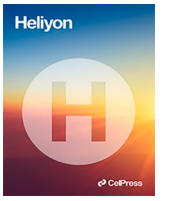
Heliyon 10 (2024) e31515

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Comprehensive pan-cancer analysis reveals the versatile role of GALNT7 in epigenetic alterations and immune modulation in cancer

Yan Liu, Yue Sun, Meixia Xiao, Shuang Li, Shengming Shi *

The First Affiliated Hospital of Huzhou Normal University, Huzhou First People's Hospital, China

A B S T R A C T

Cancer is a leading cause of mortality globally, characterized by intricate molecular alterations, including epigenetic changes such as glycosylation. This study presents a comprehensive pan-cancer analysis of Polypeptide N-Acetylgalactosaminyltransferase 7 (GALNT7), an enzyme involved in mucin-type O-linked protein glycosylation. GALNT7 has previously been linked to various cancers, but a unified analysis across cancer types is lacking.

Leveraging data from TCGA, GTEx, and other sources, we scrutinized GALNT7's expression, prognostic relevance, links to immune-related genes, immune cell infiltration, and its involvement in tumor genetic heterogeneity across 33 cancer types. GALNT7 exhibited diverse expression patterns across cancer types, showcasing its potential as an oncogenic factor, with its expression levels linked to both positive and negative prognoses, highlighting the context-specific nature of its role in cancer progression.

We delved into the intricate interplay between GALNT7 and immune genes, unveiling positive and negative correlations, underscoring complex interactions in the tumor microenvironment. GALNT7 was found to impact immune cell infiltration, which could have implications for treatment strategies. Additionally, GALNT7 displayed associations with genetic tumor aspects, encompassing genomic instability, DNA repair issues, and genetic mutations, hinting at its pivotal role in shaping the genetic landscape of diverse cancers.

Enrichment analysis uncovered potential functions of GALNT7 beyond glycosylation, such as its participation in signaling pathways and its association with various diseases, notably cancer.

This comprehensive analysis elucidates the multifaceted role of GALNT7 in cancer biology, underlining its potential as a therapeutic target and biomarker across various cancer types. These findings provide valuable insights for future research and the development of personalized cancer treatment strategies.

1. Introduction

As of today, cancer remains the second leading cause of mortality in humans [1]. The onset and progression of cancer exhibit a multitude of characteristics, one of which is epigenetic changing [2–4]. Glycosylation is an epigenetic modification, which not only directly influences cell growth and survival but also promotes tumor-induced immune modulation and eventual metastasis [5].

Cancer cells frequently express glycoproteins and glycolipids with altered glycan structures compared to normal cells. These cancer-associated glycans influence tumor cell signaling, adhesion, motility, invasion, and immune surveillance. Specific glycan changes associated with cancer include increased β1,6-branching, truncated O-glycans, accumulation of sialic acids, and expression of fetal antigens. Glycosylation alterations in multiple proteins have been observed in various cancers, including breast cancer, pancreatic cancer, liver cancer, colorectal cancer, and gastric cancer, among others. These modifications have a significant impact on cancer

* Corresponding author.
E-mail address: 1114880505@qq.com (S. Shi).

https://doi.org/10.1016/j.heliyon.2024.e31515
Received 2 March 2024; Received in revised form 16 May 2024; Accepted 16 May 2024
Available online 21 May 2024

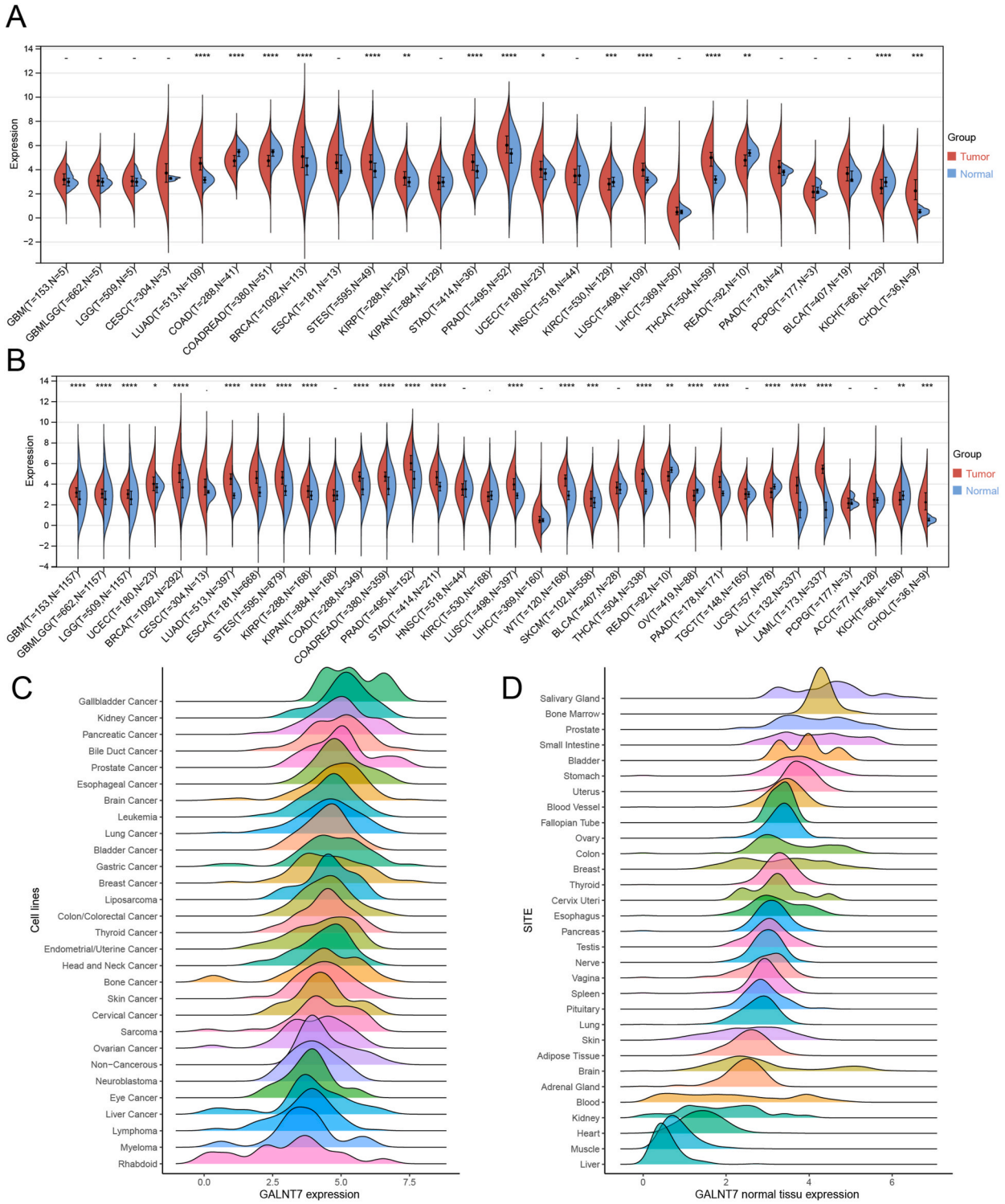


Fig. 1. The expression levels of GALNT7 in cancer and non-cancerous conditions. The differential expression of GALNT7 in cancer and normal tissues compared through TCGA (A) and TCGA as well as GTEx databases (B). (C) Expression profile of GALNT7 in cancer cell lines. (D) Expression profile of GALNT7 in normal tissue organs. TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression.

proliferation and treatment [6–10]. Furthermore, glycosylation holds immense potential in cancer therapeutics, serving as a target for cancer drug treatment [11,12].

Polypeptide N-Acetylgalactosaminyltransferase 7 (GALNT7), the enzyme encoded by this gene governs the pivotal initiation phase of mucin-type O-linked protein glycosylation, overseeing the transfer of N-acetylgalactosamine to serine and threonine amino acid residues. GALNT7 has been identified as a potential biomarker for prostate cancer, and its overexpression has been shown to promote prostate cancer growth [13,14]. In colorectal cancer, GALNT7 serves as a microRNA target and promotes the progression and development of cancer [15,16]. Additionally, in glioma, cervical carcinoma, esophageal squamous cell carcinoma, and nasopharyngeal carcinoma, GALNT7 appears to play a pro-carcinogenic role [17–20]. To summarize, GALNT7 demonstrates associations with the onset and advancement of diverse cancers, underscoring its potential as an oncogenic factor. Nevertheless, it is important to note the absence of a comprehensive pan-cancer analysis for GALNT7 at present. We assert that conducting a comprehensive and exhaustive pan-cancer analysis of GALNT7 is imperative to foster deeper insights and promote further exploration in this domain.

2. Result

2.1. GALNT7 is abnormally expressed in various cancers

We initially retrieved the expression profiles of GALNT7 in 33 different types of cancers from the TCGA database. The abbreviation list of cancers is in Supplementary Table 1. As depicted in Fig. 1A, upon contrasting the expression levels between cancerous and normal tissues, we observed that GALNT7 exhibited significantly elevated expression in LUAD, BRCA, ESCA, STES, KIRP, STAD, PRAD, UCEC, LUSC, THCA, and CHOL. Conversely, its expression was diminished in COAD, READ, KIRC, and KICH.

