

Original Article



Decoding LY6G6D in colorectal cancer: Unraveling biomarker potential and therapeutic insights

Mehwish Naqvi^{1,*,#}, Waqas Ahmad Abbasi^{1,2,#}, Muhammad Kaleem Samma^{1,*}, Asif Ali Gabol¹, Cynthia Innocent¹, Sana Iqbal¹, Samana Fatima¹, Kashif Ali¹

¹ Department of Biosciences, Faculty of Life Sciences, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Karachi 75600, Pakistan

² Dow Medical College, Dow University of Health Sciences, Karachi 74200, Pakistan

Article Info

Abstract



Article history:

Received: December 27, 2023

Accepted: March 02, 2024

Published: June 30, 2024

Use your device to scan and read the article online



Colorectal cancer (CRC) poses a significant global health challenge with high morbidity and mortality rates. This study investigates the role of LY6G6D, a member of the LY6/uPAR superfamily, in CRC. Employing a bioinformatic approach, we analyzed LY6G6D expression across different cancer types, compared it with known oncogenes in CRC, explored the involved genomic alterations, and assessed associated clinicopathological characteristics. LY6G6D exhibited aberrant expression, particularly elevated in CRC adenocarcinoma and highly specific to tumor tissues when compared with other oncogenes, despite its comparatively low frequency of genomic alteration. Subsequently, tumor immune infiltration analysis revealed distinct associations, primarily indicating a negative correlation, suggesting immune down-regulation. Survival analysis in context of LY6G6D was conducted with Kaplan-Meier (KM) curves, indicating a 10% risk of disease recurrence in the case of elevated expression. Additionally, we constructed a 3D protein model of LY6G6D through ab-initio approach. The protein model was validated, followed by conservation analysis and active site identification. Active site identification of LY6G6D's final predicted model revealed some similar sites that were estimated to be conserved. Target-guided drug molecules were collected and molecular docking was executed, proposing Cardigin (Digitoxin) and Manzamine A as potential therapeutic candidates. In conclusion, LY6G6D emerges as a significant biomarker for diagnostic and therapeutic applications in CRC, highlighting its multifaceted role in tumorigenesis. The proposed drugs present avenues for further investigations.

Keywords: Bioinformatics analysis, Biomarker, Colorectal cancer, LY6G6D, Molecular docking, Oncogene

1. Introduction

Colorectal cancer (CRC) is a highly malignant neoplasm, ranking third in global incidence and holding the second position in mortality among all cancers [1]. Despite the upward trajectory in death rates, CRC incidence and mortality are avoidable through measures like screening, surveillance, and preventive treatment options.

This underscores the need for less or non-invasive procedures, through the exploration of potential biomarkers that can detect the disease at early stages and monitor tissue changes before advanced progression [2]. These biomarkers not only serve as diagnostic tools but may also facilitate the development of targeted therapeutic approaches [3, 4]. In the context of CRC, various biomarkers have been investigated, such as some non-coding RNA, serum tumor markers, inflammatory indicators, exosomes, and DNA methylation, all of which have shown potential in the detection and monitoring of CRC.

Among these biomarkers, *LY6G6D*, a fusion of lymphocytic antigen 6 proteins from the G6D family, emerges as a potential candidate. *LY6G6D* is a membrane-anchored

protein, particularly found on the cell surface [5]. Generally, abnormal mitotic division in CRC suggests structural changes in the epithelial lining that complement the elevation of mucin levels, subsequently leading to tumor progression [6].

While prior studies have delved into certain aspects of the *LY6G6D* in CRC, a significant gap persists, notably marked by the absence of a comprehensive bioinformatic analysis and exploration of broad in silico therapeutic implications [7, 8]. Emphasizing not just the in-silico molecular landscape but also the potential therapeutics, our research seeks to surpass the limitations of previous studies. Therefore, our aim is to discover and validate the role of *LY6G6D* as a potential biomarker meanwhile identifying its potential therapeutics as well.

2. Material and Methods

2.1. Analysis of *LY6G6D* expression variation

The Tumor Immune Estimation Resource (TIMER) (<https://cistrome.shinyapps.io/timer/>; accessed on 12 July 2023) database incorporates the cancer genome atlas

* Corresponding author.

E-mail address: kaleem.samma@szabist.edu.pk; kaleemsamma@gmail.com (M. K. Samma).

These authors contributed equally

Doi: <http://dx.doi.org/10.14715/cmb/2024.70.6.3>

(TCGA) data, providing valuable insights for investigating expression patterns in various tumor and normal tissues. To visually represent the distribution of LY6G6D expression levels, we employed box plots. These plots provided an overview of how LY6G6D expression levels are distributed within the tissues. The significance of LY6G6D expression was assessed using the Wilcoxon test between tumor and normal tissues. Furthermore, we analysed the association of the mRNA expression level of LY6G6D with clinicopathological characteristics in CRC using the University of Alabama at Birmingham CANcer data analysis Portal (UALCAN) (<https://ualcan.path.uab.edu/>; accessed on 12 July 2023). LY6G6D expression based on various patient characteristics, such as age, cancer subtypes, and stages was analysed, respectively.

2.2. Validation of LY6G6D gene expression

LY6G6D expression in different cancer types was validated using two different databases, Gene Expression Profiling Interactive Analysis (GEPIA) (<http://gepia.cancer-pku.cn/>; accessed on 12 July 2023) and UALCAN (<https://ualcan.path.uab.edu/>; accessed on 12 July 2023). The GEPIA database is a publicly accessible repository that allows the analysis of RNA expression data derived from 9736 tumor and 8587 normal samples. Furthermore, UALCAN facilitates expression analysis using TCGA datasets across 24 different cancer types.

