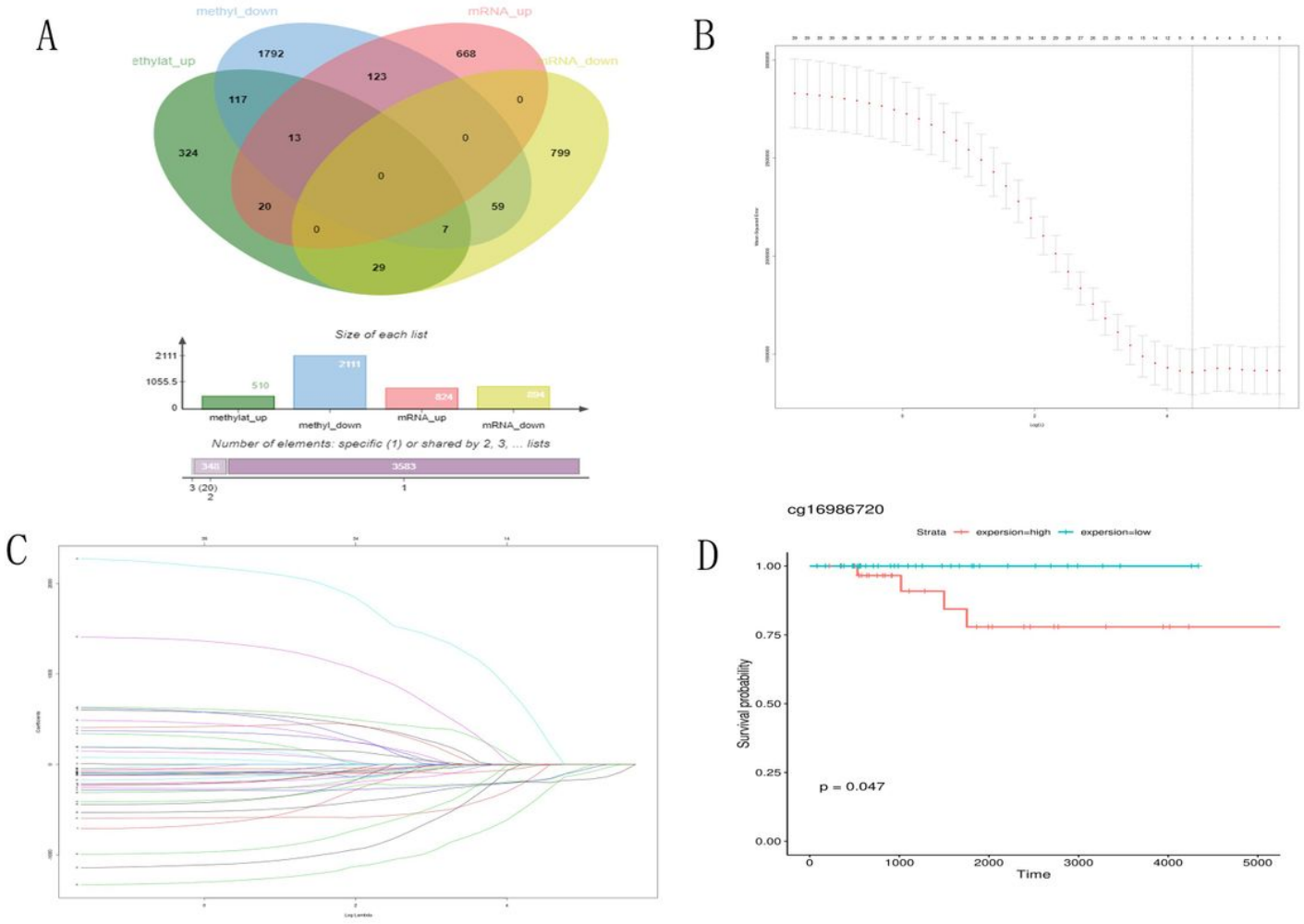

Figure 3

The construction of dmCpG-based prognostic signature. (A) Venn analysis, 84 hypomethylated and upregulated genes were identified, including seven oncogenes, 121 hypermethylated and downregulated genes were identified. (B-C) LASSO regression analysis of CpG sites in the regions of promoter, with the selection of tuning parameter (lambda) and dynamic LASSO coefficient profiling. (D) Kaplan–Meier survival curve of DLG2-cg16986720 classified in hypomethylated and hypermethylated groups.