Given the limited availability of normal tissue samples in the TCGA database, which may potentially lead to statistical biases when

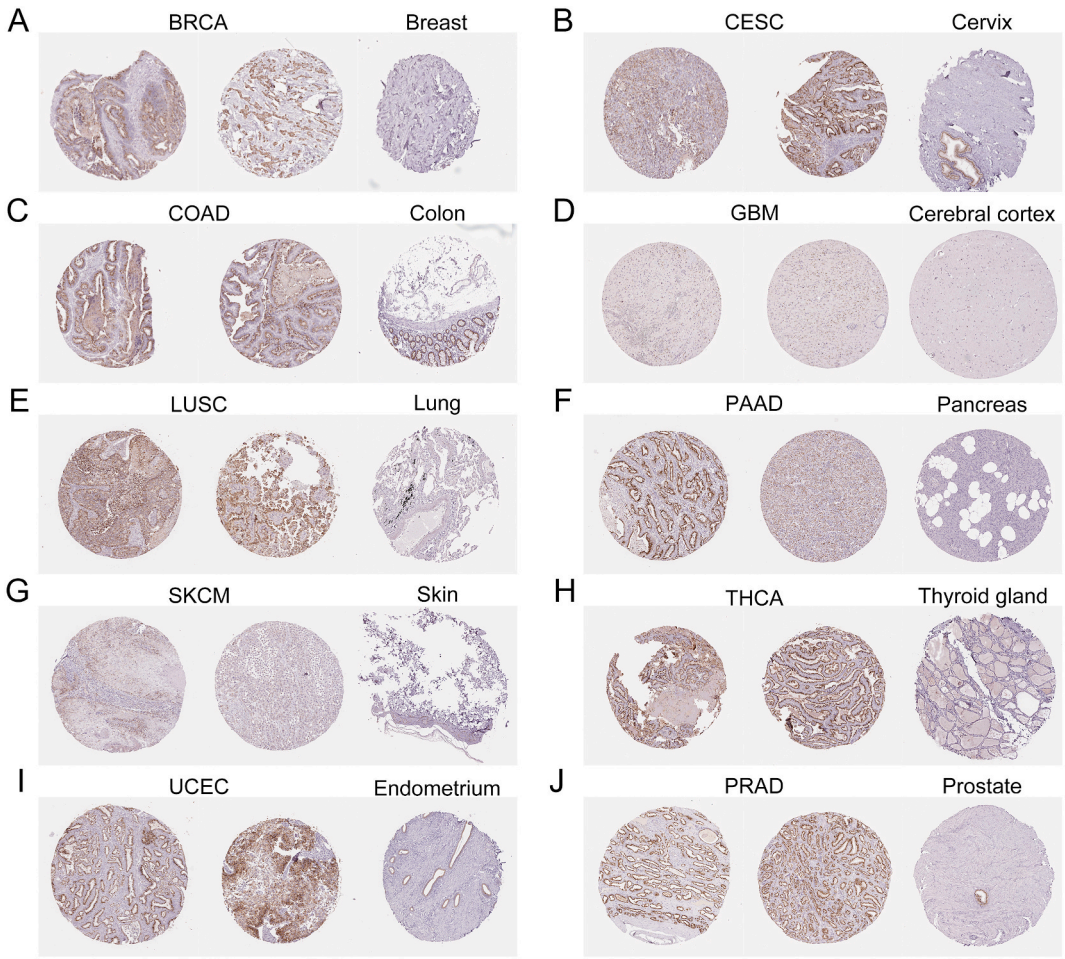


Fig. 2. Immunohistochemistry of GALNT7. Figures A to J represent the immunohistochemical results for BRCA, CESC, COAD, GBM, LUSC, PAAD, SKCM, THCA, UCEC, and PRAD. The two images on the left show immunohistochemistry of cancer tissues, while the image on the right corresponds to immunohistochemistry of the respective normal tissues.

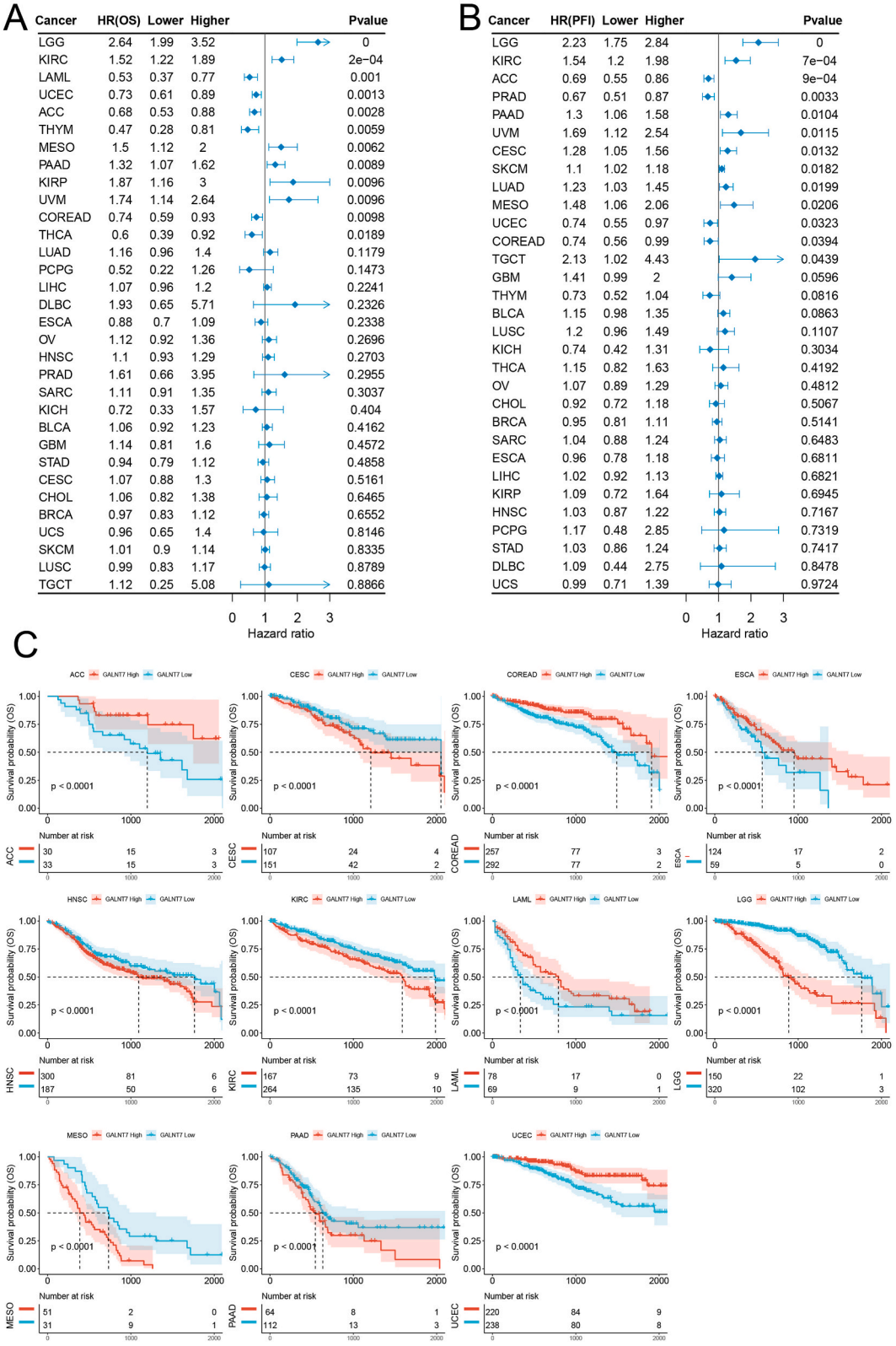


Fig. 3. The relationship between GALNT7 and cancer prognosis. (A–B) Forest plots for GALNT7 in relation to OS and PFI. (C) Kaplan-Meier survival curves in different cancers. OS, overall survival; PFI, progression free interval.

