# Comprehensive analysis of Distal-Less homeobox family gene expression in colon cancer

**Yongcheng Chen**
Sun Yat-sen University
**Dongbing Li**
ChosenMed Technology (Beijing) Co., Ltd.
**Dongliang Wang**
ChosenMed Technology (Beijing) Co., Ltd.
**Hui Peng** ( phui@mail.sysu.edu.cn )
Sun Yat-sen University

**Research Article**

**Keywords:** Colon cancer, TCGA, DLX, prognosis, immune infiltration

**Posted Date:** July 5th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1723161/v1

# Abstract

**Background**. The Distal-Less homeobox (DLX) gene family plays an important role in several tumors. However, the expression pattern, prognostic and diagnostic value, possible regulatory mechanisms, and the relationship between DLX family and immune infiltration in colon cancer (COAD) have not been systematically reported.

**Methods**. We used Wilcoxon rank sum test and t-test to assess DLX gene family expression between COAD tissues and unpaired normal colon tissues, cBioPortal to analyze DLX gene family variants, R (version 3.6.3) to analyze DLX gene expression in COAD and the relationship between DLX gene family expression and clinical features and correlation heat map, the survival package [version 3.2-10] and the Cox regression module to assess the prognostic value of the DLX gene family, the pROC package [version 1.17.0.1] to analyze the diagnostic value of the DLX gene family, the R ( version 3.6.3) to analyze the possible regulatory mechanisms of DLX gene family members and related genes, the GSVA package [version 1.34.0] to analyze the relationship between the DLX gene family and immune infiltration, and the ggplot2 [version 3.3.3], and survminer package [version 0.4.9] and clusterProfiler package [version 3.14.3] for visualization.

**Results**. DLX 2/3/4/5/6 were significantly upregulated in COAD patients. The expression of DLX family was associated with M stage, pathologic stage, primary therapy outcome, residual tumor, lymphatic invasion, T stage, N stage, age, perineural invasion, and history of colon polyps. DLX2/5 were independently correlated with the prognosis of COAD in multivariate analysis. DLX1/2/3/4/5/6 were involved in the development and progression of COAD by participating in immune infiltration and pathways, including breast cancer, gastric cancer, Hippo signaling pathway, Wnt signaling pathway, signaling pathways regulating pluripotency of stem cells, basal cell carcinoma, melanoma, and staphylococcus aureus infection.

**Conclusion**. The DLX gene family can be used as potential diagnostic or prognostic biomarkers and therapeutic targets for COAD.


# 1. Introduction

Colon cancer is a widely known tumor whose incidence is on the rise. Approximately 10% of all cancer deaths are caused by colon cancer and related complications [1]. Colon adenocarcinoma (COAD) is the most common, accounting for 98% of colon cancer cases [2]. Colon cancer has a high recurrence rate after treatment, with 42% of patients recurring within 5 years and a median time from recurrence to death of 12 months [3]. Unfortunately, about 20% of COAD patients are diagnosed with stage IV each year [4]. Therefore, exploring novel molecular markers is of great clinical significance to improve the diagnosis and treatment of COAD.

The Distal-Less homeobox (DLX) gene is a homolog of *Drosophila Distal-less* (*Dll*) and consists of six members, including DLX1, DLX2, DLX3, DLX4, DLX5, and DLX6 [5]. DLX1 can be used to identify prostate

cancer (PCa) for early diagnosis [6]. Overexpression of DLX2 is associated with poor prognosis in hepatocellular carcinoma (HCC) [7]. High expression of DLX2 was a poor prognostic marker for patients with glioblastoma multiforme (GBM) [8]. DLX3 is a key regulator of the STAT3 signaling network that maintains skin homeostasis [9]. DLX4 can be used as a prognostic marker for HCC [10]. DLX5 is a potential diagnostic biomarker and therapeutic target for oral squamous cell carcinoma (OSCC) [11]. DLX6 promoted cell proliferation and survival in OSCC [12]. To our knowledge, no studies have systematically assessed the role of the DLX family in COAD using bioinformatics methods.

In this study, we aimed to investigate the expression level, clinical significance, and relationship between DLX family and immune infiltration in COAD to establish an adequate scientific basis for clinical decision making and risk management.

## 2. Materials And Methods

# 2.1 cBioPortal analysis

The cBio Cancer Genomics Portal (cBioPortal) (http://cbioportal.org) was applied to study mutations in DLX genes in COAD. Queries for visualization and analysis were performed by entering (1) cancer type: colon adenocarcinoma; (2) 2 selected studies: colon adenocarcinoma (CaseCCC, PNAS 2015), colon cancer (CPTAC-2 Prospective, Cell 2019); (3) Molecular profile: mutations, structural variants, and copy number alterations; (4) selection of patients/case sets: all samples (139); (5) input genes: DLX1[ENSG00000144355], DLX2[ENSG00000115844], DLX3[ENSG00000064195], DLX4[ENSG00000108813], DLX5[ENSG00000105880], and DLX6[ENSG00000006377]. After submission of queries, accessions were made including origin studies, mutation profiles, mutation number, overall survival status, overall survival (months), disease-free status, and disease-free period (months) tracks.

# 2.2 Dysregulation of DLX genes in COAD

Software was R (version 3.6.3) (statistical analysis and visualization). R package was mainly ggplot2 [version 3.3.3] (for visualization). Molecules were DLX1, DLX2, DLX3, DLX4, DLX5 and DLX6. Disease was COAD. UCSC XENA (https://xenabrowser.net/datapages/) RNAseq data were uniformly processed by the Toil process [13] into TPM format for TCGA and GTEx. Data for COAD were extracted from TCGA and corresponding normal tissue data were extracted from GTEx. RNAseq data were in TPM (transcripts per million reads) format and log2 transformed for expression comparisons between samples. The data filtering condition is to retain paired samples. Significance markers: ns, $P \geq 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

# 2.3 Correlation heat map

Correlation between every two genes of DLX was assessed using a Pearson's correlation coefficient. Software was R (version 3.6.3). R package was mainly ggplot2 [version 3.3.3]. Disease was COAD. Data were obtained from the TCGA COAD project for RNAseq data in level 3 HTSeq-FPKM format. RNAseq data in FPKM (Fregments Per Kilobase per Million) format were converted to TPM format and then log2

transformed. The filtering condition was to remove the data from the control/normal group (not all items have control/normal group).

## 2.4 Association of DLX gene expression with clinical features of TCGA-COAD

Software was R (version 3.6.3). R package was the basic R package. Molecules were DLX1, DLX2, DLX3, DLX4, DLX5 and DLX6. The grouping condition is the median. Disease was COAD. Data were obtained from the TCGA COAD project for RNAseq data in level 3 HTSeq-FPKM format. RNAseq data in FPKM format were converted to TPM format and then log2 transformed.

## 2.5 Survival analysis

Software was R (version 3.6.3). R packages were survminer package [version 0.4.9] (for visualization) and survival package [version 3.2–10] (for statistical analysis of survival data). Molecules were DLX1, DLX2, DLX3, DLX4, DLX5 and DLX6. Subgroups were 0–50 and 50–100. Prognosis types were overall survival (OS), progression free interval (PFI), and disease specific survival (DSS). Disease was COAD. Data were obtained from the TCGA COAD project for RNAseq data in level 3 HTSeq-FPKM format. RNAseq data in FPKM format were converted to TPM format and then log2 transformed. Supplementary data was prognostic data from the reference [14]. The filtering condition was to remove the data for control/normal (not all items have control/normal) and keep the data for clinical information.