2.3. Comparison of LY6G6D with oncogenes

The GEPIA database was utilized (<http://gepia.cancer-pku.cn/>; accessed on 13 July 2023) to compare the expression levels of LY6G6D with established oncogenes in normal and tumor tissues among various significantly expressed cancer types. Beside that, we depicted the expression of LY6G6D in comparison to other oncogenes in CRC through box plots using UALCAN (<https://ualcan.path.uab.edu/>; accessed on 14 July 2023). Additionally, to explore the likely impact of the LY6G6D mutation in CRC, we conducted an analysis along with other oncogenes using the cBioPortal database (<https://www.cbioportal.org/>) utilising colorectal adenocarcinoma TCGA PanCancer Atlas dataset comprising of 526 samples/patients. This tool helped us in the exploration of various mutation types.

2.4. LY6G6D expression and tumor immune infiltration correlation

The TIMER tool (<https://cistrome.shinyapps.io/timer/>; accessed on 15 July 2023) was used to systematically analyze immune infiltration levels among various types of cancer samples using the TCGA data. Correlation between LY6G6D expression and the infiltration levels of different types of immune cells (B cells, CD4+T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells) were explored.

2.5. Survival analysis of LY6G6D

To assess the impact of LY6G6D expression on the survival of CRC patients, we conducted KM (Kaplan-Meier) analysis. Patient survival data were obtained from the GEPIA database (<http://gepia.cancer-pku.cn/>; accessed on 15 July 2023) to plot the survival curves for both overall survival (OS) and disease-free survival (DFS). Kaplan-Meier curves were utilized to compare two groups of patients, and hazard ratios (HR), along with their corresponding

95% confidence intervals and log-rank p-values, were computed.

2.6. Sequence data and template search

The amino acid (aa) sequence of LY6G6D (133 aa) having accession number A0A1L6Z9X4 was downloaded from the Universal Protein Resource (UniProt) (<http://www.uniprot.org/>; accessed on 15 August 2023) database. To find a suitable template for the LY6G6D protein, we conducted searches using Basic Local Alignment Search Tool (BLAST; BLASTP) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>; accessed on 15 August, 2023) and SWISS-MODEL (<https://swissmodel.expasy.org/> accessed on 15 August 2023). We employed the SWISS-MODEL template library (SMTL) to apply both BLAST and HHblits for the LY6G6D amino acid sequence. The distant homologs were identified as likely template structures [9].

2.7. Protein modeling and validation

The LY6G6D protein reference sequence (NP_067069.2) was used for template-free (de novo or ab initio) prediction of 3-Dimensional (3D) structures of the LY6G6D protein utilising the Robetta server (<https://rosetta.bakerlab.org/>; accessed on August 16, 2023). Five distinct model structures were produced on the Robetta server. We first compared these five models among themselves and then with the LY6G6D AlphaFold 3D model to determine the best model. The predicted model quality was assessed using various validation tools. The stereochemical properties of the model were assessed with a Ramachandran plot using PROCHECK (<https://saves.mbi.ucla.edu/>; accessed on 01 September 2023). Additionally, we verified the X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, and other theoretical calculations using protein structure analysis (ProSA) (<https://prosa.services.came.sbg.ac.at/prosa.php/>; accessed on 01 September 2023), and the overall quality factor was checked using ERRAT (<https://saves.mbi.ucla.edu/>; accessed on 01 September 2023).

2.8. Energy minimization

We conducted energy minimization and refinement of the selected model using the Galaxy Web Server (<https://galaxy.seoklab.org/>; accessed on 03 September 2023). The quality of the energy-minimised model was then reassessed using PROCHECK, ProSA, and ERRAT.

2.9. Evolutionary conservation analysis

The conservation exploration of our final model protein (LY6G6D) was performed using the ConSurf server (https://consurf.tau.ac.il/consurf_index.php/; accessed on 05 September 2023). The analysis utilised the following parameters: Chain identifier: A; homologous search algorithm: PSI-BLAST; number of iterations: 3; E-value cutoff: 0.0001; protein database: UniProt Reference Clusters-90 (UniRef); reference sequence: closest; number of reference sequences selected: 150; maximum sequence identity: 95%; minimum identity for counterparts: 35%; alignment method: MAFFT-L-INS-i; calculation method: Bayesian; evolutionary substitution model: best model (standard).

2.10. Identification of active sites

The binding sites of the LY6G6D model were identi-

fied using the Computed Atlas of Surface Topography of Protein (CASTp) (<http://sts.bioe.uic.edu/castp/>; accessed on 05 September 2023) server, which utilises the pocket algorithm based on alpha shape theory (Holm et al., 2008).

2.11. Collection of drug agents

We collected target-guided drug molecules from two sources: the Drug SIGnatures DataBase (DSigDB) (<https://dsigdb.tanlab.org/DSigDBv1.0/>; accessed on 05 September 2023) (comprising of 151 drug molecules, Set-A) [10] and published articles (55 drug molecules, Set-B). These collected drugs were used to explore potential candidate drugs that could interact with our specified target, LY6G6D, by molecular docking. Detailed lists of the A and B sets are given in the supplementary file.

2.12. Molecular docking

Molecular docking analysis of LY6G6D was conducted with sets (A+B) in order to find potential drugs for CRC treatment. The 3D structure of our final modelled LY6G6D was used as a target. We retrieved the 3D structures of 206 drug molecules from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>; accessed on 8 September 2023). The starting directory was set to the appropriate folder prior to docking. The protein molecule was loaded into the workspace of AutoDock 1.5.7 (accessed on 11 September 2023). Subsequently, the structure of LY6G6D was prepared by adding Kollman charges and hydrogen atoms. The target (LY6G6D) protein model and ligands were saved in PDBQT format [11]. The grid box parameter was set to default (0.375 Å), and the centre grid box values were 25.856, -11.586, and 1.510. The grid points on the maps were set with dimensions of 50 along the x-axis, 52 along the y-axis, and 42 along the z-axis. The map contained a total of 116,229 grid points. Autodock Vina was then used, considering receptors as rigid molecules during docking, while ligands were flexible to determine the most suitable pose. The binding affinity scores between the ligands and LY6G6D (receptor) were calculated. Discovery Studio Visualizer 2019 and ChimeraX were used to analyse the docked complexes.