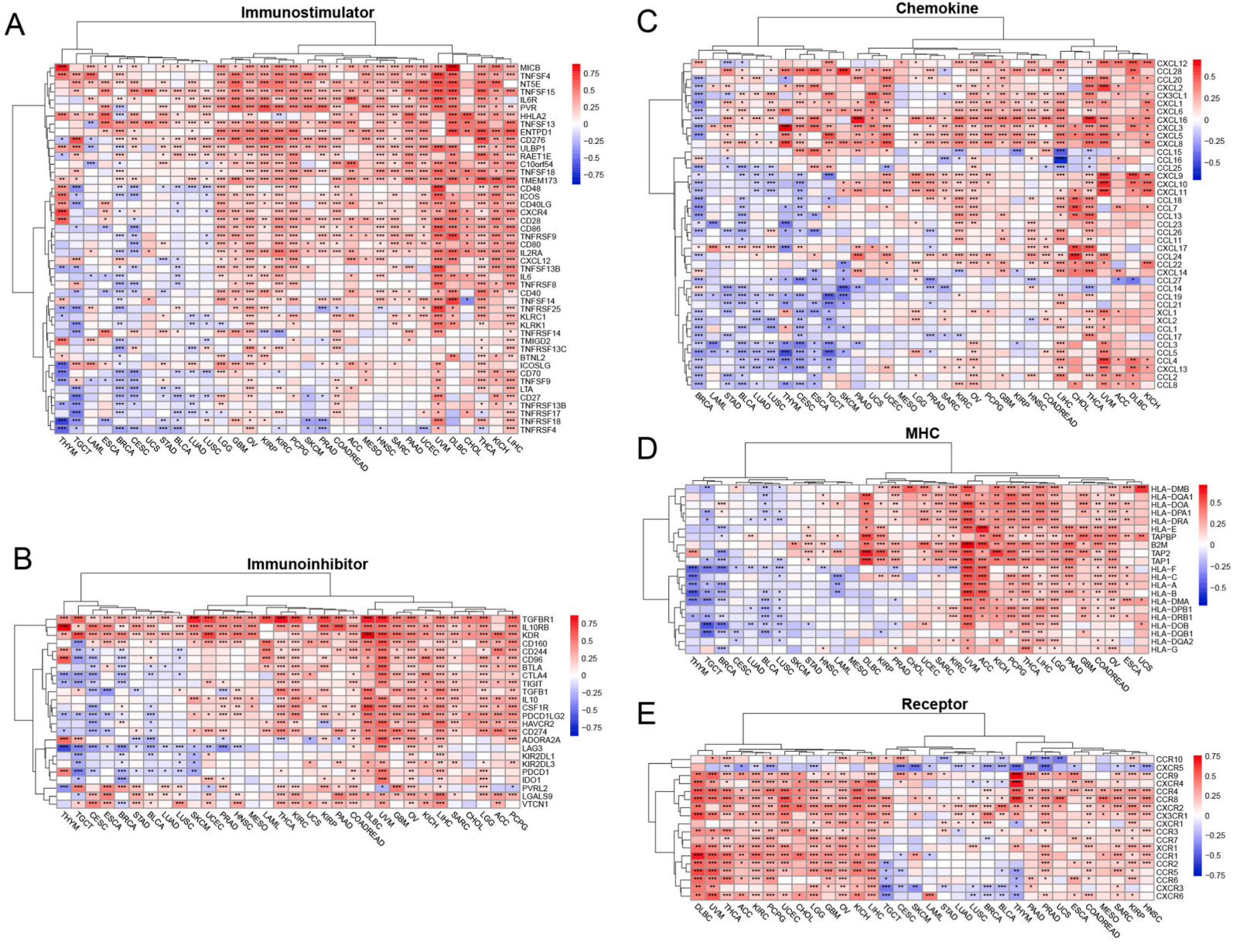


Fig. 4. The correlation between GALNT7 and immune genes. Figures A to E represent the correlation heatmaps of GALNT7 with immune stimulatory genes, immune inhibitory genes, chemokines, MHC, and immune receptors, respectively. In the heatmaps, red indicates a positive correlation, while blue represents a negative correlation. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. MHC, major histocompatibility complex.

contrasting normal and cancerous tissues, we augmented our analysis by incorporating normal tissue expression matrices from the GTEX database, merging them with the TCGA data for differential analysis. Furthermore, both GTEX and TCGA data have been corrected for batch effects, allowing for direct comparisons. As illustrated in Fig. 1B, we identified further instances of aberrant GALNT7 expression. Notably, GALNT7 displayed significantly elevated expression in GBM, LGG, UCEC, BRCA, LUAD, ESCA, STES, KIRP, COAD, PRAD, STAD, LUSC, WT, SKCM, THCA, PAAD, ALL, LAML, and CHOL, while its expression remained relatively low in KICH, UCS, and READ compared to normal tissue in cancerous tissues. It is worth noting that the comparison in COAD shifted from relatively low expression to relatively high expression after including more normal tissue samples, while KIRC showed no statistically significant difference. Next, through an analysis of CCLE data, we observed a relatively consistent distribution of GALNT7 expression across various cancer cell lines (Fig. 1C). By exclusively analyzing GTEX data, we found that GALNT7 exhibited significantly low expression levels in liver, muscle, and heart tissues, while demonstrating higher expression levels in salivary gland, bone marrow, and prostate tissues (Fig. 1D). Considering that both TCGA and GTEX datasets pertain to mRNA levels, we employed the HPA database to validate the protein expression of GALNT7 in cancerous and normal tissues. Fig. 2A–J presents the immunohistochemical results of GALNT7 across various cancers. Consistent with mRNA-level observations, we noted that GALNT7 exhibited higher expression at the protein level in cancer.

In summary, Fig. 1 illustrates the expression profiles of GALNT7 in cancerous tissues, normal tissues, and cell lines. Our findings indicate a tendency for GALNT7 to exhibit elevated expression levels in cancer, suggesting a potential oncogenic role for GALNT7.

2.2. GALNT7 is associated with prognosis in various cancers

In this study, prognostic information was acquired from TCGA, and subsequent survival analysis was conducted based on GALNT7 expression levels, with patients stratified into high and low expression groups using the median as the cutoff point. The relationship between GALNT7 expression and patient Overall Survival (OS) and Progression-Free Interval (PFI) was visually illustrated in Fig. 3A–B. The findings revealed distinct patterns across different cancer types.

In several cancer types, including LGG, KIRC, MESO, PAAD, KIRP, and UVM, patients with high GALNT7 expression exhibited relatively favorable prognoses, while contrasting trends were observed in LAML, UCEC, ACC, THYM, COREAD, and THCA. These observations were consistent with the forest plot analysis for PFI, which showed that high GALNT7 expression was associated with worse prognosis in LGG, KIRC, PAAD, UVM, CESC, SKCM, LUAD, MESO, and TGCT, but with better outcomes in ACC, PRAD, UCEC, and COREAD. Importantly, high GALNT7 expression was statistically correlated with poorer OS and PFI in LGG, KIRC, UVM, PAAD, and MESO.

Furthermore, it was noted that in LUAD, LGG, SKCM, and PAAD, GALNT7 expression was not only markedly upregulated in cancerous tissues but was also associated with worse prognoses. This observation warrants further investigation by experts in the relevant field.

Subsequently, the study employed the optimal cutoff method to stratify each type of cancer based on GALNT7 expression levels and conducted survival difference analysis. The use of cutoff values allowed for the identification of distinct risk groups within each cancer type, emphasizing the significance of GALNT7 in 11 different types of cancer (Fig. 3C). These findings suggest that GALNT7 may play a crucial role in the prognosis of these cancers, providing valuable insights for future research and potential therapeutic interventions.

3. The relationship between GALNT7 and immune genes

Currently, the scientific literature offers a limited number of articles discussing the interplay between GALNT7 and the immune system. Nonetheless, we contend that GALNT7's significance in the realm of tumor immunity has been notably underestimated. In light of this, we conducted an extensive examination of the association between GALNT7 and various immune-related genes across 33 distinct cancer types.

Fig. 4A visually represents a correlation heatmap elucidating the connections between GALNT7 and genes responsible for stimulating the immune response. Notably, our analysis revealed a substantial positive correlation in several cancer types, namely LIHC, KICH, THCA, UVM, GBM, LGG, and OV, with a contrasting negative correlation pattern predominating in BRCA, ESCA, and TGCT. Additionally, we observed consistent positive correlations between GALNT7 and TMEM173, TNFSF18, ENTPD1, IL6R, TNFSF15, NT5E, and TNFSF4 across most cancer types.

Fig. 4B delineates the relationship between GALNT7 and genes associated with immune inhibition. Interestingly, we found a prevailing positive correlation in UVM, GBM, OV, LIHC, THCA, and KIRC, with statistically significant correlations in OV and LIHC. Conversely, negative correlations predominated in CESC, ESCA, BRCA, and TGCT.

Subsequently, we conducted an in-depth examination of GALNT7's correlation with chemokine genes, major histocompatibility complex (MHC)-related genes, and immune receptor genes, as illustrated in Fig. 4C, D, and 4E, respectively. Our analysis unveiled varying patterns of correlation in different cancer types. Multiple chemokine genes displayed positive correlations with GALNT7 in LIHC, CHOL, THCA, UVM, ACC, DLBA, and KICH, whereas negative correlations were noted in ESCA, CESC, THYM, LUSC, LUAD, BLCA, STAD, and BRCA, with a particularly pronounced negative trend in BRCA. Likewise, GALNT7 exhibited positive correlations with MHC-related genes in UVM, ACC, KICH, PCPG, THCA, LIHC, LGG, PAAD, GBM, COREAD, and OV, while negative correlations were evident in BRCA, TGCT, BLCA, LUSC, and THYM. Furthermore, Fig. 4E indicated a predominantly positive correlation between GALNT7 and immune receptor genes across most tumors, with the exception of BRCA, BLCA, CESC, TGCT, and THYM, where negative trends were observed. Notably, CXCR5 displayed a consistent negative correlation with GALNT7 in most cancer types. It is worth mentioning that, unfortunately, we did not identify a strong association between GALNT7 expression and prognosis in these specific

tumor types.

3.1. GALNT7 and immune infiltration

Based on the aforementioned analysis of GALNT7 and immune-related genes, we have identified a potential correlation between GALNT7 and immunity in cancer. Furthermore, we employed various algorithms to further analyze the relationship between GALNT7 and immune infiltration. ESTIMATE (Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data) is a computational algorithm used in cancer immunology research to estimate the composition of stromal and immune cells within tumor tissues based on gene expression data. It was developed to gain insights into the tumor microenvironment and understand how the immune system interacts with cancer cells. Fig. 5A displays the correlation between immune infiltration scores calculated by the ESTIMATE algorithm and GALNT7. In STAD, UCEC, CESC, PRAD, BLCA, and LUSC, GALNT7 exhibits a negative correlation with all three immune scores. In BRCA, LAML, LUAD, KIRP, and TGCT, GALNT7 is negatively correlated with immune scores and ESTIMATE scores. In LIHC and LGG, GALNT7 shows a positive correlation trend with all three immune scores.