## 2.6 Univariate and multivariate Cox regression analysis

Software was R (version 3.6.3). R package was survivor package [version 3.2–10]. Statistical method was Cox regression module. Prognosis type was OS. Included variables were DLX1, DLX2, DLX3, DLX4, DLX5, and DLX6. Disease was COAD. Data were obtained from the TCGA COAD project for RNAseq data in level 3 HTSeq-FPKM format. RNAseq data in FPKM format were converted to TPM format and then log2 transformed. Supplementary data was prognostic data from the reference [15]. The filtering condition was to remove the data for control/normal (not all items have control/normal) and keep the data for clinical information.

## 2.7 ROC curve analysis

Software was R (version 3.6.3). R packages were pROC package [version 1.17.0.1] (for analysis) and ggplot2 package [version 3.3.3]. Molecules were DLX1, DLX2, DLX3, DLX4, DLX5, and DLX6. Clinical Variables were tumor and normal. Disease was COAD. UCSC XENA (https://xenabrowser.net/datapages/) RNAseq data are uniformly processed by the Toil process [13] into TPM format for TCGA and GTEx. Data for COAD were extracted from TCGA and corresponding normal tissue data were extracted from GTEx. The RNAseq data were in TPM format and log2 transformed for expression comparison between samples. Data were not filtered. The horizontal coordinates were the false positive rate (FPR) and the vertical coordinates were the true positive rate (TPR).

## 2.8 Correlation analysis for genes associated with DLX genes

Software was R (version 3.6.3). R package was stat package [version 3.6.3] (base package). Molecules were DLX1, DLX2, DLX3, DLX4, DLX5, and DLX6. Disease was COAD. Data were obtained from the TCGA COAD project for RNAseq data in level 3 HTSeq-FPKM format. RNAseq data in FPKM format were converted to TPM format and then log2 transformed. The data filtering condition was to remove the control/normal group (not all projects had control/normal groups).

## 2.9 Functional enrichment analysis of genes associated with DLX genes

Software was R (version 3.6.3). R packages were mainly ggplot2 package [version 3.3.3] and clusterProfiler package [version 3.14.3].

## 2.10 Correlation between the expression of DLX genes in COAD and immune cells

Software was R (version 3.6.3). R package was GSVA package [version 1.34.0][16]. Immuno-infiltration algorithm was ssGSEA (built-in algorithm of the GSVA package). Molecules were DLX1, DLX2, DLX3, DLX4, DLX5, and DLX6. Immune cells were 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, and Treg. Disease was COAD. Data were obtained from the TCGA COAD project for RNAseq data in level 3 HTSeq-FPKM format. RNAseq data in FPKM format were converted to TPM format and then log2 transformed. The data filtering condition was to remove the control/normal group (not all projects had control/normal groups). Markers for 24 immune cells were obtained from the reference [17].

## 2.11 Statistical analysis

All statistical analyses were performed using R (v.3.6.3). The Wilcoxon rank sum test, chi-square test, and Fisher exact test were used to analyze the relationship between clinical characteristics and DLXs. P values less than 0.05 were considered statistically significant.

# 3. Results

## 3.1 DLX gene alterations and mRNA expression in COAD

The cBioPortal online tool was used to analyze the gene expression of DLX genes in COAD patients. Alterations in the DLX genes in COAD ranged from 0.7–3% (Fig. 1). The mutation data, CNA (copy

number alteration) data, and deep deletion from the 2 studies are depicted in Fig. 2. The analysis of DLX gene expression was performed based on 41 COAD tumor tissue samples and 41 paired samples of normal colon tissues (Fig. 3**)**. The results showed that the expression level of DLX2 in COAD tumor samples was significantly higher than that of DLX2 in normal colon tissues (0.317 ± 0.448 vs. 0.133 ± 0.101, P = 0.015), the expression level of DLX3 in COAD tumor samples was significantly higher than that of DLX3 in normal colon tissues (0.928 ± 1.246 vs. 0.288 ± 0.165, P = 0.003), the expression level of DLX4 in COAD tumor samples was significantly higher than that of DLX4 in normal colon tissues (0.726 ± 0.525 vs. 0.166 ± 0.083, P < 0.001), the expression level of DLX5 in COAD tumor samples was significantly higher than that of DLX5 in normal colon tissues (0.413 ± 0.415 vs. 0.211 ± 0.216, P = 0.005), the expression level of DLX6 in COAD tumor samples was significantly higher than that of DLX6 in normal colon tissues (0.511 ± 0.693 vs. 0.071 ± 0.076, P < 0.001). There was no significantly. However, there was no significant difference between DLX6 expression in COAD tumor samples and in normal colon tissue (0.392 ± 0.643 vs. 0.337 ± 0.263, P = 0.641). We examined the correlation between DLX genes using Pearson correlation analysis. As shown in Fig. 4, there was no significant correlation between DLX1 and DLX3, DLX1 and DLX6, and there was a significant positive correlation between other DLX genes.

## 3.2 Relationship between DLX gene expression and clinical characteristics and prognosis of COAD patients

Clinical characteristics data and gene expression data for 478 COAD tumor samples were downloaded from the TCGA database (**Table S1**). DLX2 expression was associated with M stage (P = 0.005), pathologic stage (P = 0.014), primary therapy outcome (P = 0.036), residual tumor (P = 0.002), and lymphatic invasion (P = 0.013) in COAD patients. DLX3 expression was associated with N stage (P < 0.001), M stage (P < 0.001), pathologic stage (P < 0.001), height (P = 0.045), and residual tumor (P < 0.001) in COAD patients. DLX5 expression was associated with T stage (P < 0.001), N stage (P < 0.001), M stage (P = 0.005), pathologic stage (P < 0.001), primary therapy outcome (P = 0.005), age (P < 0.001), perineural invasion (P = 0.023), lymphatic invasion (P < 0.001), and history of colon polyps (P = 0.009). However, the expression of DLX1, DLX4 and DLX6 did not significantly correlate with the clinical characteristics of COAD patients.

As shown in Fig. 5 **and Figure S1**, high DLX1 expression was associated with the PFI (HR: 1.50; 95% CI: 0.47–0.91; P = 0.024) of COAD; high DLX2 expression was associated with the OS (HR: 1.74; 95% CI: 1.17–2.59; P = 0.006), PFI (HR: 1.66; 95% CI: 1.16–2.37; P = 0.005), and DSS (HR: 2.02; 95% CI: 1.20–3.38; P = 0.008) of COAD; high DLX3 expression was associated with the OS (HR: 1.58; 95% CI: 1.07–2.35; P = 0.022), PFI (HR: 1.64; 95% CI: 1.15–2.34; P = 0.006), and DSS (HR: 1.98; 95% CI: 1.19–3.30; P = 0.009) of COAD; high DLX2 expression was associated with the OS (HR: 1.74; 95% CI: 1.17–2.59; P = 0.006), PFI (HR: 1.66; 95% CI: 1.16–2.37; P = 0.005), and DSS (HR: 2.02; 95% CI: 1.20–3.38; P = 0.008) of COAD; high DLX4 expression was associated with the PFI (HR: 1.54; 95% CI: 1.09–2.19; P = 0.015) of COAD; high DLX5 expression was associated with the PFI (HR: 1.69; 95% CI: 1.18–2.41; P = 0.004) and DSS (HR: 1.86; 95% CI: 1.11–3.11; P = 0.019) of COAD. However, high DLX6 expression was not significantly associated with the prognosis of COAD.