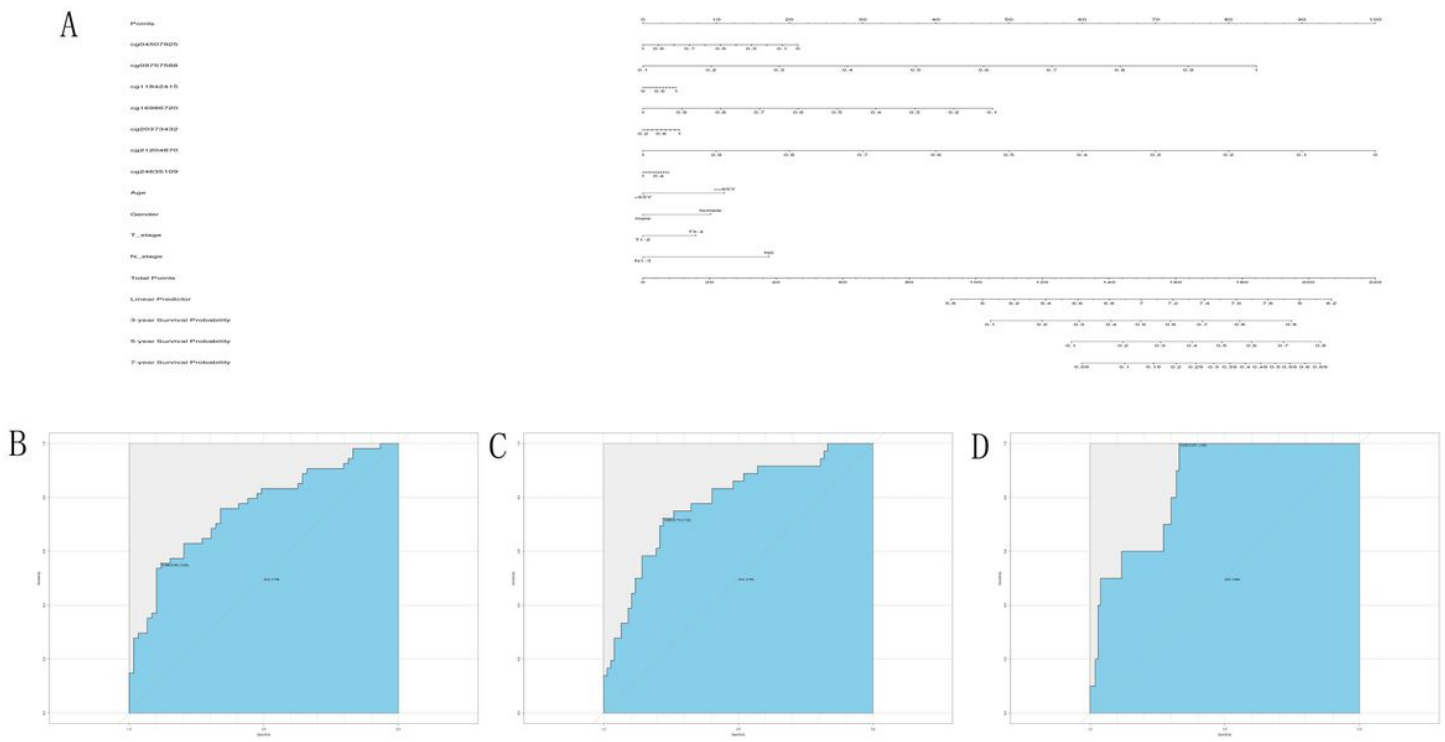

Figure 4

Nomogram of the TCGA cohort. (A) The nomogram of 3-, 5-, and 7-year OS. (B–D) The calibration curves of 3-, 5-, and 7-year OS of TCGA cohort. The blue dot and light gray line represent the ideal prediction model, and the black solid line represents the observed model. ROC curve analysis was for distinguishing normal and cancer tissues. Independent prognostic prediction analysis. (C–E) The 5-, 7-, and 10-year ROC curves of risk score,age, TNM (tumor, node, metastasis).