TIMER (Tumor Immune Estimation Resource) is a computational algorithm used in cancer immunology research to estimate the abundance of immune cell populations within tumor tissues based on gene expression data. Fig. 5B represents the relationship between GALNT7 and immune cell infiltration scores obtained using the TIMER algorithm. The results indicate that in THYM, KIRC, LIHC, LGG, PCPG, OV, and KICH, GALNT7 exhibits a positive correlation trend with various immune cell types. Conversely, in BRCA and CESC, the trend is negative. $CD8^+$ T cells and neutrophils show a positive correlation trend with GALNT7 in multiple tumors.

In addition to the TIMER algorithm, we utilized the EPIC algorithm to compute the correlation between GALNT7 and various immune cells. As depicted in Fig. 5C, NK cells, B cells, and macrophages exhibit a negative correlation trend with GALNT7 across most tumors. Notably, $CD4^+$ T cells display a significant positive correlation with GALNT7 in tumors such as UVM, KICH, THCA, PRAD, and UCS. $CD4^+$ T cells are closely associated with the efficacy of immunotherapy [21].

3.2. GALNT7 and tumor heterogeneity

HRD, MSI, LOH, and tumor purity are key markers used in cancer research to evaluate genetic and molecular aspects of tumors. They have significant implications for treatment decisions, patient prognosis, and our understanding of the underlying biology of

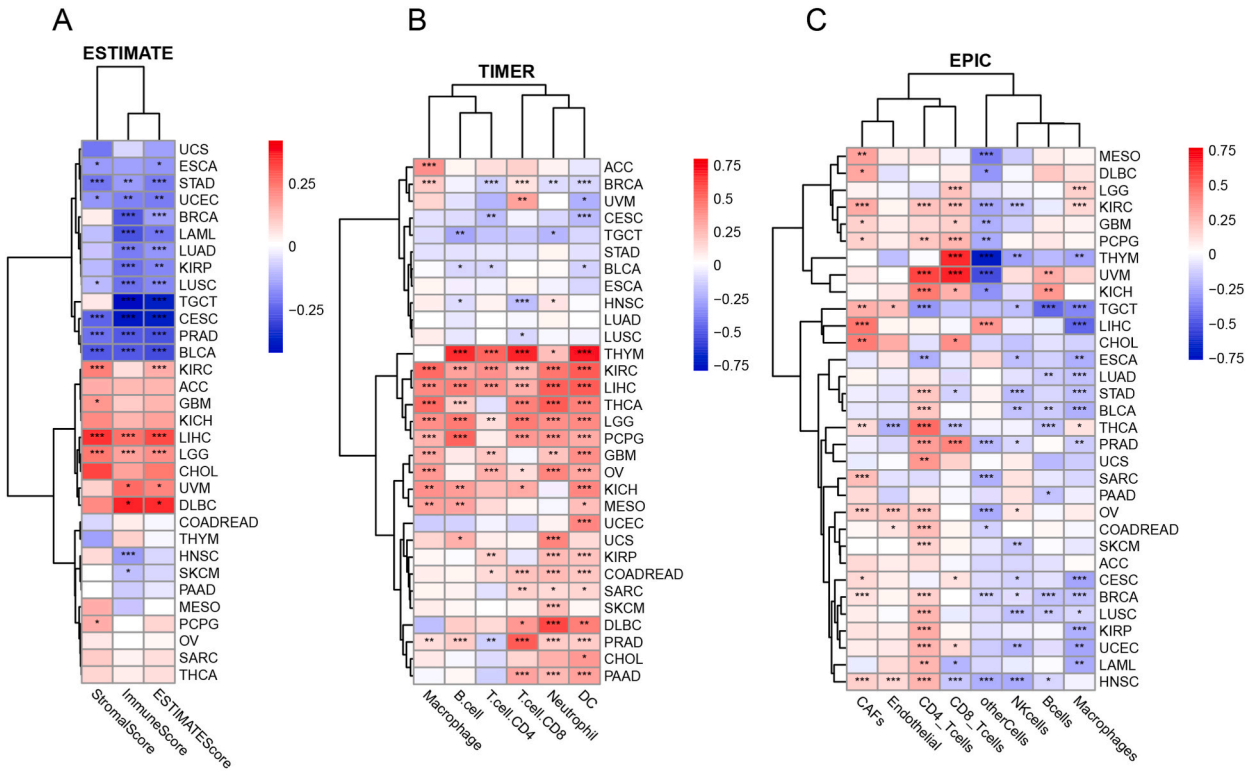


Fig. 5. The correlation between GALNT7 and immune infiltration. Figures A–C represent the correlation heatmaps of GALNT7 with tumor-infiltrating cells calculated through the ESTIMATE algorithm, TIMER algorithm, and EPIC algorithm, respectively. In these heatmaps, red signifies a positive correlation, while blue indicates a negative correlation. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. ESTIMATE, estimation of stromal and immune cells in malignant tumor tissues using expression data; TIMER, tumor immune estimation resource.

different cancer types.

HRD is a measure of genomic instability, particularly in the context of DNA repair mechanisms. As illustrated in Fig. 6A, the correlation between GALNT7 expression levels and HRD is presented. We observed a significant correlation in 19 different tumors. Notably, in BRCA, the correlation coefficient is R = −0.43, while in CHOL, it is R = 0.40.

MSI is a marker for DNA mismatch repair (MMR) deficiency. Tumors with MSI exhibit alterations in the lengths of microsatellite repeats due to the accumulation of errors in DNA replication. MSI-high tumors are often associated with a better response to immunotherapy, as they tend to have a higher mutation load, making them more recognizable to the immune system. Fig. 6B illustrates the correlation between GALNT7 and Microsatellite Instability (MSI). We observed a significant correlation in 13 different tumors, with a positive correlation in 8 of them and a negative correlation in 5. However, in DLBC, the correlation coefficient is particularly notable at R = −0.54, while in other tumors, the correlation values are approximately around 0.2. Therefore, we may focus on

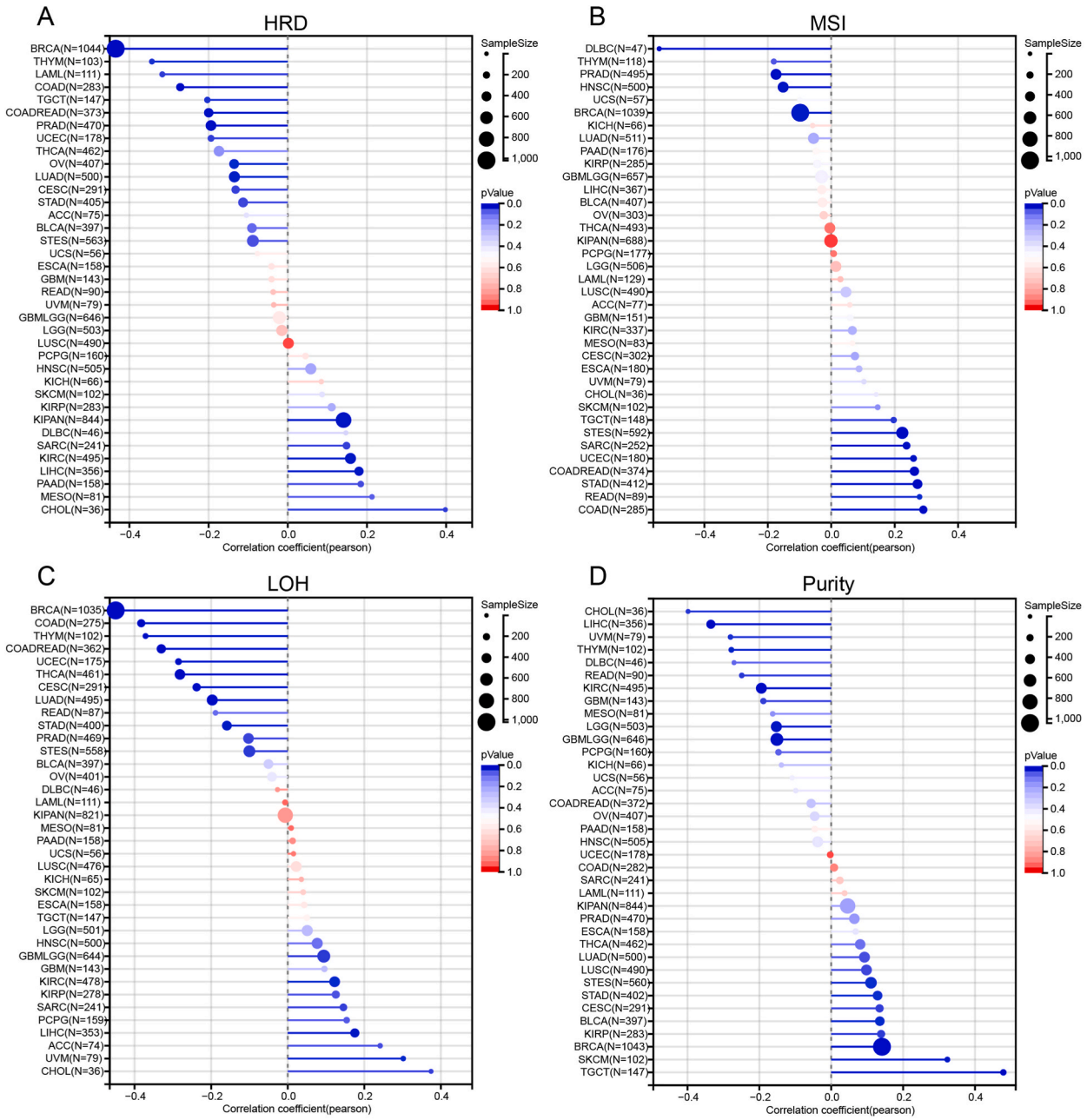


Fig. 6. The relationship between GALNT7 and various tumor heterogeneity indicators. A-D respectively display the associations of GALNT7 with HRD, MSI, LOH, and tumor purity. The x-axis represents the correlation coefficient (R), the color indicates the p-value (blue for $P < 0.05$), and the dot size corresponds to sample size. HRD, homologous recombination deficiency; MSI, microsatellite instability; LOH, loss of heterozygosity.