As shown in Table 1, univariate cox regression analysis for OS showed that DLX2 (HR: 1.738; 95%CI: 1.168–2.586, P = 0.006) and DLX3 (HR: 1.583; 95%CI: 1.068–2.347, P = 0.022) were negative predictors of OS outcome in COAD patients, and DLX1 (HR: 1.500; 95%CI: 1.055–2.131, P = 0.024), DLX2 (HR: 1.661; 95%CI: 1.165–2.369, P = 0.005, DLX3 (HR: 1.641; 95%CI: 1.153–2.335, P = 0.006), DLX4 (HR: 1.544; 95%CI: 1.086–2.194, P = 0.015), and DLX5 (HR: 1.687; 95%CI: 1.180–2.410, P = 0.004) were negative predictors of PFS outcome in COAD patients. DLX2 (HR: 1.541; 95%CI: 1.011–2.348, P = 0.044) was independently correlated with OS of COAD in multivariate analysis, and DLX5 (HR: 1.539; 95%CI: 1.072–2.210, P = 0.019) was independently correlated with PFS of COAD in multivariate analysis.

Table 1
Univariate and multivariate Cox regression analyses with DLX genes and prognosis of COAD patients.

| | | | Univariate analysis | | Multivariate analysis | |
|---|---|---|---|---|---|---|
| | **Characteristics** | **Total(N)** | **HR (95% CI)** | **P value** | **HR (95% CI)** | **P value** |
| Overall survival | DLX1 (Low vs. High) | 477 | 1.245 (0.844–1.836) | 0.268 | | |
| | DLX2 (Low vs. High) | 477 | 1.738 (1.168–2.586) | 0.006 | 1.541 (1.011–2.348) | 0.044 |
| | DLX3 (Low vs. High) | 477 | 1.583 (1.068–2.347) | 0.022 | 1.325 (0.870–2.018) | 0.19 |
| | DLX4 (Low vs. High) | 477 | 1.437 (0.973–2.124) | 0.068 | 1.123 (0.736–1.714) | 0.592 |
| | DLX5 (Low vs. High) | 477 | 1.465 (0.988–2.171) | 0.058 | 1.333 (0.892–1.990) | 0.161 |
| | DLX6 (Low vs. High) | 477 | 0.951 (0.646–1.402) | 0.801 | | |
| Progress free interval | DLX1 (Low vs. High ) | 477 | 1.500 (1.055–2.131) | 0.024 | 1.262 (0.867–1.838) | 0.225 |
| | DLX2 (Low vs. High ) | 477 | 1.661 (1.165–2.369) | 0.005 | 1.299 (0.874–1.930) | 0.196 |
| | DLX3 (Low vs. High ) | 477 | 1.641 (1.153–2.335) | 0.006 | 1.321 (0.906–1.927) | 0.149 |
| | DLX4 (Low vs. High ) | 477 | 1.544 (1.086–2.194) | 0.015 | 1.244 (0.855–1.809) | 0.254 |
| | DLX5 (Low vs. High ) | 477 | 1.687 (1.180–2.410) | 0.004 | 1.539 (1.072–2.210) | 0.019 |
| | DLX6 (Low vs. High ) | 477 | 0.939 (0.664–1.329) | 0.723 | | |

As shown in Fig. 6, it can be obtained that in predicting normal and tumor outcomes, variable DLX1 has some accuracy in predicting (AUC = 0.893, CI = 0.867–0.920), variable DLX2 has some accuracy in predicting (AUC = 0.731, CI = 0.691–0.771), variable DLX3 has a lower accuracy (AUC = 0.561, CI = 0.512–0.611), variable DLX4 had some accuracy in predictive ability (AUC = 0.834, CI = 0.802–0.867), variable DLX5 had low accuracy in predictive ability (AUC = 0.590, CI = 0.546–0.635), and variable DLX6 had had poor diagnostic efficacy (AUC = 0.486, CI = 0.439–0.534).

## 3.3 The function of genes associated with DLX genes

The top 10 significantly associated genes for each DLX gene are shown in the single gene co-expression heat map (Fig. 7). Genes significantly associated with DLX1 include DLX2, KLF14, CHRND, KCNN1, IGDCC3, ARHGAP36, NCAN, TFAP2B, CNPY1, and CACNG7. Genes significantly associated with DLX2 include DLX1, CNPY1, CHRND, NEUROD1, IGDCC3, TNFRSF19, KLF14, NELL2, HS3ST4, and SLC38A8. Genes significantly associated with DLX3 include NOTUM, NKD1, APCDD1, ADAMTSL2, MYH7B, PRR9, LRRC43, CAB39L, ABCC2, and DLX4. Genes significantly associated with DLX4 include DLX3, TTLL4, DNMT3B, CDK5R1, IGF2BP1, STK36, UNK, AMER3, PHF12, and WNT3. Genes significantly associated with DLX5 include DYNC1I1, DLX6, RASL11B, ID4, SP7, AMBN, KRT31, MYL3, VENTX, and ISM1. Genes significantly associated with DLX6 include DLX5, TRIM71, SH3GL2, SLC46A1, DYNC1I1, PGBD5, GAL, COCH, AXIN2, and CKB. The top 30 genes significantly associated with each DLX gene (147 in total) were analyzed for GO and KEGG enrichment (**Table S2**). The top five biological processes, including pattern specification process, regionalization, skeletal system morphogenesis, endochondral bone morphogenesis, and hippocampus development; the top five cytological components, including integral component of synaptic membrane, intrinsic component of synaptic membrane, hippocampal mossy fiber to CA3 synapse, integral component of postsynaptic specialization membrane, and intrinsic component of postsynaptic specialization membrane; the significantly related molecular function including motor activity, microtubule binding, cadherin binding, and cell adhesion molecule binding, are shown in Fig. 8 and **Table S3**. The significantly related pathways, including breast cancer, gastric cancer, Hippo signaling pathway, Wnt signaling pathway, signaling pathways regulating pluripotency of stem cells, basal cell carcinoma, melanoma, and staphylococcus aureus infection, were shown in Fig. 9 and **Table S3**.

## 3.4 Correlation of DLX gene expression and immune cells in COAD

As shown in Fig. 10, there was a correlation between DLX gene expression and immune cells in COAD. DLX1 gene expression was positively correlated with some TIICs, including aDC, Cytotoxic cells, DC, Eosinophils, iDC, Macrophages, Mast cells, Neutrophils, NK CD56dim cells, NK cells, Tem, TFH, Tgd, Th1 cells, and TReg, and negatively correlated with Th17 cells. DLX2 gene expression was positively correlated with Mast cells and TFH, and negatively correlated with pDC and Th17 cells. DLX3 gene expression was negatively correlated with some TIICs, including aDC, CD8 T cells, Cytotoxic cells, DC, Macrophages, Neutrophils, T cells, T helper cells, Th1 cells, Th2 cells, and TReg. DLX4 gene expression was positively correlated with NK cells, and negatively correlated with some TIICs, including Cytotoxic cells, DC, Macrophages, pDC, Th1 cells, and Th2 cells. DLX5 gene expression was positively correlated

with some TIICs, including B cells, CD8 T cells, DC, iDC, Macrophages, Mast cells, Neutrophils, NK cells, pDC, Tem, TFH, Tgd, and TReg, and negatively correlated with Th17 cells and Th2 cells. DLX6 gene expression was negatively correlated with some TIICs, including aDC, Cytotoxic cells, DC, Macrophages, Neutrophils, NK CD56dim cells, T cells, Tem, and Th1 cells.