3. Results

3.1. LY6G6D expression in cancer and normal tissues

To examine the potential role of LY6G6D in cancer, we checked its expression levels in various tumors and normal tissues using the TIMER database. Our analysis indicated significant upregulation of LY6G6D expression in colon adenocarcinoma (COAD), rectum adenocarcinoma (READ), bladder carcinoma (BLCA), breast carcinoma (BRCA), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), prostate adenocarcinoma (PRAD), skin cancer melanoma (SKCM), and thyroid carcinoma (THCA).

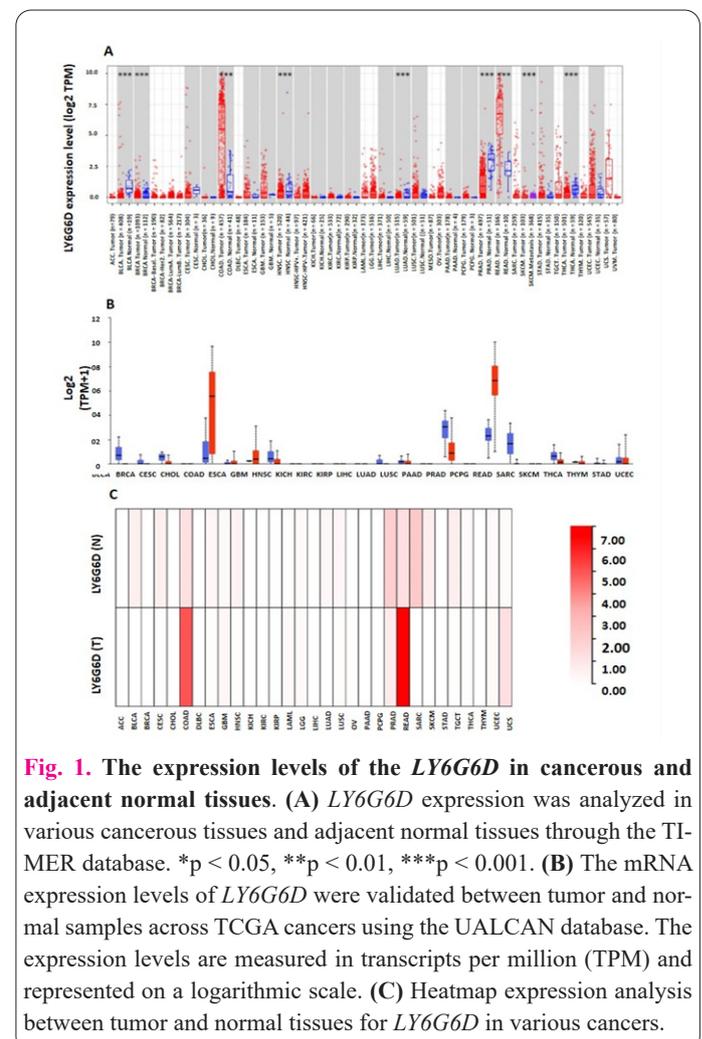
Conversely, downregulation of LY6G6D was observed in specific subtypes of breast carcinoma: breast cancer basal-like (BRCA Basal), breast cancer human epidermal growth factor receptor 2 (BRCA HER2), breast cancer luminal A (BRCA LumA), and breast cancer luminal B (BRCA LumB). In addition, we also noticed downregulation of LY6G6D in cervical squamous cell carcinoma (CESE), cholangiocarcinoma (CHOL), esophageal carcinoma (ESCA), glioblastoma multiform (GBM), head and neck squamous cell carcinoma with different molecu-

lar tumor subtypes based on human papilloma virus status (HNSC-HPV+, HNSC-HPV-), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), brain lower grade glioma (LGG), liver hepatocellular carcinoma (LIHC), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), prostate adenocarcinoma (PRAD), sarcoma (SARC), stomach adenocarcinoma (STAD), thymoma (THYM), uterine corpus endometrial carcinoma (UCEC), uterine carcinosarcoma (UCS), and uveal melanoma (UVM) (Fig. 1A).

The upregulation of LY6G6D in nine different cancer types suggests that it may be a common occurrence with a role in cancer development. Interestingly, we observed that LY6G6D expression in CRC (COAD and READ) was markedly higher as compared to other cancer types and adjacent normal tissues. To validate the expression of LY6G6D in CRC, we further utilised the UALCAN database. Our validation confirmed the upregulation of LY6G6D expression in CRC (Fig. 1B). Additionally, the heatmap expression analysis of *LY6G6D* between tumor and normal tissues revealed distinct expression in COAD and READ (Fig. 1C). Overall, these findings support the notion that upregulation of LY6G6D may contribute to the development and progression of different cancer types, with particular emphasis on CRC.