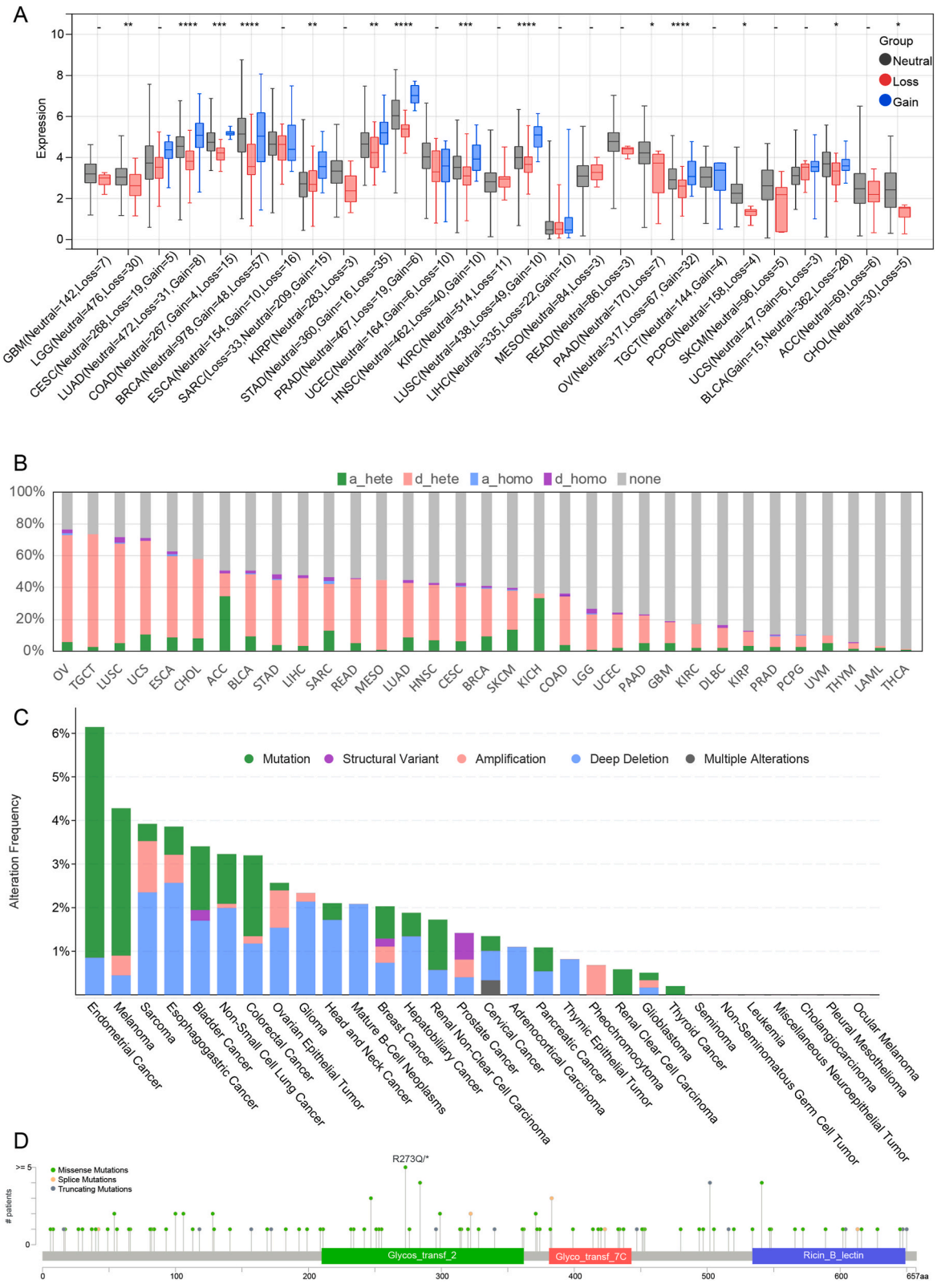


Fig. 7. The relationship between GALNT7 and gene mutations. (A) Differential expression of GALNT7 among different CNV mutations in cancer; (B) The CNV mutation frequency of GALNT7 in different cancers; (C) The frequency of mutation types of GALNT7 in different cancers; (D) Information about the mutation sites of GALNT7.

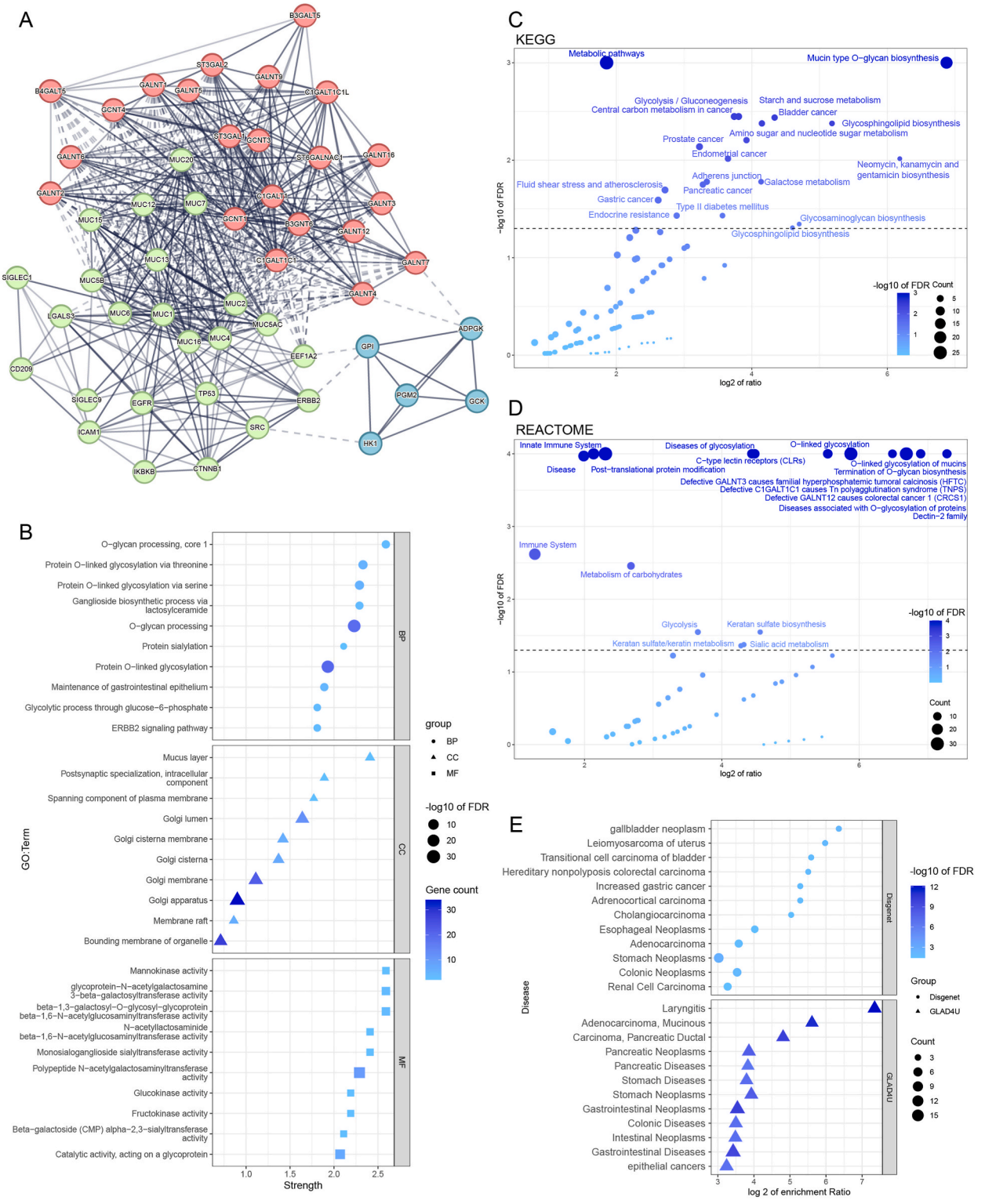


Fig. 8. Results of GALNT7 enrichment analysis. (A) A protein-protein interaction network obtained using GALNT7 as a keyword. (B) GO analysis results based on the PPI network. (C–D) KEGG and REACTOME results based on the PPI network. (E) Enriched diseases related to the PPI network. PPI, protein-protein interaction; GO, gene ontology; KEGG, kyoto encyclopedia of genes and genomes; FDR, false discovery rate.

exploring the relationship between GALNT7 and MSI in DLBC.

