medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Profiling Neoadjuvant Therapy Response in Rectal Cancer Using Publicly Available Transcriptomic RNA-seq Datasets

Aleksandra Stanojevic[1], Rafael Stroggilos[2], Mladen Marinkovic[3,4], Ana Djuric[1] Suzana Stojanovic-Rundic[3,4], Radmila Jankovic[1], Sergi Castellvi-Bel[5,6], Remond J.A. Fijneman[7], Antonia Vlahou[2], Jerome Zoidakis[2,8], Milena Cavic[1]

[1]Department of Experimental Oncology, Institute for Oncology and Radiology of Serbia (IORS), Belgrade, Serbia

[2]Department of Biotechnology, Biomedical Research Foundation, Academy of Athens (BRFAA), Athens, Greece

[3]Clinic for Radiation Oncology, Institute for Oncology and Radiology of Serbia (IORS), Belgrade, Serbia

[4]Faculty of Medicine, University of Belgrade, Belgrade, Serbia

[5]Gastroenterology Department, Fundació Clínic per la Recerca Biomèdica-Institut d'Investigacions Biomèdiques August Pi i Sunyer (FRCB-IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Hospital Clínic, University of Barcelona, Barcelona, Spain

[6]Faculty of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Spain

[7]Department of Pathology, The Netherlands Cancer Institute, Amsterdam, the Netherlands

[8]Department of Biology, National and Kapodistrian University of Athens, Greece

ABSTRACT

Neoadjuvant chemoradiotherapy followed by total mesorectal excision is standard for locally advanced rectal cancer, but response varies and current markers are insufficient. This study integrates public bulk RNAseq data to identify predictive features of response. *TRIM54* and *PABPC4* were up-regulated in the responder group, while *ADSS1* and *MGAT1* were up-regulated in non-responder group. *ARMC2* was identified as a predictive biomarker up-regulated in pathological complete response. Responder group showed enrichment of NK cells and CD4+ lymphocytes, while immune precursors were linked to poor outcome. Transcription factor analysis revealed *SP1* and *NFKB* activations in the non-responder group and *TCF15* in responder group. *SMAD3* and *RDXANK* were associated with complete regression, while *MYC* was dominant in incomplete regression. These findings provided insight into mechanisms underlying therapy response. To our knowledge, this is the first meta-analysis using high-throughput sequencing data, providing a valuable starting point for future rectal cancer research.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

INTRODUCTION

Colorectal cancer is the third most prevalent form of cancer globally in both sexes with nearly two million newly diagnosed cases in 2022, according to Globocan statistics [1]. Although colorectal cancer has been considered as a one disease entity for a long time, it is now acknowledged that colon cancer and rectal cancer represent two distinct types of disease at pathological, clinical and molecular levels [2–4]. Increasing evidence suggests that these two cancer types should be further divided based on their molecular profile on both genomic and transcriptomic levels thus classifying them into different molecular subtypes while considering their inherent heterogeneity [3–6]. Rectal cancer (RC) represents approximately one-third of total colorectal cancer cases [7,8] while locally advanced rectal cancer (LARC) is characterized as stage II (T3/4N0M0) and III (T1-4 N+ M0) [9].

Since 2004, neoadjuvant chemoradiotherapy (nCRT) has been the standard of care for locally advanced rectal cancer (LARC). This treatment typically consists of chemotherapy based on 5-fluorouracil combined with radiation therapy at a dose of 50.4–54 Gy, followed by total mesorectal excision (TME), with or without adjuvant chemotherapy [8]. Approximately 10–20% of patients achieve a pathological complete response (pCR) after TME, defined as the absence of residual tumor cells in the surgical specimen (ypT0N0) [10–12]. However, around 35% of patients show a poor or no response to nCRT, and a smaller subset, estimated at 2–4%, may even experience distant progression during treatment[13–15].

In some cases, patients achieve a clinical complete response (cCR) after nCRT. This is defined as the absence of residual tumor upon clinical reevaluation, which includes digital rectal examination, pelvic magnetic resonance imaging (MRI), and proctoscopy. For patients with cCR, a non-operative management strategy known as the "watch-and-wait" approach may be considered, involving close follow-up and reserving surgery only in cases of local tumor regrowth. However, the concordance between cCR and pCR remains low [16,17].

To improve cCR rates following neoadjuvant treatment, current research directions focus on several key strategies: prolongation of the interval between the completion of neoadjuvant therapy and surgery, escalation of radiotherapy dose, and intensification of systemic treatment within the framework of total neoadjuvant therapy (TNT) [18–22]. Furthermore, the incorporation of immunotherapy in the management of mismatch repair-deficient (dMMR) rectal cancer has expanded the therapeutic landscape and opened new possibilities for personalized treatment [23]. Within the organ preservation paradigm, particularly in patients demonstrating a clinical complete response, the watch-and-wait approach aims to maintain oncologic safety while preserving quality of life. These evolving therapeutic modalities underscore the increasing need for refined patient stratification and risk-adapted treatment strategies to optimize outcomes.

Despite significant variability in patient responses, neoadjuvant chemotherapy remains a conventional treatment protocol. At the molecular level, currently, there is still a lack of clinically validated predictive features that can distinguish treatment response and enable improved quality of life. Identifying predictive biomarkers will enable improved patient selection and enhance treatment optimization opportunities. Furthermore, the watch-and-wait approach still has limited potential, as the primary objective of organ preservation may be compromised by the potential for tumor regrowth, distant metastasis and ultimate local failure [24]. Discovery of predictive biomarkers may improve patient selection with clinical complete responses that could

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

be eligible for a watch-and-wait approach and prevent TME or, at the very least, prolong the interval between treatment and surgery.

It is widely recognized that the cellular response to DNA damage induced by irradiation is complex and involves multiple forms of cell death including apoptosis, mitotic devastation, senescence, necrosis, autophagy and stress signaling pathways, whereas chemotherapy serves to sensitize tumor cells to irradiation. Furthermore, it has been reported how radiation affects tumor cells and the surrounding during treatment but there is a lack of evidence regarding the impact of transcriptomic and proteomic profiles of tumor and tumor-microenvironment (TME) on tumor's therapeutic response [25–27].

Transcriptomic profiling of tissue provides characterization of molecular features and enables insight into characteristics that can be used for better understanding of mechanism behind response. LARC was profiled through various omics approaches and technologies [28–32]. Research conducted by our team demonstrated that patients with varying responses exhibit distinct proteomics profiles [33] as well as hematological features [34] that can be used for their stratification. By applying a radiomics approach, radiomic features showed differences between treatment outcomes [35]. Numerous studies examining LARC tissue have been performed at various levels but most suffer from the limitation of small sample size, resulting in limited statistical power and usually the lack of validation of results obtained. Conversely, the integration of studies conducted at the same omics level may enhance statistical power and increase likelihood of significant biological findings.

This study aims to integrate publicly available bulk RNAseq data and as a step beyond the state of the art in this disease context, to perform a comprehensive characterization of LARC at the trascriptome level, towards the identification of predictive biomarkers. To date, meta-analyses concerning LARC were conducted using microarray transcriptomic data[36]; to the best of our knowledge this represents the first study that uses publicly available transcriptomic data derived from Next Generation Sequencing methodologies, offering comprehensive insights.

MATERIALS AND METHODS

Study cohort

A comprehensive dataset search strategy was employed, encompassing the Gene Expression Omnibus and PubMed repositories using a variety of keywords as shown in Figure 1 conducted on 31/07/2024. Diverse range of keywords were used to include all articles/datasets containing transcriptome RNAseq data on pretreatment biopsy specimens along with data on therapeutic response to therapy after surgery of LARC patients. Studies examining other anatomical segments of the colorectum, as well as those conducted on animal models or cell lines were excluded. The inclusion and exclusion criteria are presented in Figure 1. Throughout the search process, overlapping results were noted among different keywords. Following extensive searches for rectal cancer studies, across Gene Expression Omnibus (GEO) and PubMed repositories five transcriptomics datasets were identified fulfilling the inclusion criteria, encompassing a total of 248 samples (GSE190826[37], GSE233517[38,39], GSE209746 [40], GSE80606 [41], Toomey et al.[42]) as detailed in Table 2-3, Figure 1. All patients were diagnosed with LARC and underwent nCRT utilizing 5-FU or Capecitabine, some patients were treated in accordance with the total neoadjuvant therapy (TNT) model which incorporates oxaliplatin, either induction or consolidation. Patient responses were evaluated after surgery using different methods as shown

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

in Table 1. The scales were standardized [43] and the patients were categorized into responders (R) and non-responders (NR) or alternatively as those exhibiting a pathological complete response (pCR) and pathological non-complete response (Non pCR) Table 1. Patients classified as responders (R) are defined as those achieving either complete response or a near-complete response to therapy. Patients exhibiting absent or minimal response were classified as nonresponders (NR). The same patient cohort was stratified into those achieving a pathological complete response (pCR), characterized by non- detectable tumor cells after surgery, and all others achieved pathological non-complete response (Non pCR).

Table 1. Harmonization of different classifications of response to nCRT

Mandard five-tiered TRG	Dworak five-tiered system	CAP System	RCPath System	Ryan modified three-tiered Mandard TRG	Responder (R)/Nonresponder (NR)	Pathological complete responders (pCR)/ Non pathological complete responders (Non pCR)
TRG 1	TRG 4	TRG 0	RCPath A	TRG 1	R	pCR
TRG 2	TRG 3	TRG 1	RCPath B		R	Non pCR
TRG 3	TRG 2	TRG 2	RCPath C	TRG 2	NR	Non pCR
TRG 4	TRG 1	TRG 3		TRG 3	NR	Non pCR
TRG 5	TRG 0				NR	Non pCR

nCRT- neoadjuvant chemoradiotherapy, TRG- tumor regression grade, R- responder, NR- non-responder, pCR - pathological complete response, Non pCR- Pathological non-complete response, RCPath- Royal College of Pathologists, CAP- College of American Pathologists

Table 2. Transcriptomics datasets characteristics.

Dataset	Number of samples included in study	Number of samples included in analysis	Response evaluation method used in study	Definition of pCR/Non pCR	Number of pCR/Non pCR	Definition of R/N
GSE190826 [37]	117	105	AJCC	0-pCR 1-Non pCR 2-Non pCR 3- Non pCR	26/79	0-R 1-NA 2-NA 3-NA
GSE209746 [40]	738	93	AJCC	0-pCR 1-Non pCR 2-Non pCR 3-Non pCR	23/70	0-R 1- R 2- NR 3- NR
GSE233517 [38,39]	22	11	Mandard scale	ypT, ypN-0 – pCR other Non pCR	3/8	R- R NR- NR
GSE80606 [41]	22	22	Mandard scale	total- pCR near total- Non pCR moderate- Non pCR minimal- Non pCR	3/19	total- R near total- R moderate- NR minimal- NR
Toomey et al.[42]	24	17	RCPath System	RCPath A- pCR RCPath B- Non pCR RCPath C- Non pCR	**5/12**	RCPath A- R RCPath B- NA RCPath C- NR

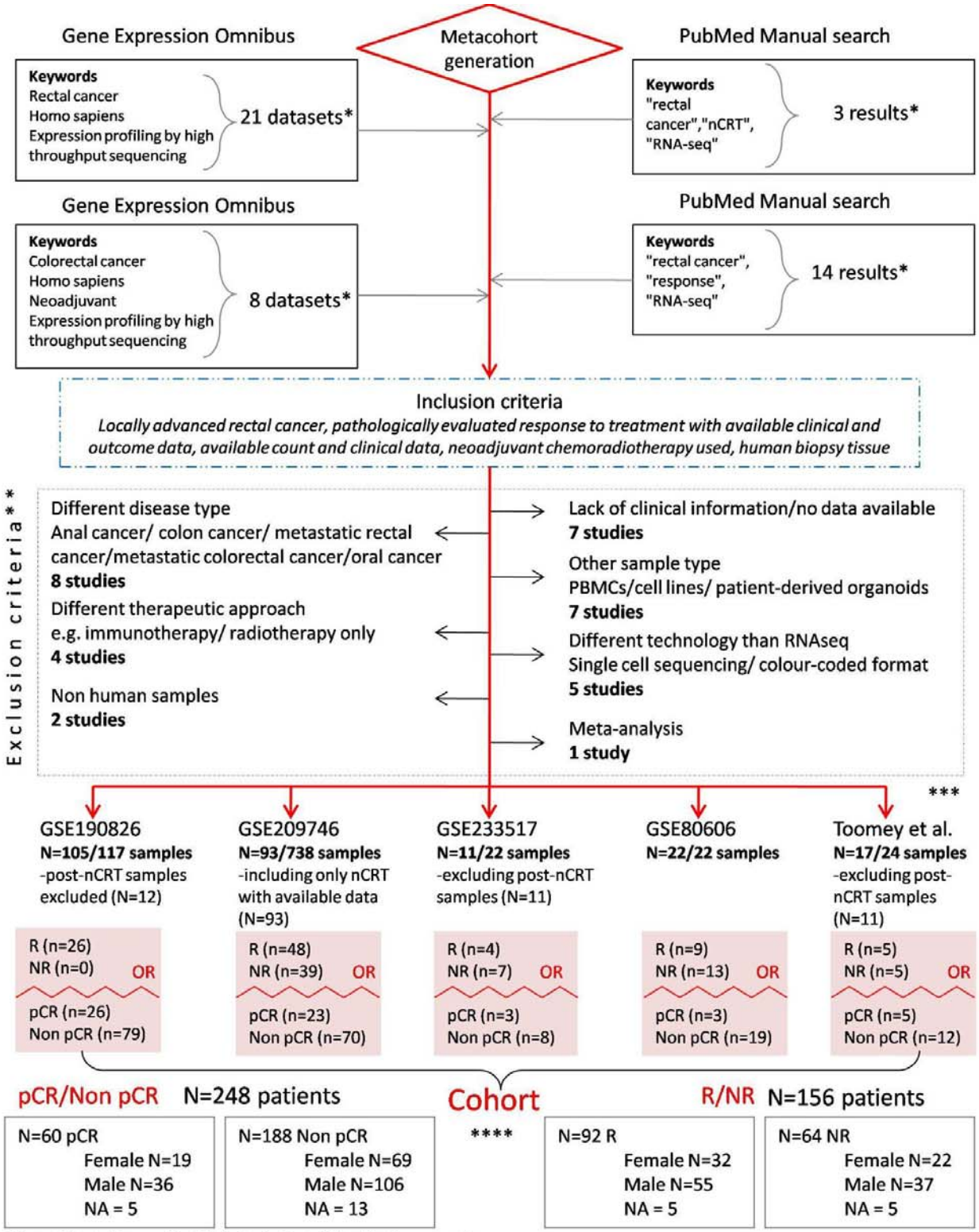
medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Sum	**923**	**248**		60/188=**248** **24%/76%**

**-patients were classified as pCR/Non pCR - pCR was defined as R, for Non pCR it was not possible to determine the R/NR status.*

R- responder, NR- non-responder, pCR - pathological complete response, Non pCR-Pathological non-complete response. RCPath- Royal College of Pathologists, AJCC-American Joint Committee on Cancer, NA- Not Applicable

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .



*Some of the studies were identified using both study identification approaches
**Some studies were excluded due to multiple criteria
***N= number of samples included in meta-analysis/ number of samples provided in study
****Only patients who met the inclusion criteria and patients in whom it was possible to determine a therapeutic response based on the available data were included in the study, which led to the difference in cohorts within individual studies and the final cohort within the meta-analysis.

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

nCRT-neoadjuvant chemoradiotherapy, NA- Not Applicable, R- responder, NR- non-responder, pCR - pathological complete response, Non pCR-Pathological non-complete response

Figure 1. Transcriptomic data search flowchart. The publicly available scientific database PubMed together with the Gene Expression Omnibus database was used to search for studies conducted on pretreatment LARC samples with the aim of determining predictive biomarkers. Studies were reviewed based on the inclusion criteria. The final group consisted of five studies where patients were divided into responders (TRG1-2, Mandard scale) and non-responders (TRG3-5, Mandard scale) or pCR (TRG1, Mandard scale) and Non pCR (TRG2-5, Mandard scale)

Table 3. Studies included in meta-analysis

Title	Study	GEO identifier	Year of publication	Type of Study	Study
Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer	Nicolas et al., 2022[37]	GSE190826	2022	Retrospective	Germa
Genomic and transcriptomic determinants of response to neoadjuvant therapy in rectal cancer	Chatila K et al., 2022 [40]	GSE209746	2022	Multi cohort retrospective	USA
YWHAZ and TBP are potential reference gene candidates for qPCR analysis of response to radiation therapy in colorectal cancer Neoadjuvant chemoradiotherapy up-regulates PD-L1 in radioresistant colorectal cancer	Kim et al., 2023 [39] Bae Uk et al., 2024 [38]	GSE233517	2023, 2024	Retrospective	Republ Korea
Genomic and transcriptomic characterisation of response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer	Toomey et al., 2020 [42]	NA	2020	Retrospective	Ireland
PSMB8 as a Candidate Marker of Responsiveness to Preoperative Radiation Therapy in Rectal Cancer Patients	Ha Jin et al., 2017 [41]	GSE80606	2017	Retrospective	Republ Korea
Prognostic genome and transcriptome signatures in colorectal cancers	Nunes et al., 2024 [44]	NA	2024	Retrospective	Sweder

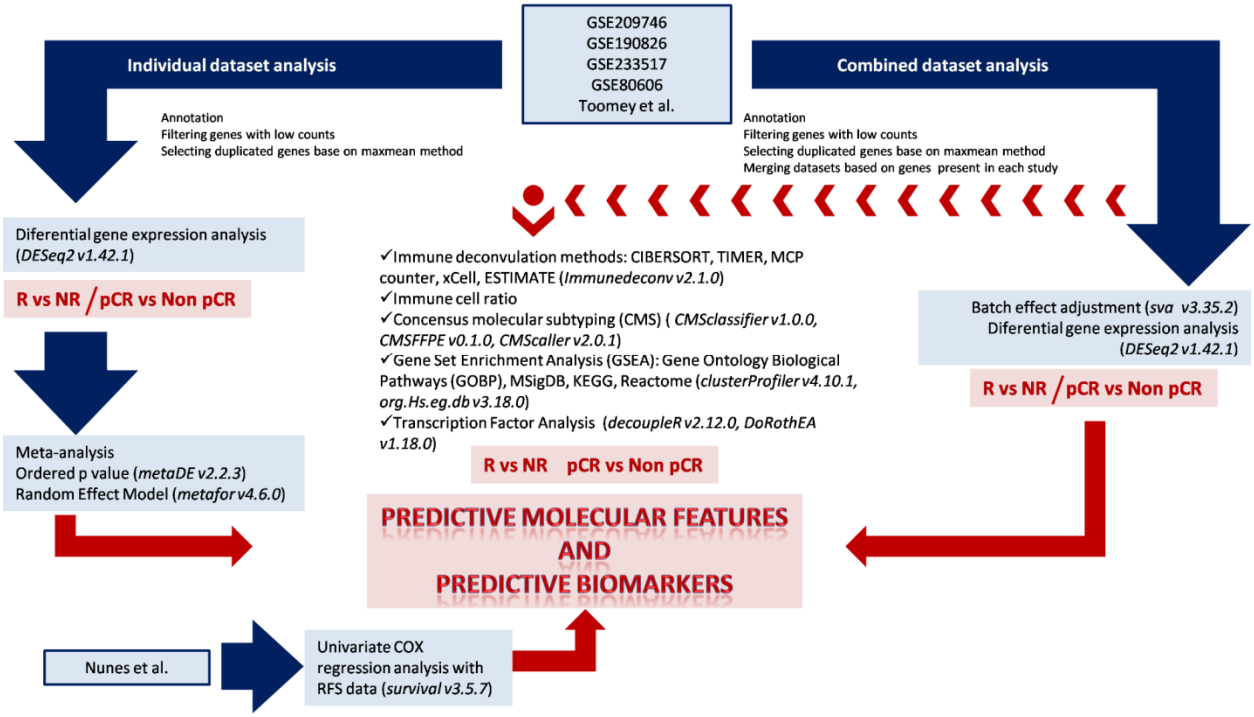
GEO- Gene Expression Omnibus

Combined batch-free dataset generation and Differential gene expression analysis

Each dataset was initially analyzed independently to identify differentially expressed genes (DEGs) between R/NR and pCR/NonpCR. Genes exhibiting a cumulative count of less than 10 across samples were excluded. Differential expression analysis was performed using the DESeq2 v1.42.1 pipeline [45], specifically developed for the analysis of RNAseq raw counts data. Statistical analysis was performed by applying the Wald statistics, while p-values were corrected with Benjamini-Hochberg adjustment method for multiple comparisons. GSE190826 was excluded from the individual R and NR comparison due to absence of NR data. Beside the identification of predictive biomarkers for R/NR, predictive biomarkers of pathological complete response were identified. The combined cohort was generated by integrating samples from five datasets based on genes present in all studies (intersection). Genes with redundant sequencing data were retained by selecting those with the highest mean counts across all samples. Given that different generations of sequencing technology were used (Illumina NextSeq 500, Illumina HiSeq 2000,

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Illumina HiSeq 2500, and Illumina HiSeq 4000), and the variation in sample and library sizes across datasets, batch effect removal was performed using the ComBat-seq algorithm (sva v3.35.2 [46]) and DESeq2 for downstream differential expression analysis. ComBat Seq is a variation of ComBat method grounded in empirical Bayes techniques and, in contrast to alternative batch adjustment methods, assumes that data follows a negative binomial distribution. In addition, this approach ensures a consistent workflow for both individual datasets and the combined cohort while maintaining the negative binomial distribution in the batch-adjusted counts data and preserving integer counts form. ComBat seq effectively reduces batch effects by eliminating technical variances and preserving biological variability (Figure 3). A comprehensive overview of the methodology employed in this meta-analysis is provided in Figure 2. After obtaining a representative combined dataset, DESeq2 pipeline with the dataset as covariance was applied in order to identify DEGs between R/NR and pCR/Non pCR.



R- responder, NR- non-responder, pCR - pathological complete response, Non pCR-Pathological non-complete response, KEGG-Kyoto Encyclopedia of Genes and Genomes, MSigDB-The Molecular Signatures Database

Figure 2. Overview of the methodological approach used to identify predictive parameters of response (R vs NR and pCR vs Non pCR) to neoadjuvant chemoradiotherapy in patients with LARC.

Meta-analysis of individual datasets

Meta-analysis approaches, ordered p-value (rOP) method and the random effect model (REM) was used to identify predictive biomarkers of response to nCRT in LARC based on statistical output of individual studies. [47] OrP method accounts for the distribution of p value for each gene across analyzed datasets, whereas REM approach assesses the LogFC by considering the distribution of effect sizes for each gene, accounting for inter- and intra-study variations, thereby

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

estimating the effect size along with the corresponding p-value and adjusted p-value. In order to consider both effects in meta-analysis setting methods, p-value and Log2FC from all analyzed genes in individual datasets were used. The estimation of between-study variance was performed employing the Restricted Maximum Likelihood (REML) method. To perform meta-analysis, R packages MetaDE v2.2.3 [48] and metafor v4.6.0[49] were used.

Validation cohort

The study of Nunes et al[44] published in August 2024 was used for independent validation, in which the level of gene expression as prognostic markers of LARC was examined. No independent study examining the predictive potential of gene expression was available at the time of validation. Regardless, it is expected that patients who have a better response to therapy have a longer survival without tumor regression and vice versa. Patients with available RNAseq data who had stage II and III rectal cancer and received nCRT were subtracted from the study. A total of 68 patients (33 female, 35 male) with a median age of 67 (range 38-93) were included in the validation study. Regression-free survival (RFS) was used as an input variable and was defined as the interval from surgery to the earliest incidence of local or distal recurrence or mortality. Univariate COX regression analysis was performed to evaluate the prognostic impact of gene expression using the R package survival v3.5.7 [50]. The results of COX analysis were used to validate the identified predictive markers, validation criteria was the direction of the hazard ratio (HR). RFS curves of predictive biomarkers were generated using the Kaplan-Meier method from the survminer v0.5.0 [51] package and patients were categorized based on the median gene expression into "low" and "high".

Biomarker discovery

To identify predictive biomarkers associated with response to neoadjuvant Chemoradiotherapy (nCRT), outputs from differential expression analysis of both the combined dataset and meta-analysis (orP and REM) settings were systematically compared. A gene was identified as a predictive marker if it fulfilled the following criteria: differentially expressed genes (DEGs) with p-value <0.05 and adjusted *p*-value <0.25 in the combined dataset, REM p-value <0.05, and orP p-value< 0.05. The predictive efficacy of identified biomarkers was confirmed or denied through the analysis of RFS data as an independent validation, employing COX regression analysis with p value<0.05 and HR with the same trend as z value of REM and LFC in DESeq2 .

Immune cells deconvolution

Immune deconvolution analysis employing MCPcounter, ESTIMATE, xCelll, CIBERSORT and TIMER method from the immunedeconv package v2.1.0 [52] by using TPMs data prior batch effect correction were conducted. MCPcounter and TIMER facilitated the quantification of immune infiltration within tumor tissue, ESTIMATE was used for the assessment of tumor purity, as well as the immune and stromal score. xCell was utilized for the relative infiltration of 64 distinct types of immune and stromal cells, alongside CIBERSORT. Furthermore, in addition to the overrepresentation of specific immune subtypes statistical significance of the immune cells ratio across different treatment outcomes was investigated. During the ratio computation, all ratios that had the value NaN, Inf and 0 were classified as NAs. Ratios exceeding 80% NAs were excluded prior to statistical analysis. Ratios demonstrating a p-value of <0.05 and an adjusted p-value of <0.25 were deemed statistically significant. The statistical significance was evaluated employing the Mann-Whitney U test.

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Consensus Molecular Subtyping

To estimate CMS (Consensus Molecular Subtyping) distribution relative to different treatment outcome machine-learning based algorithms CMSclassifier v1.0.0 [5], CMScaller v2.0.1 [53] and CMSFFPE v0.1.0 [54](specifically developed for FFPE samples) were employed. To ensure the reliability of the results, algorithm outputs were only assessed for samples exhibiting comparable classifications across different algorithms. All methods reported both predicted and nearest classification, while CMSclassifier additionally used two methods, RF (random forest) and SS (single sample) predictor. The difference in distribution of CMS categories among treatment outcomes was evaluated by using the Chi-square/Fisher exact test statistics.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was conducted by using molecular signature database libraries C3 (Gene Ontology Biological Processes), KEGG, Canonical Pathways (Reactome subset) alongside MSigDB databases utilizing packages clusterProfiler v4.10.1[55] and org.Hs.eg.db v3.18.0 [56]. Pathways were deemed significant when exhibiting adjusted p-value<0.05, FDR<0.25 and NES>1.5 for those upregulated in pCR and R groups, and NES < -1.5 for those upregulated in Non pCR and NR groups.

Analysis of transcription factor activity

Transcription factors (TFs) target genes were retrieved from CollecTRI- derived regulons, composed of data from 12 different sources. An analysis of correlation between treatment response and transcription factor activity was conducted employing the decoupleR package v2.12.0 [57] which uses methods based on sign and weight of network interaction. Data input included DESeq2 statistical output, LogFC/se, derived from a combined dataset between R/NR and pCR/Non pCR. To determine the transcription factor activity score a univariate linear model (ULM) was used, while viper and norm_fgsea and norm_wmeans and norm_wsum served as additional validation chose based on Yashar et al results [58]. Transcriptional factors with statistically significant scores through all methods were considered as significant. In order to determine the involvement of activated transcription factors in signaling pathways, an overrepresentation analysis was performed using the Gene Ontology database of biological processes as a reference. Finally, the correlation of activated TFs and miRNAs, which proved to be potential therapeutic targets in rectal cancer, was investigated using multiMiR v1.24.0 package [59].

RESULTS

We assembled an RNAseq cohort of 248 publicly available transcriptomics data from patients diagnosed with rectal cancer categorized as pathological complete responders (pCR, n = 60) or non-complete pathological responders (Non pCR, n = 188), including 156 samples classified as responders (R, n = 92) and non-responders (NR, n = 64). Patients characteristics are shown in table 4. No differences were observed in the distribution of sex among response groups, as tested with Pearson Chi-square test and pCR/Non pCR groups. The cohort's age followed normal distribution among R/NR and pCR/Non pCR. A principal component analysis (PCA) plot showing the effectiveness of neutralizing the batch effect associated with the study variable is shown in Figure 3.