# 4. Discussion

DLX1 was significantly upregulated in PCa tissues and cells [18]. DLX2 was significantly upregulated in HCC tissues and cell lines [7, 19]. DLX2 expression in gastric cancer was significantly correlated with tumor size, depth of infiltration, lymph node metastasis, and tumor-lymph node-metastasis stage [20]. DLX4 was significantly upregulated in nasopharyngeal carcinoma (NPC) cell lines [21]. DLX4 expression was significantly elevated in HCC tissues and correlated significantly with tumor size, histopathological classification, and serum alpha-fetoprotein (AFP) [10]. DLX5 was upregulated in OSCC tissues and cell lines and was significantly associated with advanced TNM staging, lymph node metastasis, poor cell differentiation, and tumor location [11]. DLX6 was significantly upregulated in oral cancer tissues and was associated with advanced tumor stage and poor prognosis [12]. In this study, DLX 2/3/4/5/6 were significantly upregulated in COAD patients. The expression of DLX family was associated with M stage, pathologic stage, primary therapy outcome, residual tumor, lymphatic invasion, T stage, N stage, age, perineural invasion, and history of colon polyps. DLX2 was independently correlated with OS of COAD in multivariate analysis, and DLX5 was independently correlated with PFS of COAD in multivariate analysis. In predicting the outcome of normal and tumor tissues, DXL1/2/4 had some accuracy in predicting normal and tumor.

miR-129-5p impedes the biological function of cancer cells by inhibiting DLX1 expression [22]. DLX1, a key target of FOXM1, promoted ovarian cancer aggressiveness by enhancing TGF-β/SMAD4 signaling [23]. Circ_HIPK3 promotes HCC progression by mediating the miR-582-3p/DLX2 pathway [19]. In tumor cells, DLX2/3/4 can be involved in the control of fenretinide (4HPR)-mediated apoptosis [24]. The homology domain protein DLX4 promoted NPC progression through upregulation of YB-1 [21]. DLX5 regulation of CCND1 affects the progression of OSCC [11]. DLX5 promotes osteosarcoma progression through activation of NOTCH signaling pathway [25]. DLX6 regulated OSCC cell proliferation through the EGFR-CCND1 axis [12]. In this study, the DLX gene family was involved in the development and progression of COAD by participating in pathways, including breast cancer, gastric cancer, Hippo signaling pathway, Wnt signaling pathway, signaling pathways regulating pluripotency of stem cells, basal cell carcinoma, melanoma, and staphylococcus aureus infection. The specific mechanisms by which the DLX gene family mediates the above pathways involved in the development of colon cancer need to be further investigated.

Immune-related mechanisms play an important role in COAD, and immunotherapeutic strategies are considered a promising direction for the treatment of COAD [26]. Another important aspect of this study was that the expression of the DLX gene family correlated with different levels of immune infiltration. In this study, the expression DLX genes were negatively correlated with some TIICs, and positively correlated

with some TIICs. The DLX gene family plays an important role in the recruitment and regulation of immune infiltrating cells in COAD.

The present study has several limitations. First, the study lacks validation by biological or molecular experiments. Secondly, COAD shows strong heterogeneity, and the mRNA expression levels in the TCGA database are the average mRNA expression levels for all cell types within the tumor. Single-cell sequencing is needed to further elucidate the role of DLX in COAD.

# 5. Conclusion

DLX 2/3/4/5/6 were significantly upregulated in COAD patients. DLX2/3/5 were associated with M stage, pathologic stage, primary therapy outcome, residual tumor, lymphatic invasion, T stage, N stage, age, perineural invasion, and history of colon polyps. DLX2/5 were independently correlated with the prognosis of COAD in multivariate analysis. DXL1/2/4 had some accuracy in predicting normal and tumor. The DLX gene family was involved in the development and progression of COAD by participating in immune infiltration and pathways, including breast cancer, gastric cancer, Hippo signaling pathway, Wnt signaling pathway, signaling pathways regulating pluripotency of stem cells, basal cell carcinoma, melanoma, and staphylococcus aureus infection. The results of this study suggested a role for DLX gene family as a potential diagnostic or prognostic biomarker and therapeutic target in COAD.

# List Of Abbreviations

| Abbreviations | Full Name |
|---|---|
| DLX | The Distal-Less homeobox |
| COAD | Colon adenocarcinoma |
| *Dll* | *Drosophila Distal-less* |
| PCa | Prostate cancer |
| HCC | Hepatocellular carcinoma |
| GBM | Glioblastoma multiforme |
| OSCC | Oral squamous cell carcinoma |
| cBioPortal | cBio Cancer Genomics Portal |
| TPM | Transcripts per million reads |
| FPKM | Fregments per kilobase per million |
| OS | Overall survival |
| PFI | Progression free interval |
| DSS | Disease specific survival |
| FPR | False positive rate |
| TPR | True positive rate |
| aDC | Activated DC |
| iDC | Immature DC |
| pDC | Plasmacytoid DC |
| Tcm | T central memory |
| Tem | T effector memory |
| Tfh | T follicular helper |
| Tgd | T gamma delta |
| CNA | Copy number alteration |
| NPC | Nasopharyngeal carcinoma |
| AFP | alpha-fetoprotein |

# Declarations

## Ethics approval

This study was approved by the ethical review committee of our institution.

### Consent for publication

All authors have read and approved the final version of the manuscript to be submitted.

### Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

### Competing interests

The authors declare that they have no conflict of interest. All authors have no financial relationships to disclosure.


### Funding

This study was supported by National Key Clinical Discipline.


### Authors' contributions

YC and HP contributed to the conception of the study; YC, DL and DW contributed to analysis and manuscript preparation; YC and DL performed the data analyses and wrote the manuscript. HP revised the manuscript.


### Acknowledgements

Nil.


### Disclosure

The authors declare that they have no conflict of interest. All authors have no financial relationships to disclosure.

### Data availability statement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

## References


1. Bray, F., 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): p. 394–424.

2. Alibolandi, M., et al., *Tetrac-conjugated polymersomes for integrin-targeted delivery of camptothecin to colon adenocarcinoma in vitro and in vivo*. Int J Pharm, 2017. **532**(1): p. 581–594.
3. Sargent, D.J., et al., *Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials*. J Clin Oncol, 2005. **23**(34): p. 8664–70.
4. Zhou, Y., et al., *Candidate genes involved in metastasis of colon cancer identified by integrated analysis*. Cancer Med, 2019. **8**(5): p. 2338–2347.
5. Cohen, S.M., et al., *Distal-less encodes a homoeodomain protein required for limb development in Drosophila*. Nature, 1989. **338**(6214): p. 432–4.
6. Leyten, G.H., et al., *Identification of a Candidate Gene Panel for the Early Diagnosis of Prostate Cancer*. Clin Cancer Res, 2015. **21**(13): p. 3061–70.
7. Liu, J., et al., *Overexpression of DLX2 is associated with poor prognosis and sorafenib resistance in hepatocellular carcinoma*. Exp Mol Pathol, 2016. **101**(1): p. 58–65.
8. Yan, Z.H., et al., *Upregulation of DLX2 confers a poor prognosis in glioblastoma patients by inducing a proliferative phenotype*. Curr Mol Med, 2013. **13**(3): p. 438–45.
9. Bhattacharya, S., et al., *DLX3-Dependent STAT3 Signaling in Keratinocytes Regulates Skin Immune Homeostasis*. J Invest Dermatol, 2018. **138**(5): p. 1052–1061.
10. Gao, Y., et al., *DLX4 as a prognostic marker for hepatocellular carcinoma*. Neoplasma, 2014. **61**(3): p. 318–23.
11. Zhang, J., et al., *Dlx5 promotes cancer progression through regulation of CCND1 in oral squamous cell carcinoma (OSCC)*. Biochem Cell Biol, 2021. **99**(4): p. 424–434.
12. Liang, J., et al., *DLX6 promotes cell proliferation and survival in oral squamous cell carcinoma*. Oral Dis, 2022. **28**(1): p. 87–96.
13. Vivian, J., et al., *Toil enables reproducible, open source, big biomedical data analyses*. Nature biotechnology, 2017. **35**(4): p. 314–316.
14. Liu, J., et al., *An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics*. Cell, 2018. **173**(2): p. 400–416.e11.
15. Liu, J., et al., *An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics*. Cell, 2018. **173**(2): p. 400–416.e11.
16. Hänzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. BMC Bioinformatics, 2013. **14**: p. 7.
17. Bindea, G., et al., *Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer*. Immunity, 2013. **39**(4): p. 782–95.
18. Sun, B., et al., *MicroRNA-539 functions as a tumour suppressor in prostate cancer via the TGF-β/Smad4 signalling pathway by down-regulating DLX1*. J Cell Mol Med, 2019. **23**(9): p. 5934–5948.
19. Zhang, H., et al., *Knockdown of circ_HIPK3 inhibits tumorigenesis of hepatocellular carcinoma via the miR-582-3p/DLX2 axis*. Biochem Biophys Res Commun, 2020. **533**(3): p. 501–509.