LOH analysis is essential in cancer genetics, as it can identify tumor suppressor genes that have undergone inactivation due to the loss of one functional copy. We found statistically significant correlations between GALNT7 and LOH in 19 different tumors, with 8 showing positive correlations and 11 exhibiting negative correlations (Fig. 6C). Notably, in BRCA, COAD, and THYM, the correlation coefficient (R) is around −0.4, particularly in the case of BRCA. In BRCA, GALNT7 not only displays a significant negative correlation with LOH but also a significant negative correlation with Homologous Recombination Deficiency (HRD). Similarly, in CHOL, GALNT7 shows significant positive correlations with both HRD and LOH.

Finally, we analyzed the correlation between GALNT7 and tumor purity (Fig. 6D). GALNT7 displayed a positive correlation with tumor purity in 10 types of tumors and a negative correlation in 9 types of tumors. Notably, CHOL exhibited a correlation of −0.40 with tumor purity, while TGCT had a correlation of 0.48. This is consistent with the results in Fig. 5A, where a positive correlation between high GALNT7 and tumor purity suggests lower immune infiltration, and the opposite holds true.

3.3. GALNT7 mutations in tumors

Tumor heterogeneity and genetic mutations are intricately interlinked in the domain of oncology. Tumor heterogeneity denotes the existence of a myriad of cell subpopulations within a neoplasm, each endowed with its unique genetic and molecular attributes. The origin of these divergent cell subpopulations can be ascribed to genetic mutations, and they wield significant influence in the trajectory of tumor genesis and advancement. As discerned from Fig. 6, an association between GALNT7 and tumor heterogeneity has been unveiled, notably within neoplasms such as BRCA and CHOL. Consequently, we undertook an examination of mutations involving GALNT7.

Firstly, we analyzed the relationship between GALNT7 and CNV mutations (Fig. 7A). We observed that samples with CNV loss generally exhibited lower GALNT7 expression, whereas patients with CNV gain displayed higher GALNT7 expression compared to the neutral group. This suggests that CNV variations in the loss group lead to a decrease in the copy number of the GALNT7 gene. This decrease in copy number is significantly correlated with the substantial reduction in GALNT7 expression, indicating that a reduced copy number may impair GALNT7 function, resulting in decreased expression levels. Conversely, CNV variations in the gain group result in an increase in the copy number of the GALNT7 gene. Corresponding to this increase in copy number, GALNT7 expression levels also significantly increase. This implies that an increased copy number may lead to the overexpression of GALNT7.

Next, we conducted an analysis of GALNT7's CNV mutations across different tumors (Fig. 7B). We found that the primary mutation type of GALNT7 is heterozygous deletion, with proportions exceeding 50 % in OV, TGCT, LUSC, UCS, ESCA, and CHOL. In ACC and KICH, heterozygous amplification of GALNT7 accounts for over 30 %. However, whether homozygous amplification or homozygous deletion, the proportions remain extremely low. Furthermore, in THCA, LAML, THYM, and UVM, the overall CNV mutation proportions all remain below 10 %.

Subsequently, we employed the cbioportal platform to conduct an in-depth examination of the frequency of alterations and the specific mutation loci within the GALNT7 gene. In general, the data indicates a relatively modest incidence of GALNT7 mutations, with frequencies surpassing 6 % documented in the context of endometrial cancer, while hovering around 4 % in melanoma, sarcoma, and esophagogastric cancer. Conversely, in the case of other malignancies, the mutation frequency consistently resides below the 4 % threshold. Notably, in CHOL, the incidence of GALNT7 mutations stands at 0 % (refer to Fig. 7C). The primary mutation archetypes predominantly encompass deep deletion events and point mutations.

Finally, we analyzed the mutation sites of GALNT7 (Fig. 7D). The results indicate that among the known mutation cases, nonsense mutations are the most prevalent, with a smaller number of splice mutations and truncation mutations. Furthermore, R273Q/* is the most frequently occurring mutation site (10 occurrences), followed by R284Q/* (4 occurrences), R284Q/* (4 occurrences), and E541K (4 occurrences).

3.4. Enrichment analysis of GALNT7

Based on the previous results, we believe that GALNT7 may be associated with tumor immunity and tumor heterogeneity. Previous studies have found that GALNT7 primarily functions in glycosylation processes, and we hypothesize that it may play a role in various cellular functions. Therefore, we further conducted enrichment analysis to identify its potential additional functions. Fig. 8A depicts the protein-protein interaction network related to GALNT7 obtained from the STRING database. These genes interact with or are similar to GALNT7, suggesting potential synergistic roles in biological processes.

We conducted a Gene Ontology (GO) analysis of the potential functions of the genes in Fig. 8A. The GO Biological Process (GOBP) results reveal that GALNT7, in addition to its involvement in glycosylation, also plays a role in processes such as Ganglioside biosynthesis, Protein sialylation, and the ERBB2 signaling pathway. Furthermore, the primary cellular localization of these genes is in the Mucus layer and the Golgi apparatus. Additionally, these proteins participate in the activation of various enzymes (Fig. 8B).

In comparison to GO analysis, both KEGG analysis and REACTOME analysis emphasize the annotation of metabolic pathways and cellular signaling pathways. Therefore, we further used these two annotation libraries to analyze the proteins in the PPI network. As shown in Fig. 8C, KEGG pathways also enrich several glycosylation-related entries. In addition, pathways related to prostate cancer, bladder cancer, pancreatic cancer, gastric cancer, endometrial cancer, and various metabolic pathways are enriched. This suggests that GALNT7 may play multiple roles in cancer beyond its involvement in glycosylation processes (Fig. 8C). Similarly, REACTOME results include various metabolic pathways, including Metabolism of carbohydrates and Glycolysis (Fig. 8D). These enrichment analysis results warrant our attention, as GALNT7 may have undiscovered functions that are worth further research and exploration. Finally,

Fig. 8E displays disease enrichment results based on the PPI network proteins. We found that multiple cancers exhibit significant statistical significance in the results.

4. Discussion

Our investigation provides compelling evidence implicating GALNT7 in cancer progression and underscores its potential as a novel therapeutic target. We observed consistent overexpression of GALNT7 across various cancer types compared to normal tissues, with higher expression levels correlating with poorer patient outcomes. The oncogenic properties of GALNT7 likely stem from its modulation of O-linked glycosylation pathways, a phenomenon well-documented in cancer biology. Aberrant glycosylation, a hallmark of cancer, contributes significantly to crucial processes including epithelial-mesenchymal transition, metastasis, and resistance to chemotherapy [22–28]. As a member of the polypeptide N-acetylgalactosaminyltransferase family, GALNT7 assumes a pivotal role in orchestrating the initial addition of galactose residues during protein O-linked glycosylation. Consequently, glycosyltransferases like GALNT7 emerge as plausible drivers of oncogenesis by reprogramming the glycoproteome. Furthermore, GALNT7 exhibits heightened expression across multiple solid tumor types as evidenced by cancer genome databases, a finding validated through our comparative analysis of cancer versus normal tissue samples, indicative of its oncogenic potential. It is worth noting that overexpressed genes often present viable targets for therapeutic intervention in cancer. Given the central role of aberrant glycosylation in cancer pathogenesis, we postulate that GALNT7, functioning as a central regulator of protein glycosylation, may act as a "master switch" modulating multiple tumorigenic pathways. Its widespread upregulation across diverse cancer types underscores the potential of GALNT7 as a universal therapeutic target in cancer, transcending the confines of individual tumor types.

In the Results section, our initial focus was on revealing the behavior of the GALNT7 gene across various cancer types. Our analysis revealed heterogeneous expression patterns of GALNT7, with heightened activity observed in certain cancers (notably excluding KICH, UCS, and READ) and diminished activity in others. To enhance the accuracy of our findings, we combined data derived from both cancerous and normal tissues. This comprehensive approach served to corroborate the elevated activity of GALNT7 across multiple cancer types. Furthermore, we extended our investigation to ascertain the protein levels of GALNT7, which were consistently elevated in cancer cells compared to their non-cancerous counterparts.

We then analyzed how GALNT7's activity is related to the prognosis (how well patients do) in different types of cancer. In some cancers, high GALNT7 activity is associated with better outcomes, while in others, it's linked to worse outcomes. Notably, in cancers like LUAD, LGG, SKCM, and PAAD, high GALNT7 activity is not only associated with cancer but also with poorer outcomes, highlighting the need for further research. An investigation has revealed a notable elevation of GALNT7 expression within PRAD tissues, in accordance with our findings [14]. However, whereas this inquiry posits that the heightened GALNT7 expression fosters PRAD proliferation, a contention incongruent with our study's findings regarding PFI. Glioma represents a malignancy of high aggression. Our investigation unveils a pronounced elevation in GALNT7 expression within both GBM and LGG, concomitant with a notably adverse prognosis. This observation finds validation in a study demonstrating that heightened GALNT7 expression facilitates the proliferation and invasion of glioma cells [29]. In CESC, research has also confirmed the oncogenic capability of GALNT7. MicroRNA-30e has been identified as a regulator of GALNT7 expression, with knockdown of GALNT7 resulting in inhibited cell proliferation while overexpression promotes cell proliferation [18]. Our study corroborates these findings by demonstrating that high GALNT7 expression in CESC is associated with poorer prognosis, aligning with the results of this research.

