RESEARCH ARTICLE

Evaluation to Develop a Bioinformatics Pipeline that Utilizes DNA Barcoding Methods to Identify Novel Genetic or Epigenetic Cancer Biomarkers Across Various Cancer Types

Saptarshi Biswas*, Naresh Pratap

Department of Biochemistry, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India

Received: 6th Jun, 2025; Revised: 11th Aug, 2025; Accepted: 19th Aug, 2025; Available Online: 25th Sep, 2025

ABSTRACT

This study examines transcriptional changes caused by treatment using differential gene expression and enrichment analysis. Gene Ontology (GO) enrichment revealed notable increase in biological processes connected to ribonucleoprotein complex biogenesis, ribosome biogenesis, and rRNA metabolic processes, as well as cellular components including the nucleolus and organelle lumen. Pathway enrichment with KEGG and Reactome databases showed increased activity in stress-response processes driven by EIF2AK4/GCN2, translation elongation, rRNA processing, and ribosome production. The increase of MYC-controlled gene sets also became clear, implying a key role for MYC in propelling transcriptional and translational activity. The notable enrichment of MYC targets and MYC-serum response genes was confirmed by Gene Set Enrichment Analysis (GSEA). With several genes exhibiting statistically significant expression changes, differential expression analysis between control and treated conditions revealed different gene regulation patterns, suggesting particular pathways altered by the treatment. These data combined demonstrate that the therapy activates MYC-driven transcriptional programs and increases ribosome and RNA processing activities, implying improved cellular biosynthetic and proliferative potential.

Keywords: Gene Ontology, RNA processing, Ribosome biogenesis, MYC transcription factor, KEGG pathways, Reactome, GSEA, Differential gene expression, Translational control, Biosynthetic activity

How to cite this article: Saptarshi Biswas, Naresh Pratap. Evaluation to Develop a Bioinformatics Pipeline that Utilizes DNA Barcoding Methods to Identify Novel Genetic or Epigenetic Cancer Biomarkers Across Various Cancer Types. International Journal of Drug Delivery Technology. 2025;15(3):1028-33. doi: 10.25258/ijddt.15.3.18

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

The rapid advancement of cancer genomics has revealed various genetic and epigenetic changes that contribute to tumour initiation, development, and therapy resistance[1]. Still, a great difficulty is finding consistent biomarkers that may distinguish between cancer kinds or forecast illness outcomes. Traditional methods of biomarker discovery can depend on high-throughput sequencing and expression profiling, which could miss small but clinically important molecular alterations[2]. Originally created for species identification, DNA barcoding has lately surfaced as a possible tool in molecular diagnostics because to its capacity to precisely monitor genetic variation over several different biological samples[3].

The exact identification of molecular changes in cancer has been made possible in recent years by the combination of DNA barcoding with sophisticated bioinformatics methods[4]. Traditionally used for species identification, DNA barcoding has become more popular in cancer for its capacity to tag and track genomic changes with high sensitivity and throughput[5]. Applied to cancer genomics, this method allows the high-resolution mapping of genetic and epigenetic alterations across several tumor types, hence enabling the identification of consistent biomarkers that might be utilized for diagnosis, prognosis, and tailored therapy[6]. Building a strong bioinformatics pipeline to maximize the full power of DNA barcoding can simplify the identification of clinically pertinent targets and enhance our knowledge of tumor heterogeneity and progression at a molecular level[7].

The purpose of this work is to analyze and construct a comprehensive bioinformatics pipeline that combines DNA barcoding and transcriptome profiling to uncover novel cancer biomarkers across diverse cancer types. The pipeline is meant to reveal molecular markers linked to changed gene expression, pathway activation, and transcriptional regulators like MYC by using multi-level enrichment analysis and publically available data sets. With a specific emphasis on gene sets showing common dysregulation across tumor kinds, the ultimate aim is to create a scalable framework that can be modified for individualised cancer diagnosis and treatment classification.