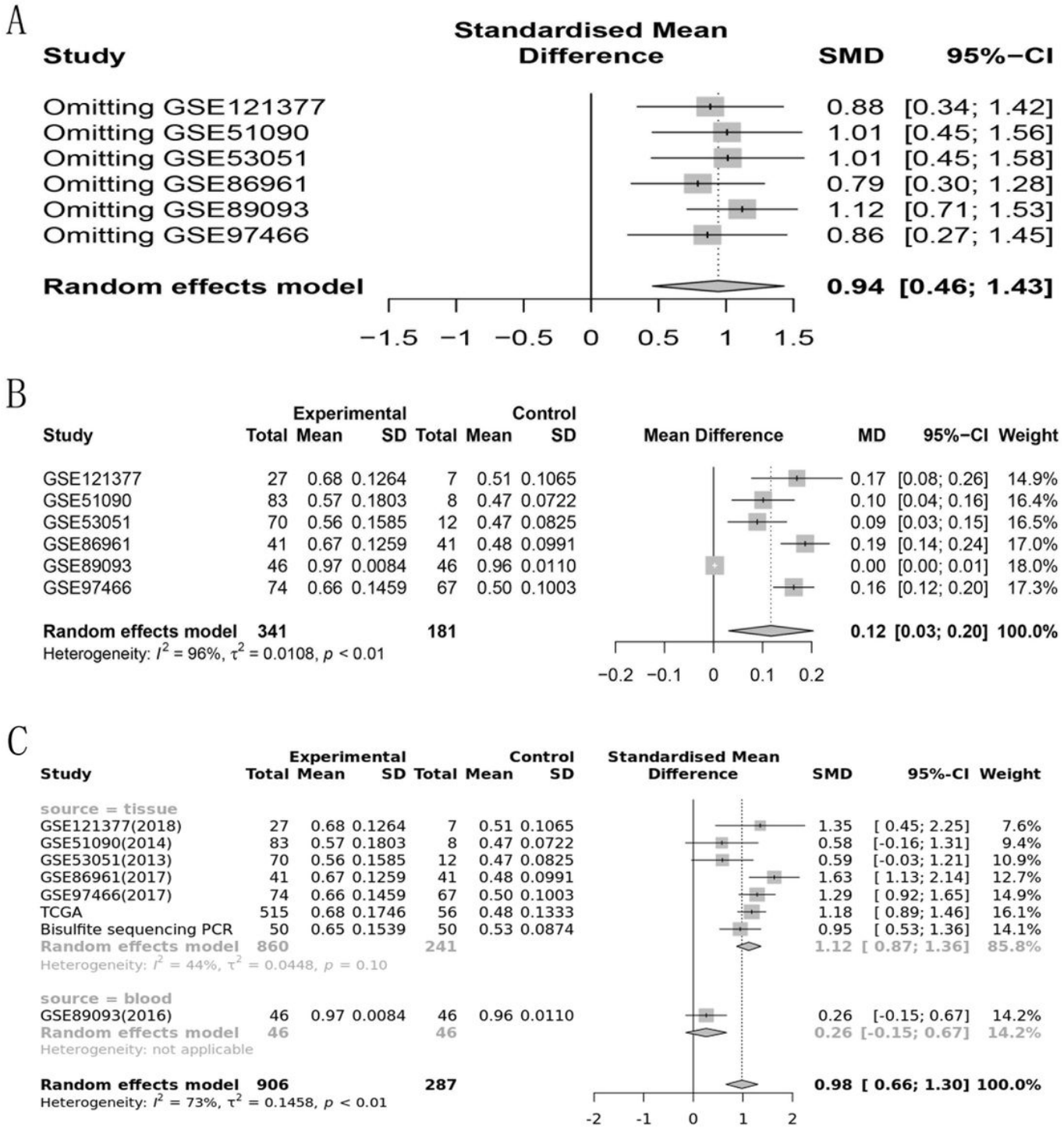

Figure 5

Meta-analysis of Gene Expression Omnibus (GEO) data. Notes. (A) Forest plot of GEO chips. The pooled standard mean deviation of 0.94 (95%: 0.46, 1.43) with great heterogeneity (I2 = 96%, p < 0.01) showed that DLG2-cg16986720 had markedly reduced in the PCT. (B) Sensitivity analysis of GEO chips. Results of subgroup analyses. Notes: a Subgroup analysis based on sample source. The tissue subgroup had no significant heterogeneity (I2 = 44%, p= 0.1). (1: tissue, 0: blood).

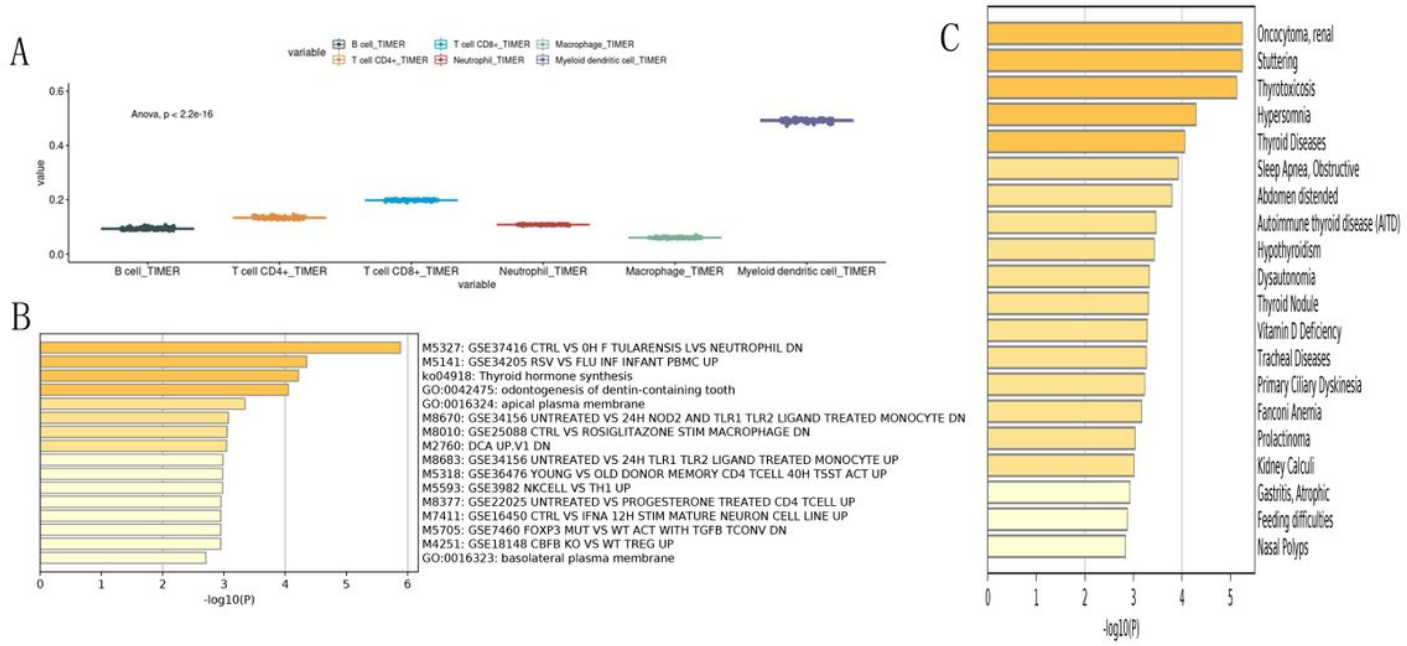

Figure 6

Immune and Functional annotation of aberrantly hypomethylated and downregulated genes methylated DMGs. (A) Immune microenvironments as shown by timer database and analysis of immune expression score in the TCGA. (B) Bar graph of enriched terms across input gene lists, colored by p-values. (C) Summary of enrichment analysis in DisGeNET.