We also studied how GALNT7 is connected to genes related to the immune system in various cancers. We found that GALNT7 has both positive and negative connections with different immune-related genes in different cancers, suggesting a complex relationship between GALNT7 and the immune system. In colorectal cancer, a glycosylation-based risk prediction model can forecast patient immune checkpoint inhibitor response, with GALNT7 emerging as a pivotal factor within this predictive framework [30]. Notably, in OV, LICH, and UVM, GALNT7 exhibits a predominantly positive and statistically significant correlation with nearly all immune-related genes. The underlying reasons for this phenomenon merit in-depth investigation. Conversely, in the case of BRCA, GALNT7 displays a negative correlation with immune-related genes. This unique characteristic of BRCA is further emphasized in subsequent analyses of tumor heterogeneity. This suggests that GALNT7 may simultaneously influence tumor heterogeneity and tumor immunity in the context of BRCA. A study has elucidated an association between GALNT7 and the estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 (HER2) in breast cancer. Furthermore, heightened GALNT7 expression correlates with favorable outcomes in recurrence-free survival and distant metastasis-free survival among HER2-dependent breast cancer patients [31]. Synthesizing our findings, a plausible link emerges: GALNT7 may exert an influence on the tumor microenvironment or patient prognosis across various subtypes of breast cancer. An additional investigation has unveiled that GALNT7 is modulated by SPDEF in luminal-type breast cancer, where it serves a dual function in both tumorigenesis and stemness processes [32].

Next, we looked at how GALNT7 is related to the presence of immune cells in tumors. In some cancers, GALNT7 was associated with fewer immune cells, and in others, it was linked to more immune cells, which might have implications for treatment and cancer behavior. In LIHC, GALNT7 exhibits not only a positive correlation with immune genes but also a significant positive correlation with various tumor immune-infiltrating cells. This suggests that in LIHC, GALNT7 may potentially influence tumor development by enriching immune cells through immune-related genes. However, disappointingly, high GALNT7 expression in LIHC does not lead to either worse or better prognosis. In the context of lung cancer, miR-30c exhibits the capacity to target GALNT7, thereby triggering the cytotoxicity of NK cells against lung cancer cells [33].

We explored how GALNT7 is associated with genetic aspects of tumors, such as genomic instability (HRD), DNA repair (MSI), and genetic mutations. GALNT7's activity was related to these markers in various ways, suggesting that it could influence the genetic characteristics of cancer. Cancer types warranting further attention include BRCA, CHOL, and TGCT. In BRCA, higher GALNT7

expression corresponds to lower tumor heterogeneity. This indicates potential for GALNT7 to serve as a therapeutic biomarker in treatments targeting DNA repair mechanisms. Conversely, in CHOL, higher GALNT7 expression signifies higher tumor heterogeneity. In CHOL, LIHC, and UVM, GALNT7 demonstrates a statistically significant negative correlation with tumor purity and an inverse relationship with immune-related genes. This connection potentially suggests that GALNT7 may be associated with tumor immune infiltration, leading to reduced tumor purity. Furthermore, GALNT7 is linked to poorer prognosis in UVM, which highlights the need for more in-depth research in this context.

We also checked for mutations in the GALNT7 gene in different types of cancer. Some cancers showed more genetic changes in GALNT7, while others had fewer changes. These genetic changes could impact how GALNT7 functions.

Lastly, we conducted an enrichment analysis to understand other potential functions of GALNT7 beyond glycosylation. We found that GALNT7 might play roles in processes like ganglioside biosynthesis, signaling pathways, and diseases such as various cancers.

Certainly, this article has certain limitations. Firstly, the study lacks in vivo or in vitro validation. Additionally, this research extensively discusses correlations, but it's crucial to note that correlation does not imply causation; it represents a trend and can only assist us in identifying potential directions. In summary, our results show that GALNT7 is involved in cancer development, immune system interactions, genetic characteristics of tumors, and possibly other functions. This suggests that GALNT7 is an important gene with various roles in cancer.

5. Method

5.1. Data acquisition and processing

We procured a standardized pan-cancer dataset from the Xena functional genomics explorer database (https://xenabrowser.net/) [34]. To ensure uniformity and standardization, we subjected the expression values to a log2(x+1) transformation. Mitigating the constraint imposed by limited sample sizes for normal tissues within the TCGA database, we supplemented our analysis with expression data sourced from the GTEX database [35]. Furthermore, the dataset extracted from Xena, having undergone rigorous batch effect correction, facilitates direct comparability and amenable analysis.

5.2. Gene expression and survival analysis

We obtained a robust prognostic dataset from the TCGA prognostic investigation conducted by Liu J et al. [36]. Subsequently, we constructed a Cox proportional hazards regression model using the coxph function available within the "survival" R package. This facilitated an in-depth exploration of the correlation between gene expression and prognostic outcomes across various tumor types. To visually represent the statistical outcomes, we created forest plots using the "forestplot" R package. These plots effectively illustrated the P values, hazard ratios (HR), and corresponding 95 % confidence intervals (CI) for each variable of interest. We obtained immunohistochemical images of GALNT7 from the HPA website (https://www.proteinatlas.org/).

5.3. Immune analysis

We conducted an analysis involving the extraction of expression data pertaining to two categories of immune checkpoint pathway genes, namely inhibitory and stimulatory, in addition to five categories of immune pathway genes encompassing chemokine, receptor, MHC, immuno-inhibitor, and immuno-stimulator genes. This analysis was based on the TCGA dataset, with a stringent filtering process that excluded all normal samples, focusing exclusively on tumor samples.

To explore the relationship between immune cells and the expression of GALNT7, we employed the deconvo_xCell method available through the R package IOBR [37,38]. It is worth noting that the ESTIMATE algorithm furnishes three distinctive scores: the immune score, which gauges the level of immune cell infiltration; the stromal score, which evaluates the stromal component's involvement in immunity; and the overall ESTIMATE score [39]. We employed the "Estimate" R package to calculate these three scores for each TCGA sample. To maintain statistical rigor and validity, we employed the false discovery rate (FDR) method for p-value correction during the correlation analysis.

5.4. Enrichment analysis and protein interaction analysis

The construction of the protein-protein interaction (PPI) network was facilitated through the utilization of the Search Tool for the Retrieval of Interacting Genes (STRING) platform, accessible at https://cn.string-db.org/. We performed enrichment analysis using the R package "clusterProfiler" for GO, KEGG, and RECTOME analyses [40]. For the genes acquired via PPI, we performed KEGG, RECTOME and GO enrichment analyses. Within each cancer type present in the TCGA dataset, we categorized the samples into high and low-expression groups, stratified according to GALNT7 expression levels. Subsequently, we employed the "limma" R package to conduct differential analysis. Specifically, we selected genes that exhibited a fold change (FC) of $\geq$2 and a p-value less than 0.05 for further enrichment analysis. For disease enrichment analysis, we used two dateset named DisGeNet and GLAD4U, respectively [41–43].

5.5. Genetic heterogeneity analysis

Next, we conducted an analysis of the relationship between GALNT7 and various tumor heterogeneity indicators, including MSI (Microsatellite Instability), HRD (Homologous Recombination Deficiency), LOH (Loss of Heterozygosity), and tumor purity. From the previous research, we obtained MSI data, HRD, LOH, and tumor purity data for each TCGA sample [44,45]. After performing log transformation on the data, we merged it with GALNT7 expression data and conducted correlation analysis.

5.6. Gene mutation analysis

For Copy Number Variation (CNV) mutation data, we have diligently leveraged GISTIC software to meticulously process the Copy Number Variation dataset at level 4. Subsequently, we seamlessly amalgamated the refined data with the GALNT7 expression dataset. This confluence enabled us to embark upon inter-cancer group comparative analysis as well as mutation proportion analysis for each distinct neoplastic entity [46]. Furthermore, our investigative arsenal included the esteemed cBioPortal, through which we accessed the precise mutational loci pertaining to GALNT7, along with their respective prevalence in the heterogeneous landscape of malignancies [47].

5.7. Statistical analysis

Given the substantial disparities in mRNA-level expression, we judiciously subjected the expression matrix to logarithmic transformation using the log2(x+1) method. For inter-group analyses, we availed ourselves of either the *t*-test or non-parametric tests. When engaging in correlation assessments, the Pearson correlation coefficient was harnessed, accompanied by p-value calibration through the Holm-Bonferroni method.

Data availability statement

The datasets analyzed during the current study are available in the TCGA (https://xenabrowser.net), GETx (https://www.genome.gov/), CCLE (https://depmap.org/portal/ccle/), XENA (https://xenabrowser.net/datapages/), cBioportal (https://www.cbioportal.org/) and HPA (https://www.proteinatlas.org/) databases.

Additional information

The authors declare no competing interests.

This research was funded by the Huzhou Science and Technology Bureau (2023 Municipal Science and Technology Plan Projects, No. 2023GY40).

CRediT authorship contribution statement

Yan Liu: Writing – original draft, Conceptualization. **Yue Sun:** Formal analysis, Data curation. **Meixia Xiao:** Software, Resources, Data curation. **Shuang Li:** Methodology, Investigation. **Shengming Shi:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the online analysis website Sangerbox (http://vip.sangerbox.com/).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31515.

References

[1] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Jemal, Cancer statistics, CA A Cancer J. Clin. 73 (2023) (2023) 17–48, https://doi.org/10.3322/caac.21763.
[2] N.N. Pavlova, J. Zhu, C.B. Thompson, The hallmarks of cancer metabolism: still emerging, Cell Metabol. 34 (2022) 355–377, https://doi.org/10.1016/j.cmet.2022.01.007.
[3] D. Hanahan, Hallmarks of cancer: new dimensions, Cancer Discov. 12 (2022) 31–46, https://doi.org/10.1158/2159-8290.Cd-21-1059.