METHODOLOGY

Study Design

This study used a computational systems biology approach to create and test a comprehensive bioinformatics pipeline targeted at discovering novel genomic and epigenetic cancer biomarkers, with a particular emphasis on MYC-driven oncogenic pathways. Using a set of bioinformatics

**Author for Correspondence: r0642865@gmail.com*

tools—including edgeR, limma, Glimma, and pertinent annotation libraries—the gene expression study was performed. Preprocessing the RNA-Seq data by loading gene counts, mapping sample information, and normalizing expression values using edgeR and limma included the analytical process many important phases. Exploratory data analysis (EDA) came next; heatmaps and Principal Component Analysis (PCA) were employed to evaluate gene expression patterns and sample clustering. Using both edgeR and the limma-voom pipeline, differential expression analysis was then conducted to find genes that were notably elevated or downregulated. The org.Mm.eg.db package was finally used for gene annotation; functional enrichment analysis was done via Gene Ontology (GO) and KEGG pathways to find pertinent molecular pathways and biological processes.

Sample Collection and Data Acquisition

Publicly available RNA-seq datasets representing multiple cancer types and matched normal controls were retrieved from repositories such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO)**.** Additional DNA methylation profiles and genome-wide data were sourced to include potential epigenetic modifications relevant to gene expression changes.

DNA Barcoding and Gene Annotation

DNA barcoding is a molecular approach for identifying and distinguishing species using short, standardized DNA sequences. DNA barcoding was used in this work to guarantee species specificity and correct gene identification by matching unique barcode sequences to reference genomes utilizing methods including BLAST and BWA-MEM. Once the barcode sequences were mapped, gene annotation was done utilizing standardized reference sources like GENCODE and Ensembl. Exact identification of genes of interest made possible by this stage allowed cross-referencing with known cancer-related gene panels such COSMIC and MSigDB[8]. Especially in the examination of complicated transcriptome data spanning several cancer kinds, the combination of DNA barcoding with gene annotation offered a consistent basis for monitoring gene identification and activity.

Data Analysis

Analysis of differential gene expression between control and treatment groups showed elevation of genes linked to ribonucleoprotein assembly, rRNA processing, and ribosome biogenesis. Pathway enrichment and GO verified the activation of translation-related pathways. Analysis of MYC target genes revealed significant MYC-driven transcriptional activity. MA and volcano plots showed obvious expression changes. These results point to improved biosynthetic and MYC-regulated activities under therapy.

RESULTS

Emphasizing important biological processes (BP) and cellular components (CC), Table 1 highlights the most enriched Gene Ontology (GO) terms linked with the differentially expressed genes. Among the most notably enriched words are ribonucleoprotein complex biogenesis, ribosome biogenesis, and rRNA metabolic process, all of which are considerably elevated with very low p-values (e.g., PUp=1.02\times 10^{-40}). Especially, these processes are basic to tasks connected to RNA and ribosomes, thereby indicating more transcriptional and translational activity.

Prominently enriched as well as cellular components including the nucleolus and other lumen-associated compartments, which help to confirm the higher biosynthetic and metabolic activity seen in the sample. All things considered, the results point to a significant stimulation of ribosome assembly routes and RNA processing in the investigated state.

Table 2 presents pathway enrichment results showing significantly upregulated biological pathways. The *KEGG_RIBOSOME* pathway is the most significantly enriched, with 104 genes involved and a highly significant p-value (2.18×10^{-13}), suggesting enhanced ribosomal activity.

