

LncRNACNVIntegrateR: a novel framework for correlating long non-coding RNAs with copy number variation abnormalities and disease progression

Neetu Tyagi^{1,2}, Shikha Roy² and Dinesh Gupta²

¹ Regional Centre for Biotechnology, Faridabad, Haryana, India

² Structural and Computational Biology Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, Delhi, India

ABSTRACT

Understanding complex biological systems and disease mechanisms necessitates the integration of multiple molecular layers, making multi-omics data integration a cornerstone of modern biomedical research. By combining datasets from different omics domains, researchers can uncover intricate molecular relationships, discover robust biomarkers, and advance precision medicine. Despite advancements in high-throughput technologies that have increased the availability of multi-omics datasets, challenges such as sample consistency and the development of reliable analytical frameworks hinder their full potential. Addressing these challenges is crucial for achieving a comprehensive understanding of biological systems and leveraging multi-omics data to drive breakthroughs in healthcare.

LncRNACNVIntegrateR is an R package that facilitates multi-omics data integration to explore the interplay between long non-coding RNAs (lncRNAs) and copy number variations (CNVs). The package integrates transcriptomic data, CNV profiles, and clinical information from matched samples, providing a complete pipeline for data preprocessing, lncRNA-CNV correlation analysis, and identification of CNV-driven prognostic signatures. Additionally, the package supports the construction of risk score models based on CNV-associated lncRNAs and functional enrichment analyses to reveal the role of corresponding target genes in disease progression. We validated LncRNACNVIntegrateR using The Cancer Genome Atlas (TCGA) Glioblastoma (GBM) and Colorectal Adenocarcinoma (COAD) datasets. The risk score models developed by the package demonstrated promising predictive performance, achieving an area under the receiver operating characteristic curve (AUC) of 0.80 for GBM and 0.71 for COAD. Functional enrichment analyses further highlighted the biological significance of the identified prognostic CNV-driven lncRNA signatures, providing insights into disease progression, risk stratification, and potential therapeutic targets to support clinical decision-making and personalized treatment approaches.

Submitted 26 April 2025

Accepted 3 September 2025

Published 16 October 2025

Corresponding author

Dinesh Gupta, dinesh@icgeb.res.in

Academic editor

Kamil Steczkiewicz

Additional Information and
Declarations can be found on
page 14

DOI [10.7717/peerj.20131](https://doi.org/10.7717/peerj.20131)

© Copyright

2025 Tyagi et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Computational Biology

Keywords Multi-omic, Survival analysis, Biomarker, Data integration, lncRNA, CNV

INTRODUCTION

Advancements in high-throughput sequencing have revolutionized biomedical research, enabling multi-level integration of genomic, transcriptomic, methylomic, and epigenomic data to unravel complex diseases and multidimensional biological patterns (Lee, 2023). To comprehensively understand complex biological processes, it is crucial to adopt an integrative approach combining multi-omics data and exploring the interrelationships among biomolecules and their functions. With the growing availability of multi-omics data from numerous samples, several tools and methods have been developed to integrate and interpret data for various applications, including disease subtyping, biomarker prediction, and data analysis (Subramanian et al., 2020). Numerous studies have shown that long non-coding RNAs (lncRNAs) are crucial regulatory molecules that can be targeted to modulate cellular physiology and functions in the pathophysiological progression of tumors, as well as in other diseases such as autoimmune diseases and cardiovascular diseases (Hajjari, Khoshnevisan & Shin, 2014; Kumar & Goyal, 2017; Lorenzen & Thum, 2016; Ng et al., 2013; Ounzain et al., 2015; Wu et al., 2015). Integrative analyses of various omics datasets have highlighted the deregulation of lncRNAs due to copy number variations (CNVs), drawing significant attention to their roles.

lncRNAs are >200-nucleotide RNA molecules that lack protein-coding potential (Beermann et al., 2016). They are key regulators in human diseases, particularly cancer, where they influence critical processes such as cell proliferation, signaling pathways, angiogenesis, and metastasis, driving various stages of cancer progression (Kondo, Shinjo & Katsushima, 2017; Ling, Fabbri & Calin, 2013). Additionally, lncRNAs can function as decoys or sponges, sequestering regulatory factors and miRNAs to modulate gene expression (Morriss & Cooper, 2017). lncRNAs regulate gene expression by modifying chromatin, influencing allelic imprinting, and controlling post-transcriptional and post-translational processes (Mercer, Dinger & Mattick, 2009; Wilusz, Sunwoo & Spector, 2009).

CNVs are genetic alteration involving duplications or deletions of DNA segments ranging from 1 kb to 3 Mb (Nakamura, 2009). In complex diseases like cancer, CNVs can activate oncogenes or delete tumor suppressor genes, affecting cellular adhesion, communication, and function (Conrad et al., 2010). CNVs contribute to individual variability and were initially considered protective by creating redundancy. However, many CNVs are now recognized as harmful, being associated with various diseases, including neurological and developmental disorders (Morrow, 2010; Valsesia et al., 2013).

By altering the expression of genes, including lncRNAs, CNVs can disrupt normal cellular functions and contribute to disease progression. The interaction between CNVs and lncRNAs is crucial not only in cancer (Ning et al., 2021; Tian & Luo, 2022; Wang et al., 2020; Zheng et al., 2020; Zhong et al., 2021) but also in various other diseases (Lu et al., 2023a, 2023b). This interplay can disrupt key genes and pathways contributing to disease development and progression and underscoring the broader impact of CNVs and lncRNAs association on human health.

The importance of novel tools for multi-omics data integration lies in their ability to address the complexities of harmonizing diverse datasets, enhancing our understanding of biological systems and improving clinical applications. For instance, the mixOmics package (Rohart et al., 2017) enables the multivariate analysis and integration of high-throughput 'omics data, facilitating the discovery of molecular relationships. TimeOmics (Bodein et al., 2022) offers a comprehensive framework for analyzing longitudinal multi-omics data, crucial for studying dynamic biological processes. Holomics (Munk et al., 2024) provides an accessible workflow for single-omics analysis, addressing challenges in data scale and complexity. MOVICS (Lu et al., 2021) focuses on cancer subtyping through multi-omics integration, advancing personalized medicine. These tools exemplify the critical advancements needed to streamline data integration and interpretation, ultimately leading to better disease understanding and therapeutic strategies.

To address the growing demand for user-friendly tools in multi-omics data integration, we developed lncRNACNVIntegrateR, an R package, to explore the relationship between lncRNAs and CNV abnormalities for disease diagnosis and prognosis. We tested lncRNACNVIntegrateR on The Cancer Genome Atlas (TCGA) datasets for Glioblastoma (GBM) and Colorectal Adenocarcinoma (COAD), identifying several key CNV-driven lncRNA prognostic biomarkers. Furthermore, we validated the risk score model, and functional enrichment analysis revealed the roles of these biomarkers in disease progression, providing insights into their prognostic potential and supporting clinical decision-making for personalized treatment.