[4] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (2011) 646–674, https://doi.org/10.1016/j.cell.2011.02.013.
[5] S.R. Stowell, T. Ju, R.D. Cummings, Protein glycosylation in cancer, Annual Review of Pathology 10 (2015) 473–510, https://doi.org/10.1146/annurev-pathol-012414-040438.
[6] H. Ma, et al., Targeting N-glycosylation of 4F2hc mediated by glycosyltransferase B3GNT3 sensitizes ferroptosis of pancreatic ductal adenocarcinoma, Cell Death Differ. 30 (2023) 1988–2004, https://doi.org/10.1038/s41418-023-01188-z.
[7] R. Gupta, R. Ponangi, K.G. Indresh, Role of glycosylation in breast cancer progression and metastasis: implications for miRNA, EMT and multidrug resistance, Glycobiology 33 (2023) 545–555, https://doi.org/10.1093/glycob/cwad046.
[8] Y. Wang, H. Chen, Protein glycosylation alterations in hepatocellular carcinoma: function and clinical implications, Oncogene 42 (2023) 1970–1979, https://doi.org/10.1038/s41388-023-02702-w.
[9] L. He, Z. Guo, W. Wang, S. Tian, R. Lin, FUT2 inhibits the EMT and metastasis of colorectal cancer by increasing LRP1 fucosylation, Cell Commun. Signal. : CCS 21 (2023) 63, https://doi.org/10.1186/s12964-023-01060-0.
[10] J. Zhang, et al., GALNT1 enhances malignant phenotype of gastric cancer via modulating CD44 glycosylation to activate the Wnt/β-catenin signaling pathway, Int. J. Biol. Sci. 18 (2022) 6068–6083, https://doi.org/10.7150/ijbs.73431.
[11] A.F. Costa, D. Campos, C.A. Reis, C. Gomes, Targeting glycosylation: a new road for cancer drug discovery, Trends in cancer 6 (2020) 757–766, https://doi.org/10.1016/j.trecan.2020.04.002.
[12] F. Rocamora, et al., Glycosylation shapes the efficacy and safety of diverse protein, gene and cell therapies, Biotechnol. Adv. 67 (2023) 108206, https://doi.org/10.1016/j.biotechadv.2023.108206.
[13] M.C. Masone, The role of GALNT7 as a potential diagnostic marker in prostate cancer, Nat. Rev. Urol. 20 (2023) 198, https://doi.org/10.1038/s41585-023-00756-9.
[14] E. Scott, et al., Upregulation of GALNT7 in prostate cancer modifies O-glycosylation and promotes tumour growth, Oncogene 42 (2023) 926–937, https://doi.org/10.1038/s41388-023-02604-x.
[15] Y. Li, et al., Long non-coding RNA-SNHG7 acts as a target of miR-34a to increase GALNT7 level and regulate PI3K/Akt/mTOR pathway in colorectal cancer progression, J. Hematol. Oncol. 11 (2018) 89, https://doi.org/10.1186/s13045-018-0632-2.
[16] Y. Zhang, et al., Long non-coding RNA CCDC144NL-AS1 promotes cell proliferation by regulating the miR-363-3p/GALNT7 axis in colorectal cancer, J. Cancer 13 (2022) 752–763, https://doi.org/10.7150/jca.65885.
[17] J.B. Wang, et al., Long non-coding RNA TP73-AS1 contributes to glioma tumorigenesis by sponging the miR-103a/GALNT7 pathway, Brain Res. 1741 (2020) 146886, https://doi.org/10.1016/j.brainres.2020.146886.
[18] H. Wu, et al., MicroRNA-30e Functions as a Tumor Suppressor in Cervical Carcinoma Cells through Targeting GALNT7, Translational oncology 10 (2017) 876–885, https://doi.org/10.1016/j.tranon.2017.08.006.
[19] Q. Lu, et al., miR-214 inhibits invasion and migration via downregulating GALNT7 in esophageal squamous cell cancer, Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 37 (2016) 14605–14614, https://doi.org/10.1007/s13277-016-5320-7.
[20] G.H. Nie, et al., GALNT7, a target of miR-494, participates in the oncogenesis of nasopharyngeal carcinoma, Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 37 (2016) 4559–4567, https://doi.org/10.1007/s13277-015-4281-6.
[21] R.E. Tay, E.K. Richardson, H.C. Toh, Revisiting the role of CD4+ T cells in cancer immunotherapy—new insights into old paradigms, Cancer Gene Ther. 28 (2021) 5–17.
[22] P. Carvalho-cruz, F. Alisson-Silva, A.R. Todeschini, W.B. Dias, Cellular glycosylation senses metabolic changes and modulates cell plasticity during epithelial to mesenchymal transition, Dev. Dynam. 247 (2018) 481–491.
[23] L.M.d. Fonseca, et al., Glycosylation in cancer: interplay between multidrug resistance and epithelial-to-mesenchymal transition? Front. Oncol. 6 (2016) 158.
[24] M.C. Lucena, et al., Epithelial mesenchymal transition induces aberrant glycosylation through hexosamine biosynthetic pathway activation, J. Biol. Chem. 291 (2016) 12917–12929.
[25] M. Detarya, et al., Emerging roles of GALNT5 on promoting EGFR activation in cholangiocarcinoma: a mechanistic insight, Am. J. Cancer Res. 12 (2022) 4140.
[26] E.-M. Beaman, D.R. Carter, S.A. Brooks, GALNTs: master regulators of metastasis-associated epithelial-mesenchymal transition (EMT)? Glycobiology 32 (2022) 556–579.
[27] J.M. Burchell, A. Mungul, J. Taylor-Papadimitriou, O-linked glycosylation in the mammary gland: changes that occur during malignancy, J. Mammary Gland Biol. Neoplasia 6 (2001) 355–364.
[28] J.M. Burchell, R. Beatson, R. Graham, J. Taylor-Papadimitriou, V. Tajadura-Ortega, O-linked mucin-type glycosylation in breast cancer, Biochem. Soc. Trans. 46 (2018) 779–788.
[29] S. Hua, et al., High expression of GALNT7 promotes invasion and proliferation of glioma cells, Oncol. Lett. 16 (2018) 6307–6314, https://doi.org/10.3892/ol.2018.9498.
[30] P. Qiu, et al., Emerging glyco-risk prediction model to forecast response to immune checkpoint inhibitors in colorectal cancer, J. Cancer Res. Clin. Oncol. 149 (2023) 6411–6434, https://doi.org/10.1007/s00432-023-04626-0.
[31] C. Berkel, E. Cacan, The expression of O-linked glycosyltransferase GALNT7 in breast cancer is dependent on estrogen-, progesterone-, and HER2-receptor status, with prognostic implications, Glycoconj. J. 40 (2023) 631–644, https://doi.org/10.1007/s10719-023-10137-4.
[32] J. Li, et al., SPDEF enhances cancer stem cell-like properties and tumorigenesis through directly promoting GALNT7 transcription in luminal breast cancer, Cell Death Dis. 14 (2023) 569, https://doi.org/10.1038/s41419-023-06098-z.
[33] F. Gao, et al., MiR-30c facilitates natural killer cell cytotoxicity to lung cancer through targeting GALNT7, Genes & genomics 45 (2023) 247–260, https://doi.org/10.1007/s13258-022-01306-0.
[34] M.J. Goldman, et al., Visualizing and interpreting cancer genomics data via the Xena platform, Nat. Biotechnol. 38 (2020) 675–678, https://doi.org/10.1038/s41587-020-0546-8.
[35] The Genotype-Tissue Expression (GTEx) project, Nat. Genet. 45 (2013) 580–585, https://doi.org/10.1038/ng.2653.
[36] J. Liu, et al., An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics, Cell 173 (2018) 400–416.e411, https://doi.org/10.1016/j.cell.2018.02.052.
[37] D. Aran, Z. Hu, A.J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape, Genome Biol. 18 (2017) 220, https://doi.org/10.1186/s13059-017-1349-1.
[38] D. Zeng, et al., IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures, Front. Immunol. 12 (2021) 687975, https://doi.org/10.3389/fimmu.2021.687975.
[39] K. Yoshihara, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612, https://doi.org/10.1038/ncomms3612.
[40] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS A J. Integr. Biol. 16 (2012) 284–287, https://doi.org/10.1089/omi.2011.0118.
[41] J. Piñero, et al., DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants, Nucleic Acids Res. 45 (2017) D833–D839, https://doi.org/10.1093/nar/gkw943.
[42] J. Piñero, et al., DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes, Database : the journal of biological databases and curation 2015 (2015) bav028, https://doi.org/10.1093/database/bav028.
[43] J. Jourquin, D. Duncan, Z. Shi, B. Zhang, GLAD4U: deriving and prioritizing gene lists from PubMed literature, BMC Genom. 13 (Suppl 8) (2012) S20, https://doi.org/10.1186/1471-2164-13-s8-s20.
[44] V. Thorsson, et al., The immune landscape of cancer, Immunity 48 (2018) 812–830.e814, https://doi.org/10.1016/j.immuni.2018.03.023.

[45] R. Bonneville, et al., Landscape of microsatellite instability across 39 cancer types, JCO precision oncology 2017 (2017), https://doi.org/10.1200/po.17.00073.
[46] C.H. Mermel, et al., GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, Genome Biol. 12 (2011) R41, https://doi.org/10.1186/gb-2011-12-4-r41.
[47] E. Cerami, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (2012) 401–404, https://doi.org/10.1158/2159-8290.Cd-12-0095.