3.2. Comparison of LY6G6D with oncogenes

We conducted a comparative analysis using the GEPIA database of *LY6G6D* expression with well-established



oncogenes in normal and tumor tissues, including Adenomatous polyposis coli (APC), Tumor protein 53 (TP53), Kirsten rat sarcoma virus (KRAS), AXIS inhibition protein 2 (AXIN2), MutL homolog 2 (MSH2), MutL homolog 1 (MSH6), and MutL homolog 1 (MLH1). We found that these oncogenes exhibited differential expression patterns across a range of cancer types (BLCA, BRCA, COAD, HNSC, LUAD, PRAD, READ, SKCM, and THCA) and adjacent normal tissues. Notably, *LY6G6D* displayed prominent expression exclusively in CRC (Fig. 2A). Furthermore, following up on the highly specific expression of *LY6G6D* in tumor tissue of CRC, unlike other oncogenes, we conducted expression ratio analyses, as illustrated in (Fig. 2B), to get a better view of the pronounced expression (READ = 5.07, COAD = 4.16). In addition to this, we also validated our findings by UALCAN (Figure I, S-1). These findings strongly suggest that *LY6G6D* may play a pivotal role as a promising biomarker for CRC.

3.3. *LY6G6D* gene mutation in CRC tumorigenesis

Furthermore, we explored the potential contribution of genomic alteration associated with *LY6G6D* and compared it with other oncogenes in CRC development. The distribution of genomic alterations in these genes revealed *LY6G6D* 1%, APC 75%, TP53 60%, KRAS 42%, AXIN2 7%, MSH2 4%, MSH6 5%, and *MLH1* 4%, (Fig. 3). These results suggests that *LY6G6D* had a low frequency of genomic alteration in comparison with other oncogenes and that it may not be a solo major reason that drives cancer.

3.4. Association of differential immune cell infiltration with *LY6G6D* expression

To look into the potential relationship between *LY6G6D* expression and immune infiltration in CRC, we utilised the TIMER database. Immune cell infiltration analysis revealed distinct associations in COAD and READ (Fig. II, S1). In COAD, *LY6G6D* expression exhibited a positive correlation with tumor purity (cor = 0.214, $P = 1.30 \times 10^{-5}$) and CD4+ T cells (cor = 0.017, $P = 7.31 \times 10^{-1}$), while showing a negative correlation with infiltrating levels of B cells (cor = -0.272, $P = 2.64 \times 10^{-8}$), CD8+ T cells (cor = -0.366, $P = 2.75 \times 10^{-14}$), macrophages (cor = -0.057, $P = 2.49 \times 10^{-1}$), neutrophil cells (cor = -0.326, $P = 1.81 \times 10^{-12}$), and dendritic cells (cor = -0.326, $P = 2.00 \times 10^{-11}$). Similarly, in READ, *LY6G6D* expression displayed a positive correlation with tumor purity (cor = 0.209, $P = 2.64 \times 10^{-8}$), CD4+ T cells (cor = 0.063, $P = 4.63 \times 10^{-1}$), and macrophages (cor = 0.002, $P = 9.83 \times 10^{-1}$), while exhibiting a negative correlation with infiltrating levels of B cells (cor = -0.003, $P = 7.03 \times 10^{-1}$), CD8+ T cells (cor = -0.12, $P = 1.58 \times 10^{-1}$), neutrophil cells (cor = -0.162, $P = 5.69 \times 10^{-2}$), and dendritic cells (cor = -0.125, $P = 1.43 \times 10^{-1}$).

Positive correlation of CD4+ T cells in COAD along with macrophages in READ contrary to negative correlation of B cells, CD8+ T cells, neutrophils, and dendritic cells in READ with addition of macrophages in COAD suggests complex interplay within the immune microenvironment, further proposing *LY6G6D*'s role in suppressing immune function, ultimately leading to the development or progression of CRC.

3.5. *LY6G6D* expression as a key indicator associated with clinicopathological features

Considering the significant upregulation of *LY6G6D* in

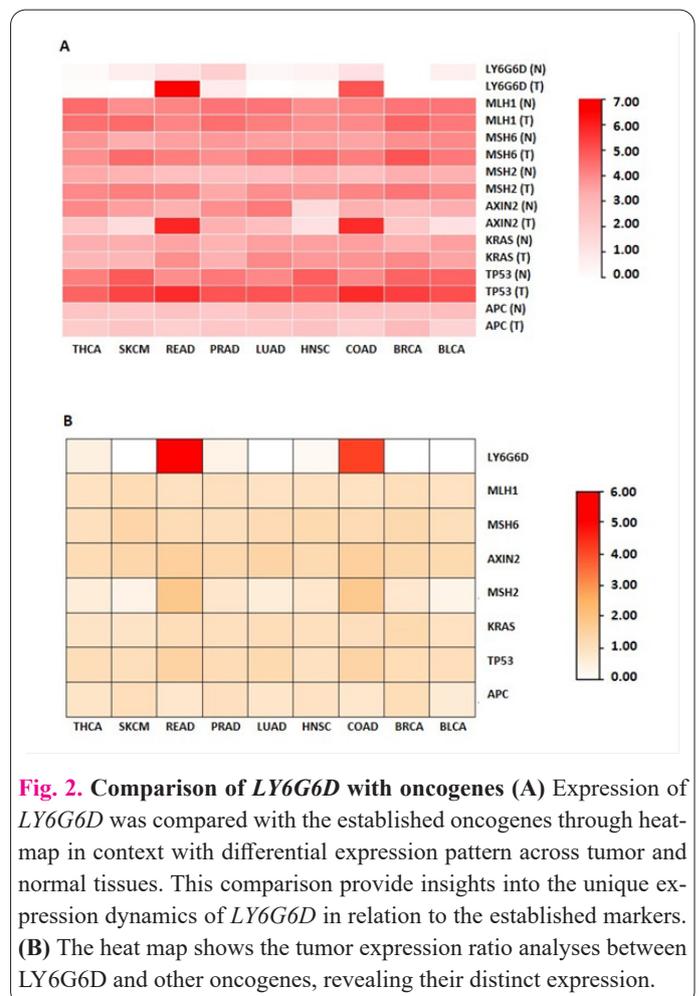


Fig. 2. Comparison of *LY6G6D* with oncogenes (A) Expression of *LY6G6D* was compared with the established oncogenes through heatmap in context with differential expression pattern across tumor and normal tissues. This comparison provide insights into the unique expression dynamics of *LY6G6D* in relation to the established markers. (B) The heatmap shows the tumor expression ratio analyses between *LY6G6D* and other oncogenes, revealing their distinct expression.