20. Tang, P., et al., *Increased expression of DLX2 correlates with advanced stage of gastric adenocarcinoma*. World J Gastroenterol, 2013. **19**(17): p. 2697–703.
21. Ling, Z., et al., *Homeodomain protein DLX4 facilitates nasopharyngeal carcinoma progression via up-regulation of YB-1*. Genes Cells, 2020. **25**(7): p. 466–474.
22. Li, Q., et al., *MiR-129-5p/DLX1 signalling axis mediates functions of prostate cancer during malignant progression*. Andrologia, 2021. **53**(11): p. e14230.
23. Chan, D.W., et al., *DLX1 acts as a crucial target of FOXM1 to promote ovarian cancer aggressiveness by enhancing TGF-β/SMAD4 signaling*. Oncogene, 2017. **36**(10): p. 1404–1416.
24. Ferrari, N., et al., *Induction of apoptosis by fenretinide in tumor cell lines correlates with DLX2, DLX3 and DLX4 gene expression*. Oncol Rep, 2003. **10**(4): p. 973–7.
25. Zhang, X., et al., *DLX5 promotes osteosarcoma progression via activation of the NOTCH signaling pathway*. Am J Cancer Res, 2021. **11**(6): p. 3354–3374.
26. Procaccio, L., et al., *Immunotherapy in Gastrointestinal Cancers*. Biomed Res Int, 2017. **2017**: p. 4346576.

# Figures


## Figure 1

mRNA expression of DLX genes in COAD in cBioPortal (RNA Seq V2 RSEM).

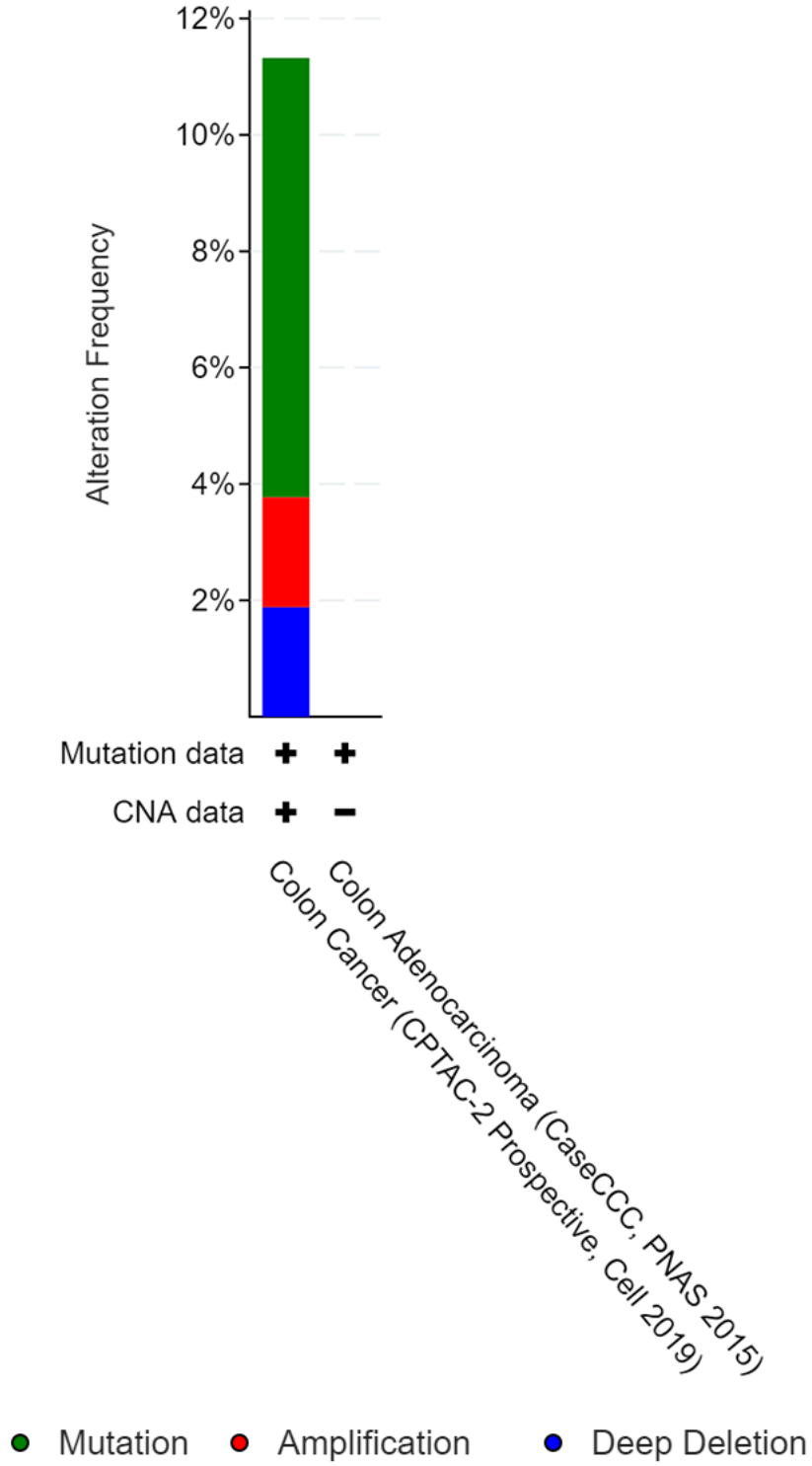
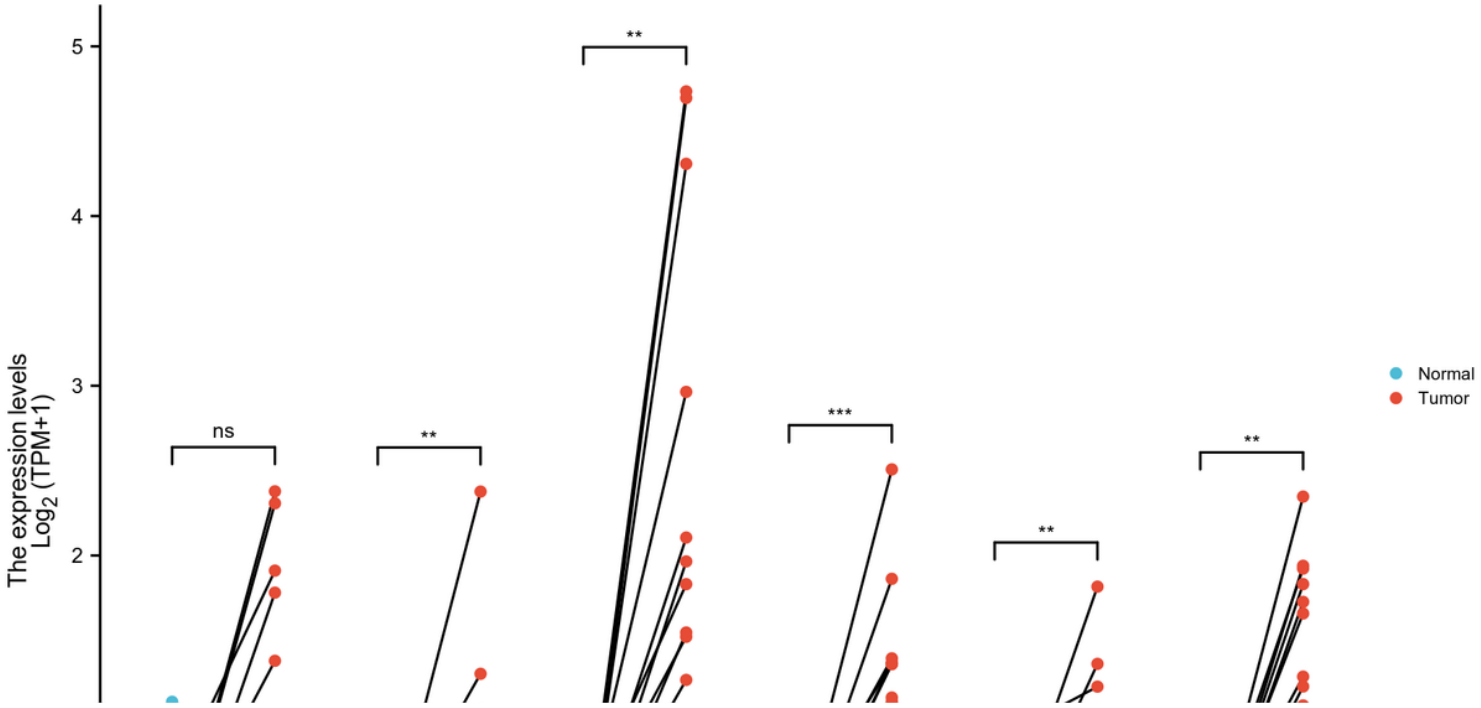