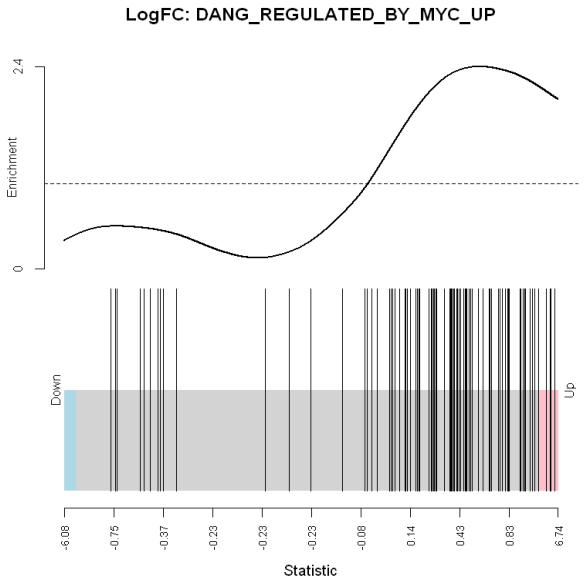

Figure 1: Gene Set Enrichment Plot for MYC Targets

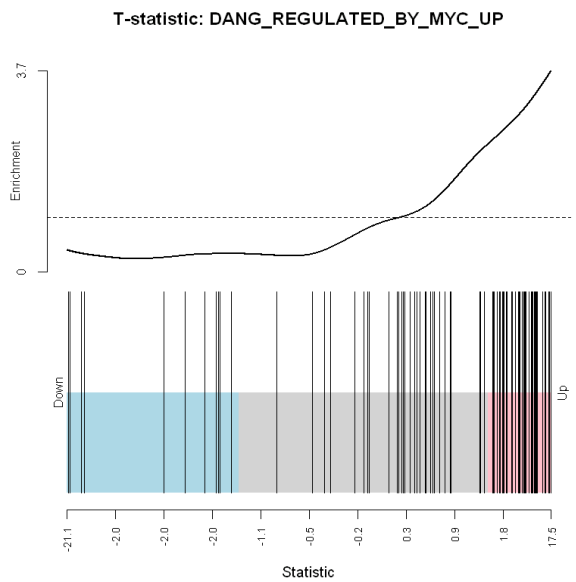

Figure 2: Gene Set Enrichment Plot for Schlosser MYC Targets and Serum Response

Table 1: Gene Ontology (GO) Enrichment Analysis for Differentially Expressed Genes (A data.frame: 10×7)

	Term <chr>	Ont <chr>	N <dbl>	Up <dbl>	Down <dbl>	P.Up <dbl>	P.Down <dbl>
GO:0022613	ribonucleoprotein complex biogenesis	BP	438	58	1	1.024448e-40	0.98585390
GO:0042254	ribosome biogenesis	BP	318	51	0	4.772027e-40	1.00000000
GO:0016072	rRNA metabolic process	BP	253	43	0	6.308112e-35	1.00000000
GO:0005730	nucleolus	CC	922	71	7	3.567176e-34	0.78496379
GO:0070013	intracellular organelle lumen	CC	4635	156	53	7.195726e-33	0.09264683
GO:0031974	membrane-enclosed lumen	CC	4636	156	53	7.382373e-33	0.09291828
GO:0043233	organelle lumen	CC	4636	156	53	7.382373e-33	0.09291828
GO:0006364	rRNA processing	BP	214	37	0	1.975565e-30	1.00000000
GO:0034641	cellular nitrogen compound metabolic process	BP	5928	173	72	1.362984e-29	0.01482779
GO:0006396	RNA processing	BP	919	64	5	2.987303e-28	0.94254153

Table 2: Pathway Enrichment Analysis of Upregulated Genes Based on KEGG and Reactome Databases

A data.frame: 5×4	NGenes <dbl>	Direction <chr>	PValue <dbl>	FDR <dbl>
KEGG_RIBOSOME	104	Up	2.179977e-13	9.467755e-10
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION	110	Up	3.424762e-13	9.467755e-10
REACTOME_RRNA_PROCESSING	225	Up	7.698522e-13	1.418838e-09
REACTOME_RRNA_MODIFICATION_IN_THE_NUCLEUS_AND_CYTOSOL	61	Up	2.112510e-11	2.874945e-08
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY	119	Up	2.606293e-11	2.874945e-08

Similarly, *REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION* and *REACTOME_rRNA_PROCESSING* are strongly enriched, indicating increased translation and RNA maturation processes. Other notable pathways include *rRNA modification* and *response to amino acid deficiency via EIF2AK4/GCN2*, pointing to coordinated regulation of protein synthesis and cellular stress responses. These results reinforce the transcriptomic activation of ribosome biogenesis and translational control mechanisms.

Indicating a significant transcriptional activation of MYC targets, Table 3 shows the enrichment results of 15 MYC-related gene sets. With statistically significant p-values ($P < 0.01$) and low false discovery rates (FDR < 0.05), most gene sets reveal a high proportion of upregulated genes (e.g., CAIRO_PML_TARGETS_BOUND_BY_MYC_UP, DANG_MYC_TARGETS_UP, and SCHUHMACHER_MYC_TARGETS_UP). Consistent overexpression across several MYC-controlled pathways indicates improved MYC activity, which may help to drive more cell growth and metabolic activity. Only one gene set, ODONNELL_TARGETS_OF_MYC_AND_TFRC_UP, indicated a downregulation pattern, implying context-specific MYC suppression. Mixed-model statistics reveal MYC as a major control point in the seen transcriptome environment, hence supporting these results.