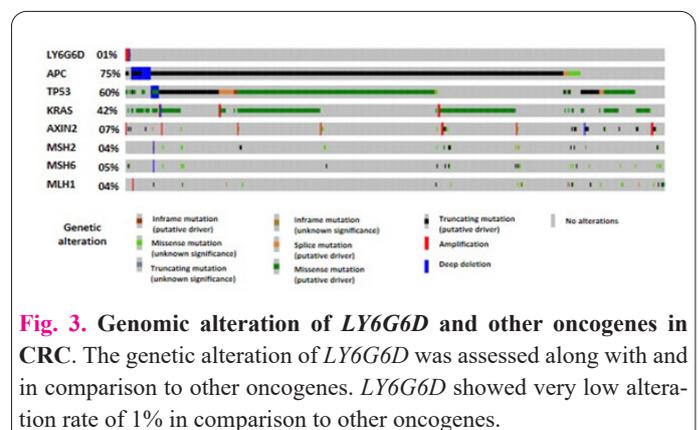


Fig. 3. Genomic alteration of *LY6G6D* and other oncogenes in CRC. The genetic alteration of *LY6G6D* was assessed along with and in comparison to other oncogenes. *LY6G6D* showed very low alteration rate of 1% in comparison to other oncogenes.

CRC compared to the adjacent normal tissues, our research aimed to investigate the potential correlation between *LY6G6D* expression and clinicopathological features, utilising the UALCAN database (Fig III, S1). The clinicopathological characteristics encompass age, histological subtypes, and cancer stages. Our investigation showed that *LY6G6D* expression levels in colon and rectal cancers were significant in patients age groups (21-40), (41-60) and (61-80). Histologically, *LY6G6D* expression was significant among both types, but higher in adenocarcinoma as compared to mucinous adenocarcinoma in both COAD and READ. Furthermore, interestingly *LY6G6D* expression was significantly upregulated among all 4 stages in CRC patients, emphasizing its prominence.

3.6. Correlation of *LY6G6D* with survival of CRC

To assess the prognostic importance of *LY6G6D* in CRC, we conducted a survival analysis using the GEPIA tool (Fig. IV, S1). The difference in OS between the high

and low expression groups was not statistically significant (p -value = 0.25). Similarly, DFS showed insignificant results as well (p = 0.62), whereas HR = 1.1 of DFS suggested a 10% increase in the risk of an event (disease recurrence) in patients with higher LY6G6D expression. These results suggest the need for further validation in larger and more diverse cohorts to better understand the significance of LY6G6D in context of survival outcomes.

3.7. LY6G6D protein structure analysis

To propose *in silico* (target-guided) efficient candidate drugs for the treatment of CRC, we employed protein modelling of LY6G6D, as its 3D model has not been determined yet. Hence, the amino acid sequence of LY6G6D was used to search for suitable templates. No similarity was found through BLASTP analysis, whereas templates generated through SWISS MODEL had <30% sequence identity. Moreover, the Alpha Fold (artificial intelligence-assisted program) generated structure had a sequence identity of 72.52%.

Due to the inaccessibility of experimentally solved homologs having more than 30% sequence identity, template-based modeling was not feasible. Thus, the complete structure of LY6G6D has been forecasted using a template-free modelling approach (*de novo* or *ab initio*) using the Robetta server. The server-generated results produced five structures, which were compared among themselves to determine the best model on the basis of validation scores by PROCHECK, ERRAT, and ProSA Web (Table 1). Validation scores suggested that model 01 was comparatively reliable (Fig. V, S1). Furthermore, the selected model 01 was also compared with the AlphaFold predicted model (Table 1), Ramachandran plot analysis proved that the AlphaFold model had a lower quality with errors in the following structural parameters: overall Ramachandran, residue properties, planar groups, and bond len/angles, whereas the Robetta-generated model 01 in comparison had residues placed in the most favoured regions with no errors in the mentioned structural parameters. Moreover, the Z scores predicted by the ProSa web for model 01 suggested that it was more suitable for subsequent investigations, as it aligns with the space occupied by NMR protein structures in the protein data bank (PDB) (Z score = -3.72), while the ERRAT score of 93.04 accounted for the overall quality factor. These findings concluded that the Robetta-generated model 01 is the best predicted structure of LY6G6D.

3.8. Energy minimization and model quality assessment

The selected model 01 was subjected to energy minimization and refinement using the Galaxy server. We

acquired five refined structures, and the best among them was selected (Fig. VI, S1) for further analysis based on the best scores: MolProbity score = 1.753, GDT-HA score = 1.0000, Clash score = 9.3, RMSD score = 0.213, Rama preferred score = 96.2, and poor rotamers core = 0.0. The refined model was further validated using PROCHECK, ProSA, and ERRAT. ProSA web presented a z-score of -3.4 (NMR spectroscopy: dark blue). The PROCHECK analysis revealed that 90.7% of the residues are located in core region, 7.5% in allowed regions and 1.9% in generally allowed regions. ERRAT calculated an overall quality factor of 93.20.