METHODS

Design and implementation

The structure and design of the lncRNACNVIntegrateR package is modular and user-centric, ensuring ease of use and flexibility for integrating and analyzing multi-omics data. The package is organized into several core functions, each performing specific tasks such as data input, preprocessing, correlation analysis, risk score model development, and functional enrichment analysis. This modular approach allows users to utilize each function independently or as a comprehensive workflow. Figure 1 illustrates the end-to-end workflow implemented in lncRNACNVIntegrateR, highlighting the major data integration and analysis steps. The analysis begins with three essential input files consisting of the following datasets:

- (1) **Expression profiles:** Provided as a raw count matrix or a tab-delimited file. Rows represent genes (including lncRNAs), and columns represent samples. The first row should include sample identifiers (header), followed by gene identifiers and their corresponding expression values.
- (2) **Absolute CNV profiles:** A matrix or tab-delimited file where rows correspond to CNV features and columns to samples. The header includes sample IDs, followed by rows

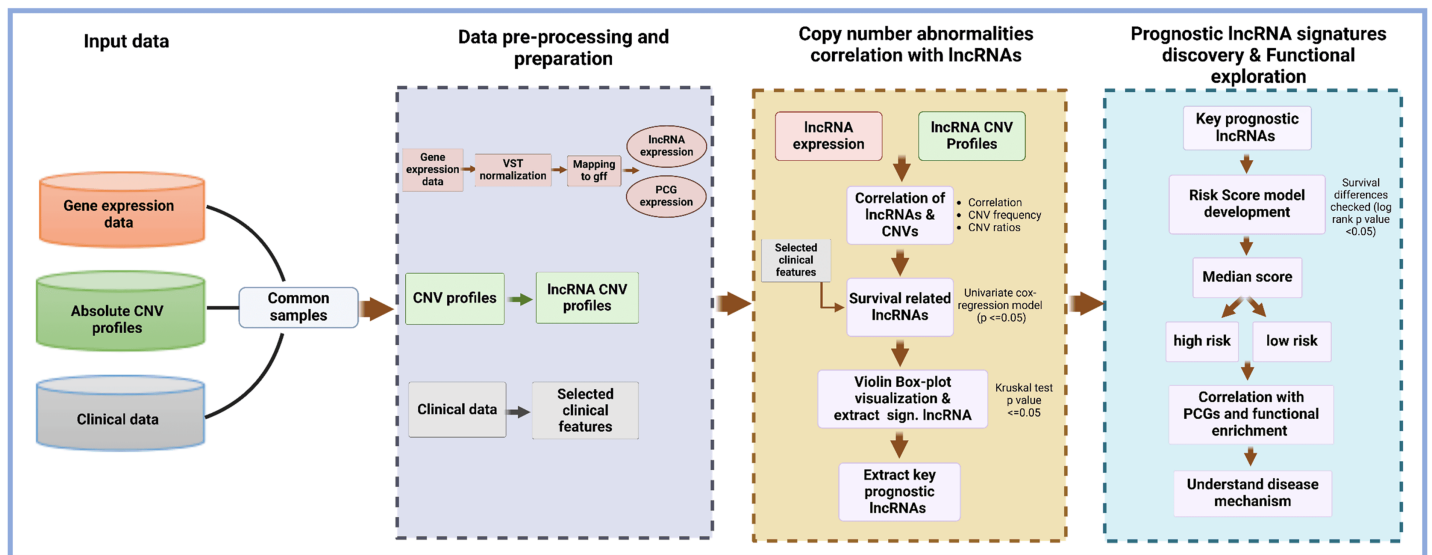


Figure 1 Overview of the lncRNACNVIntegrateR workflow. A schematic representation of the analytical workflow, starting from raw multi-omics input and proceeding through integration, biomarker discovery, risk modeling, and functional interpretation.

Full-size DOI: 10.7717/peerj.20131/fig-1

with CNV identifiers and their absolute copy number values (indicating amplifications or deletions).

- (3) **Clinical data:** A tab-delimited file with clinical annotations. Rows represent samples and columns represent clinical variables, including at minimum: Sample_ID, Survival_Time, and Status.

These datasets are processed through sequential stages of data harmonization, correlation analysis, survival modelling, and downstream functional exploration.

Core functionalities of the package

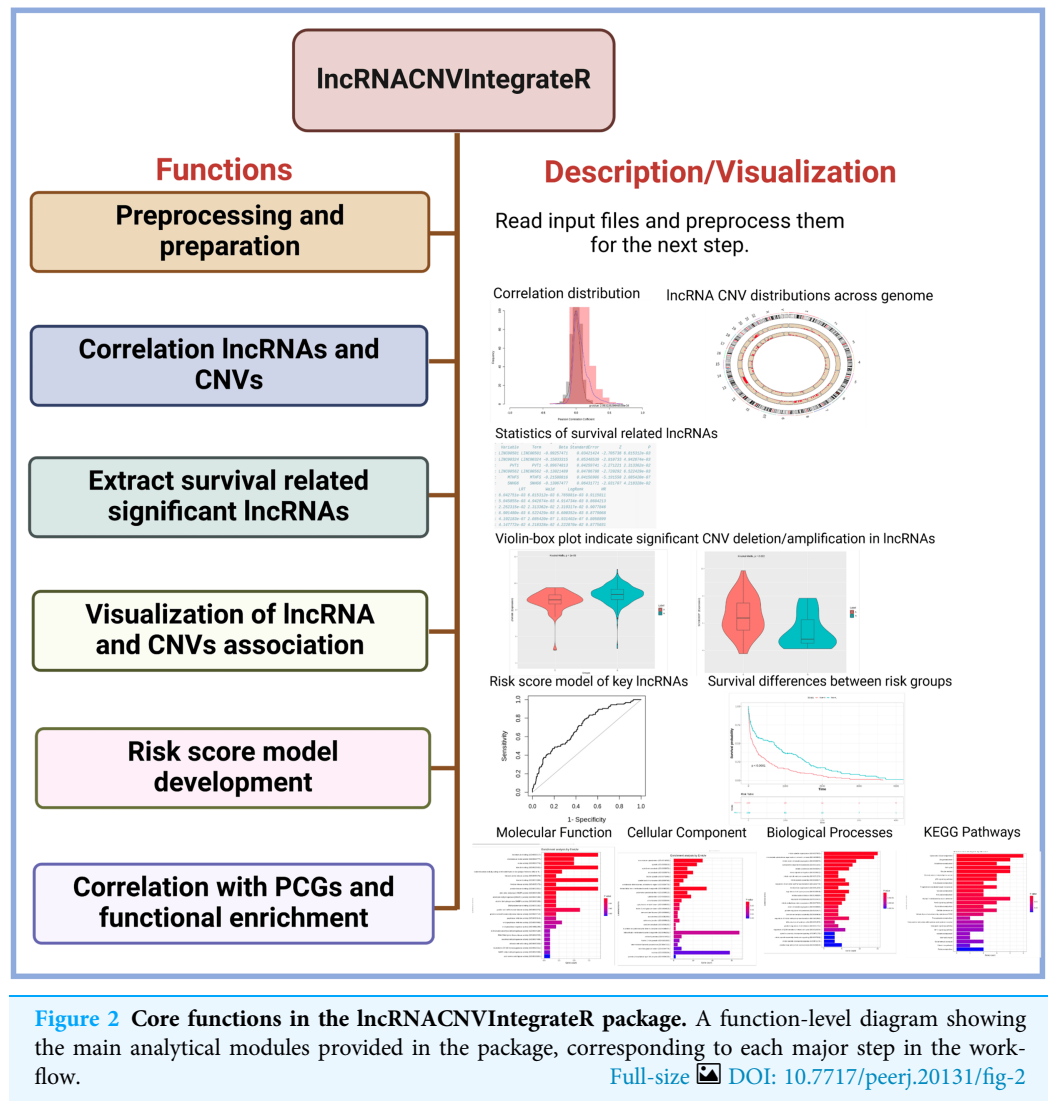
The lncRNACNVIntegrateR package provides a comprehensive suite of functions, each corresponding to a key step in the analysis pipeline (Fig. 1). These include:

Data preprocessing: The function automatically harmonizes sample IDs across the expression, CNV, and clinical datasets by identifying the common samples shared among all three. It then proceeds with downstream analysis using only these matched samples, ensuring consistency and compatibility across data types.

Correlation analysis: Computes statistical associations between lncRNA expression and CNV status to identify CNV-driven lncRNAs.

Survival association testing: Evaluates the prognostic significance of candidate lncRNAs using clinical survival data.

Signature identification: Filters for the most informative CNV-deregulated lncRNAs based on statistical significance and biological relevance.