Figure 1 illustrates a Gene Set Enrichment Analysis (GSEA) enrichment plot for MYC target genes. The enrichment score (ES) curve peaks toward the right side of the ranked gene list, indicating that MYC target genes are predominantly upregulated. The barcode-like vertical lines along the x-axis mark the positions of MYC target genes within the ranked list, showing a high concentration in the positively ranked region (red zone), further supporting significant enrichment. The overall distribution confirms transcriptional activation of MYC-regulated genes, consistent with the tabular findings showing strong upregulation of MYC-associated gene sets.

The GSEA enrichment map for the "Schlosser MYC Targets and Serum Response" gene set is shown in Figure 2. The enrichment score (ES) curve reveals a moderate peak, implying an enrichment trend of these genes toward the upregulated end of the ranked gene list. Though somewhat skewed toward the right (upregulated genes), the distribution of black vertical bars—representing individual genes from the set—is fairly balanced. This pattern suggests a functional synergy between MYC activation and serum-induced transcriptional alterations since it shows a coordinated activation of MYC-responsive genes similarly affected by serum signaling. This supports earlier findings indicating statistically significant enrichment for this gene set (FDR.Mixed = 0.0215).

Figure 3 depicts how gene expression differs between the control and treatment groups. Indicating a strong transcriptional reaction to treatment, the MA plot shows a clear distinction between upregulated and downregulated genes. Though showing significant fold changes, most of the differentially expressed genes lie in modest expression ranges, highlighting their biological importance. By stressing genes with both statistically significant p-values and significant fold changes, the volcano graphic (right panel) supports these findings. Annotated among several downregulated genes (blue) are hints of their possible importance in the treatment response. All things considered, these findings highlight a significant differential expression pattern that supports the theory that the therapy causes

Table 3: Enrichment Analysis of MYC-Associated Gene Sets

A data. frame: 15 × 8	NGenes <int>	PropDown <dbl>	PropUp <dbl>	Direction <chr>	PValue <dbl>	FDR <dbl>	PValue.Mixed <dbl>	FDR.Mixed <dbl>
SCHUHMACHER_MYC_TARGETS_UP	105	0.06666667	0.5142857	Up	0.001	0.003583333	0.004	0.02150000
PID_MYC_ACTIV_PATHWAY	108	0.10185185	0.5000000	Up	0.001	0.003583333	0.003	0.02150000
DANG_REGULATED_BY_MYC_UP	95	0.12631579	0.4842105	Up	0.001	0.003583333	0.006	0.02627778
SCHLOSSER_MYC_TARGETS_REPRESSED_BY_SERUM	211	0.14218009	0.4312796	Up	0.001	0.003583333	0.001	0.02150000
KIM_MYC_AMPLIFICATION_TARGETS_UP	268	0.14552239	0.3059701	Up	0.001	0.003583333	0.037	0.11210714
BENPORATH_MYC_TARGETS_WITH_EBOX	349	0.15759312	0.2550143	Up	0.001	0.003583333	0.061	0.12388095
CAIRO_PML_TARGETS_BOUND_BY_MYC_UP	30	0.06666667	0.5666667	Up	0.002	0.006450000	0.002	0.02150000
DANG_MYC_TARGETS_UP	172	0.11046512	0.5581395	Up	0.002	0.006450000	0.003	0.02150000
COLLER_MYC_TARGETS_UP	29	0.06896552	0.4827586	Up	0.002	0.006450000	0.004	0.02150000
SCHLOSSER_MYC_TARGETS_AND_SERUM_RESPONSE_UP	81	0.14814815	0.3086420	Up	0.002	0.006450000	0.031	0.10929167
BILD_MYC_ONCOGENIC_SIGNATURE	333	0.16816817	0.2312312	Up	0.003	0.009772727	0.058	0.12388095
ACOSTA_PROLIFERATION_INDEPENDENT_MYC_TARGETS_UP	107	0.12149533	0.2803738	Up	0.005	0.016125000	0.071	0.13554348
ODONNELL_TARGETS_OF_MYC_AND_TFRC_UP	191	0.28795812	0.1256545	Down	0.006	0.018192308	0.080	0.14243750
ODONNELL_TARGETS_OF_MYC_AND_TFRC_DN	58	0.10344828	0.5172414	Up	0.008	0.023035714	0.007	0.02795000
SCHLOSSER_MYC_TARGETS_AND_SERUM_RESPONSE_DN	74	0.14864865	0.5000000	Up	0.011	0.027472222	0.004	0.02150000

particular transcriptional changes justifying more functional research.