3.9. Evolutionary conservation analysis

The final refined model of LY6G6D was introduced to the ConSurf server to estimate the evolutionary conservation of each amino acid in the LY6G6D protein sequence [12]. The computed ConSurf score was projected on the surface of the protein, and colours were assigned to LY6G6D's amino acids accordingly, following a colouring-code scheme representing average, variable, and conserved regions (Fig.VII, S1). We identified a prominent distribution of conserved regions in our modelled LY6G6D, which shows potentially important sites for interactions.

3.10. Identification of active sites

To predict the active binding sites of LY6G6D, we utilised the CASTp server. The PDB file of final refined model 01 was uploaded, and the radius probe was set to 1.2 (Å). From the 32 predicted cavities, the largest cavity, encompassing a surface area of 124.677 Å² and a volume of 37.830 Å³, was chosen (Fig.VII, S1). The residues within the selected pocket were: Leu 13, Ala 16, Ala 17, Gly 19, Leu 98, Gly 99, Asp 100, Leu 101, Cys 102, Asn 103, Ser 108, Val 110, Ala 111, Pro 112, Ile 115, and Leu 116.

3.11. Exploring candidate drugs by molecular docking analysis

Molecular docking was performed between LY6G6D and the selected 206 drugs (S-2). The resultant binding energies were observed from the log output file of each complex, which was then compiled. The complex output (in PDBQT format) contained all the poses. After identifying the top-scoring complexes, poses were split. Among the 206 drugs, Cardigin and Manzamine A significantly showed the lowest binding affinity, with scores of -8.4 and -8.2 kcal/mol, respectively. Therefore, we concluded that the top-ordered two drugs (Cardigin and Manzamine A) were the candidate drugs in our study for LY6G6D. We also examined their complete interaction profile (Fig.4), accounting for specific interacting residues

Table 1 Comparison of LY6G6D protein models.

Modelling server	Model #	PROCHECK % (Core region)	ERRAT	ProSA
Robetta	01	81.3% (No Error in overall Ramachandran)	93.04	-3.72
	02	77.6 % (Error in overall Ramachandran)	93.91	-3.24
	03	80.4 % (Error in overall Ramachandran)	83.76	-3.17
	04	79.4 % (Error in overall Ramachandran)	75.89	-3.15
	05	70.1 % (Error in overall Ramachandran)	84.74	-3.28
Alpha Fold	01	86.0% (Error in overall Ramachandran)	89.10	-2.59

having the presence of conventional hydrogen bonds, carbon hydrogen bonds, alkyl and pi-alkyl interactions, using Discovery Studio Visualizer 2019. To further visualize the high-dimensional 3D view of the docked ligand at the pocket, Chimera X, a powerful molecular visualization tool was used (Fig.5).

4. Discussion

CRC remains a significant global health concern due to its high morbidity and mortality rates, necessitating the determination of effective biomarkers and therapeutics [13]. Advancements in big data have provided extensive datasets, making bioinformatic analysis crucial in oncology research [14].

Using the bioinformatics approach, we analyzed *LY6G6D* expression across various cancers and observed significant upregulation in CRC tissues, emphasizing its diagnostic relevance. Our findings align well with the recent reports, supporting the oncogenic role of *LY6G6D* in CRC [7, 15]. Notably, *LY6G6D* also exhibited correlation and broad applicability across multiple age groups, all cancer stages and types, validating its utility as a robust diagnostic marker.

Moreover, our survival analysis indicated a 10% increased risk of recurrence in patients exhibiting higher expression, providing valuable prognostic insights. Previously, the prognostic power of *LY6G6D* has been reported for uterine and pancreatic cancers [16, 17], whereas few *LY6* gene family members have been reported to encompass overall poor prognosis [18]. This study uncover insights by reporting the survival estimates of increased *LY6G6D* expression in CRC, which were unaddressed before.

The investigation into the relationship between *LY6G6D* expression and immune infiltration in CRC revealed distinct associations, especially the negative correlation with key immune effectors, particularly CD8⁺ T cells, prompt questions about *LY6G6D*'s potential immunomodulatory role, proposing it as a target for immunotherapy. Furthermore, this nuanced response of immune infiltration suggests economical genomic instability [19, 20]. Subsequent evidence from our comparative analysis with other oncogenes revealed a low frequency of genomic alteration in *LY6G6D*, specifically 1%, within CRC. Despite low genomic alteration in comparison to other oncogenes, *LY6G6D* showed exceptionally pronounced expression. This discrepancy suggests a divergence between the two parameters. While *LY6G6D* presents itself as a potential tumorigenesis target in CRC, extensive in-depth investigations into the molecular mechanisms is warranted.