Risk score modelling: Constructs and validates a prognostic model using selected lncRNA signatures.

Functional enrichment analysis: Investigates the downstream mRNA targets of identified lncRNAs to interpret their indirect roles in biological pathways and processes.

Each of these functionalities is implemented as a standalone function, allowing users to use each function in a standalone mode in their data analysis (as shown in Fig. 2). The structure and design of the lncRNACNVIntegrateR package is modular and user-centric, ensuring ease of use and flexibility for integrating and analyzing multi-omics data.

The lncRNACNVIntegrateR package processes the three primary input files: expression profiles, absolute CNV profiles, and clinical data (as discussed above) through various dedicated functions. It facilitates comprehensive analysis and integration of the data, encompassing several key tasks. These tasks include data preprocessing and preparation, evaluating the correlation between lncRNA expression and CNVs, and identifying

survival-associated lncRNAs. The package further identifies significant CNV-deregulated lncRNA signatures, constructs a risk score model for these signature lncRNAs, and performs functional enrichment analysis on the target genes of the identified lncRNA signatures. Detailed descriptions of the functions provided by this package are outlined below.

Input functions

To streamline data handling, the package provides dedicated input functions for each data type:

- `input_expr` for loading raw gene expression count matrices (whole transcriptome data),
- `input_cnv` for importing absolute CNV profiles, and
- `input_clin` for clinical data, including survival time and event status.

In addition, the package supports annotation using the GTF file *via* the `input_gtf` function (sourced from GENCODE version 22) and retrieves a curated list of lncRNAs using the `input_lncRNAs` function. The `'input_lncRNA_positions'` function helps incorporate genomic positions of lncRNAs for downstream analyses and is available in the example `extdata` directory of the package.

These datasets serve as inputs for the analysis and are passed into the package using the corresponding input functions. For the GBM and COAD case study, gene expression and clinical data were obtained from TCGA *via* the GDC portal (<https://gdc.cancer.gov/>), while absolute CNV data were downloaded from the UCSC Xena Browser (<http://xena.ucsc.edu/>). These datasets serve as inputs for the analysis and are passed into the package using the corresponding input functions.

Pre-processing

The `'preprocessing_and_preparation'` function conducts preprocessing on individual data types. The first step involves identifying common samples across the three data types, extracting the relevant information from each data type for these common samples, and then proceeding with the preprocessing. For gene expression data, first, data normalization is performed using the DESeq2 Variance Stabilizing Transformation (VST) (Love, Huber & Anders, 2014). Further, the lncRNAs and protein-coding genes (PCGs) are mapped using GTF information (obtained from gencode version 22, <https://www.encodeproject.org/files/gencode.v22.annotation>). Genes categorized in the GTF file under labels such as `"processed_transcripts,"` `"3prime_overlapping_ncRNA,"` `"sense_intronic,"` `"sense_overlapping,"` `"antisense,"` and `"lincRNA"` are recognized as lncRNAs, whereas those labeled as `"protein_coding"` are identified as PCGs. Following the extraction of lncRNAs and PCGs expression profiles, the CNV profiles for the lncRNAs are then extracted from the total absolute CNV data. Important parameters, including survival-related information, are extracted from the clinical data.

Expression data undergoes lncRNA and PCG annotation based on the GTF file. Genes annotated as `"processed_transcript,"` `"3prime_overlapping_ncRNA,"` `"sense_intronic,"` `"sense_overlapping,"` `"antisense,"` and `"lincRNA"` are considered lncRNAs, while those

labeled “protein_coding” are designated as PCGs. Expression data is normalized using the VST method from the DESeq2 package (Love, Huber & Anders, 2014). Simultaneously, lncRNA-specific CNV profiles are extracted from the absolute CNV dataset. Clinical parameters such as survival time and status are also extracted at this stage.

lncRNA-CNV correlation

After the data preprocessing and preparation step, the processed lncRNA expression and CNV profiles are utilized to compute the Pearson correlation coefficient (PCC) for each lncRNA using the ‘Correlation_lncRNAs_and_CNVs’ function. The correlation distribution is then plotted against a random distribution to evaluate any correlation. Subsequently, prognosis-related lncRNA signatures are extracted using a univariate Cox proportional risk regression model with a p -value cut-off <0.05 under function ‘extract_survival_related_significant_lncRNA’. Significant lncRNAs exhibiting amplification or deletion are further identified using a violin-box plot based on the Kruskal-Wallis test (cut-off p -value < 0.05).

Risk score model construction

The list of screened lncRNAs displaying significant CNV amplification/deletion, identified through the Kruskal-Wallis test in the ‘visualization_of_lncRNAs_and_CNVs_association’ function, is refined using a p -value cut-off <0.05 . Key prognostic lncRNAs (pr-lncRNAs) are selected to construct a risk score model using the PredictABEL (a R package available at <https://cran.r-project.org/package=PredictABEL>, version 1.2-4). With the identified key lncRNAs, the model can stratify patients into two risk groups using their respective risk scores. This functionality is implemented through the ‘Risk_score_model_development’ function. A survival plot is created to determine if these two groups differ in survival probabilities (p -value < 0.05).

The list of screened lncRNAs showing significant CNV amplification/deletion, determined by the Kruskal-Wallis test, was further refined using a multivariate Cox regression model with a p -value < 0.05 . Key prognostic lncRNAs (pr-lncRNAs) were chosen to build a risk score model using the PredictABEL R package (<https://cran.r-project.org/package=PredictABEL>). Based on these key lncRNAs, this model can classify patients into two risk categories using their respective risk scores. A survival plot was created to determine if these two groups differ in survival probabilities (p -value < 0.05).

Functional enrichment analysis

To explore the function of the identified pr-lncRNAs, the tool extracts highly correlated genes associated with these pr-lncRNAs. To analyze the functional enrichment of PCGs correlated with these lncRNAs, we utilize the enrichR package (Kuleshov et al., 2016) within the ‘Correlation_with_PCGs_and_functional_enrichment’ function.

RESULTS

To demonstrate the utility of lncRNACNVIntegrateR, we performed an analysis of GBM and COAD cancer datasets. The gene expression and clinical data sets were obtained from TCGA, and absolute CNV calls were sourced from UCSC Xena, respectively.

Case study 1: identification of CNV-driven lncRNA signatures in glioblastoma using lncRNACNVIntegrateR

To demonstrate the utility of lncRNACNVIntegrateR, we analyzed multi-omics data from GBM, the most aggressive primary brain tumor with a 5-year survival rate of 5.8% (Zhu *et al.*, 2021). The lncRNACNVIntegrateR pipeline was applied to integrate gene expression, CNV, and clinical data from the GBM dataset, accessed *via* the ‘input_expr’, ‘input_cnv’, and ‘input_clin’ functions. Data preprocessing was performed using the ‘preprocessing_and_preparation’ function, which extracted lncRNA and PCG expression profiles mapped to a GTF reference file.

The ‘Correlation_lncRNAs_and_CNVs’ function was used to evaluate associations between lncRNA expression and CNV profiles. This analysis revealed statistically significant correlations ($p < 0.05$), which were visualized as correlation distribution plots. The Pearson correlation distribution is shown in Fig. 3A, and the genome-wide distribution of lncRNA CNVs is presented in Fig. 3B. Next, the ‘extract_survival_related_significant_lncRNA’ function was applied to identify lncRNAs significantly associated with patient survival. The ‘visualization_of_lncRNA_and_CNVs_association’ function further confirmed CNV alterations, amplifications and deletions for these lncRNAs, as illustrated by representative violin plots.

By applying the Kruskal-Wallis test to compare CNV groups (diploid *vs.* amplified/deleted), and selecting those with statistically significant differences, we identified 21 lncRNAs significantly associated with both CNV alterations and patient survival. These include C1orf229, C3orf35, C9orf147, FENDRR, FTX, HCG18, LINC00115, LINC00265, LINC00271, LINC00324, LINC00354, LINC00641, LINC00645, LINC00667, LINC00676, LINC00858, LINC00907, LINC00963, LINC00970, MIAT, NEAT1, PCA3, RNU6ATAC35P, SNHG6, SNHG7, and SNX29P2.