**Figure 2**

Percentage of DLX genes in COAD cases calculated using the cancer type summary in cBioPortal.




## Figure 3

The mRNA levels of DLX genes between COAD tissue and unpaired normal stomach tissue in TCGA. ns: not significant, P≥0.05; **, P<0.01; ***, P<0.001.

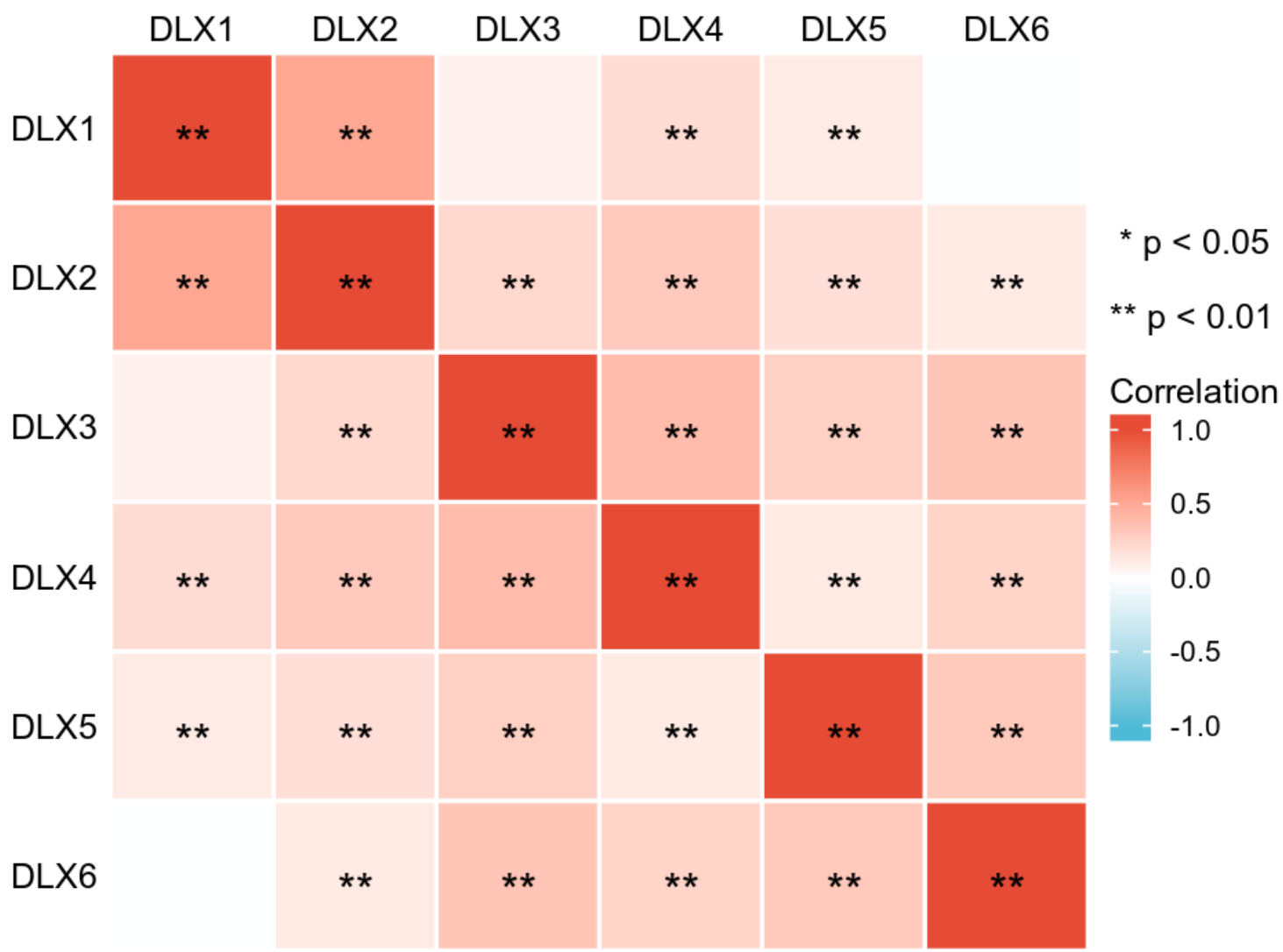
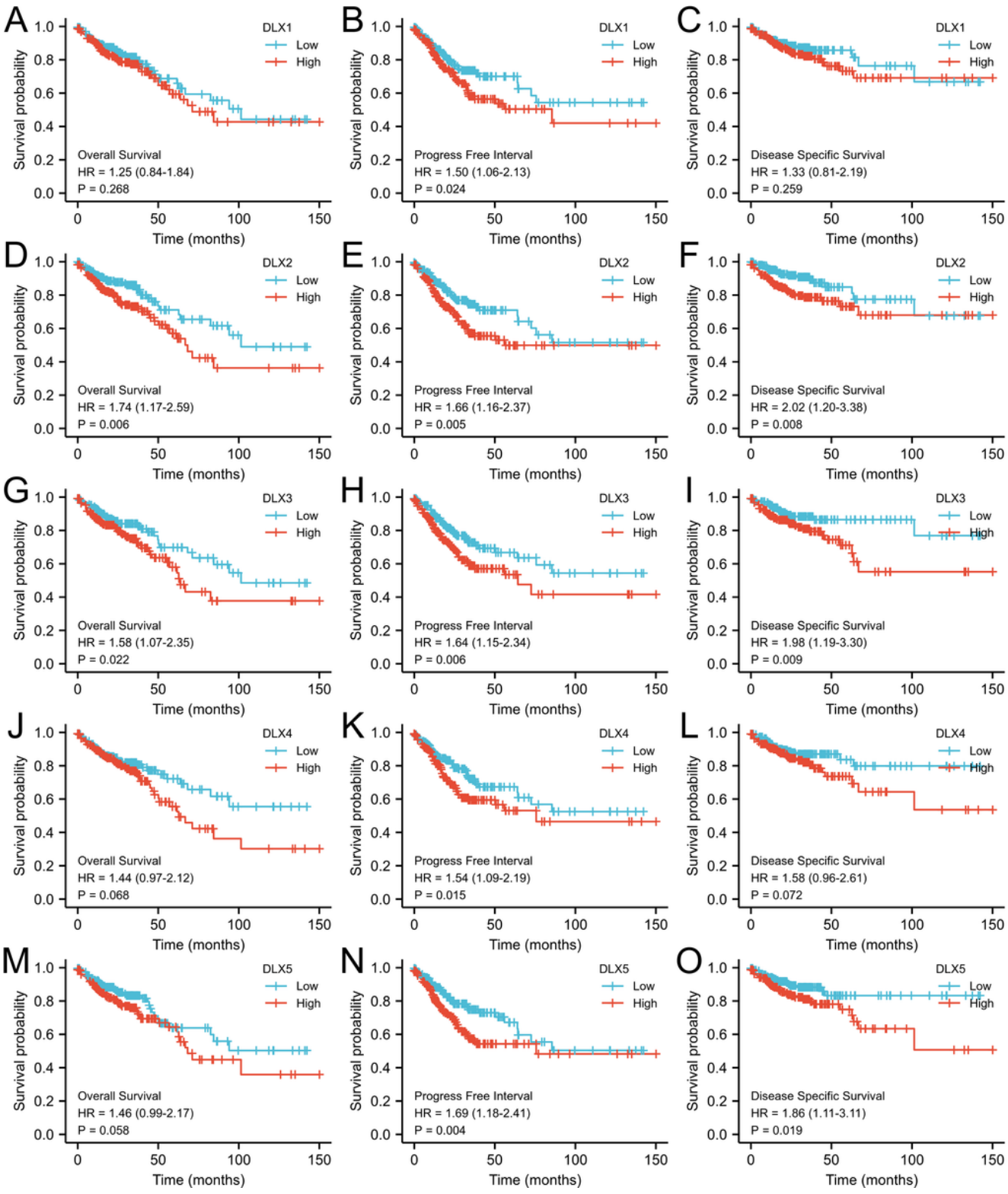