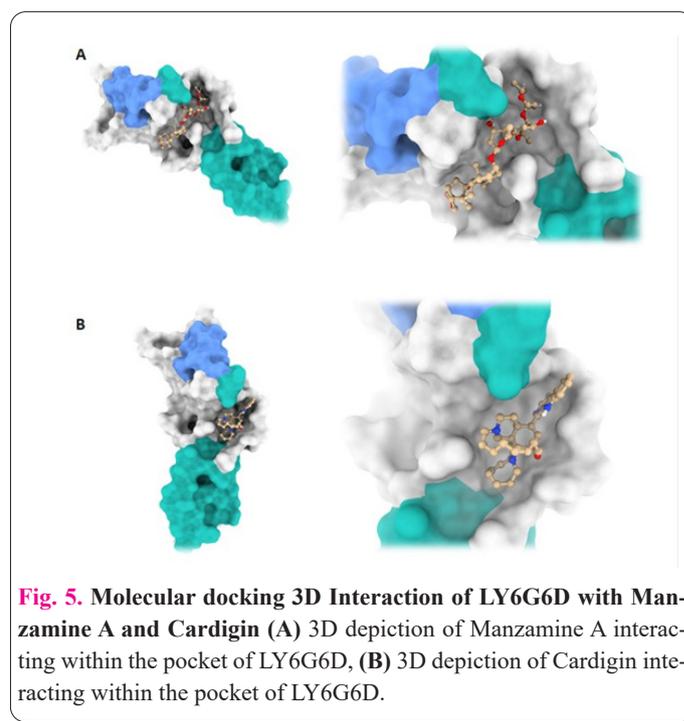
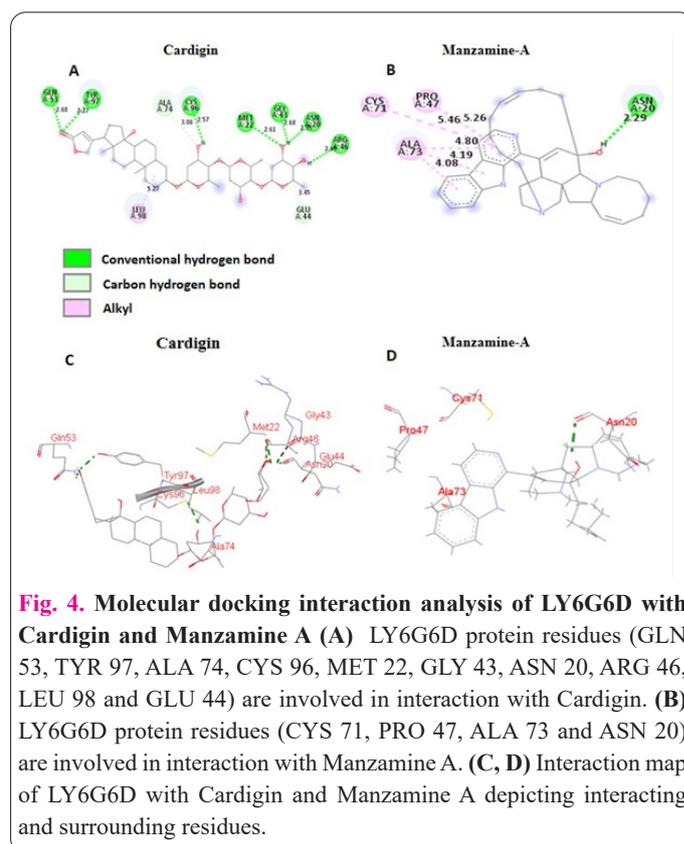
In our pursuit of therapeutic insights, a 3D protein model for *LY6G6D* was constructed through ab-initio approach and molecular docking was executed on the final refined model. Notably, the identified active sites of the model included some common residues that were marked conserved through our conservation analysis, thus validating them as functionally important.

We identified the top two candidates, Cardigin which is also referred to as Digitoxin and Manzamine A, which exhibited strong binding affinities of -8.4 and -8.2 kcal/mol. Manzamine A, reported for its anti-proliferative effects on colorectal and pancreatic cancer cells, is said to trigger apoptotic cell death and disrupt epithelial-mesenchymal transition [21-23]. Meanwhile, Cardigin has demonstrated effectiveness against cervical cancer, with FDA approval

for various diseases, suggesting it to be a promising repurposable agent for CRC treatment [24]. While Manzamine A is not yet FDA-approved, Cardigin's (Digitoxin) established safety profile, positions it well for clinical repurposing. Despite computational insights, further molecular level analysis is crucial before clinical interventions.

5. Conclusion

This study provides a considerable analysis of *LY6G6D* as a potential biomarker and therapeutic target in CRC. Moreover, in our investigation of candidate drugs, Cardigin (Digitoxin) and Manzamine A highlighted promising binding affinities and interactions with prospects for appli-



cation in CRC treatment. Overall, in silico analyses offered insights into both the structural and functional aspects of *LY6G6D*, serving as a foundation for future experimental validations.

Conflict of interests

The authors declared no conflict of interest.

Author's contributions

Mehwish Zehra: Concept design, Literature search, bioinformatic analysis; **Waqas Ahmad Abbasi:** Molecular docking, Interpretation of results, Manuscript writing; **Muhammad Kaleem Samma:** Critical appraisal, Manuscript writing, Interpretation of results; **Asif Ali Gabol:** Image editing and compilation; **Cynthia Innocent:** Final Draft preparation; **Sana Iqbal:** Bioinformatic analysis, Final review; **Samana Fatima:** Manuscript writing; **Kashif Ali:** Critical appraisal and final review

*All authors have read the final draft.

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71 (3): 209-249. doi: 10.3322/caac.21660
- Yang H, Li X, Meng Q, Sun H, Wu S, Hu W, Liu G, Li X, Yang Y, Chen R (2020) CircPTK2 (hsa_circ_0005273) as a novel therapeutic target for metastatic colorectal cancer. *Mol Cancer* 19: 1-15. doi: 10.1186/s12943-020-1139-3
- Luo X-J, Zhao Q, Liu J, Zheng J-B, Qiu M-Z, Ju H-Q, Xu R-H (2021) Novel genetic and epigenetic biomarkers of prognostic and predictive significance in stage II/III colorectal cancer. *Mol Ther* 29 (2): 587-596. doi: 10.1016/j.ymthe.2020.12.017
- Shetu SA, James N, Rivera G, Bandyopadhyay D (2023) Molecular research in pancreatic cancer: Small molecule inhibitors, their mechanistic pathways and beyond. *Curr Issues Mol Biol* 45 (3): 1914-1949. doi: 10.3390/cimb45030124
- Mallya M, Campbell RD, Aguado B (2006) Characterization of the five novel Ly-6 superfamily members encoded in the MHC, and detection of cells expressing their potential ligands. *Protein Sci* 15 (10): 2244-2256. doi: 10.1110/ps.062242606
- Coleman OI, Haller D (2021) Microbe-mucus interface in the pathogenesis of colorectal cancer. *Cancers (Basel)* 13 (4): 616. doi: 10.3390/cancers13040616
- Corrales L, Hipp S, Martin K, Sabarth N, Tirapu I, Fuchs K, Thaler B, Walterskirchen C, Bauer K, Fabits M (2022) LY6G6D is a selectively expressed colorectal cancer antigen that can be used for targeting a therapeutic T-cell response by a T-cell engager. *Front Immunol* 13: 1008764. doi: 10.3389/fimmu.2022.1008764
- Wang P, Sun LL, Clark R, Hristopoulos M, Chiu CP, Dillon M, Lin W, Lo AA, Chalsani S, Das Thakur M (2022) Novel Anti-LY6G6D/CD3 T-Cell-Dependent Bispecific Antibody for the Treatment of Colorectal Cancer. *Mol Cancer Ther* 21 (6): 974-985. doi: 10.1158/1535-7163.MCT-21-0599
- Dilly S, Garnier M, Solé M, Bailly R, Taib N, Bestel I (2020) In silico identification of a key residue for substrate recognition of the riboflavin membrane transporter RFVT3. *J Chem Inf Model* 60 (3): 1368-1375. doi: 10.1021/acs.jcim.9b01020
- Yoo M, Shin J, Kim J, Ryall KA, Lee K, Lee S, Jeon M, Kang J, Tan AC (2015) DSigDB: drug signatures database for gene set analysis. *Bioinformatics* 31 (18): 3069-3071. doi: 10.1093/bioinformatics/btv313
- El-Hachem N, Haibe-Kains B, Khalil A, Kobeissy FH, Nemer G (2017) AutoDock and AutoDockTools for protein-ligand docking: beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) as a case study. *Methods Mol Biol* (1598): 391-403. doi: 10.1007/978-1-4939-6952-4_20
- Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 44 (W1): W344-W350. doi: 10.1093/nar/gkw408
- Debela DT, Muzazu SG, Heraro KD, Ndalama MT, Mesele BW, Haile DC, Kitui SK, Manyazewal T (2021) New approaches and procedures for cancer treatment: Current perspectives. *SAGE Open Med* 9: 20503121211034366. doi: 10.1177/20503121211034366
- Nagarajan N, Yapp EK, Le NQK, Kamaraj B, Al-Subaie AM, Yeh H-Y (2019) Application of computational biology and artificial intelligence technologies in cancer precision drug discovery. *Bio-med Res Int* 2019. doi: 10.1155/2019/8427042
- Giordano G, Pancione M (2023) MHC class III lymphocyte antigens 6 as endogenous immunotoxins: Unlocking immunotherapy in proficient mismatch repair colorectal cancer. *WIREs Mech Dis*: e1631. doi: 10.1002/wsbm.1631
- Rathbun LA, Magliocco AM, Bamezai AK (2023) Human LY6 gene family: potential tumor-associated antigens and biomarkers of prognosis in uterine corpus endometrial carcinoma. *Oncotarget* 14: 426. doi: 10.18632/oncotarget.28409
- Russ E, Bhuvaneshwar K, Wang G, Jin B, Gage MM, Madhavan S, Gusev Y, Upadhyay G (2021) High mRNA expression of LY6 gene family is associated with overall survival outcome in pancreatic ductal adenocarcinoma. *Oncotarget* 12 (3): 145. doi: 10.18632/oncotarget.27880
- Upadhyay G (2019) Emerging role of lymphocyte antigen-6 family of genes in cancer and immune cells. *Front Immunol* 10: 819. doi: 10.3389/fimmu.2019.00819
- Karn T, Jiang T, Hatzis C, Sängner N, El-Balat A, Rody A, Holtrich U, Becker S, Bianchini G, Pusztai L (2017) Association between genomic metrics and immune infiltration in triple-negative breast cancer. *JAMA Oncol* 3 (12): 1707-1711. doi: 10.1001/jamaoncol.2017.2140
- Safonov A, Jiang T, Bianchini G, Györfly B, Karn T, Hatzis C, Pusztai L (2017) Immune gene expression is associated with genomic aberrations in breast cancer. *Cancer Res* 77 (12): 3317-3324. doi: 10.1158/0008-5472.CAN-16-3478
- Guzmán EA, Johnson JD, Linley PA, Gunasekera SE, Wright AE (2011) A novel activity from an old compound: Manzamine A reduces the metastatic potential of AsPC-1 pancreatic cancer cells and sensitizes them to TRAIL-induced apoptosis. *Invest New Drugs* 29: 777-785. doi: 10.1007/s10637-010-9422-6
- Islam MA, Hossen MB, Horaira MA, Hossen MA, Kibria MK, Reza MS, Tuly KF, Faruqe MO, Kabir F, Mahumud RA (2023) Exploring Core Genes by Comparative Transcriptomics Analysis for Early Diagnosis, Prognosis, and Therapies of Colorectal Cancer. *Cancers (Basel)* 15 (5): 1369. doi: 10.3390/cancers15051369
- Lin L-C, Kuo T-T, Chang H-Y, Liu W-S, Hsia S-M, Huang T-C (2018) Manzamine A exerts anticancer activity against human colorectal cancer cells. *Mar Drugs* 16 (8): 252. doi: 10.3390/md16080252
- Gan H, Qi M, Chan C, Leung P, Ye G, Lei Y, Liu A, Xue F, Liu D, Ye W (2020) Digitoxin inhibits HeLa cell growth through the induction of G2/M cell cycle arrest and apoptosis in vitro and in vivo. *Int J Oncol* 57 (2): 562-573. doi: 10.3892/ijo.2020.5070