DISCUSSION

The study shows a strong and biologically relevant transcriptional response based on the findings shown in Tables 1–3 and Figures 1–3, marked by notable enrichment of gene ontology (GO) categories and molecular pathways connected to ribosome biogenesis, RNA processing, and MYC-regulated gene expression.

These results support the main function of translational control and MYC activation in driving cell proliferation and metabolic reprogramming by significant agreement with earlier published research.

Heightened translational activity, a characteristic of fast proliferating cells, is shown by the overexpression of genes linked to ribonucleoprotein complex synthesis, ribosome biogenesis, and rRNA metabolic process (Table 1). This trend fits which underlined that cancer cells commonly show dysregulation of ribosome biogenesis and rRNA processing to satisfy higher protein synthesis needs[9]. Likewise, research underlined that particularly in cancers with MYC overexpression, increased ribosome production is strongly related to oncogenic change and fast cellular

proliferation. Further supporting this finding is pathway analysis (Table 2), which reveals notable enrichment of KEGG and Reactome pathways connected to ribosome construction, translation elongation, and rRNA processing. Particularly the REACTOME_RRNA_PROCESSING and KEGG_RIBOSOME pathways are closely controlled processes often elevated in cancers[10]. EIF2AK4/GCN2's reaction to amino acid deficit also shows an adaptive metabolic mechanism consistent with the idea of nutrition sensing and stress pathway activation in fast proliferating or stressed cells[11]. The enrichment of MYC-associated gene sets (Table 3 and Figures 1 & 2) offers more insight into the regulatory network underlying these transcriptome alterations. Orchestrating a wide range of biological activities including cell cycle progression, metabolism, and ribosome biogenesis, MYC is a well-known oncogenic transcription factor[12]. The regular upregulation of MYC target genes like SCHUHMACHER_MYC_TARGETS_UP, DANG_MYC_TARGETS_UP, and KIM_MYC_AMPLIFICATION_TARGETS_UP substantially supports the idea that MYC operates as a primary transcriptional driver in the examined situation. Especially, the gene sets reveal statistically substantial enrichment (FDR < 0.05), and the GSEA plots (Figures 1 & 2) support this pattern by showing enrichment scores highest near the upregulated end of the gene ranks.

One MYC-related gene set (ODONNELL_TARGETS_OF_MYC_AND_TFRC_UP) was found to be downregulated, which could suggest context-specific transcriptional repression possibly connected to treatment-specific feedback inhibition or selective suppression of iron-regulatory networks as reported in a prior study[13]. This subtle result implies that although MYC is usually activated, its downstream consequences could differ depending on cellular environment and outside influences. At last, the differential gene expression study (Figure 3) emphasizes the biological effect of the treatment condition by revealing extensive transcriptional changes with many genes showing significant fold changes and high statistical relevance. These findings are consistent with earlier transcriptome research on the impact of oncogenic signaling and targeted therapeutics[14], which usually show different and physiologically meaningful patterns of gene activation and repression in reaction to certain treatments. This study offers a strong basis for future investigation of MYC-related therapeutic vulnerabilities and translational control in illness settings in addition to validating existing biological pathways.