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Table 4. Patient characteristics stratified by therapy response

		Patients with pathological complete response pCR (%)	Patients with Non pathological complete response Non pCR (%)	Pearson Chi-square test *p*-value	Responders (R) (%)	Non-r (NR) (
Patients, n (%)		60 (24)	188 (76)		92 (59)	64 (41
Sex						
	Male	36 (60)	106 (56)		55 (60)	37 (58
	Female	19 (32)	70 (37)	*ns*	32 (35)	22 (34
	NA	5 (8)	12 (7l)		5 (5)	5 (8)
Therapy						
	50.4 Gy + 5-FU	34 (57)	93 (49)		44 (48)	29 (45
	50.4 Gy + Capecitabine	/	3 (2)		/	/
	50.4 Gy + 5-FU/ Capecitabine	5 (8)	12 (6)	*ns*	5 (5)	5 (8)
	50.4 Gy + 5-FU/Ox	4 (7)	22 (12)		4 (4)	/
	Consolidation CRT	6 (10)	11 (6)		10 (11)	4 (6)
	Induction CRT	11 (18)	47 (25)		29 (32)	26 (41
Age						
	Median	60	61	*ns*	58	56
	Range	32-86	26-82		32-86	26-82

Legend: ns- not significant, p value>0.05

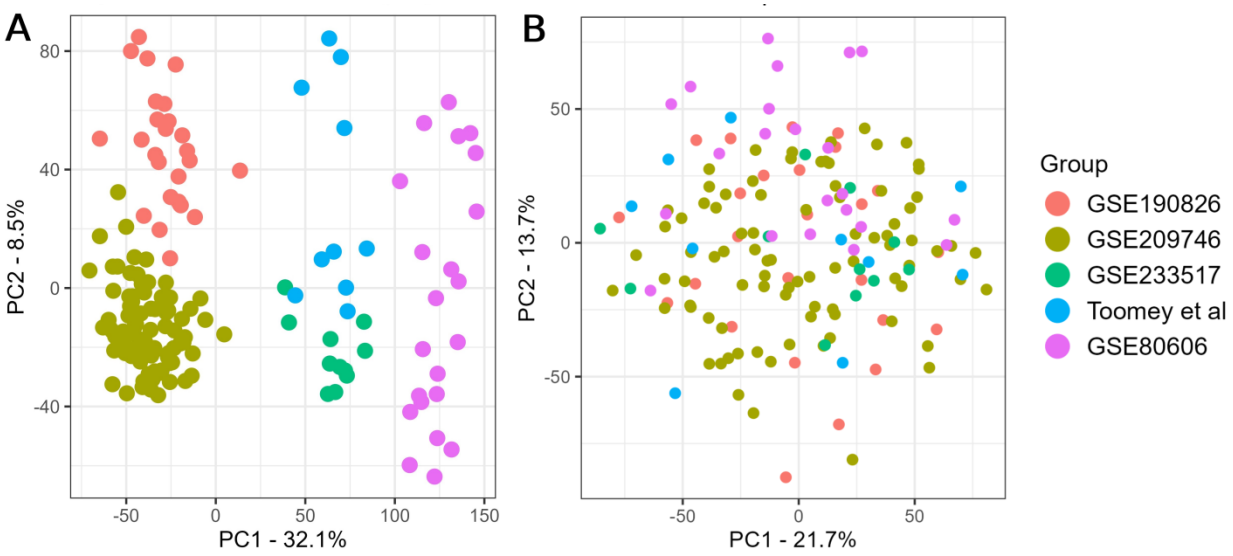


Figure 3. PCA plot before (A) and after (B) batch correction. PCA, Principal component analysis

Identification of predictive biomarkers of response to nCRT in LARC

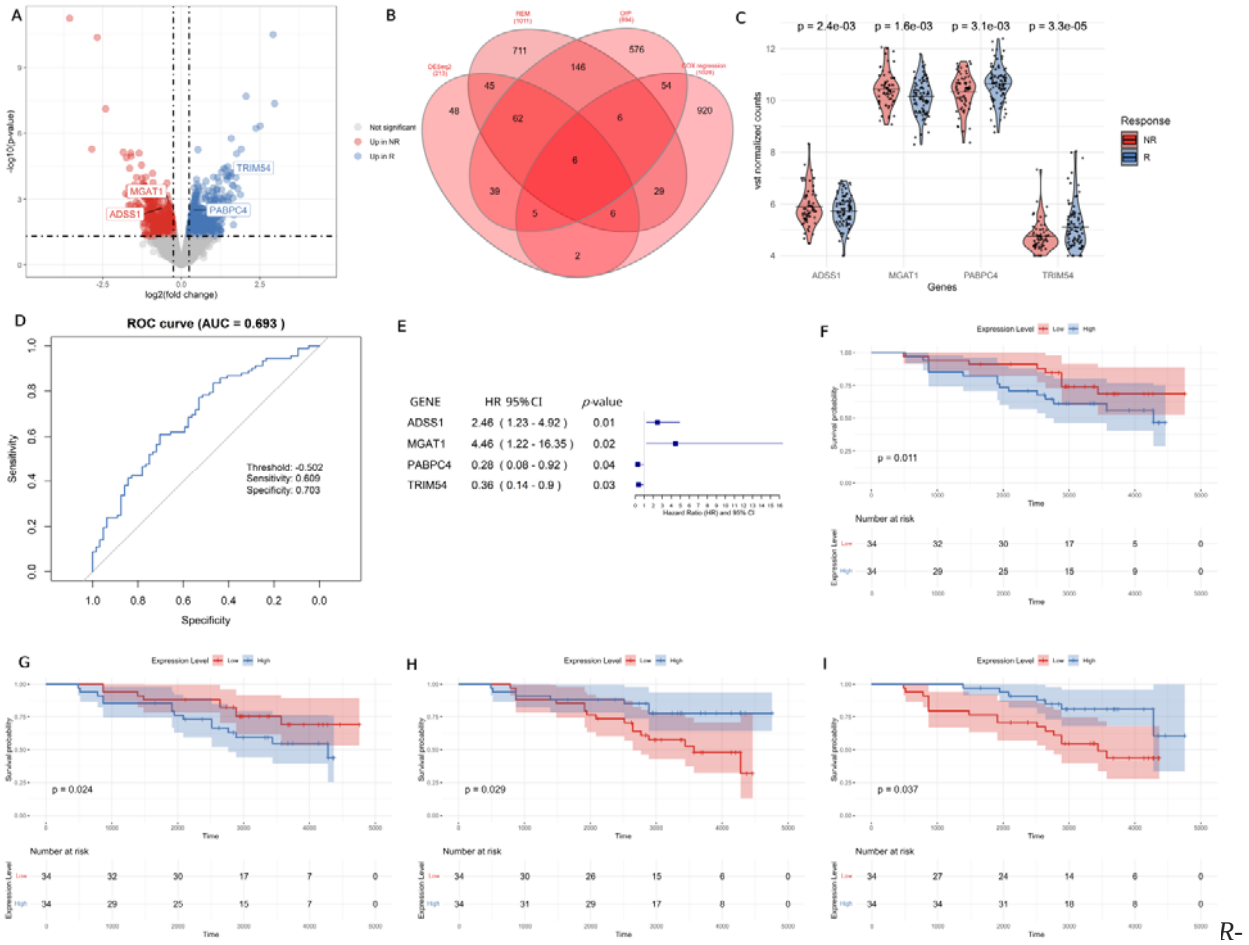
An initial screening for profiling gene expression alterations between response groups was performed both at the individual dataset level, and at the batch-free combined datasets. The combined cohort was composed of four individual datasets containing 17342 and 17510 common genes intersecting between all samples with available R/NR and pCR/Non pCR metadata. Results of differential expression analysis for the individual datasets are shown in Table 5. Figures 4 and 5 demonstrate results from the analysis performed using the combined cohort.

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Table 5. Results of individual studies and combined dataset

Dataset	Number of samples	Number of common genes	p < 0.05 and Log2FC < -1	p < 0.05 and Log2FC < 0	p < 0.05 and Log2FC > 1	p < 0.05 Log2FC
	R vs NR		*Up in NR*	*Up in NR*	*Up in R*	*Up in R*
GSE233517	11 (4 R, 7 NR)	37734	330	489	262	357
GSE209746	87 (48 R, 39 NR)	37734	219	1338	172	1532
GSE80606	22 (9 R, 13 NR)	37734	323	490	231	570
Toomey et al	10 (5 R, 5 NR)	33908	211	240	156	215
Combined dataset	156 (92 R, 64 NR)	17342	81	700	97	757
	pCR vs Non pCR		*Up in Non pCR*	*Up in Non pCR*	*Up in pCR*	*Up in pC*
GSE233517	11 (8 pCR, 3 Non pCR)	37734	233	427	212	276
GSE209746	93 (23 pCR, 70 Non pCR)	37734	195	1344	234	2208
GSE80606	22 (3 pCR, 19 Non pCR)	37734	2022	2599	479	921
Toomey et al	17 (5 pCR, 12 Non pCR)	33908	364	656	343	948
GSE190826	104 (26 pCR, 78 Non pCR)	34784	330	487	262	674
Combined dataset	248 (60 pCR, 188 Non pCR)	17510	15	329	71	1316

R- responder, NR- non-responder, pCR - pathological complete response, Non pCR-Pathological non-complete response, Log2FC- Log2 fold change

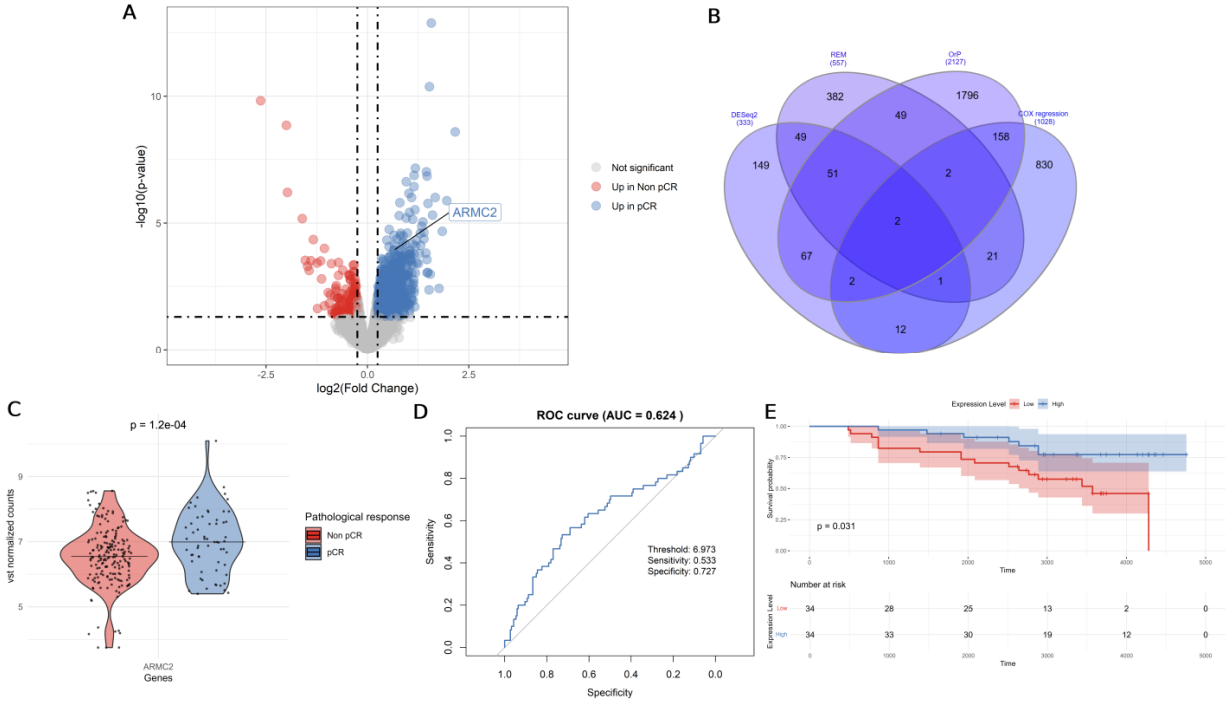


R-responder, NR- non-responder, REM- Random Effect Model, OrP- Order p value, HR- Hazard Ratio

Figure 4 Predictive biomarkers of poor/good treatment response A.Volcano plot of DEGs in R/NR. B. Venn Diagram [60] of REM, OrP, DESeq2 and COX regression significant results

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

C.Boxplot of gene expression identified as predictive markers of good response to therapy along with p values from the combined dataset. D. ROC curve obtained from VST normalized gene expression of predictive biomarkers of response to therapy. The combined score was obtained by adding the expression levels of genes overexpressed in good response to therapy, from which the expression level of genes overexpressed in poor therapeutic response was subtracted. E. Forest plot based on univariate COX regression analysis using RFS data and expression level of predictive biomarkers. F/G/H/I. Kaplan-Meier curve of mean expression of *ADSS1*, *MGAT1*, *TRIM54* and *PABPC4* across samples using RFS data. DEGs, differentially expressed genes.



pCR - pathological complete response, Non pCR-Pathological non-complete response, REM- Random Effect Model, OrP- Ordered p- value

Figure 5. Predictive biomarker of pathological complete treatment response A. Volcano plot of DEGs. B. Venn Diagram [60] of REM, OrP, DESeq2 and COX regression significant results C. Boxplot of gene expression identified as predictive markers of pathological complete response to therapy along with p values from the combined dataset. D. ROC curve obtained from VST normalized gene expression of predictive biomarkers of response to therapy. E. Kaplan-Meier curve of mean expression of *ARMC2* across samples using RFS data.

In order to discover predictive markers of response to neoadjuvant chemoradiotherapy we integrated the differential gene expression lists of the combined and the individual datasets using a combination of ordered p-value and a random effect model-based *on* rank aggregation (Supplementary Data 1). Combining studies increased group heterogeneity and statistical power while technical variation was a challenge. On the other hand, by applying meta-analysis, each study is analyzed separately, on the basis of which p-values and effect sizes are estimated for each gene [61]. By combining these two methods, the advantages of both were used in the identification of predictive biomarkers of response. In order to achieve the maximum number of

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

patients in the discovery cohort, we used a dataset for validation that did not contain information on response to therapy but did have information on time to disease recurrence after surgery. It was expected that patients who had a pathological complete response or a better response would have a longer period of relapse of the disease compared to patients in whom this response was absent. After the analysis of the results, 6 genes were identified that showed significance in the prediction of good and bad therapeutic response. Based on the prognostic potential of the identified genes, it was established that two genes have the potential of a good prognostic factor and are predictive markers of a good response (*TRIM54* and *PABPC4*), while two genes were concluded to be poor prognostic markers and predictors of a poor therapeutic response (*ADSS1* and *MGAT1*). The two markers had opposite predictive and prognostic effects so they were not identified as promising candidates (Table 6, Figure 4). Beyond the predictive biomarkers of R/NR, predictive biomarkers of pathological complete response were investigated. By using the same meta-analysis approach and combining dataset statistics together with RFS data as a source of independent validation two genes were identified as significant but only *ARMC2* were identified as predictive biomarker upregulated in patients with pathological complete response and good prognostic marker in all investigated methods (Table 7, Figure 5).

Table 6. Predictive biomarker of treatment response

		Discovery							Validation			
		Combined dataset			Random Effect Model			Ordered p value	Univariate COX regression analysis			
GENE ID	**Gene Name**	**Log2FC**	**p value**	**p adj**	**z value**	**95% (CI)**	**p value**	**p value**	**HR**	**95% CI**	**p value**	**Up in R/NR**
TRIM54	Tripartite motif-containing protein 54	1.478	$3.31x10^{-5}$	0.029	2.992	0.427-2.048	$2.77x10^{-3}$	$2.97x10^{-2}$	0.359	0.143-0.900	$2.89x10^{-2}$	Up in R
PABPC4	Poly(A) binding protein cytoplasmic 4	0.364	$3.07x10^{-3}$	0.249	2.917	0.067-0.340	$3.54x10^{-3}$	$4.00x10^{-2}$	0.280	0.085-0.924	$3.67x10^{-2}$	Up in R
MGAT1	Alpha-1,3-Mannosyl-Glycoprotein 2-Beta-N-Acetylglucosaminyltransferase	-0.383	$1.56x10^{-3}$	0.190	-3.138	-0.347 - -0.080	$1.70x10^{-3}$	$7.81x10^{-3}$	4.464	1.218-16.351	$2.39x10^{-2}$	Up in NR
ADSS1	Adenylosuccinate synthase 1	-0.596	$2.42x10^{-3}$	0.237	-3.232	-0.963- -0.236	$1.23x10^{-3}$	$3.93x10^{-2}$	2.457	1.228-4.918	$1.11x10^{-2}$	Up in NR

R- responder, NR- non-responder, HR- hazard ratio, CI-confidence interval, Log2FC- Log2 fold change, p adj- adjusted p value

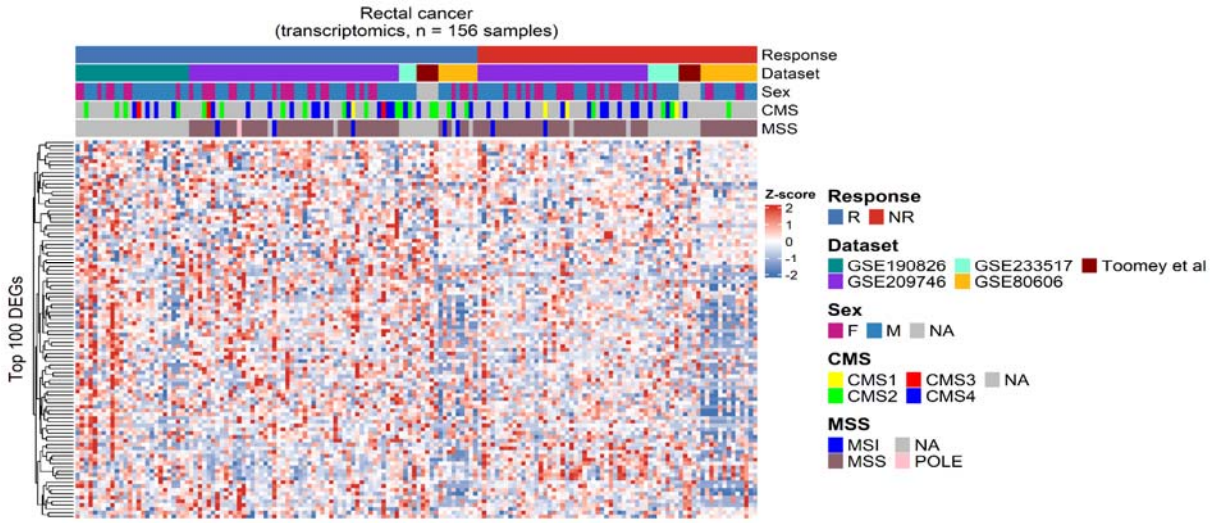
Table 7. Predictive biomarker of pathological complete treatment response

	Discovery			Validation	
	Combined dataset	Random Effect Model	Ordered p value	Univariate COX regression analysis	

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

GENE ID	Gene Name	LogFC	p value	p adj	z value	95% (CI)	p value	p value	HR	95% CI	p value	Up in pCR/Non pCR
ARMC2	Armadillo Repeat Containing 2	0.651	1.16 x10^{-4}	0.037	2.711	0.133-0.826	6.71 x10^{-3}	1.21 x10^{-2}	0.293	0.096-0.894	0.031	Up in pCR

pCR - pathological complete response, Non pCR-Pathological non-complete response, HR- hazard ratio, CI-confidence interval, Log2FC- Log2 fold change, p adj- adjusted p value



R- responder, NR- non-responder, MSI- mismatch instability, MSS- mismatch stability, NA- not applicable, F-female, M-male, CMS-consensus molecular subtype, POLE- DNA polymerases □, DEGs- differentially expressed genes

Figure 6. A. Heatmap of Top100 DEGs between R/NR.

Investigation of expression patterns of the top 100 DEGs showed difficulties in predictive capacity concerning heatmap. Covariates such as sex, CMS and microsatellite stability/instability (MSS/MSI) were employed (Figure 6). Further investigation using immune deconvolution methods, gene set enrichment analysis and molecular classification was performed in order to understand underlying mechanisms which effect will be seen further on phenotypic profile.

Variability in immune landscape as predictor of therapy response

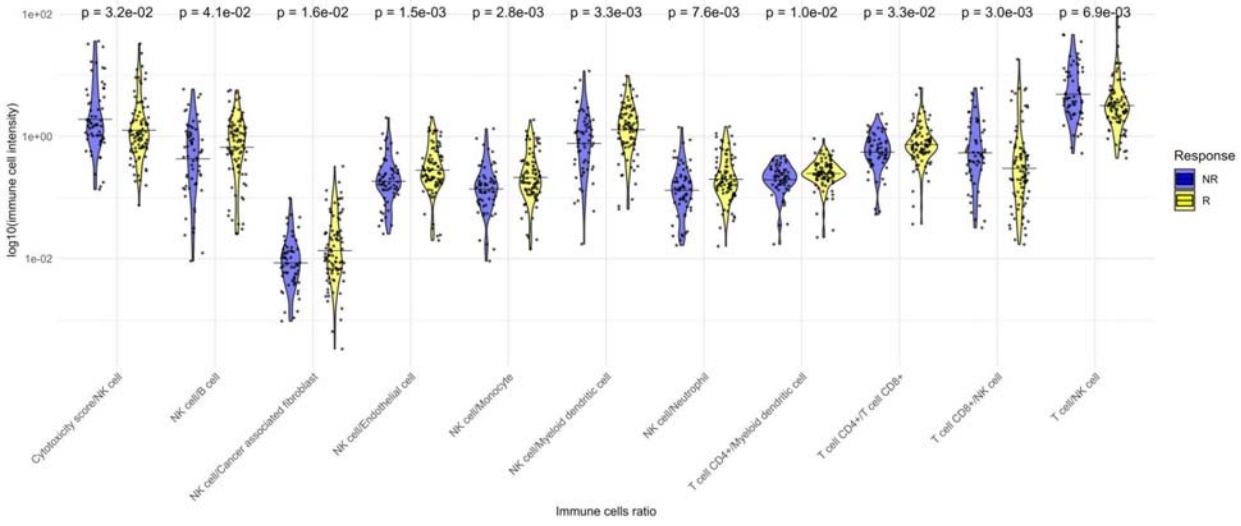
Immune cell deconvolution from bulk RNA sequencing data was executed utilizing various deconvolution methodologies, including CIBERSORT, xCell, TRIMER, and MCP-counter. The immune deconvolution approaches revealed discrepancies between responders and nonresponders regarding the presentation of immune cells prior to therapy, suggesting activation of the immune system in both protective and suppressive modalities. Notably, the responder cohort exhibited a markedly higher proportion of CD4+ T cells, monocytes, and resting NK cells, thereby enhancing the antitumor environment and facilitating improved treatment outcomes (Table 8). Furthermore, overrepresentation of macrophages, activated myeloid dendritic cells, hematopoietic stem cells, NK T cells, and common lymphoid progenitors were common among patients experiencing poor treatment outcome [62]. However, employed immune deconvolution methods did not yield identifiable features that would enable prediction of pathological complete response .

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Table 8. Immune subtypes and immune cell ratio with predictive R/NR capacity

Immune cell subtype	Deconvolution Method	p value	p adjusted	Log2 Fold Change	Up in R/NR
Myeloid dendritic cell activated	xCell	0.034	0.239	-0.039	Up in NR
T cell CD4+ (non-regulatory)	xCell	0.030	0.239	2.84E-04	Up in R
Common lymphoid progenitor	xCell	0.004	0.070	-0.062	Up in NR
Hematopoietic stem cell	xCell	0.037	0.239	-0.026	Up in NR
Macrophage	xCell	0.022	0.239	-0.008	Up in NR
T cell NK	xCell	0.001	0.021	-0.075	Up in NR
T cell CD4+	TIMER	0.006	0.037	0.039	Up in R
NK cell	MCP counter	0.001	0.009	1.555	Up in R
T cell CD4+/Myeloid dendritic cell	TIMER	0.010	0.153	0.056	Up in R
T cell CD4+/T cell CD8+	TIMER	0.033	0.248	0.278	Up in R
NK cells resting/Macrophages M1	CIBERSORT	0.001	0.127	2.168	Up in R
NK cell/Endothelial cell	MCP counter	0.002	0.035	0.130	Up in R
NK cell/Monocyte	MCP counter	0.003	0.035	0.114	Up in R
NK cell/Macrophage/Monocyte	MCP counter	0.003	0.035	0.114	Up in R
T cell CD8+/NK cell	MCP counter	0.003	0.035	-0.057	Up in NR
NK cell/Myeloid dendritic cell	MCP counter	0.003	0.035	0.459	Up in R
T cell/NK cell	MCP counter	0.007	0.059	-2.379	Up in NR
NK cell/Neutrophil	MCP counter	0.008	0.059	0.080	Up in R
NK cell/Cancer associated fibroblast	MCP counter	0.016	0.109	0.016	Up in R
Cytotoxicity score/NK cell	MCP counter	0.032	0.195	-1.832	Up in NR
NK cell/B cell	MCP counter	0.041	0.222	0.229	Up in R

Furthermore, ratios of various immune cells were examined. An increased quantity of NK cells, in comparison to other immune cells, correlates with favorable treatment outcomes, including Endothelial cells, Monocytes, Myeloid dendritic cells, Neutrophils, Cancer-associated fibroblasts, B cells, and resting NK cells/Macrophages. The overrepresentation of CD4+ T cells relative to Myeloid dendritic cells and CD8+ T cells also contributes to improved treatment outcomes. Conversely, poor treatment outcomes were characterized by an overrepresentation of T cells, CD8+ T cells, and cytotoxicity score over NK cells (Table 8, Figure 7)**.**



medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Figure 7. Log 10 Immune cell intensity of significant immune cell ratio between R and NR

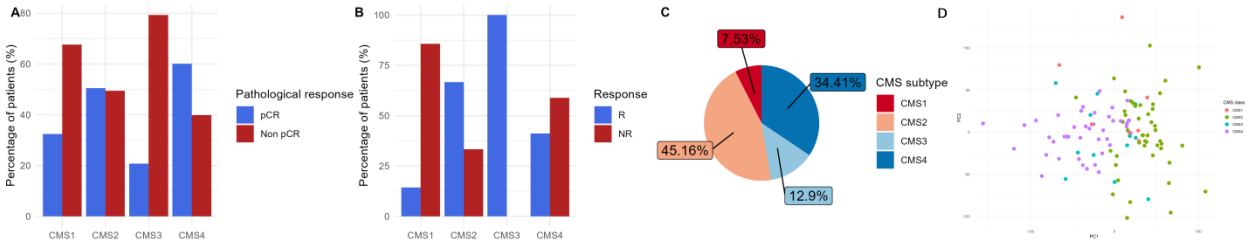
Molecular subtype-based characterization of tumor response heterogeneity

Different algorithms were used for CMS sample classification. The results were compared and only samples that had consistency in classification between different algorithms were used further for analysis. 51 R/NR and 93 pCR/Non pCR CMS categories were successfully identified (Supplementary Data 2). There is a trend towards a difference in CMS distribution between R and NR (p=0.077). The CMS2 and CMS4 subtypes (Table 7) had higher prevalence R/NR and pCR/Non pCR than the CMS1 and CMS3 subtypes. The PCA plot clearly demonstrates distinct molecular clustering for CMS2 and CMS4 based on transcriptomic profiles (Figure 8D). Patients within the NR category were predominantly classified as CMS4, whereas those in the R category were categorized as either CMS2 or CMS4. Among other CMS categories CMS1 were dominantly characterized as NR, while none of NR were classified as CMS3 (Table 9, Figure 8). Considering pathological complete response most of the patients were identified as CMS2 or 4. Non pCR were dominant over pCR in the CMS3 subtype (Table 9, Figure 8).

Table 9. Distribution of samples among CMS subtypes

	R	NR	pCR	Non pCR
Total	34 (100)	17 (100)	24 (100)	69 (100)
CMS1	1 (2.94)	3 (17.65)	1 (4.17)	6 (8.70)
CMS2	16 (47.06)	4 (23.53)	11 (45.83)	31 (44.93)
CMS3	3 (8.82)	0 (0.00)	1 (4.17)	11 (15.94)
CMS4	14 (41.18)	10 (58.82)	11 (45.83)	21 (30.43)

R- responder, NR- non-responder, pCR - pathological complete response, Non pCR-Pathological non-complete response, CMS- consensus molecular subtype



R- responder, NR- non-responder, pCR - pathological complete response, Non pCR-Non- pathological complete response, CMS- consensus molecular subtype

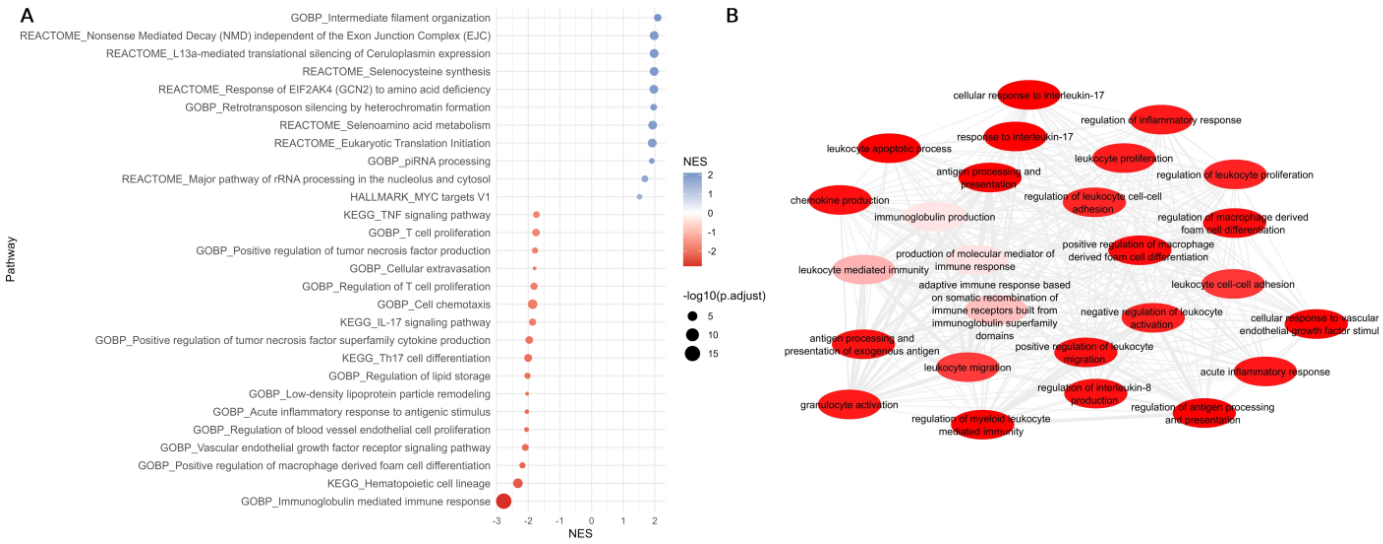
Figure 8. CMS distribution A. and B. Percentage of patients per CMS category scaled by the total number of patients per group (pCR/Non pCR and R/NR). C. CMS subtype distribution of rectal cancer samples. D. PCA plot of transcriptomic data based on different CMS

Gene Set Enrichment Analysis highlights differentially activated pathways

Various signaling pathways were enriched in the nonresponder group, primarily related to immune response, necrosis, cell metastasis, movement through extracellular matrix, angiogenesis and vessels development, lipids and lipoprotein metabolism, signal transduction, and many bacterial/viral infection responses (Figure 9). The NR group was characterized by a very complex list of immune-related pathways and Revigo (http://revigo.irb.hr/) [63] was utilized to condense these pathways by eliminating redundant GO terms together with Cytoscape [64] (Figure

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

9B). Patients with favourable treatment response showed enrichment in signaling pathways related to cytoskeleton organization, control of transcription, translation, signaling pathways that reduce the occurrence of errors, control of ribosomal RNA processing, translational imprinting, as well as pathways related to piRNA processing, transposons and retrotransposons (Supplementary Data 3, Figure 9). Enriched pathways among patients with pathological complete response showed activation of immune related signaling pathways particularly those related to the activity of B cells and NK cells. The gene set enrichment analysis highlighted the significance of protein synthesis and translation control in patients with Non pCR. This observation is supported by the fact that aside from protein folding control and translation control, enriched signaling pathways include many of those related to energy metabolism. Furthermore, many signaling pathways related to protein synthesis include pathways and processes in endoplasmic reticulum that are usually producing membrane bound proteins as well as proteins that are part of lysosomes, endosomes and secretory vesicles (Supplementary Data 4).



NES- normalized enrichment score

Figure 9. Representative signaling pathways differentially activated between R and NR. A. Selected pathway from Gene Set Enrichment analysis upregulated in R (NES>1.5, blue) and NR (NES< -1.5, red). B. Overview of Immune related pathways enriched in NR group

Identification of key transcription factors of therapeutic sensitivity and resistance

Since transcription factors can regulate a large number of genes and in that way affect various biological processes, the subsequent phase of the analysis involved a comprehensive examination of transcription factors (TFs). The regulatory functions of TFs exhibit both positive and negative impacts within the responder (R) and non-responder (NR) groups. As illustrated in the volcano plot (Figure 10D,E,F), the number of genes regulated in favor of a particular outcome relative to others dictates their overall significance score. The ULM algorithm was employed to determine the activity of transcription factors based on therapeutic response (Figure 10A), while additional algorithms VIPER, wsum, wmean, and FGSEA were used to validate the obtained findings. A

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

total of 104 TFs were identified as statistically significant (p-value ≤ 0.05) within our studied cohort, which includes 6 enriched in R and 98 in NR. Out of all identified significant TFs, 42 of them were validated by incorporating additional methodologies for evaluating transcription factor activity (wmean, wsum, fgsea, and viper) (Figure 10A/B/C) (Supplementary Data 5). Among NR, the highest statistical significance was found for *SP1* (Figure 11D) and *NFKB* (Figure 10E) while *TCF15* (Figure 10R) was dominant among R.

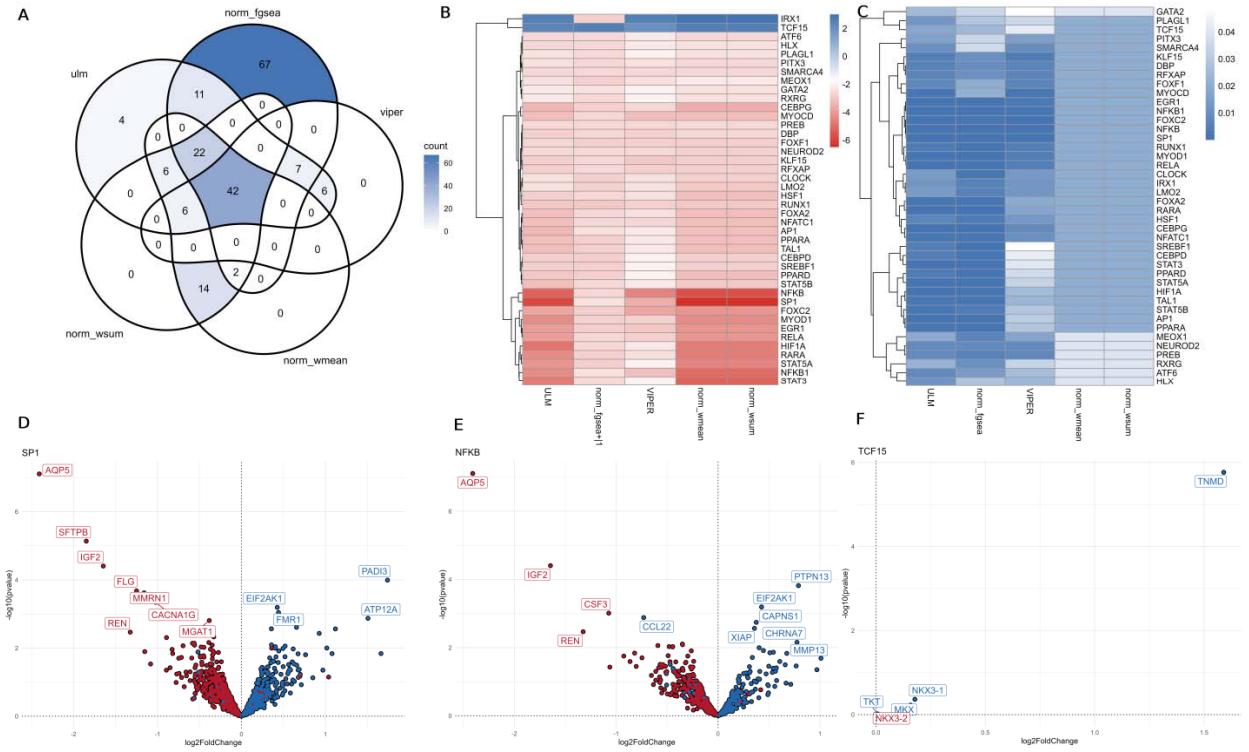


Figure 10. Results of TFs activity among R/NR groups A. Cross section of transcription factors identified by ULM, wmax,wmean, VIPER and FGSEA within R/NR group. B. Score and C. p value of confirmed transcription factors. D/E/F. Volcano plot of DEGs (log2FoldChange vs -log10 P value) with top TFs (D. *SP1*, E. *NFKB* and F. *TCF15*) target genes colored based on TFs effect on them (red is activating effect and blue is suppressive effect)

An analysis of transcription factors' activity within the cohort pCR/Non pCR identified 93 TFs with statistically significant activity; of these, 72 TFs had increased activity toward pathological complete response. Through the incorporation of supplementary validation methods, a total of 29 transcription factors were confirmed (Supplementary Data 5, Figure 11A/B/C), 26 of which were activated in favor of a pathological complete response . *SMAD3* and *RFXANK* (Figure 11 F/G) had the most significant effects in favor of pCR. While the most notable effect in favor of Non pCR had *MYC* (Figure 11E). TF upregulated in pCR had a strong specificity toward activating genes upregulated in pCR.

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

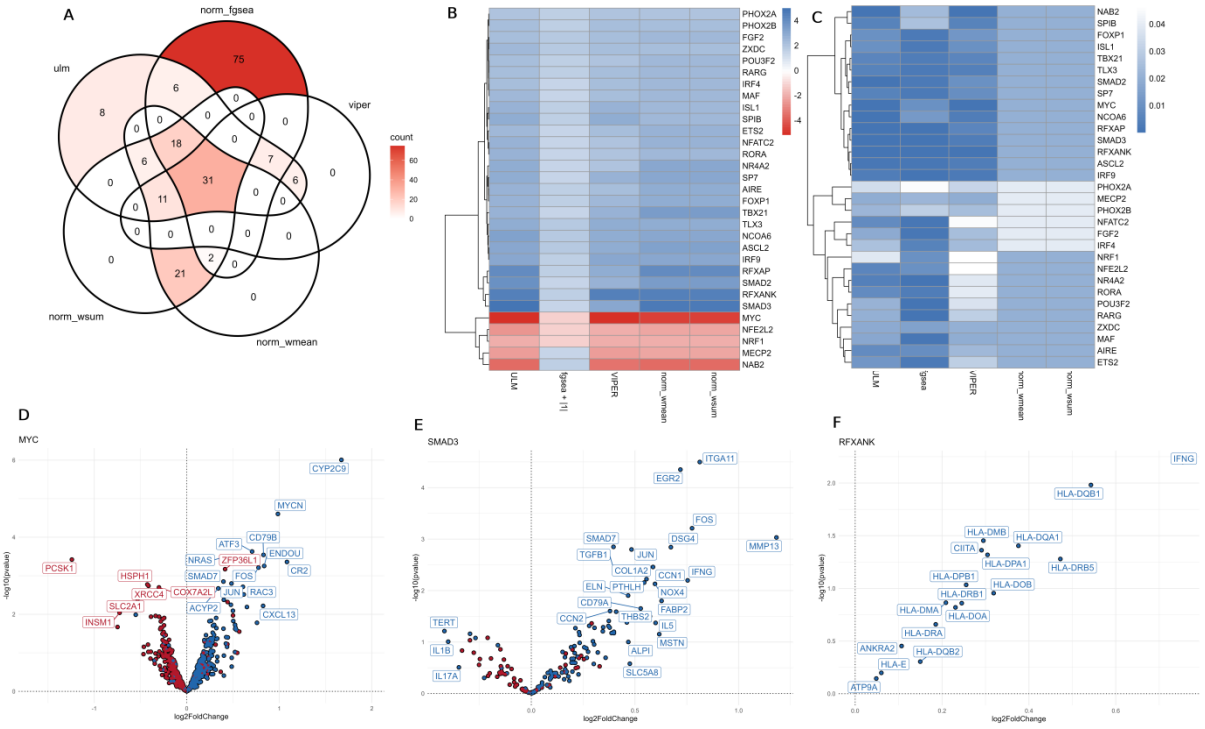


Figure 11. Results of TFs activity among pCR/Non pCR groups A. Cross section of transcription factors identified by ULM, wmax, wmean, VIPER and FGSEA B/C. Score (B) and log10 p-value (C) of overlapped transcription factors. E/F/G. Volcano plot of DEGs (log2FoldChange vs -log10 P value) with top TFs based on score (D. *MYC*, E. *SMAD3*, F. *RFXANK*); target genes colored based on TFs effect on them (red is activating effect and blue is suppressive effect).

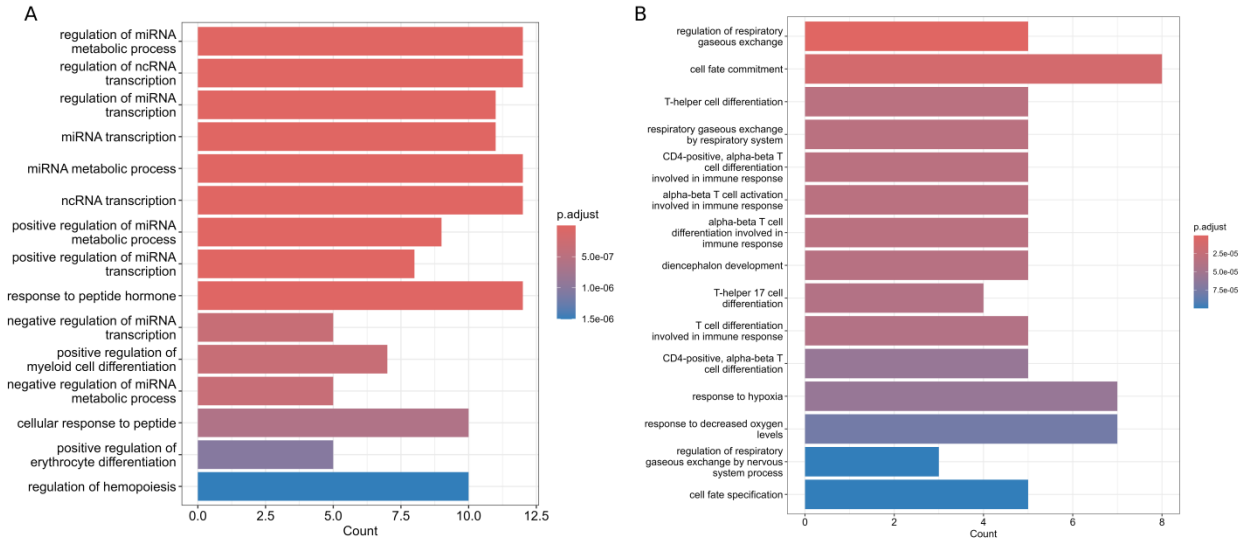


Figure 12. Top 15 signaling pathways overrepresented among TFs differentially activated between R and NR (A) and pCR and Non pCR (B)

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Gene Ontology database was used as a source of signaling pathways and genes involved, used to identify overrepresented pathways among confirmed TFs. Analysis of TFs activated in NR highlights the importance of signaling pathways related to miRNA and ncRNA metabolism and activity (Figure 12A). Furthermore, TFs activated in pCR were part of signaling pathways related to immune response, particularly, CD4+ T cell activation, Th17 immune response and alpha-beta T-cells (Figure 12B).

The review article by Imedio et al. [65] provided an overview of miRNAs identified as possible therapeutic targets in rectal cancer. List of miRNAs includes miR-450-5p[66], miR-577[67], miR-375-3p[68], miR-139-5p [69] with therapeutic impact on chemotherapy, miR-130a[70], miR-145[71], miR-122-5p [72] with therapeutic impact on radiotherapy as well as miR-31[73], miR-129-5p[74], miR-144 [75], miR-381[76] and miR 19a-3p[77]. A list of potential targets was used to detect interactions with significantly activated transcription factors. Numerous transcription factors that have been demonstrated to be significant in response to therapy interact with the majority of the highlighted miRNAs (Figure 13). The exact direction of regulation is not known, but the results indicate a clear connection between the activation of transcription factors in R/NR and the regulation of miRNAs, whose potential as therapeutic targets has been reported [65–69,72,74,77].

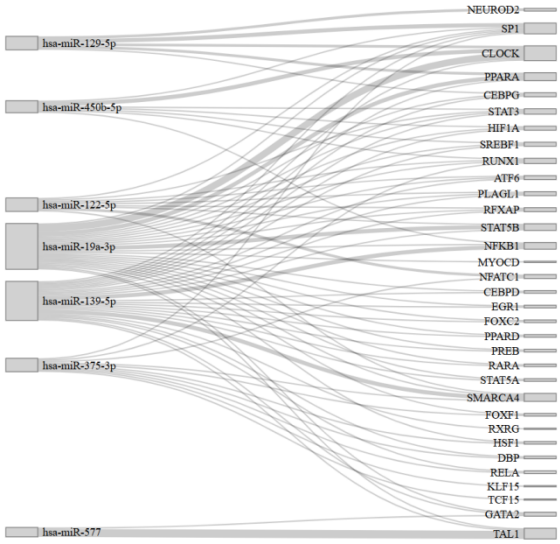


Figure 13. Activated TFs in NR and their targets, miRNAs that have been shown as predictive markers of response to nCRT

DISCUSSION

Locally advanced rectal cancer is one of the most commonly diagnosed forms of rectal cancer, with 10–20% of patients achieving a pathological complete response following standard nCRT [8]. Routine clinical practice and currently available methodologies are insufficient for identifying these patients and providing them with optimized care. Nonetheless, a sizable portion of patients experience suboptimal therapeutic outcomes. This issue is even more critical considering that recent modifications in neoadjuvant treatment — mainly the introduction of additional chemotherapy in the neoadjuvant setting as part of the total neoadjuvant therapy (TNT) approach, and the escalation of radiation doses — are resulting in a greater number of patients

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

achieving a clinical complete response[18–20,22]. Nowadays, in light of these changes in LARC treatment, it is crucial to avoid both overtreatment and undertreatment in order to achieve the best treatment outcomes while preserving patients' quality of life.

Currently, there are no clinically confirmed predictive markers for response to nCRT, and the precise criteria for identifying patients eligible for the *watch-and-wait* strategy remain unclear. The rationale for this phenomenon is attributed to the often limited patient cohort in studies, alongside the variation in treatment of LARC patients concerning radiation dosage and administration techniques. Furthermore, it is common for available published studies to lack supplementary clinical data and additional omics levels that would facilitate the establishment of a biomarker panel at appropriate levels via a multi-omics approach. Additionally, a significant challenge arises from the population specificity of the identified biomarkers, as their validation across diverse populations is notably absent [78–80]. Meta-analysis of transcriptional profiles of LARC have provided insight into predictive biomarkers and biological features associated with different treatment outcomes. Combining individual studies revealed higher sample heterogeneity due to different populations and increased statistical power [61]. It is thus important to synthesize the knowledge obtained from available studies and to proceed to a prospective, multicenter, multinational validation study.

Among the genes examined, two genes were identified with predictive potential of good response to therapy including *TRIM54* and *PABPC4*, up-regulated in responder group, while *ADSS1* and *MGAT1*, up-regulated in non-responder group were identified as predictive biomarkers of poor treatment outcome (Figure 4, Table 5). Notably, these genes have not been previously analyzed as predictive biomarkers for response to neoadjuvant chemoradiotherapy (nCRT) in rectal cancer; however, positive effects of their expression have been documented in other malignancies including colorectal cancer and some of them were recognized as prognostic biomarkers according The Human Protein Atlas (HPA) [81].

TRIM54 functions as ubiquitinase associated with FSP1, key ferroptosis regulator, and promotes its ubiquitination-mediated degradation [82]. *TRIM54* is gene coding inhibitor of *FSP1* thus could promote ferroptosis in rectal cancer and increase cell death rate during radiotherapy resulting in better treatment outcome. The role of TRIM54 as a direct regulator of ferroptosis was described in hepatocellular carcinoma (HCC), where it was shown that Sorafenib modulates FSP1 activity through TRIM54, ultimately inducing ferroptosis [82]. Although the mechanism of action of TRIM54 is poorly investigated one study on HCC reported that overexpression of TRIM54 lead to Axin1-mediated activation of the Wnt/β-catenin signaling pathway [83], which is known, poor prognostic sign in colorectal cancer and is associated with more aggressive tumors [84]. Although shown to be a poor prognostic factor, *TRIM54* expression in the context of enhancing response to chemotherapy or radiotherapy has not yet been investigated. *PABPC4*, gene coding the poly(A) binding protein cytoplasmic 4, is important RNA-binding protein with role of regulating gene expression. It was not investigated as a predictive biomarker but its role in better survival of CRC patients stage II and stage III was reported [85]. According to the HPA data in the READ validation dataset, it was shown that LARC patients with higher expression have better survival (p=0.036) [81]. MGAT1 protein coded by *MGAT1* gene, called *mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase* is an enzyme involved in the process of N-glycosylation, and according to HPA [81], it is a poor prognostic biomarker in stage II and III rectal cancer, which is consistent with the obtained results. *ADSS1*, adenylosuccinate synthase 1, protein product has a role in AMP synthesis thus highly energy-consuming cells like cancer cells

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

can use it to proliferate effectively [86]. ADSS1 is highly abundant in cardiac and heart tissue, controlling purine de novo biosynthesis but it has low cancer expression level [81]. According to HPA data, *ADSS1* expression level correlates with long-term survival over 3 years in patients with stage II and III rectal cancer. 5-year survival in the group of patients with high expression is 29% vs. 53% [81]. Furthermore, predictive biomarkers of pathological complete response genes were investigated, resulting in the identification of *ARMC2*, up-regulated in pCR (Figure 5, Table 6). *ARMC2*, Armadillo repeat containing 2, has a protein product with a role in maintenance of the axonemal central pair complex which is essential for the functionality of cells with cilia[87]. HPA data from both TCGA (p=0.026) and Validation cohort (p=3.3 × 10□□) confirmed that *ARMC2* has a prognostic role in rectal cancer stage II/III [81]. Each of these genes cannot be considered as perfect biomarker but in combination with other monitoring methods including hematological parameters, clinical data, radiomics features or biomarkers on proteomic level, it could be complementary and thus contribute to better stratification of patients for watch-and-wait approach, better local disease control, organ preservation and offer comprehensive disease monitoring.

The examination of phenotypic characteristics in relation to the transcriptional profile is crucial for elucidating the mechanism underlying the response to neoadjuvant chemoradiotherapy (nCRT). Additionally, certain characteristics derived from bioinformatics algorithms may serve as predictive biomarkers of response, subject to validation through molecular genetics methods that would allow their direct measurement [88–90]. Phenotypic profiling has enabled the identification of immune cell subtypes, the activity of distinct signaling pathways, and the functionality of transcription factors, which vary according to therapeutic response.

Immune deconvolution methods demonstrated that the responder group had a significantly higher proportion of CD4+T cells, monocytes and NK cells which affect the tumor microenvironment and facilitate improved treatment outcomes (Table 8). Existing literature indicates that NK cells are triggered by radiation therapy thus their increased presence in the tumor microenvironment plays a crucial role in improved treatment outcomes and prolonged survival [91]. Previously conducted *in vitro* and *in vivo* studies highlighted that ionizing radiation increases expression of stress-induced NKG2D ligands like MICA/B and ULBP1–3, making tumor cells more susceptible to NK-cell–mediated killing. Radiotherapy can synergize with NK-based immune responses leading to improved antitumor defense[92,93]. Analyzing the immune cell ratio confirmed that the presence of NK cells in relation to Macrophages M1, endothelial cells, monocytes and macrophages, dendritic cells, B cells, neutrophils and cancer-associated fibroblasts had predictive potential of a good response to therapy in LARC patients. The increased level of CD4+ T lymphocytes relative to myeloid dendritic cells and CD8+ T lymphocytes also had the same effect (Table 8, Figure 7).

Furthermore, available results showed that macrophages, myeloid dendritic cells, hematopoietic stem cells and common lymphoid progenitors are overrepresented within the group of patients with unfavourable treatment outcome (Table 8) [94]. Most of the cells dominant in patients with a poor therapeutic outcome are precursors in the differentiation of immune cells with specific proinflammatory and tumor suppressor functions. The presence of myeloid progenitor cells and generally immune cell precursors in the circulation and at the tumor site was described as early as the early 2000s, although this is not expected under normal conditions. These cells are thought to have a high potential in suppressing the immune system. It has been shown that these cells can produce anti-inflammatory cytokines including IL-10 and TGFβ and thus reduce the infiltration

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

of T lymphocytes [95]. In addition, examination of the ratio of immune cells in patients with a poor therapeutic outcome showed an increased presence of CD8+ T lymphocytes compared to natural killer cells (Table 8, Figure 7). As a consequence of the effect of chemoradiotherapy, there may be a process of reducing the expression of MHC class I molecules, which leads to the inability to activate CD8+ T lymphocytes and thus the inability to attack tumor cells, making them ineffective. Together with the decrease in the amount of NK cells, the possibility of anti-tumor defense is absent, which can lead to a poor outcome of the therapy [96].

Patients who had favourable response to therapy had enriched signaling pathways related to cytoskeleton organization, transcriptional and translational control, signaling pathways that reduce error occurrence, regulation of ribosomal RNA processing, translational imprinting, alongside pathways related to piRNA processing, transposons and retrotransposons (Figure 9). The enriched signaling pathways imply that treatment-naive tissue has a specific phenotypic profile, resulting in variable responses to therapy among patients. Furthermore, the responder group is predominantly characterized as CMS2 and CMS4, where CMS2 is associated with a proliferative environment and MYC and WNT activation which is in direct correlation with the high transcriptional rate within the cells (Table 9, Figure 8) [97]. The gene set enrichment analysis identifies signaling pathways enriched within non-responder groups related to immune response, necrosis, cell metastasis and movement through extracellular matrix, angiogenesis and vessels development, lipids and lipoprotein metabolism as well as signal transduction (Figure 9). Collectively, these findings suggest a highly active microenvironment, which correlates with the observation that a majority of patients in the non-responder group are categorized as CMS4, recognized as the most aggressive CMS subtype with enhanced epithelial-mesenchymal transition (EMT) activation, angiogenesis, and matrix remodeling (Table 9, Figure 8) [97,98].

Enriched pathways among pCR groups showed activation of immune related signaling pathways especially those related to activity of B cells and NK cells. Patients with pCR are mainly distributed among two molecular categories CMS4 and CMS2. Gene set enrichment analysis highlighted the importance of protein synthesis and translation control in patients with Non pCR. Activation of signaling pathways related to protein synthesis indicates that the key in difference in treatment response could be on the protein level. Controlling translation and protein degradation, cancer cells can promote oncogenic phenotype through selective translation of mRNA transcripts that can be crucial for tumorigenesis, therapy resistance and metastasis, or degrading transcripts of mRNA that protein product would promote cell death [99]. This is supported by the fact that apart from protein folding control, translation control, enriched signalling pathways include many of them related to energy metabolism. Bearing in mind that protein-synthesis is the most energy consuming process, which is supported by our results, showing that rectal cancer cells in patients with non pathological response have a high rate of energy production that is needed for translation control. In addition, our data support that many alterations in protein synthesis are mainly related to proteins processed at endoplasmic reticulum, proteins that are usually membrane bound, as well as part of lysosomes, endosomes and secretory vesicles. Those proteins could act as mediators to the tumor microenvironment supporting the protective role of cancer cells, but also could be used as a good secretory biomarker or drug target. Molecular classification of patients with Non pCR indicated that most of them are classified as CMS2 and then CMS4, groups that are classified with high proliferative rate and EMT and angiogenesis, followed by CMS3, with highest prevalence of *KRAS* mutations and metabolic deregulation (Table 9, Figure 8) [97]. EMT has been recognized as a key

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

mechanism contributing to resistance to chemoradiotherapy (CRT) in various cancers, including rectal cancer [100,101].

Transcription factor activity analysis identified 42 transcription factors exhibiting alterations in responders/non-responders, confirmed by various models. Prominent transcription factors highlighted the importance of immune response modulation, inflammation in patients with poor therapeutic response (Figure 12). Furthermore, the identified transcription factors are involved in the processes of proliferation, differentiation, cellular response to stress, hypoxia, DNA damage, oncogenesis and tumor growth. Among these, *NF-KB* and *SP1* demonstrated the highest score enriched in the NR group and *RCF15* in the R group (Figure 10). It is shown that TFs *NFKB* and *STAT3* are key factors that enable cancer cells to avoid apoptosis and regulate angiogenesis and invasiveness [102]. The activation of these two transcription factors is critical in linking immunity and tumorigenesis, inducing the expression of proteins associated with angiogenesis, hypoxia, chemokines, and immunosuppressive cytokines [102]. NF-KB promotes anti-apoptotic signaling (e.g., upregulation of Bcl-2, Survivin), promotes radioresistance through the STING–IFNAR1 axis in colorectal cancer, promotes immune evasion and tumor-promoting inflammation [103,104]. In research analyzing the effects of FoxO3, a transcription factor regulated by SP1, findings indicated that SP1 enhances FoxO expression which contributes to CRC progression, while knockdown of FoxO3 resulting in lower activity of SP1 lead to reduced cell migration and proliferation ability [105]. Studies have shown that miR-375-3p can target SP1 and thereby increase the sensitivity of CRC cells to 5-FU-based chemotherapy [106]. Overrepresentation analysis highlighted involvement of activated TFs in signaling pathways related to miRNA and ncRNA activation and metabolism. Literature data highlight that miRNAs can be used as therapeutic targets. Our results indicate that miR-129-5p, miR-450b-5p, miR-122-5p, miR-19a-3p, miR-139-5p, miR-375-3p as potential therapeutic targets are regulated by transcription factors activated in patients with different therapeutic responses (Figure 13) [65,107–111].

The analysis of transcription factors activity within the pCR/Non pCR cohort identified 29 transcription factors mostly enriched within the pCR group. *SMAD3* and *RFXANK* had the most significant effects in favor of pCR. While the most notable effect in favor of Non pCR had *MYC* (Figure 11). *RDXANK* is a key regulator of MHC II complex expression that is required for antigen presentation and activation of CD4+ T lymphocytes [112]. Although the representation of immune subtypes in patients with complete regression was not significant when it comes to CD4+ T lymphocytes, they were noted to be differentially activated in patients with a better therapeutic outcome. This confirms the positive effect of direct or indirect activity of CD4+ T lymphocytes in response to therapy. *MYC* is a known oncogene and regulator of cell proliferation as well as a key regulator of cell growth while increased activation of *SMAD3* by TGFβ allows inhibition of *MYC* transcription leading to inhibition of proliferation, differentiation and cell death [113]. Confirmed response modulators are involved in immune response, cell differentiation according to overrepresentation analysis (Figure 12). The modulation of transcription factor activity presents a valuable avenue for in vitro research with potential translational implications. Transcription factors regulate numerous target genes, rendering expression mimicry challenging; however, enhancing expression in patients with phenotypes predisposed to poor therapeutic responses may significantly improve treatment outcomes, thereby contributing substantially to therapeutic efficacy.

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Our research group has previously investigated various levels of omics methodologies with the objective of predicting response to neoadjuvant chemoradiotherapy in LARC. Proteomic analysis using DIA-MS/MS [33], initial level of hematological parameters [34] and a machine learning model based on radiomics of pretreatment magnetic resonance imaging (MRI) 3D T2W contrast sequence scans in conjunction with clinical parameters [35] revealed moderate to high predictive potential in term of response to nCRT in LARC patients. The results shown in our studies as well as the results shown in other studies at other molecular and clinical levels indicate that one approach is probably not enough to effectively predict the response to therapy or pre select patients for a watch-and-wait approach. Further studies are necessary to include well-defined cohorts with detailed clinical data and well-molecularly profiled samples (genomic, transcriptomic, proteomic level) with radiomics data, hematological and pathological characteristics that would be used together through machine learning methods to construct a predictive model. In this way, each of the mentioned factors that has the potential of prediction would be additionally supported by other predictive parameters.

Although meta-analysis methods can elucidate significant biological factors that influence different phenotypes, they have limitations. The neutralization of the batch effect in order to reduce the technical variability or unavailability of raw counts undoubtedly affects the biological variability, which is not fully observed by any existing method. Therefore, the results obtained through meta-analyses should be further validated by *in vitro, in vivo and ex vivo* experiments. Additionally, FFPE or tissue biopsy represents only a small portion of the tumor microenvironment, even though it is a gold standard in tumor diagnosis. Furthermore, as the tumor cells from LARC can invade the muscularis propria, which also secretes biological molecules and is not accessible during routine diagnostic analysis, we need to approach other body fluids as complementary sources of information.

CONCLUSION

The meta-analysis enabled the profiling of the response to neoadjuvant chemoradiotherapy in a geographically heterogeneous but clinically uniform group of patients. It provided insight into the characteristics that enable a better understanding and prediction of the response to therapy. By applying bioinformatics methods and integrating different algorithms, the transcriptional profile of patients with LARC was characterized in detail. In this study, important characteristics of tumors in patients with a poor therapeutic outcome were identified, which can be further investigated through functional studies to design target therapies that would improve the treatment of patients. Markers of a good therapeutic response were also identified, which represent an important contribution to designing a panel of characteristics that would be used for highly effective selection of patients for a wait-and-watch approach to treatment. As the first study of this type, we believe it represents a significant contribution to the current state of the art and a basis for further research, emphasizing the importance of integrating clinical data with molecular data in order to effectively subgroup patients for tailoring treatment depending on expected outcome.

Acknowledgements

This study was funded by the Horizon Europe STEPUPIORS Project (HORIZON-WIDERA-2021-ACCESS-03, European Commission, Agreement No. 101079217). The study was

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

supported by the COST action TRANSLACORE (Translational control in Cancer European Network, European Cooperation in Science and Technology, Action No. CA21154) and the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Agreement No. 451-03-136/2025-03/200043).

Author Contributions

A.S. Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing-original draft, Writing-review and editing; R.S. Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing-review and editing, Supervision; M.M. Conceptualization, Data Curation, Investigation, Methodology, Writing-original draft, Writing-review and editing, Supervision; A.D. Funding acquisition, Investigation, Project Administration, Writing-review and editing; S.S-R. Supervision, Writing-review and editing; R.J. Funding acquisition, Project Administration, Supervision, Writing-review and editing; S.C.B Funding acquisition, Project Administration, Supervision, Writing-review and editing; R.F. Funding acquisition, Project Administration, Supervision, Writing-review and editing; A.V. Funding acquisition, Investigation, Methodology, Project Administration, Supervision, Writing-review and editing; J.Z. Conceptualization, Funding acquisition, Investigation, Methodology, Project Administration, Supervision, Writing-review and editing; M.C Conceptualization, Funding acquisition, Investigation, Methodology, Project Administration, Supervision, Writing-original draft, Writing-review and editing. All authors have read and agreed to the published version of the manuscript.

Competing Interests

All authors declare no financial or non-financial competing interests.

Data availability statement

The datasets analyzed in this study are publicly available in the Gene Expression Omnibus (GEO) repository under the accession numbers GSE190826, GSE209746, GSE80606, and GSE233517 (https://www.ncbi.nlm.nih.gov/geo/). An additional dataset used in this study was obtained directly from the corresponding author of the study by Toomey et al. (DOI: 10.3390/cancers12071808) and is available from the authors upon reasonable request.

References

1. Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**, 229–263 (2024).
2. Paschke, S. *et al.* Are Colon and Rectal Cancer Two Different Tumor Entities? A Proposal to Abandon the Term Colorectal Cancer. *Int J Mol Sci* **19**, (2018).
3. Nunes, L. *et al.* Prognostic genome and transcriptome signatures in colorectal cancers. *Nature* **633**, 137–146 (2024).
4. Cornish, A. J. *et al.* The genomic landscape of 2,023 colorectal cancers. *Nature* **633**, 127–136 (2024).
5. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350–6 (2015).

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

6. Dunne, P. D. & Arends, M. J. Molecular pathological classification of colorectal cancer-an update. *Virchows Arch* **484**, 273–285 (2024).

7. Glynne-Jones, R. *et al.* Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* **28**, iv22–iv40 (2017).

8. Scott, A. J. *et al.* Management of Locally Advanced Rectal Cancer: ASCO Guideline. *Journal of Clinical Oncology* **42**, 3355–3375 (2024).

9. O'Sullivan, B. *et al.* The TNM classification of malignant tumours—towards common understanding and reasonable expectations. *Lancet Oncol* **18**, 849–851 (2017).

10. Gambacorta, M. A. *et al.* Timing to achieve the highest rate of pCR after preoperative radiochemotherapy in rectal cancer: a pooled analysis of 3085 patients from 7 randomized trials. *Radiotherapy and Oncology* **154**, 154–160 (2021).

11. Maas, M. *et al.* Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* **11**, 835–844 (2010).

12. Probst, C. P. *et al.* Extended Intervals after Neoadjuvant Therapy in Locally Advanced Rectal Cancer: The Key to Improved Tumor Response and Potential Organ Preservation. *J Am Coll Surg* **221**, 430–40 (2015).

13. Siddiqui, M. R. S. *et al.* Defining response to radiotherapy in rectal cancer using magnetic resonance imaging and histopathological scales. *World J Gastroenterol* **22**, 8414–8434 (2016).

14. Oh, B. Y. *et al.* Are We Predicting Disease Progress of the Rectal Cancer Patients without Surgery after Neoadjuvant Chemoradiotherapy? *Cancer Res Treat* **50**, 634–645 (2018).

15. Park, H. Predictive factors for early distant metastasis after neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *World J Gastrointest Oncol* **13**, 252–264 (2021).

16. Dossa, F., Chesney, T. R., Acuna, S. A. & Baxter, N. N. A watch-and-wait approach for locally advanced rectal cancer after a clinical complete response following neoadjuvant chemoradiation: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* **2**, 501–513 (2017).

17. Jankowski, M. *et al.* Watch-and-wait strategy in rectal cancer: Is there a tumour size limit? Results from two pooled prospective studies. *Radiotherapy and Oncology* **160**, 229–235 (2021).

18. Bahadoer, R. R. *et al.* Short-course radiotherapy followed by chemotherapy before total mesorectal excision (TME) versus preoperative chemoradiotherapy, TME, and optional adjuvant chemotherapy in locally advanced rectal cancer (RAPIDO): a randomised, open-label, phase 3 trial. *Lancet Oncol* **22**, 29–42 (2021).

19. Conroy, T. *et al.* Neoadjuvant chemotherapy with FOLFIRINOX and preoperative chemoradiotherapy for patients with locally advanced rectal cancer (UNICANCER-PRODIGE 23): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* **22**, 702–715 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

20. Yang, Y. *et al.* Preoperative Volumetric Modulated Arc Therapy With Simultaneous Integrated Boost for Locally Advanced Distal Rectal Cancer. *Technol Cancer Res Treat* **18**, 1533033818824367 (2019).

21. Stojanovic-Rundic, S. *et al.* Dose Escalation in Neoadjuvant Chemoradiotherapy for Rectal Cancer: Short-Term Efficacy and Toxicity of VMAT-SIB vs. 3D-CRT. *Medicina (Kaunas)* **61**, (2025).

22. Macchia, G. *et al.* Time to surgery and pathologic complete response after neoadjuvant chemoradiation in rectal cancer: A population study on 2094 patients. *Clin Transl Radiat Oncol* **4**, 8–14 (2017).

23. Cercek, A. *et al.* PD-1 Blockade in Mismatch Repair–Deficient, Locally Advanced Rectal Cancer. *New England Journal of Medicine* **386**, 2363–2376 (2022).

24. Socha, J., Glynne-Jones, R. & Bujko, K. Oncological risks associated with the planned watch-and-wait strategy using total neoadjuvant treatment for rectal cancer: A narrative review. *Cancer Treat Rev* **129**, 102796 (2024).

25. Lawrence, T. S., Blackstock, A. W. & McGinn, C. The mechanism of action of radiosensitization of conventional chemotherapeutic agents. *Semin Radiat Oncol* **13**, 13–21 (2003).

26. RALLIS, K. S., LAI YAU, T. H. & SIDERIS, M. Chemoradiotherapy in Cancer Treatment: Rationale and Clinical Applications. *Anticancer Res* **41**, 1–7 (2021).

27. Liu, Y. *et al.* Molecular mechanisms of chemo□ and radiotherapy resistance and the potential implications for cancer treatment. *MedComm (Beijing)* **2**, 315–340 (2021).

28. Hu, T. *et al.* Magnetic resonance imaging□based radiomics analysis for prediction of treatment response to neoadjuvant chemoradiotherapy and clinical outcome in patients with locally advanced rectal cancer: A large multicentric and validated study. *MedComm (Beijing)* **5**, (2024).

29. Li, M. *et al.* Predicting response to neoadjuvant chemoradiotherapy in rectal cancer: from biomarkers to tumor models. *Ther Adv Med Oncol* **14**, (2022).

30. Chatila, W. K. *et al.* Genomic and transcriptomic determinants of response to neoadjuvant therapy in rectal cancer. *Nat Med* **28**, 1646–1655 (2022).

31. Feng, L. *et al.* Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit Health* **4**, e8–e17 (2022).

32. Wang, J. *et al.* Identification of proteomic markers for prediction of the response to 5-Fluorouracil based neoadjuvant chemoradiotherapy in locally advanced rectal cancer patients. *Cancer Cell Int* **22**, 117 (2022).

33. Stanojevic, A. *et al.* Data-Independent Acquisition Mass Spectrometry Analysis of FFPE Rectal Cancer Samples Offers In-Depth Proteomics Characterization of the Response to Neoadjuvant Chemoradiotherapy. *Int J Mol Sci* **24**, 15412 (2023).

34. Marinkovic, M. *et al.* Exploring novel genetic and hematological predictors of response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Front Genet* **14**, (2023).

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

35. Marinkovic, M. *et al.* Performance and Dimensionality of Pretreatment MRI Radiomics in Rectal Carcinoma Chemoradiotherapy Prediction. *J Clin Med* **13**, 421 (2024).

36. Momma, T. *et al.* Validation of Gene Expression-Based Predictive Biomarkers for Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Cancers (Basel)* **13**, 4642 (2021).

37. Nicolas, A. M. *et al.* Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer. *Cancer Cell* **40**, 168-184.e13 (2022).

38. Bae, S. U. *et al.* Neoadjuvant chemoradiotherapy up-regulates PD-L1 in radioresistant colorectal cancer. *Clin Transl Radiat Oncol* **51**, 100906 (2025).

39. Kim, S. *et al.* YWHAZ and TBP are potential reference gene candidates for qPCR analysis of response to radiation therapy in colorectal cancer. *Sci Rep* **13**, 12902 (2023).

40. Chatila, W. K. *et al.* Genomic and transcriptomic determinants of response to neoadjuvant therapy in rectal cancer. *Nat Med* **28**, 1646–1655 (2022).

41. Ha, Y. J. *et al.* PSMB8 as a Candidate Marker of Responsiveness to Preoperative Radiation Therapy in Rectal Cancer Patients. *International Journal of Radiation Oncology*Biology*Physics* **98**, 1164–1173 (2017).

42. Toomey, S. *et al.* Genomic and Transcriptomic Characterisation of Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Cancers (Basel)* **12**, 1808 (2020).

43. Micev, M. *et al.* Histopathological evaluation of response to neoadjuvant chemoradiotherapy in locally advanced rectal carcinoma. *Acta Chir Iugosl* **63**, 59–73 (2016).

44. Nunes, L. *et al.* Prognostic genome and transcriptome signatures in colorectal cancers. *Nature* **633**, 137–146 (2024).

45. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

46. Leek, J., Johnson, E., Parker, H., Jaffe, A. & Storey, J. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

47. Song, C. & Tseng, G. C. Hypothesis Setting and Order Statistic for Robust Genomic Meta-Analysis. *Ann Appl Stat* **8**, 777–800 (2014).

48. Ma, T., Wang, X., Li, J. & Tseng, G. C. MetaDE: Transcriptome meta-analysis for differentially expressed gene detection. *GitHub repository* (2017).

49. Viechtbauer, W. Conducting Meta-Analyses in *R* with the **metafor** Package. *J Stat Softw* **36**, (2010).

50. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model*. (Springer New York, New York, NY, 2000). doi:10.1007/978-1-4757-3294-8.

51. Kassambara, A., Kosinski, M. & Biecek, P. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.5.0. *https://rpkgs.datanovia.com/survminer/index.html* (2024).

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

52. Sturm, G., Finotello, F. & List, M. Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. in 223–232 (2020). doi:10.1007/978-1-0716-0327-7_16.

53. Eide, P. W., Bruun, J., Lothe, R. A. & Sveen, A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci Rep* **7**, 16618 (2017).

54. de Back, T. R. *et al.* A consensus molecular subtypes classification strategy for clinical colorectal cancer tissues. *Life Sci Alliance* **7**, (2024).

55. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))* **2**, 100141 (2021).

56. Carlson, M. org.Hs.eg.db: Genome wide annotation for Human. R package version 3.18.0. *Bioconductor* (2017).

57. Badia-i-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances* **2**, (2022).

58. Yashar, W. M. *et al.* Predicting transcription factor activity using prior biological information. *iScience* **27**, 109124 (2024).

59. Ru, Y. *et al.* The multiMiR R package and database: integration of microRNA–target interactions along with their disease and drug associations. *Nucleic Acids Res* **42**, e133–e133 (2014).

60. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169 (2015).

61. Chang, L.-C., Lin, H.-M., Sibille, E. & Tseng, G. C. Meta-analysis methods for combining multiple expression profiles: comparisons, statistical characterization and an application guideline. *BMC Bioinformatics* **14**, 368 (2013).

62. Sartorius, D. *et al.* Implications of Rectal Cancer Radiotherapy on the Immune Microenvironment: Allies and Foes to Therapy Resistance and Patients' Outcome. *Cancers (Basel)* **15**, 5124 (2023).

63. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* **6**, e21800 (2011).

64. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–504 (2003).

65. Imedio, L. *et al.* MicroRNAs in Rectal Cancer: Functional Significance and Promising Therapeutic Value. *Cancers (Basel)* **12**, 2040 (2020).

66. Jin, Y. *et al.* miR-450b-5p Suppresses Stemness and the Development of Chemoresistance by Targeting *SOX2* in Colorectal Cancer. *DNA Cell Biol* **35**, 249–256 (2016).

67. Jiang, H., Ju, H., Zhang, L., Lu, H. & Jie, K. microRNA□577 suppresses tumor growth and enhances chemosensitivity in colorectal cancer. *J Biochem Mol Toxicol* **31**, (2017).

68. Xu, F. *et al.* MicroRNA□375□3p enhances chemosensitivity to 5□fluorouracil by targeting thymidylate synthase in colorectal cancer. *Cancer Sci* **111**, 1528–1541 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

69. Liu, H. *et al.* miR-139-5p sensitizes colorectal cancer cells to 5-fluorouracil by targeting NOTCH-1. *Pathol Res Pract* **212**, 643–649 (2016).

70. Ha Thi, H. T., Kim, H.-Y., Kim, Y.-M. & Hong, S. MicroRNA-130a modulates a radiosensitivity of rectal cancer by targeting SOX4. *Neoplasia* **21**, 882–892 (2019).

71. Zhu, Y. *et al.* miR-145 Antagonizes SNAI1-Mediated Stemness and Radiation Resistance in Colorectal Cancer. *Molecular Therapy* **26**, 744–754 (2018).

72. Ge, Y. *et al.* MiR-122-5p increases radiosensitivity and aggravates radiation-induced rectal injury through CCAR1. *Toxicol Appl Pharmacol* **399**, 115054 (2020).

73. Mu, J., Wang, X. & Sun, P. Expression of miR-31 in rectal cancer patients and its effect on proliferation ability of rectal cancer cells SW837. *Eur Rev Med Pharmacol Sci* **22**, 8675–8681 (2018).

74. Wan, P. *et al.* miR□129□5p inhibits proliferation, migration, and invasion in rectal adenocarcinoma cells through targeting E2F7. *J Cell Physiol* **235**, 5689–5701 (2020).

75. Cai, S.-D. *et al.* MicroRNA-144 inhibits migration and proliferation in rectal cancer by downregulating ROCK-1. *Mol Med Rep* **12**, 7396–7402 (2015).

76. Zhang, Y. *et al.* UBE2C promotes rectal carcinoma via miR-381. *Cancer Biol Ther* **19**, 230–238 (2018).

77. Yu, F.-B. *et al.* MiR-19a-3p regulates the Forkhead box F2-mediated Wnt/β-catenin signaling pathway and affects the biological functions of colorectal cancer cells. *World J Gastroenterol* **26**, 627–644 (2020).

78. Tahmasebi, H., Trajcevski, K., Higgins, V. & Adeli, K. Influence of ethnicity on population reference values for biochemical markers. *Crit Rev Clin Lab Sci* **55**, 359–375 (2018).

79. Gijsberts, C. M. *et al.* The ethnicity-specific association of biomarkers with the angiographic severity of coronary artery disease. *Neth Heart J* **24**, 188–98 (2016).

80. Taube, S. E. *et al.* A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *J Natl Cancer Inst* **101**, 1453–63 (2009).

81. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (1979)* **347**, (2015).

82. Liu, M. *et al.* Sorafenib induces ferroptosis by promoting TRIM54-mediated FSP1 ubiquitination and degradation in hepatocellular carcinoma. *Hepatol Commun* **7**, (2023).

83. Zhu, J. *et al.* Targeting TRIM54/Axin1/β-Catenin Axis Prohibits Proliferation and Metastasis in Hepatocellular Carcinoma. *Front Oncol* **11**, 759842 (2021).

84. Zhao, H. *et al.* Wnt signaling in colorectal cancer: pathogenic role and therapeutic target. *Mol Cancer* **21**, 144 (2022).

85. Liu, D. *et al.* Cytoplasmic poly(A) binding protein 4 is highly expressed in human colorectal cancer and correlates with better prognosis. *J Genet Genomics* **39**, 369–74 (2012).

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

86. Wen, H. Y., Xia, Y., Young, M. E., Taegtmeyer, H. & Kellems, R. E. The adenylosuccinate synthetase-1 gene is activated in the hypertrophied heart. *J Cell Mol Med* **6**, 235–43 (2002).

87. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* **54**, 1.30.1-1.30.33 (2016).

88. Ben-Hamo, R. *et al.* Predicting and affecting response to cancer therapy based on pathway-level biomarkers. *Nat Commun* **11**, 3296 (2020).

89. Merrick, B. A., London, R. E., Bushel, P. R., Grissom, S. F. & Paules, R. S. Platforms for biomarker analysis using high-throughput approaches in genomics, transcriptomics, proteomics, metabolomics, and bioinformatics. *IARC Sci Publ* 121–42 (2011).

90. Dalkılıç, S. *et al.* Bioinformatics analysis of colorectal cancer transcriptomic data reveals novel prognostic signature and potential biomarker genes. *Scand J Gastroenterol* **60**, 42–53 (2025).

91. Zheng, W., Ling, S., Cao, Y., Shao, C. & Sun, X. Combined use of NK cells and radiotherapy in the treatment of solid tumors. *Front Immunol* **14**, (2024).

92. Lee, Y. S., Heo, W., Nam, J., Jeung, Y. H. & Bae, J. The combination of ionizing radiation and proteasomal inhibition by bortezomib enhances the expression of NKG2D ligands in multiple myeloma cells. *J Radiat Res* **59**, 245–252 (2018).

93. Liu, T. *et al.* Harnessing NKG2D CAR-T cells with radiotherapy: a novel approach for esophageal squamous cell carcinoma treatment. *Front Immunol* **16**, (2025).

94. Sartorius, D. *et al.* Implications of Rectal Cancer Radiotherapy on the Immune Microenvironment: Allies and Foes to Therapy Resistance and Patients' Outcome. *Cancers (Basel)* **15**, (2023).

95. Väyrynen, J. P. *et al.* Prognostic significance of myeloid immune cells and their spatial distribution in the colorectal cancer microenvironment. *J Immunother Cancer* **9**, e002297 (2021).

96. Watson, N. F. S. *et al.* Immunosurveillance is active in colorectal cancer as downregulation but not complete loss of MHC class I expression correlates with a poor prognosis. *Int J Cancer* **118**, 6–10 (2006).

97. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* **21**, 1350–6 (2015).

98. Yan, J., Wan, P., Choksi, S. & Liu, Z.-G. Necroptosis and tumor progression. *Trends Cancer* **8**, 21–27 (2022).

99. Xu, Y. & Ruggero, D. The Role of Translation Control in Tumorigenesis and Its Therapeutic Implications. *Annu Rev Cancer Biol* **4**, 437–457 (2020).

100. Zhou, S. *et al.* The role of epithelial-mesenchymal transition in regulating radioresistance. *Crit Rev Oncol Hematol* **150**, 102961 (2020).

101. Bhangu, A. *et al.* The role of epithelial mesenchymal transition and resistance to neoadjuvant therapy in locally advanced rectal cancer. *Colorectal Dis* **16**, O133-43 (2014).

medRxiv preprint doi: https://doi.org/10.1101/2025.07.18.25331799; this version posted July 23, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

102. Fan, Y., Mao, R. & Yang, J. NF-κB and STAT3 signaling pathways collaboratively link inflammation to cancer. *Protein Cell* **4**, 176–185 (2013).

103. Wu, X., Sun, L. & Xu, F. NF-κB in Cell Deaths, Therapeutic Resistance and Nanotherapy of Tumors: Recent Advances. *Pharmaceuticals (Basel)* **16**, (2023).

104. Liang, B. *et al.* Antagonistic roles of cGAS/STING signaling in colorectal cancer chemotherapy. *Front Oncol* **14**, 1441935 (2024).

105. Yu, Y. *et al.* SP1 upregulated FoxO3a promotes tumor progression in colorectal cancer. *Oncol Rep* **39**, 2235–2242 (2018).

106. Xu, X. *et al.* miR-375-3p suppresses tumorigenesis and partially reverses chemoresistance by targeting YAP1 and SP1 in colorectal cancer cells. *Aging* **11**, 7357–7385 (2019).

107. De Palma, F. D. E. *et al.* The Role of Micro-RNAs and Circulating Tumor Markers as Predictors of Response to Neoadjuvant Therapy in Locally Advanced Rectal Cancer. *Int J Mol Sci* **21**, (2020).

108. Wu, F. *et al.* Screening of MicroRNA Related to Irradiation Response and the Regulation Mechanism of miRNA-96-5p in Rectal Cancer Cells. *Front Oncol* **11**, 699475 (2021).

109. Dayde, D., Tanaka, I., Jain, R., Tai, M. C. & Taguchi, A. Predictive and Prognostic Molecular Biomarkers for Response to Neoadjuvant Chemoradiation in Rectal Cancer. *Int J Mol Sci* **18**, (2017).

110. Wada, Y. *et al.* Circulating miRNA Signature Predicts Response to Preoperative Chemoradiotherapy in Locally Advanced Rectal Cancer. *JCO Precis Oncol* **5**, (2021).

111. Machackova, T., Prochazka, V., Kala, Z. & Slaby, O. Translational Potential of MicroRNAs for Preoperative Staging and Prediction of Chemoradiotherapy Response in Rectal Cancer. *Cancers (Basel)* **11**, 1545 (2019).

112. Krawczyk, M., Masternak, K., Zufferey, M., Barras, E. & Reith, W. New functions of the major histocompatibility complex class II-specific transcription factor RFXANK revealed by a high-resolution mutagenesis study. *Mol Cell Biol* **25**, 8607–18 (2005).

113. Li, Y. *et al.* Carcinoembryonic Antigen Interacts with TGF-β Receptor and Inhibits TGF-β Signaling in Colorectal Cancers. *Cancer Res* **70**, 8159–8168 (2010).