A risk score model was then developed using the ‘Risk_score_model_development’ function. The model achieved a strong predictive performance with an AUC of 0.794 (95% CI [0.722–0.867]), as shown in Fig. 3C. Further, Kaplan-Meier survival analysis based on the expression of these 21 lncRNA signatures stratified GBM patients into high-risk and low-risk groups, demonstrating a statistically significant difference in overall survival ($p < 0.05$, Fig. 3D).

Using the full functionality of lncRNACNVIntegrateR, integration analysis of GBM datasets identified 21 lncRNA signatures for GBM as described above. Among these, LINC00115, LINC00324, LINC00641, LINC00963, MIAT, NEAT1, SNHG6, and SNHG7 play significant roles in glioma progression and therapy, suggesting their potential as biomarkers and therapeutic targets for glioblastoma (Chen, Lin & Li, 2023; Chen *et al.*, 2020; Liang *et al.*, 2022; Nie *et al.*, 2021; Rao *et al.*, 2024; Tang *et al.*, 2019; Yang *et al.*, 2020; Ye *et al.*, 2020). Thereby validating the findings and demonstrating the tool’s reliability in this study. In addition, several other lncRNAs identified in this study, including LINC00271, LINC00265, LINC00858, and LINC00667, have also been implicated in the pathogenesis of various cancer types beyond GBM. The risk score model achieved an exceptional AUC of 0.794 with a 95% Confidence Interval [0.722–0.867]. The survival

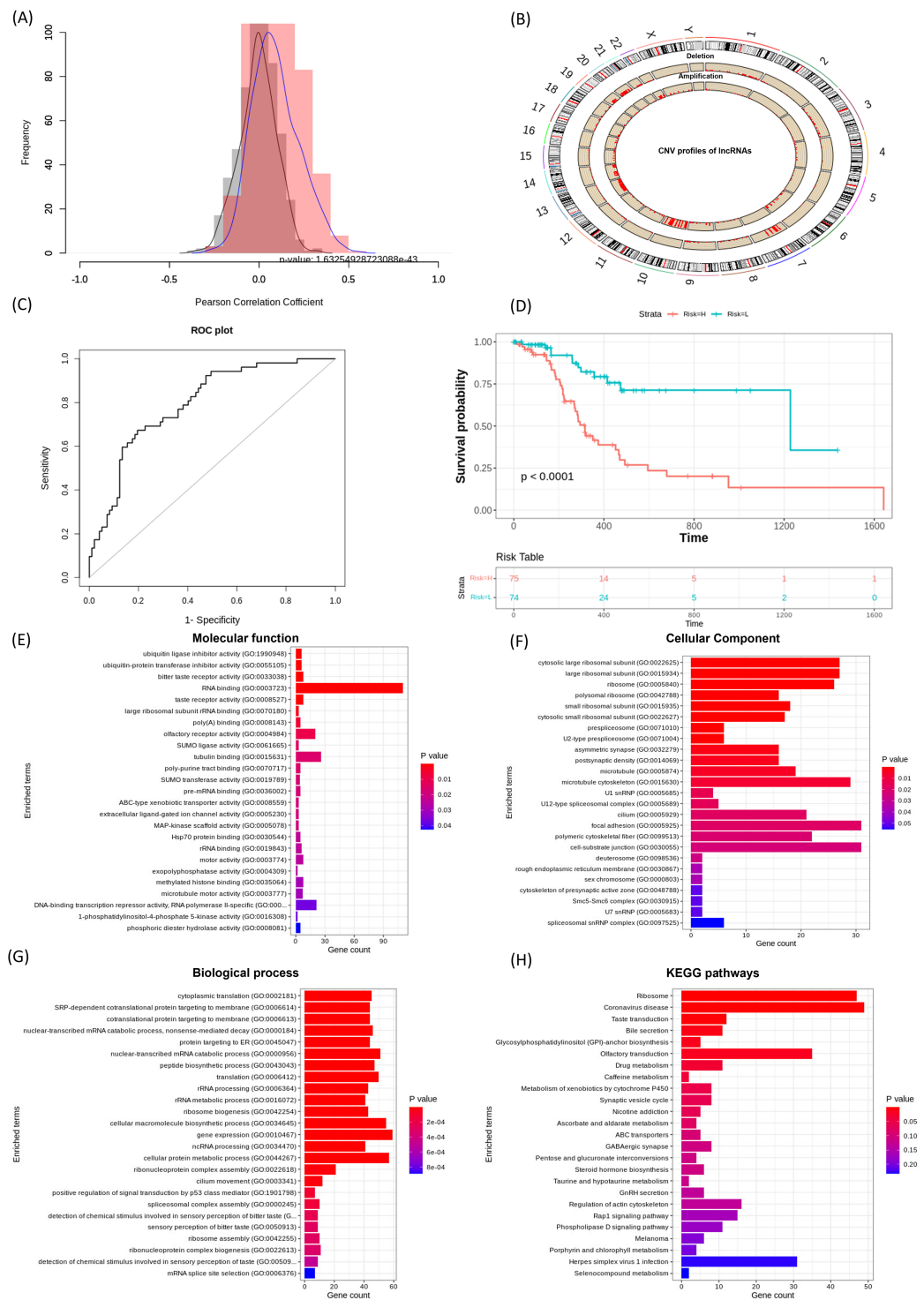


Figure 3 Performance of lncRNACNVIntegrateR on the GBM dataset. (A)–(H) illustrate the analysis workflow: (A) Correlation between lncRNAs and CNVs, (B) CNV distribution across the genome, (C) risk score plot for the final lncRNA signatures influenced by significant CNV changes, (D) survival analysis comparing high-risk and low-risk patient groups based on the risk score model, and (E)–(H) functional enrichment analysis linking lncRNAs to disease mechanisms.

Full-size DOI: 10.7717/peerj.20131/fig-3

analysis also effectively distinguished high-risk and low-risk groups and verified the model's accuracy.

Functional enrichment analysis (using function 'Correlation_with_PCGs_and_functional_enrichment') of the top correlated PCGs with the identified lncRNAs revealed significant enrichment in Gene Ontology (GO) categories, providing valuable insights into their potential biological functions. Among these, several key Biological Processes (BP) were identified to be enriched, which include ncRNA processing, p53-mediated signal transduction, nonsense-mediated mRNA decay, and regulation of gene expression, as illustrated in Fig. 3G. Corresponding Cellular Component (CC) and Molecular Function (MF) categories were also enriched, supporting the regulatory influence of these lncRNAs on diverse cellular mechanisms. Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis demonstrated that these lncRNA-targeted genes are involved in critical signaling pathways, such as the Rap1 signaling pathway, regulation of the actin cytoskeleton, and ABC transporter pathways (Fig. 3H). Collectively, these results highlight the involvement of lncRNA-associated PCGs in essential oncogenic processes, reinforcing their potential as biomarkers and therapeutic targets in GBM.

Case study 2: identification of CNV-driven lncRNA signatures in colon adenocarcinoma using lncRNACNVIntegrateR

To further evaluate and demonstrate the utility of lncRNACNVIntegrateR, we analyzed multi-omics data for COAD dataset, the predominant subtype of colorectal cancer, which ranks as the fourth most diagnosed cancer and second leading cause of cancer-related deaths in the United States (Benson *et al.*, 2018). Cancer prevention, screening, and treatment advances have significantly reduced CRC incidence and mortality rates (Edwards *et al.*, 2010). However, the prognosis for patients with advanced colon cancer remains poor (Sadanandam *et al.*, 2013), with 90% of these cases being COAD (Hajmanoochehri *et al.*, 2014).