**Figure 4**

Correlation between every two genes of DLX genes in COAD.




## Figure 5

Survival analysis results for overall survival (OS), progression free intervival (PFI), and disease specific survival (DSS). OS of (A) DLX1, (D) DLX2, (G) DLX3, (J) DLX4, and (M) DLX5; PFI of (B) DLX1, (E) DLX2, (H) DLX3, (K) DLX4, and (N) DLX5; DSS of (C) DLX1, (F) DLX2, (I) DLX3, (L) DLX4, and (O) DLX5.

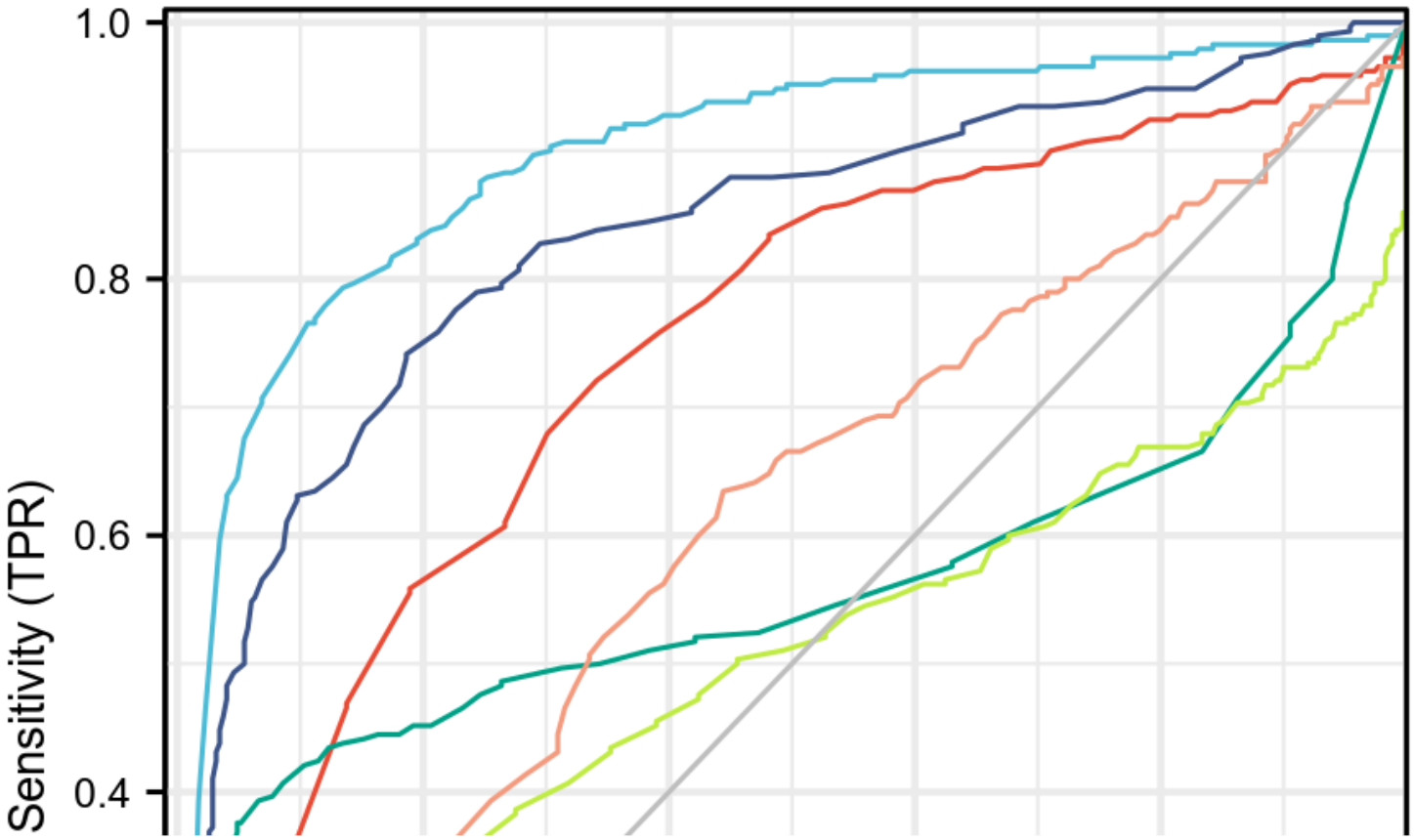
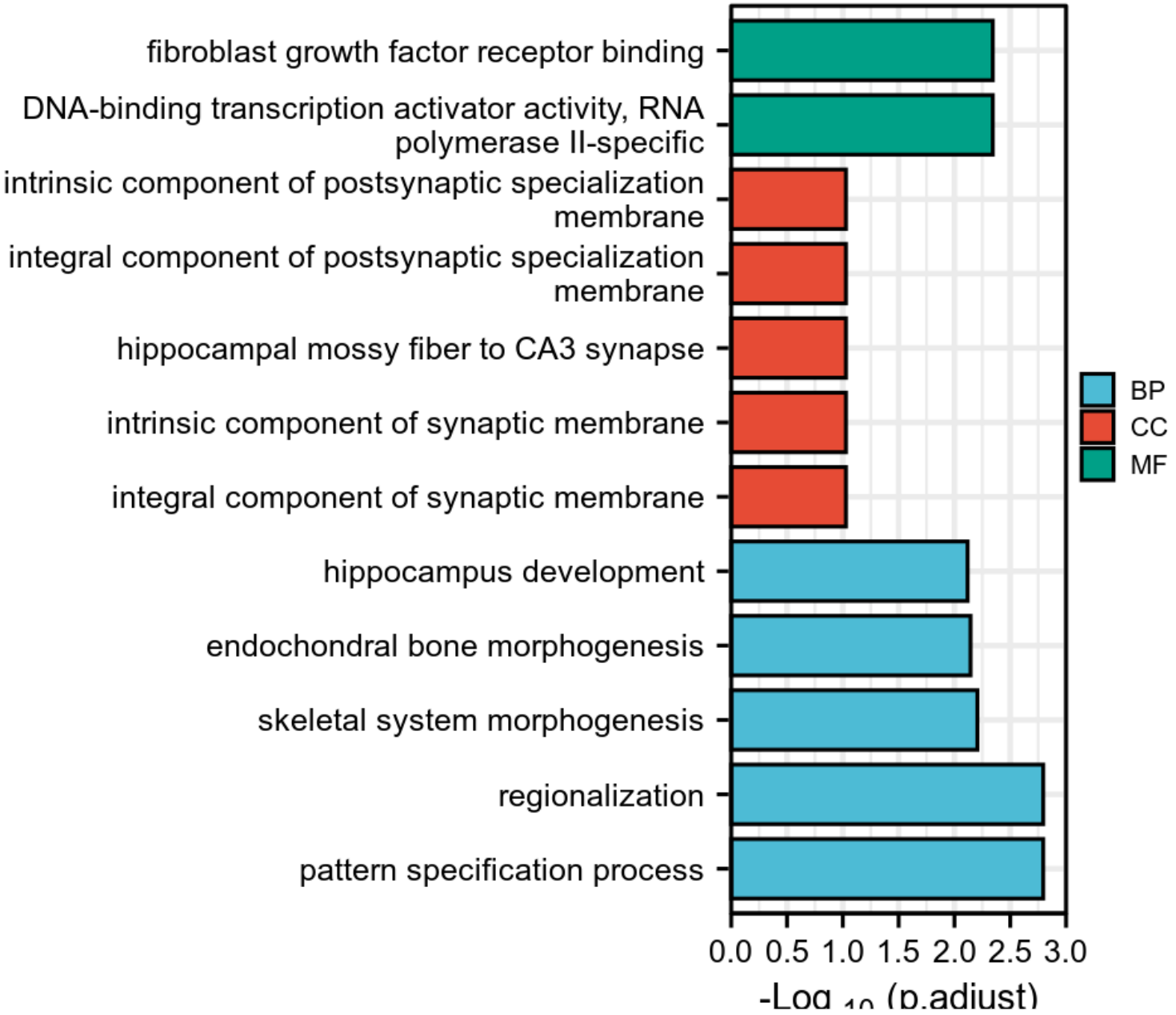