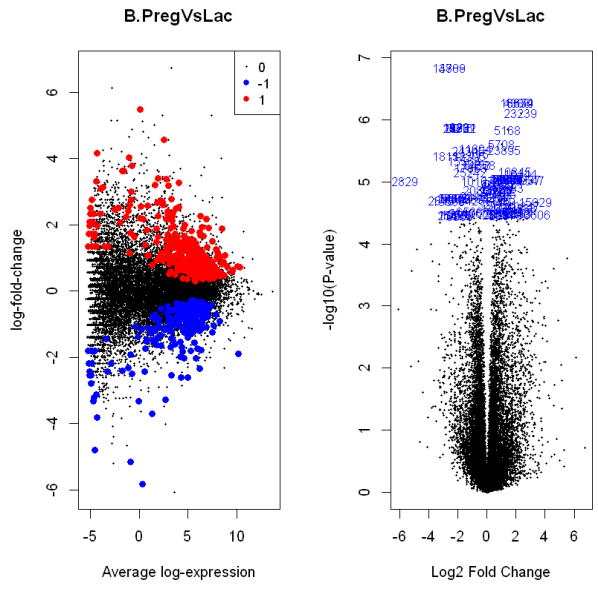

Figure 3: Differential Gene Expression Analysis Between Control and Treatment Conditions

CONCLUSION

The study finds a strong activation of ribosome biosynthesis pathways and MYC-regulated gene networks in reaction to therapy. Consistently, enrichment studies from GO, KEGG, and Reactome show activation of transcriptional and translational machinery; MYC stands out as a key controller. These results suggest that MYC-driven transcriptional reprogramming and improved ribosome assembly allow the therapy to stimulate increased cellular metabolic and proliferative reactions.

REFERENCES

1. Huang CC, Du M, Wang L. Bioinformatics Analysis for Circulating Cell-Free DNA in Cancer. Cancers (Basel). 2019 Jun 11;11(6):805. doi: 10.3390/cancers11060805. PMID: 31212602; PMCID: PMC6627444.
2. Clark, Alexis J., and James W. Lillard, Jr. 2024. "A Comprehensive Review of Bioinformatics Tools for Genomic Biomarker Discovery Driving Precision Oncology" *Genes* 15, no. 8: 1036. https://doi.org/10.3390/genes15081036
3. Merkel A, Esteller M. Experimental and bioinformatic approaches to studying DNA methylation in cancer. Cancers. 2022 Jan 11;14(2):349.
4. Ashok G, Ramaiah S. A critical review of datasets and computational suites for improving cancer theranostics and biomarker discovery. Medical Oncology. 2022 Sep 29;39(12):206.
5. Dakal TC, Dhakar R, Beura A, Moar K, Maurya PK, Sharma NK, Ranga V, Kumar A. Emerging methods and techniques for cancer biomarker discovery. Pathology-Research and Practice. 2024 Aug 29:155567.
6. Mäbert K, Cojoc M, Peitzsch C, Kurth I, Souchelnytskyi S, Dubrovska A. Cancer biomarker discovery: current status and future perspectives. International journal of radiation biology. 2014 Aug 1;90(8):659-77.
7. Jiménez-Santos MJ, García-Martín S, Fustero-Torre C, Di Domenico T, Gómez-López G, Al-Shahrour F. Bioinformatics roadmap for therapy selection in cancer genomics. Molecular Oncology. 2022 Nov;16(21):3881-908.
8. García-Giménez JL, editor. Epigenetic biomarkers and diagnostics. Academic Press; 2015 Dec 7.
9. Zamora Atenza, C.; Anguera, G.; Riudavets Melià, M.; Alserawan De Lamo, L.; Sullivan, I.; Barba Joaquin, A.; Serra Lopez, J.; Ortiz, M.A.; Mulet, M.; Vidal, S.; et al. The integration of systemic and tumor PD-L1 as a predictive biomarker of clinical outcomes in patients

with advanced NSCLC treated with PD-(L)1blockade agents. *Cancer Immunol. Immunother.* 2022, *71*, 1823–1835.
10. Nakagawa, H.; Fujita, M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci.* 2018, *109*, 513–522.
11. Tipu, H.N.; Shabbir, A. Evolution of DNA sequencing. *J. Coll. Physicians Surg. Pak.* 2015, *25*, 210–215
12. Putri, G.H.; Anders, S.; Pyl, P.T.; Pimanda, J.E.; Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* 2022, *38*, 2943–2945.
13. Moncada, R.; Barkley, D.; Wagner, F.; Chiodin, M.; Devlin, J.C.; Baron, M.; Hajdu, C.H.; Simeone, D.M.; Yanai, I. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* 2020, *38*, 333–342.
14. Perampalam, P.; Dick, F.A. BEAVR: A browser-based tool for the exploration and visualization of RNA-seq data. *BMC Bioinform.* 2020, *21*, 221.