Effective prognostic stratification is essential for improving patient outcomes in COAD. Using lncRNACNVIntegrateR, we integrated gene expression, CNV, and clinical data from COAD dataset. Significant correlations between lncRNA expression and CNV profiles were observed ($p < 0.05$), as shown in the correlation distribution (Fig. 4A) and the genome-wide CNV distribution plot (Fig. 4B). Applying this framework, the analysis identified 14 distinct CNV-driven lncRNA signatures, including SNHG7, TMEM191A, PCAT6, MIR210HG, MIAT, LINC00847, LINC00486, LINC00115, GAS5, FENDRR, FAM85B, C7orf66, C10orf95-AS1, and BCYRN1. A risk score model achieved an AUC of 0.707 (95% CI [0.631–0.782]), demonstrating reliable prognostic performance (Fig. 4C). Further, based on these 14 lncRNA signatures, the stratification of COAD patients into high- and low-risk groups (Fig. 4D), with Kaplan-Meier survival analysis confirming significant survival differences.

Eight lncRNAs, including SNHG7, PCAT6, MIR210HG, MIAT, LINC00115, GAS5, FENDRR, and BCYRN1 have been previously linked to colorectal cancer progression,

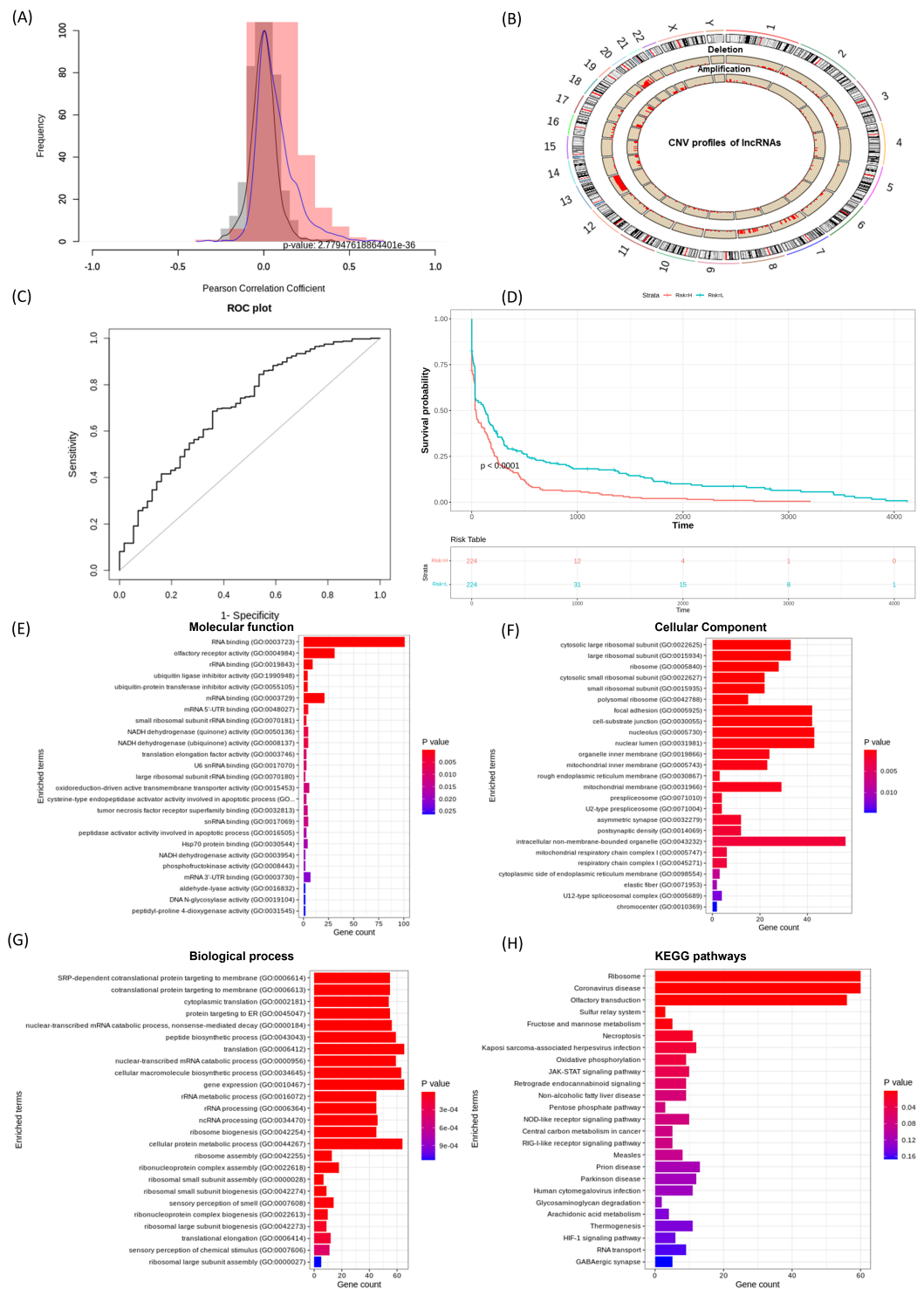


Figure 4 IncRNACNVIntegrateR performance on the COAD dataset. (A–F) The analysis workflow: (A) lncRNA-CNV correlation, (B) CNV distribution across the genome, (C) risk score plot for the final lncRNA signatures driven by significant CNV changes, (D) survival analysis comparing high-risk and low-risk patient groups based on the risk score model, and (E–H) functional enrichment analysis to associate lncRNAs with disease mechanisms. [Full-size !\[\]\(ba1b80118482ccef74a5d718ca4d7242_img.jpg\) DOI: 10.7717/peerj.20131/fig-4](https://doi.org/10.7717/peerj.20131/fig-4)

prognostic implications, and therapeutic response (*Pan et al., 2023; Han et al., 2024; Ruan et al., 2019; Liu et al., 2018*). *Feng et al. (2020), Xie et al. (2022a), Yang et al. (2023), Gu et al. (2018)* corroborating the tool's ability to identify clinically relevant biomarkers. These results underscore lncRNACNVIntegrateR's efficacy in uncovering CNV-driven lncRNA signatures with prognostic and therapeutic potential in COAD.

Figures 4E–4H showcases the results of functional enrichment analysis for the top correlated genes associated with these signature lncRNAs, emphasizing their roles in COAD tumorigenesis. Functional enrichment analysis revealed significant involvement of the correlated genes in key biological processes such as ncRNA processing, nonsense-mediated mRNA decay, translation, and gene expression. KEGG pathway analysis further highlighted enrichment in critical cancer-related pathways, including JAK-STAT signaling, necroptosis, viral infection, and central carbon metabolism in cancer. These findings demonstrate the utility of lncRNACNVIntegrateR in identifying prognostic lncRNA signatures and uncovering their functional roles in colorectal adenocarcinoma, supporting their relevance in disease progression and potential therapeutic targeting.

DISCUSSION

The lncRNACNVIntegrateR package offers a robust, user-friendly framework for integrating CNV and lncRNA expression data to identify prognostic biomarkers. We demonstrated its performance on two common cancers like GBM and COAD. By automating multi-omics data analysis, the tool streamlines the identification of CNV-driven lncRNA signatures, risk score modeling, and functional enrichment, making it accessible to researchers with varying computational expertise. Its ability to uncover both established and novel lncRNAs underscores its potential to advance biomarker discovery.

In this study, we leveraged the full capabilities of the lncRNACNVIntegrateR package to systematically analyze the interplay between CNVs and lncRNA expression in GBM and COAD datasets. GBM is one of the most aggressive and treatment-resistant forms of brain cancer; similarly, CAOD is also the most common form of Colorectal Cancer. Through lncRNACNVIntegrator analysis, we identified 21 lncRNAs exhibiting CNV-associated expression alterations in GBM and 14 in COAD.

Among these, several lncRNAs have been previously implicated in both the diseases. Their recurrence in our study provides biological validation of our approach and highlights the robustness of lncRNACNVIntegrateR in discovering biologically relevant signatures. The remaining lncRNAs, which have not yet been characterized in GBM and COAD, may serve as novel candidates for future functional validation.

Compared to existing methods, lncRNACNVIntegrateR demonstrates superior accuracy and functionality by systematically identifying CNV-driven lncRNAs and constructing risk models integrating genomic and transcriptomic layers. *Lei et al. (2025)* successfully built a prognostic model using cuproptosis-associated lncRNAs based on co-expression with known cuproptosis genes; however, its focus remained limited to a specific cell death pathway. Moreover, the associations were inferred solely from transcriptomic data, without investigating potential genomic drivers such as CNVs that

may underlie lncRNA dysregulation. In another study, although the artificial intelligence prognostic signature (AIPS) model demonstrated strong prognostic performance and integrated gene-level CNV and methylation data, it primarily focused on protein-coding genes (Jiang *et al.*, 2024). It did not systematically investigate CNV-driven regulatory mechanisms involving lncRNAs or construct a structured multi-omics framework centered on non-coding RNA regulation. Another study constructed a prognostic model for GBM using four m6A-related lncRNAs (*AC005229.3*, *SOX21-AS1*, *AL133523.1*, and *AC004847.1*) identified *via* LASSO regression (Xie *et al.*, 2022b).

In context of COAD, Lin, Liu & Xu (2025) identified CNV-driven lncRNAs and constructed a ceRNA network elucidating their involvement in essential biological processes. LINC00941 amplification was notably linked to poor prognosis, and potential drug targets were proposed. However, no prognostic risk score model was developed to evaluate the clinical utility of these CNV-lncRNAs (Lin, Liu & Xu, 2025). Another study established a prognostic model based on 19 HRR-related lncRNAs using TCGA-COAD data, which effectively stratified patients by survival risk and demonstrated strong predictive power. Functional enrichment further linked the model to the mitogen-activated protein kinase (MAPK) signaling pathway and homologous recombination deficiency (HRD), emphasizing its potential for precision medicine in colon cancer (Tang *et al.*, 2022).

While numerous studies have developed prognostic models in GBM and COAD, mostly focused on specific transcriptomic features such as cuproptosis, amino acid metabolism, m6A modification, or homologous recombination repair without examining the genomic mechanisms regulating lncRNAs. Notably, none have systematically investigated CNV-mediated regulation of lncRNAs or provided dedicated tools for such integrative analysis. Addressing this gap, we developed lncRNACNVIntegrateR, which identifies CNV-driven lncRNAs and leverages them for robust prognostic modelling. By enabling user-friendly multi-omics integration and downstream analysis, the tool advances biomarker discovery and supports broader applications in precision oncology.

This study has certain limitations. The evaluation of lncRNACNVIntegrateR was limited to publicly available tumor datasets from GBM and COAD, restricting its current assessment to cancer-specific contexts. Although the tool is designed for broader use, its performance on non-tumor datasets has not yet been examined. Moreover, while the current framework focuses on identifying CNV-driven lncRNAs and associated prognostic signatures, incorporating advanced CNV visualization tools and additional regulatory layers could further improve its interpretability and expand its applications.

CONCLUSION

Multi-omics dataset integration tools are crucial for comprehensively studying complex biological systems and identifying potential biomarkers. To meet the demand for user-friendly tools, we developed an R package, lncRNACNVIntegrateR, that integrates transcriptome and CNV data, enabling researchers to identify significant CNV-driven lncRNAs. This package facilitates exploring and developing personalized treatment strategies and clinical decision-making by providing deeper insights into the interaction

between lncRNAs and CNVs. We validated lncRNACNVIntegrateR by analyzing two TCGA cancer datasets, which enabled the shortlisting of promising biomarkers, demonstrated effective risk stratification, and provided robust survival analysis. The lncRNACNVIntegrateR can identify several reported as well as novel biomarkers in the datasets, and hence, the identified potential biomarkers may also be validated experimentally. Based on its robust validations, lncRNACNVIntegrateR serves as a valuable tool for integrating transcriptome and CNV datasets, enabling the discovery of novel and meaningful correlations between CNVs and long non-coding RNAs (lncRNAs) with significant prognostic potential.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the bioinformatics infrastructure grant from the Department of Biotechnology, Government of India (no. BT/PR40151/BTIS/137/5/2021). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Department of Biotechnology, Government of India: BT/PR40151/BTIS/137/5/2021.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Neetu Tyagi conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Shikha Roy conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Dinesh Gupta conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

TCGA-GBM datasets are available at: <https://portal.gdc.cancer.gov/projects/TCGA-GBM>.

UCSC-Xena datasets are available at:

<https://portal.gdc.cancer.gov/projects/TCGA-COAD>.

lncRNACNVIntegrateR is available at GitHub, and Zenodo:

- <https://github.com/tbgicgeb/lncRNACNVIntegrateR>,

- tbgicgeb. (2025). tbgicgeb/lncRNACNVIntegrateR: v1.0.0 lncRNACNVIntegrateR (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.15861906>.

Code is available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.20131#supplemental-information>.

REFERENCES

- Beermann J, Piccoli M-T, Viereck J, Thum T. 2016. Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. *Physiological Reviews* **96**(4):1297–1325 DOI [10.1152/physrev.00041.2015](https://doi.org/10.1152/physrev.00041.2015).
- Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen Y-J, Ciombor KK, Cohen S, Cooper HS, Deming D, Engstrom PF, Garrido-Laguna I, Grem JL, Grothey A, Hochster HS, Hoffe S, Hunt S, Kamel A, Kirilcuk N, Krishnamurthi S, Messersmith WA, Meyerhardt J, Miller ED, Mulcahy MF, Murphy JD, Nurkin S, Saltz L, Sharma S, Shibata D, Skibber JM, Sofocleous CT, Stoffel EM, Stotsky-Himelfarb E, Willett CG, Wuthrick E, Gregory KM, Freedman-Cass DA. 2018. NCCN Guidelines insights: colon cancer, version 2.2018. *Journal of the National Comprehensive Cancer Network: JNCCN* **16**(4):359–369 DOI [10.6004/jnccn.2018.0021](https://doi.org/10.6004/jnccn.2018.0021).
- Bodein A, Scott-Boyer M-P, Perin O, Lê Cao K-A, Droit A. 2022. timeOmics: an R package for longitudinal multi-omics data integration. *Bioinformatics* **38**(2):577–579 DOI [10.1093/bioinformatics/btab664](https://doi.org/10.1093/bioinformatics/btab664).
- Chen B, Lin P, Li N. 2023. Targeting the LINC00324/miR-16-5p/SEPT2 signaling cascade is effective to reverse malignant phenotypes in glioblastoma. *Anti-Cancer Agents in Medicinal Chemistry* **23**(13):1535–1544 DOI [10.2174/1871520623666230228122519](https://doi.org/10.2174/1871520623666230228122519).
- Chen Y, Yuan S, Ning T, Xu H, Guan B. 2020. SNHG7 facilitates glioblastoma progression by functioning as a molecular sponge for MicroRNA-449b-5p and thereby increasing MYCN expression. *Technology in Cancer Research & Treatment* **19**(3):1533033820945802 DOI [10.1177/1533033820945802](https://doi.org/10.1177/1533033820945802).
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, MacArthur DG, MacDonald JR, Onyiah I, Pang AWC, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME, The Wellcome Trust Case Control Consortium. 2010. Origins and functional impact of copy number variation in the human genome. *Nature* **464**:704–712 DOI [10.1038/nature08516](https://doi.org/10.1038/nature08516).
- Edwards BK, Ward E, Kohler BA, Ehemann C, Zauber AG, Anderson RN, Jemal A, Schymura MJ, Lansdorp-Vogelaar I, Seeff LC, van Ballegooijen M, Goede SL, Ries LAG. 2010. Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* **116**(3):544–573 DOI [10.1002/cncr.24760](https://doi.org/10.1002/cncr.24760).
- Feng W, Li B, Wang J, Zhang H, Liu Y, Xu D, Cheng K, Zhuang J. 2020. Long non-coding RNA LINC00115 contributes to the progression of colorectal cancer by targeting miR-489-3p via the PI3K/AKT/mTOR pathway. *Frontiers in Genetics* **11**:567630 DOI [10.3389/fgene.2020.567630](https://doi.org/10.3389/fgene.2020.567630).
- Gu L, Lu L, Zhou D, Liu Z. 2018. Long noncoding RNA BCYRN1 promotes the proliferation of colorectal cancer cells via up-regulating NPR3 expression. *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology* **48**(6):2337–2349 DOI [10.1159/000492649](https://doi.org/10.1159/000492649).

- Hajjari M, Khoshnevisan A, Shin YK. 2014.** Molecular function and regulation of long non-coding RNAs: paradigms with potential roles in cancer. *Tumor Biology* **35(11)**:10645–10663 DOI [10.1007/s13277-014-2636-z](https://doi.org/10.1007/s13277-014-2636-z).
- Hajmanoochehri F, Asefzadeh S, Kazemifar AM, Ebtehaj M. 2014.** Clinicopathological features of colon adenocarcinoma in Qazvin, Iran: a 16 year study. *Asian Pacific Journal of Cancer Prevention: APJCP* **15(2)**:951–955.
- Han L, Li Z, Zhang P, Sheng M, Wang W, Sun Y, Sun D. 2024.** LncRNA PCAT6 is a predictor of poor prognosis of colorectal cancer. *Journal of Gastrointestinal Oncology* **15(1)**:190–202 DOI [10.21037/jgo-23-910](https://doi.org/10.21037/jgo-23-910).
- Jiang Q, Yang X, Deng T, Yan J, Guo F, Mo L, An S, Huang Q. 2024.** Comprehensive machine learning-based integration develops a novel prognostic model for glioblastoma. *Molecular Therapy Oncology* **32(3)**:200838 DOI [10.1016/j.omton.2024.200838](https://doi.org/10.1016/j.omton.2024.200838).
- Kondo Y, Shinjo K, Katsushima K. 2017.** Long non-coding RNAs as an epigenetic regulator in human cancers. *Cancer Science* **108(10)**:1927–1933 DOI [10.1111/cas.13342](https://doi.org/10.1111/cas.13342).
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, Mcdermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. 2016.** Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research* **44(W1)**:W90-7 DOI [10.1093/nar/gkw377](https://doi.org/10.1093/nar/gkw377).
- Kumar MM, Goyal R. 2017.** LncRNA as a therapeutic target for angiogenesis. *Current Topics in Medicinal Chemistry* **17(15)**:1750–1757 DOI [10.2174/1568026617666161116144744](https://doi.org/10.2174/1568026617666161116144744).
- Lee J-Y. 2023.** The principles and applications of high-throughput sequencing technologies. *Development & Reproduction* **27(1)**:9–24 DOI [10.12717/DR.2023.27.1.9](https://doi.org/10.12717/DR.2023.27.1.9).
- Lei B, Zhan A, You G, Wu H, Chen S, Zhang D, Liu Z, Zheng N. 2025.** A novel cuproptosis-associated LncRNA model predicting prognostic and immunotherapy response for glioma. *Discover Oncology* **16(1)**:1089 DOI [10.1007/s12672-025-02912-6](https://doi.org/10.1007/s12672-025-02912-6).
- Liang J, Liu C, Xu D, Xie K, Li A. 2022.** LncRNA NEAT1 facilitates glioma progression via stabilizing PGK1. *Journal of Translational Medicine* **20(1)**:80 DOI [10.1186/s12967-022-03273-2](https://doi.org/10.1186/s12967-022-03273-2).
- Lin Q, Liu H, Xu Q. 2025.** Identification of dysregulated competitive endogenous RNA network driven by copy number variation in colon adenocarcinoma. *Computer Methods in Biomechanics and Biomedical Engineering* 1–12 DOI [10.1080/10255842.2025.2498718](https://doi.org/10.1080/10255842.2025.2498718).
- Ling H, Fabbri M, Calin GA. 2013.** MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nature Reviews Drug Discovery* **12(11)**:847–865 DOI [10.1038/nrd4140](https://doi.org/10.1038/nrd4140).
- Liu Z, Wang H, Cai H, Hong Y, Li Y, Su D, Fan Z. 2018.** Long non-coding RNA MIAT promotes growth and metastasis of colorectal cancer cells through regulation of miR-132/Derlin-1 pathway. *Cancer Cell International* **18(1)**:59 DOI [10.1186/s12935-017-0477-8](https://doi.org/10.1186/s12935-017-0477-8).
- Lorenzen JM, Thum T. 2016.** Long noncoding RNAs in kidney and cardiovascular diseases. *Nature Reviews Nephrology* **12(6)**:360–373 DOI [10.1038/nrneph.2016.51](https://doi.org/10.1038/nrneph.2016.51).
- Love MI, Huber W, Anders S. 2014.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15(12)**:1–21 DOI [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8).
- Lu Y, Fang Q, Qi M, Li X, Zhang X, Lin Y, Xiang Y, Fu Q, Wang B. 2023a.** Copy number variation-associated lncRNAs may contribute to the etiologies of congenital heart disease. *Communications Biology* **6(1)**:189 DOI [10.1038/s42003-023-04565-z](https://doi.org/10.1038/s42003-023-04565-z).

- Lu X, Meng J, Zhou Y, Jiang L, Yan F. 2021. MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. *Bioinformatics* 36(22–23):5539–5541 DOI 10.1093/bioinformatics/btaa1018.
- Lu Y, Zhou Y, Guo J, Qi M, Lin Y, Zhang X, Xiang Y, Fu Q, Wang B. 2023b. Integrated analysis of copy number variation-associated lncRNAs identifies candidates contributing to the etiologies of congenital kidney anomalies. *Communications Biology* 6(1):735 DOI 10.1038/s42003-023-05101-9.
- Mercer TR, Dingler ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics* 10(3):155–159 DOI 10.1038/nrg2521.
- Morriss GR, Cooper TA. 2017. Protein sequestration as a normal function of long noncoding RNAs and a pathogenic mechanism of RNAs containing nucleotide repeat expansions. *Human Genetics* 136(9):1247–1263 DOI 10.1007/s00439-017-1807-6.
- Morrow EM. 2010. Genomic copy number variation in disorders of cognitive development. *Journal of the American Academy of Child and Adolescent Psychiatry* 49(11):1091–1104 DOI 10.1016/j.jaac.2010.08.009.
- Munk K, Ilina D, Ziembra L, Brader G, Molin EM. 2024. Holomics—a user-friendly R shiny application for multi-omics data integration and analysis. *BMC Bioinformatics* 25(1):93 DOI 10.1186/s12859-024-05719-4.
- Nakamura Y. 2009. DNA variations in human and medical genetics: 25 years of my experience. *Journal of Human Genetics* 54(1):1–8 DOI 10.1038/jhg.2008.6.
- Ng S-Y, Lin L, Soh BS, Stanton LW. 2013. Long noncoding RNAs in development and disease of the central nervous system. *Trends in Genetics* 29(8):461–468 DOI 10.1016/j.tig.2013.03.002.
- Nie J, Feng Y, Wang H, Lian X-Y, Li Y-F. 2021. Long non-coding RNA SNHG6 supports glioma progression through upregulation of Notch1, Sox2, and EMT. *Frontiers in Cell and Developmental Biology* 9:707906 DOI 10.3389/fcell.2021.707906.
- Ning J, Wang F, Zhu K, Li B, Shu Q, Liu W. 2021. Characterizing the copy number variation of non-coding RNAs reveals potential therapeutic targets and prognostic markers of LUSC. *Frontiers in Genetics* 12:779155 DOI 10.3389/fgene.2021.779155.
- Ounzain S, Burdet F, Ibberson M, Pedrazzini T. 2015. Discovery and functional characterization of cardiovascular long noncoding RNAs. *Journal of Molecular and Cellular Cardiology* 89(4):17–26 DOI 10.1016/j.yjmcc.2015.09.013.
- Pan D, Chen K, Chen P, Liu Y, Wu Y, Huang J. 2023. Downregulation of LncRNA SNHG7 sensitizes colorectal cancer cells to resist anlotinib by regulating miR-181a-5p/GATA6. *Gastroenterology Research and Practice* 2023(5):6973723 DOI 10.1155/2023/6973723.
- Rao AKDM, Ramasamy D, Thangarajan R, Mani S. 2024. The lncRNA MIAT is regulated by NFYA to promote glioblastoma multiforme through the miR-221-3p/SIRT1 axis. *Gene Reports* 34:101883 DOI 10.1016/j.genrep.2024.101883.
- Rohart F, Gautier B, Singh A, Lê Cao K-A. 2017. mixOmics: an R package for ‘omics feature selection and multiple data integration. *PLOS Computational Biology* 13(11):e1005752 DOI 10.1371/journal.pcbi.1005752.
- Ruan Z, Xu Z, Li Z, Lv Y. 2019. Integral analyses of survival-related long non-coding RNA MIR210HG and its prognostic role in colon cancer. *Oncology Letters* 18(2):1107–1116 DOI 10.3892/ol.2019.10435.
- Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschleger S, Ostos LCG, Lannon WA, Grotzinger C, Del Rio M, Lhermitte B, Olshen AB, Wiedenmann B, Cantley LC, Gray JW, Hanahan D. 2013. A colorectal cancer classification system that

- associates cellular phenotype and responses to therapy. *Nature Medicine* **19**(5):619–625
DOI [10.1038/nm.3175](https://doi.org/10.1038/nm.3175).
- Subramanian I, Verma S, Kumar S, Jere A, Anamika K. 2020.** Multi-omics data integration, interpretation, and its application. *Bioinformatics and Biology Insights* **14**:7–9
DOI [10.1177/1177932219899051](https://doi.org/10.1177/1177932219899051).
- Tang X, Lin Y, He J, Luo X, Liang J, Zhu X, Li T. 2022.** Establishment and validation of a prognostic model based on HRR-related lncRNAs in colon adenocarcinoma. *World Journal of Surgical Oncology* **20**(1):74 DOI [10.1186/s12957-022-02534-0](https://doi.org/10.1186/s12957-022-02534-0).
- Tang J, Yu B, Li Y, Zhang W, Alvarez AA, Hu B, Cheng S-Y, Feng H. 2019.** TGF- β -activated lncRNA LINC00115 is a critical regulator of glioma stem-like cell tumorigenicity. *EMBO Reports* **20**(12):e48170 DOI [10.15252/embr.201948170](https://doi.org/10.15252/embr.201948170).
- Tian J, Luo B. 2022.** Identification of three prognosis-related differentially expressed lncRNAs driven by copy number variation in thyroid cancer. *Journal of Immunology Research* **2022**(9):1–18 DOI [10.1155/2022/9203796](https://doi.org/10.1155/2022/9203796).
- Valsesia A, Macé A, Jacquemont S, Beckmann JS, Kutalik Z. 2013.** The growing importance of CNVs: new insights for detection and clinical interpretation. *Frontiers in Genetics* **4**:92
DOI [10.3389/fgene.2013.00092](https://doi.org/10.3389/fgene.2013.00092).
- Wang A, Chen J, Zhou K, He Q, Ji K, Ji X, Wu X, Zhang J, Bu Z, Ji J. 2020.** Identification of lncRNA prognostic biomarkers associated with copy number variants in gastric cancer. 1–21
DOI [10.21203/rs.3.rs-32433/v1](https://doi.org/10.21203/rs.3.rs-32433/v1).
- Wilusz JE, Sunwoo H, Spector DL. 2009.** Long noncoding RNAs: functional surprises from the RNA world. *Genes & Development* **23**(13):1494–1504 DOI [10.1101/gad.1800909](https://doi.org/10.1101/gad.1800909).
- Wu G-C, Pan H-F, Leng R-X, Wang D-G, Li X-P, Li X-M, Ye D-Q. 2015.** Emerging role of long noncoding RNAs in autoimmune diseases. *Autoimmunity Reviews* **14**(9):798–805
DOI [10.1016/j.autrev.2015.05.004](https://doi.org/10.1016/j.autrev.2015.05.004).
- Xie J, Wang J-J, Li Y-J, Wu J, Gu X-J, Yang X-R. 2022a.** lncRNA GAS5 suppresses colorectal cancer progress by target miR-21/LIFR axis. *Evidence-Based Complementary and Alternative Medicine: ECAM* **2022**(19):3298939 DOI [10.1155/2022/3298939](https://doi.org/10.1155/2022/3298939).
- Xie P, Yan H, Gao Y, Li X, Zhou D-B, Liu Z-Q. 2022b.** Construction of m6A-related lncRNA prognostic signature model and immunomodulatory effect in glioblastoma multiforme. *Frontiers in Oncology* **12**:Cd013579 DOI [10.3389/fonc.2022.920926](https://doi.org/10.3389/fonc.2022.920926).
- Yang J, Hu Y, Tan Z, Zhang F, Huang W, Chen K. 2023.** The lncRNA FENDRR inhibits colorectal cancer progression via interacting with and triggering GSTP1 ubiquitination by FBX8. *Heliyon* **9**(12):e23161 DOI [10.1016/j.heliyon.2023.e23161](https://doi.org/10.1016/j.heliyon.2023.e23161).
- Yang J, Yu D, Liu X, Changyong E, Yu S. 2020.** LINC00641/miR-4262/NRGN axis confines cell proliferation in glioma. *Cancer Biology & Therapy* **21**(8):758–766
DOI [10.1080/15384047.2020.1776581](https://doi.org/10.1080/15384047.2020.1776581).
- Ye F, Xu R, Ge Y, Zheng Y, Liu X, Deng P, Xu X. 2020.** LINC00963 confers oncogenic properties in glioma by regulating the miR-506/BCAT1 axis. *Cancer Management and Research* **12**:2339–2351 DOI [10.2147/CMAR.S246332](https://doi.org/10.2147/CMAR.S246332).
- Zheng M, Hu Y, Gou R, Nie X, Li X, Liu J, Lin B. 2020.** Identification three lncRNA prognostic signature of ovarian cancer based on genome-wide copy number variation. *Biomedicine and Pharmacotherapy* **124**(2):109810 DOI [10.1016/j.biopha.2019.109810](https://doi.org/10.1016/j.biopha.2019.109810).
- Zhong Q, Lu M, Yuan W, Cui Y, Ouyang H, Fan Y, Wang Z, Wu C, Qiao J, Hang J. 2021.** Eight-lncRNA signature of cervical cancer were identified by integrating DNA methylation, copy

number variation and transcriptome data. *Journal of Translational Medicine* **19(1)**:1–16

[DOI 10.1186/s12967-021-02705-9](https://doi.org/10.1186/s12967-021-02705-9).

Zhu X, Pan S, Li R, Chen Z, Xie X, Han D, Lv S, Huang Y. 2021. Novel biomarker genes for prognosis of survival and treatment of glioma. *Frontiers in Oncology* **11**:667884

[DOI 10.3389/fonc.2021.667884](https://doi.org/10.3389/fonc.2021.667884).