**Figure 6**

ROC curves of DLX genes in COAD and normal colon tissues. The area under the ROC curve is between 0.5 and 1. The closer the AUC is to 1, the better the diagnosis. the AUC is between 0.5 and 0.7 with low accuracy, the AUC is between 0.7 and 0.9 with some accuracy, and the AUC is above 0.9 with high accuracy.

## Figure 7

Heatmap plot of top 10 correlated genes to DLX genes. (A) DLX1, (B) DLX2, (C) DLX3, (D) DLX4, (E) DLX5, and (F) DLX6.




## Figure 8

GO analysis of genes associated with DLX genes.

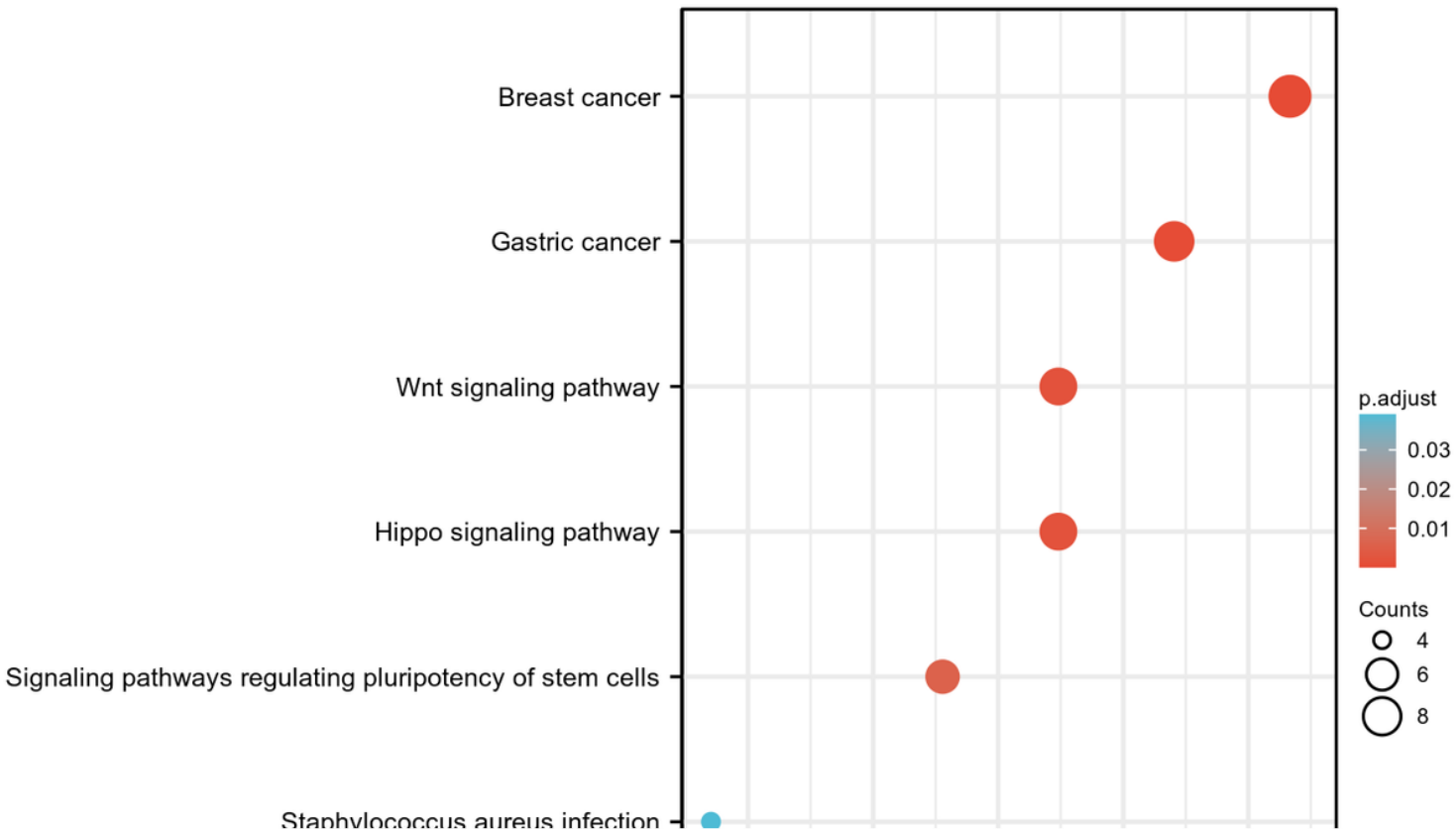



**Figure 9**

KEGG analysis of genes associated with DLX genes.

**Figure 10**

Correlation between the expression of each DLX gene and the 24 TIICs of COAD (lollipop plot). In the color bar, the darker the color, the smaller the p-value, indicating a higher statistical significance. The bubble size represents the correlation value, the larger the bubble, the larger the correlation value.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.tiff
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx