Am J Clin Exp Immunol 2026;15(3):244-254
www.ajcei.us /ISSN:2164-7712/AJCEI0173195

Original Article

Cancer-associated mutational patterns reshape miRNA regulatory regions and potentially disrupt miRNA-gene interactions

Julia Ran[1,3*], Xiwen Zhang[1,4*], Joyce Tang[1,5], Yongsheng Bai[1,2]

[1]Eastern Michigan University, Ypsilanti, MI 48197, USA; [2]Next-Gen Intelligent Science Training, Ann Arbor, MI 48105, USA; [3]Unionville High School, Kennett Square, PA 19348, USA; [4]Oakville Trafalgar High School, Oakville, ON L6J 3L6, Canada; [5]Carmel High School, Carmel, IN 46032, USA. [*]Equal contributors.

Received April 9, 2026; Accepted May 29, 2026; Epub June 25, 2026; Published June 30, 2026

Abstract: MicroRNAs (miRNAs) are key post-transcriptional regulators frequently dysregulated in cancer, yet how somatic mutational patterns influence miRNA regulatory regions remains unclear. In this study, we integrated mutation profiles, transition-transversion (Ti/Tv) ratios, and miRNA-gene interaction analyses across multiple TCGA cancer subtypes to investigate how mutation patterns may affect miRNA binding regions. miRNA interactions were strongly enriched in 3' untranslated regions (3'UTRs), with fewer interactions observed in coding sequences (CDS). Across cancers, miRNA regions exhibited moderate GC content (~51%), which typically predicts elevated Ti/Tv ratios; however, most cancer cohorts showed markedly reduced Ti/Tv ratios, indicating a shift toward transversion-heavy mutational processes. Because transversions generally produce greater nucleotide class changes than transitions, this shift may increase the likelihood of disrupting sequence-specific miRNA binding. Supporting this, we identified missense mutations overlapping predicted miRNA binding sites within coding regions, highlighting a potential intersection between somatic mutations and post-transcriptional regulation. Together, these findings suggest that cancer-associated mutational patterns may reshape miRNA regulatory regions and alter miRNA-gene interactions, providing insight into how somatic mutations impact gene regulation beyond protein-coding changes.

Keywords: microRNA, somatic mutation, transition-transversion ratio, GC content, miRNA binding, pan-cancer analysis, TCGA

Introduction

Cancer arises from the accumulation of somatic mutations in oncogenes and tumor suppressor genes that disrupt normal cellular regulation. In addition to protein-coding genes, noncoding regulatory elements such as microRNAs (miRNAs) play a critical role in cancer biology. miRNAs are small noncoding RNAs that regulate gene expression post-transcriptionally, primarily by binding to the 3' untranslated region (3'UTR) of target mRNAs, leading to mRNA degradation or translational repression [1-6]. Through these mechanisms, miRNAs influence key cellular processes including proliferation, apoptosis, and tumor progression, and have emerged as important biomarkers and potential therapeutic targets [7, 8]. Distinct miRNA expression signatures have also been identified across multiple human cancers, further supporting their diagnostic and prognostic relevance [9]. Depending on their targets and cellular context, miRNAs may function as either oncogenic miRNAs (onco-miRNAs) or tumor suppressor miRNAs, contributing to cancer progression through dysregulation of post-transcriptional regulatory networks.

Although miRNA binding predominantly occurs in 3'UTRs, increasing evidence indicates that miRNAs can also interact with coding sequences (CDS), where they may exert additional regulatory effects [10-13]. Because miRNA-gene interactions rely on sequence-specific complementarity, particularly within defined binding regions, they are highly sensitive to nucleotide changes, suggesting that mutations within noncoding regulatory regions, including untranslat-

https://doi.org/10.62347/FGPO8785

ed regions and RNA-associated elements, may contribute substantially to cancer biology despite being overlooked in traditional protein-centric analyses [14].

Cancer progression is driven by somatic mutations, including single-nucleotide variants (SNVs) and copy number alterations (CNAs) [15]. SNVs can be classified as transitions or transversions, and the transition-transversion (Ti/Tv) ratio is commonly used to characterize mutational patterns. In human miRNA regions, which exhibit moderate GC content (~51%), transitions are typically more frequent, resulting in expected Ti/Tv ratios ranging from 2.59 to 2.95 [15, 16]. However, cancer-associated mutational processes can alter these patterns, often increasing the prevalence of transversions, which introduce more substantial nucleotide changes than transitions.

Despite extensive research on miRNAs and cancer mutations, how cancer-associated mutational patterns influence miRNA regulatory regions remains poorly understood. In particular, it is unclear whether shifts in mutation types affect miRNA binding regions in both 3'UTRs and coding sequences, and whether such changes may disrupt miRNA-gene interactions. Additionally, the extent to which somatic mutations overlap with miRNA binding sites has not been systematically examined across five TCGA cancer subtypes: breast invasive carcinoma (BRCA), bladder urothelial carcinoma (BLCA), kidney renal clear cell carcinoma (KIRC), lung adenocarcinoma (LUAD), and stomach adenocarcinoma (STAD).

To address these gaps, this study integrates mutational signatures (including Ti/Tv ratios and GC content) with miRNA-gene interaction analyses across multiple cancer subtypes. By examining the distribution of mutations and their overlap with predicted miRNA binding regions in both 3'UTRs and coding sequences, we aim to determine how somatic mutational patterns may reshape miRNA regulatory landscapes in cancer.

Methods

Data sources and cancer cohorts

Data used in this study were obtained from The Cancer Genome Atlas (TCGA) through cBioPortal and from previously published supplementary datasets in Shen et al. [17]. Significant genes and their associated miRNA-gene interactions were extracted from Supplementary Table 1 of Shen et al. [17]. Six cancer subtypes were initially considered for analysis: breast invasive carcinoma (BRCA), bladder urothelial carcinoma (BLCA), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), lung adenocarcinoma (LUAD), and stomach adenocarcinoma (STAD). KIRP was included in miRNA-gene interaction analyses where sufficient data were available but was excluded from GC content and Ti/Tv analyses due to insufficient mutation counts that precluded reliable statistical comparison. Final cohort selection for each analysis is described in the relevant subsection below. For each cancer subtype, somatic mutation data, including mutation type and genomic coordinates, were retrieved using cBioPortal. All analyses were performed using publicly available datasets, and no additional patient-level filtering or exclusion criteria were applied.

Mutation classification and Ti/Tv analysis

Somatic mutation data for each cancer subtype were obtained from cBioPortal and Supplementary Table 4 of Shen et al. [17]. Genetic alterations were categorized into single-nucleotide variants (SNVs) and copy number alterations (CNAs). Only SNVs were included in transition-transversion (Ti/Tv) analysis. SNVs were classified as either transitions (purine-to-purine or pyrimidine-to-pyrimidine substitutions) or transversions (purine-to-pyrimidine substitutions) based on the reported mutation type. The total number of transitions and transversions was recorded for each cancer subtype and for each miRNA strand (3p and 5p). The Ti/Tv ratio was calculated for each group as the number of transitions divided by the number of transversions. Ratios were computed separately for each cancer subtype and strand to allow comparison of mutation patterns across datasets. Because transition-transversion classification applies only to single-nucleotide substitutions, Ti/Tv analysis was restricted to SNVs. The distribution of significant CNA and SNV mutation types across cancer cohorts is provided in Supplementary Figures 1 and 2, respectively.

MiRNA binding site analysis in CDS and 3'UTRS

miRNA-gene pairs associated with each cancer subtype were obtained from Supplementary Table 1 of Shen et al. [17]. For each gene-miRNA pair, predicted binding interactions were identified using DIANA-microT, which evaluates sequence complementarity, binding energy, and evolutionary conservation. Predicted miRNA binding sites were categorized based on genomic location as either within the coding sequence (CDS) or the 3' untranslated region (3'UTR) of the target gene. For each cancer subtype, interactions were stratified by mutation type (single-nucleotide variants [SNVs] or copy number alterations [CNAs]), and the total number of CDS and 3'UTR interactions was quantified. To enable cross-cancer comparisons independent of total interaction burden, CDS and 3'UTR interaction frequencies were also converted into proportional distributions within each cancer and mutation class (Supplementary Table 1). These data were used to compare the relative enrichment of miRNA binding across genomic regions and mutation types.

Gene-miRNA interaction categories were classified according to the oncogenic or tumor-suppressive status of both the gene and associated miRNA. Categories included onco-miRNA-oncogene (OO), tumor suppressor miRNA-tumor suppressor gene (TT), tumor suppressor miRNA-oncogene (TO), and onco-miRNA-tumor suppressor gene (OT) interactions.

Overlap of miRNA binding sites with somatic mutations

Predicted CDS miRNA binding site coordinates obtained from DIANA-microT were originally annotated in hg38 and were converted to hg19 using the UCSC liftOver tool to ensure compatibility with somatic missense mutation coordinates retrieved from cBioPortal. Converted CDS coordinates were then compared with hg19 mutation coordinates to identify genomic overlaps between predicted miRNA binding regions and somatic missense mutations.

GC content analysis

miRNA sequences associated with significant genes in each cancer subtype were obtained from Supplementary Table 1 of Shen et al. [17] and retrieved from miRBase.org. Sequence characteristics including accession number, chromosomal location, and 5p and 3p sequences were recorded for each miRNA (Supplementary Table 3). GC content was calculated for each individual miRNA sequence as the proportion of guanine and cytosine nucleotides relative to total sequence length, and values were averaged across miRNAs within each cancer subtype and strand (5p and 3p) separately (Supplementary Table 4). Cancer subtypes were included if they had sufficient miRNA sequence representation across both GC content and Ti/Tv analyses. Subtypes with no available data in either analysis (LIHC, PRAD, KICH, THCA) or data in only one analysis (HNSC, LUSC) were excluded, yielding a final cohort of five cancer subtypes: BRCA, BLCA, KIRC, LUAD, and STAD.

Statistical analysis

Differences in GC content across cancer types were assessed separately for 5p and 3p strands using the Kruskal-Wallis test, a non-parametric test appropriate for comparing distributions across multiple groups without assuming normality. Significant overall differences were followed by pairwise Dunn's post-hoc tests with Bonferroni correction, which compared GC content between each pair of cancer types individually while adjusting the significance threshold to account for the increased probability of false positives when conducting multiple comparisons simultaneously. All statistical analyses were performed in Python 3 using the scipy and scikit-posthocs libraries.

Pairwise differences in Ti/Tv ratios between cancer types were assessed using Fisher's exact test applied to 2×2 contingency tables of transition and transversion counts. This test evaluates whether the proportion of transitions relative to transversions differs significantly between two groups by calculating the exact probability of the observed distribution. Analyses were performed separately for 3p and 5p strands. KIRC was included in analyses but interpreted cautiously due to limited mutation counts ($n = 13$ for 3p, $n = 12$ for 5p). All statistical analyses were performed in Python 3 using the scipy library. Statistical significance was defined as $P < 0.05$.

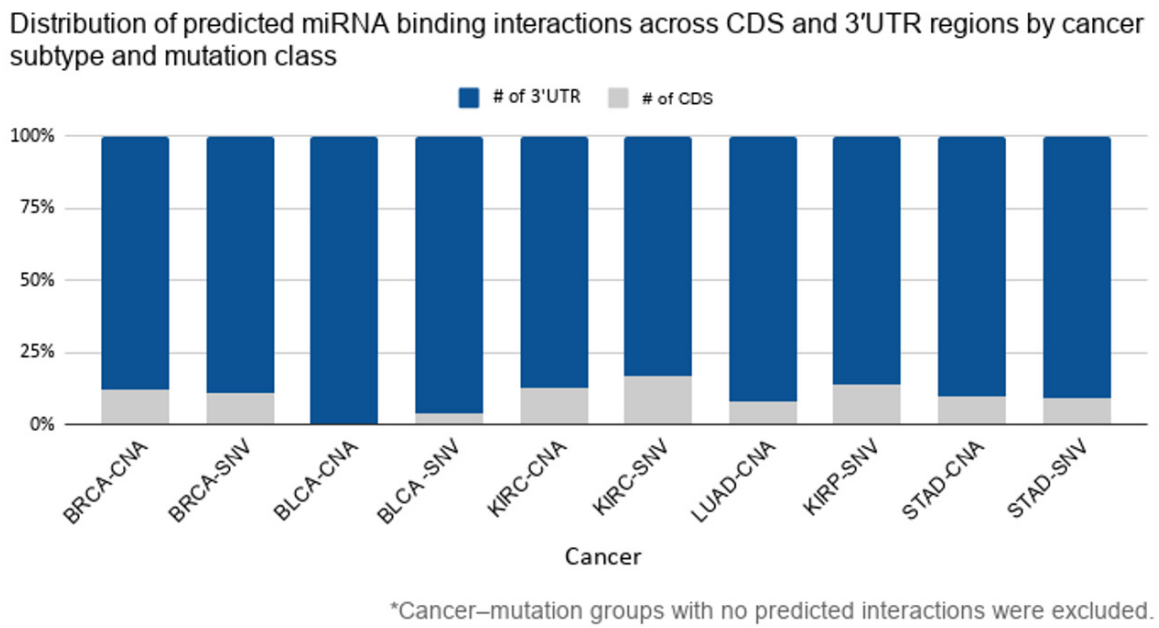


Figure 1. Proportional distribution of predicted miRNA binding interactions within coding sequences (CDS) and 3' untranslated regions (3'UTR) across cancer subtypes and mutation classes (CNA and SNV). Across nearly all analyzed cohorts, predicted interactions were strongly enriched in 3'UTRs relative to CDS regions, indicating preservation of canonical miRNA targeting architecture despite mutational variation. Cancer-mutation groups lacking detectable interactions were excluded.

Results

Overall amino acid mutation frequencies across the analyzed cancer cohorts are provided in Supplementary Table 2.

miRNA & gene interactions in CDS & 3'UTR

Across all analyzed cancer subtypes, predicted miRNA-gene interactions were consistently enriched in 3' untranslated regions (3'UTRs) relative to coding sequences (CDS). In every cancer cohort, total 3'UTR interactions substantially exceeded CDS interactions, supporting the established predominance of 3'UTR-mediated miRNA regulation. BRCA and LUAD exhibited the highest overall interaction counts, with 1,852 and 1,980 total 3'UTR interactions, respectively, while BLCA showed the fewest interactions overall. Although CDS interactions were detected in all major cohorts except for limited subsets, they remained consistently less frequent than 3'UTR interactions across cancers. Among the analyzed cohorts, KIRC demonstrated the highest relative proportion of CDS interactions compared with 3'UTR interactions, suggesting modest variability in coding-region targeting frequency between cancer types (**Figure 1**). Overall, these findings indicate that while miRNA targeting remains strongly concentrated in 3'UTRs, CDS interactions are present across multiple cancers and may represent an additional layer of regulatory complexity.

Across most cancer-mutation groups, CDS/3'UTR ratios remained consistently low, reinforcing the dominant enrichment of predicted miRNA interactions within canonical 3'UTR regions. However, several notable deviations were observed. LUAD-CNA demonstrated one of the highest total interaction burdens overall, with 168 CDS and 1,980 3'UTR interactions, suggesting a comparatively broad regulatory landscape despite preserved 3'UTR predominance (**Table 1**). KIRC-SNV also exhibited an elevated CDS proportion relative to several other cohorts, indicating that renal clear cell carcinoma may retain comparatively greater coding-sequence miRNA interaction potential under SNV-associated alterations. In contrast, BLCA-CNA showed extremely limited interaction counts overall, with only a single 3'UTR interaction and no detectable CDS interactions, suggesting minimal predicted miRNA regulatory involvement within this mutation class.

Table 1. Summary of predicted miRNA binding interactions across coding sequence (CDS) and 3' untranslated region (3'UTR) regions by cancer subtype and mutation type

Cancer	Total CDS Interactions	Total 3'UTR Interactions	CDS:UTR3 Ratio	CDS %
BRCA	233	1852	0.126	11.1
BLCA	9	219	0.041	3.9
KIRC	158	776	0.204	16.9
LUAD	168	1980	0.085	7.8
KIRP	83	522	0.159	13.7
STAD	88	874	0.101	9.1

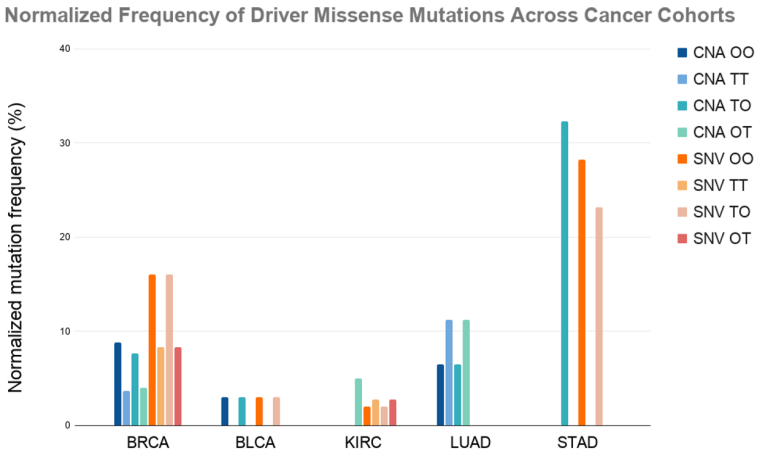


Figure 2. Normalized frequencies of miRNA-gene interaction categories across cancer cohorts. Bars represent normalized frequencies of interactions classified according to the oncogenic or tumor-suppressive status of both the miRNA and associated target gene, including onco-miRNA-oncogene (OO), tumor suppressor miRNA-tumor suppressor gene (TT), tumor suppressor miRNA-oncogene (TO), and onco-miRNA-tumor suppressor gene (OT) categories. Distinct interaction-category distributions were observed across cancer types, suggesting variability in cancer-associated regulatory network organization.

Additionally, LUAD-SNV and KIRP-CNA lacked detectable predicted interactions entirely and were therefore excluded from proportional analyses. These outliers highlight that while the overall architecture of miRNA targeting remains strongly 3'UTR-centered, the total burden and relative contribution of CDS interactions can vary substantially depending on both cancer subtype and mutation class.

Distribution of MiRNA-gene interaction

Distinct distributions of miRNA-gene interaction categories were observed across cancer cohorts (**Figure 2**). Several cancers demonstrated relatively higher frequencies of oncogenic interaction classes, whereas others exhibited broader distributions across both oncogenic and tumor-suppressive interaction categories. STAD showed elevated frequencies in multiple oncogenic interaction groups, while BRCA and LUAD displayed more distributed interaction profiles across categories. These findings suggest that miRNA-associated regulatory interaction patterns vary between cancer types and may reflect differences in cancer-specific regulatory network organization.

Overlap of CDS miRNA sites with missense mutations

To investigate whether somatic missense mutations may intersect predicted miRNA regulatory elements within coding regions, genomic coordinates of predicted CDS miRNA binding sites were compared with missense mutation positions across analyzed cancer cohorts. Among all predicted CDS miRNA-gene interactions, only three overlapping events were identified, indicating that direct intersection between coding mutations and predicted CDS miRNA recognition elements is uncommon.

Three overlap events were identified across distinct cancer subtypes (**Table 2**; **Figure 3**). In KIRC, a predicted CDS miRNA binding site within DIS3 overlapped a missense mutation at chr13:73,340,159. In LUAD, FLT3 contained a predicted CDS binding site overlapping a missense mutation at chr13:28,623,625. In STAD, PTPRT demonstrated overlap between a predicted CDS miRNA recognition region and a missense mutation at chr20:40,790,029. The-

Table 2. Predicted CDS miRNA binding sites overlapping somatic missense mutations across cancer cohorts*

Cancer	Gene	miRNA	Mutation Type	CDS Binding Site Coordinates	Missense Mutation Coordinates
KIRC	DIS3	hsa-mir-653-3p	Missense	chr13:73340148-73340175	chr13:73340159
LUAD	FLT3	hsa-mir-548e-5p	Missense	chr13:28623608-28623626	chr13:28623625
STAD	PTPRT	hsa-mir-3922-3p	Missense	chr20:40790018-40790035	chr20:40790029

*Predicted CDS miRNA binding sites and missense mutation positions were standardized to hg19 coordinates prior to overlap analysis. Predicted coding-sequence (CDS) miRNA binding coordinates were compared with somatic missense mutation positions after genomic coordinate harmonization to hg19. Only gene-miRNA pairs demonstrating direct genomic overlap between the mutation coordinate and predicted CDS miRNA recognition interval are shown.

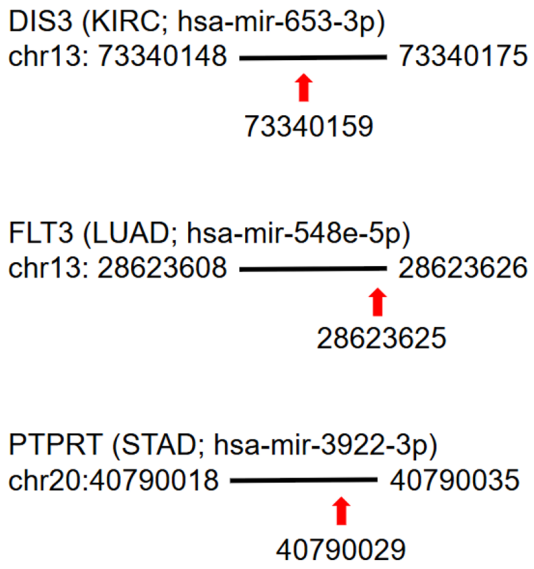


Figure 3. Schematic representation of somatic missense mutations overlapping predicted CDS miRNA binding sites. Schematic representation of the three identified cases in which somatic missense mutation coordinates fell within predicted coding-sequence (CDS) miRNA binding regions following genomic coordinate harmonization to hg19. Horizontal bars represent predicted CDS miRNA recognition intervals, and vertical markers indicate missense mutation positions. Equal bar lengths are schematic and do not represent absolute genomic distance.

se events represent candidate loci where coding mutations may coincide with predicted post-transcriptional regulatory regions.

Although rare relative to the broader predominance of 3'UTR miRNA targeting, these overlap events highlight potential sites where somatic coding mutations may affect both protein sequence and predicted miRNA-mediated regulation. While computational overlap alone does not confirm functional disruption, these candidate loci suggest a possible intersection between mutational and post-transcriptional regulatory mechanisms that may warrant future experimental validation.

GC content

GC content of miRNA-associated sequences ranged from approximately 42-52% across all five cancer types and both strands, consistent with the established baseline of approximately 51% for human miRNAs [18]. Mean GC content for the 5p strand ranged from 43.92% (BLCA) to 50.66% (BRCA), and for the 3p strand from 42.48% (BLCA) to 51.68% (STAD) (**Table 3**). Results for BLCA should be interpreted cautiously given the limited number of miRNA sequences available for that cohort (n = 8 per strand). Kruskal-Wallis testing revealed statistically significant differences in GC content across cancer types for both the 5p strand ($H = 11.668$, $P = 0.020$) and 3p strand ($H = 11.951$, $P = 0.018$). However, pairwise Dunn's post-hoc tests with Bonferroni correction identified only one significant pairwise difference: BRCA versus LUAD on the 5p strand ($P = 0.019$). No significant pairwise differences were detected on the 3p strand (**Figure 4**).

Transition-transversion ratio

Ti/Tv ratios varied substantially across cancer types and between strands (**Table 4**; **Figure 5**). On the 3p strand, STAD exhibited the highest Ti/Tv ratio (3.25), indicating strong transition dominance, while LUAD showed the lowest ratio (0.29), reflecting a predominance of transversions. On the 5p strand, STAD again showed the highest ratio (2.31), while KIRC showed the lowest (0.71). Fisher's exact tests revealed significant pairwise differences in Ti/Tv ratios on the 3p strand between LUAD and STAD ($P < 0.0001$), BLCA and LUAD ($P = 0.0003$), BRCA and LUAD ($P = 0.005$), and LUAD and KIRC ($P = 0.005$, interpreted cautiously due to small

Table 3. GC content summary

Cancer	N (5p)	Mean 5p (%)	SD 5p	N (3p)	Mean 3p (%)	SD 3p
BRCA	296	50.66	10.98	296	49.70	11.58
BLCA*	8	43.92	5.59	8	42.48	3.13
KIRC	101	49.26	10.06	100	47.69	11.70
LUAD	282	48.42	10.98	282	47.75	11.26
STAD	53	50.04	10.64	53	51.68	9.12

*BLCA counts are limited (n = 8 per strand) and should be interpreted cautiously.

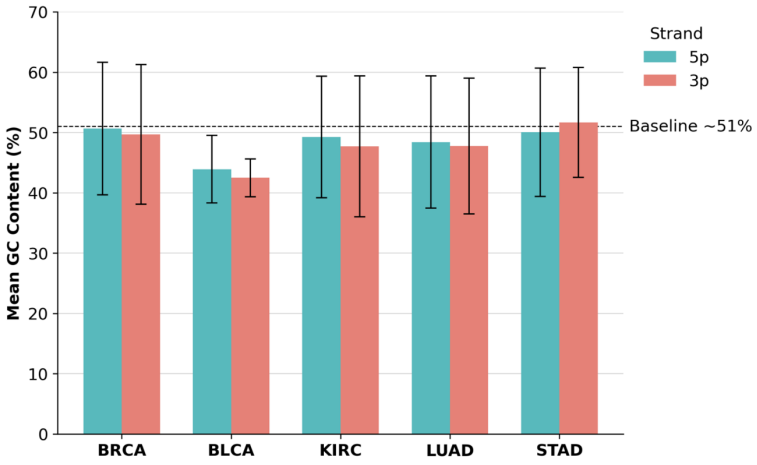


Figure 4. Mean GC content of miRNA-associated sequences across five cancer subtypes. Bars represent mean GC content (%) for 5p (teal) and 3p (salmon) miRNA strands within each cancer type. Error bars indicate one standard deviation. The dashed line represents the established baseline GC content of approximately 51% for human miRNAs [15]. Cancer subtypes with insufficient data were excluded. BRCA: n = 296; BLCA: n = 8; KIRC: n = 100-101; LUAD: n = 282; STAD: n = 53.

counts for KIRC). On the 5p strand, significant differences were identified between LUAD and STAD (P = 0.005) and BRCA and STAD (P = 0.028). Notably, KIRC exhibited marked strand asymmetry, with a 3p Ti/Tv ratio of 2.25 compared to 0.71 on the 5p strand, suggesting strand-specific differences in mutational pressure. With the exception of STAD, all cancer cohorts fell below the expected baseline Ti/Tv range of 2.59-2.95 for healthy human miRNA regions, indicating a consistent shift toward transversion-enriched mutational patterns in cancer-associated miRNA loci.

Discussion

Our integrated analysis suggests that although miRNA sequence architecture remains broadly conserved across cancers, cancer-specific somatic mutational processes substantially reshape mutational signatures and may alter regulatory landscapes within miRNA-associated regions. By combining sequence composition (GC content), mutational dynamics (Ti/Tv ratios), predicted miRNA-gene interaction distributions, and overlap between coding-sequence miRNA binding sites and somatic missense mutations, this study provides a comparative framework linking stable molecular architecture with tumor-specific genomic disruption. Because miRNAs regulate pathways involved in tumorigenesis, metastasis, and interpatient heterogeneity, disruption of miRNA-associated regulatory regions may have broader implications for cancer progression and therapeutic response [19].

GC content remained relatively conserved across analyzed cancer cohorts (~42-52%), closely approximating the established baseline of ~51% for human miRNAs reported by Guo et al. [18]. This conservation is biologically expected, as GC-rich stem regions are essential for maintaining the hairpin secondary structure required for miRNA maturation, Dicer processing, and RISC loading [18, 20-24]. Although Kruskal-Wallis testing detected overall statistical differences across cancers, the near absence of significant pairwise differences suggests that most variation is minor relative to the broader structural conservation of miRNA loci. These findings indicate that despite extensive genomic instability in cancer, miRNA regions remain under selective pressure to preserve core sequence architecture, supporting GC content as a conserved biochemical baseline rather than a primary driver of inter-cancer mutational divergence.

In contrast to this conserved baseline, Ti/Tv ratios varied substantially across cancer types and frequently deviated from the expected healthy-miRNA range of 2.59-2.95 [18]. LUAD demonstrated pronounced transversion enrichment, particularly on the 3p strand, consistent with tobacco-associated G→T transversion sig-

Table 4. Ti/Tv counts and ratios

Cancer	3p Ti	3p Tv	3p Ti/Tv	5p Ti	5p Tv	5p Ti/Tv
BRCA	20	17	1.18	16	19	0.84
BLCA	27	16	1.69	29	26	1.12
KIRC*	9	4	2.25*	5	7	0.71
LUAD	9	31	0.29	29	38	0.76
STAD	39	12	3.25	37	16	2.31

*KIRC counts are limited (n < 20 per strand) and should be interpreted cautiously.

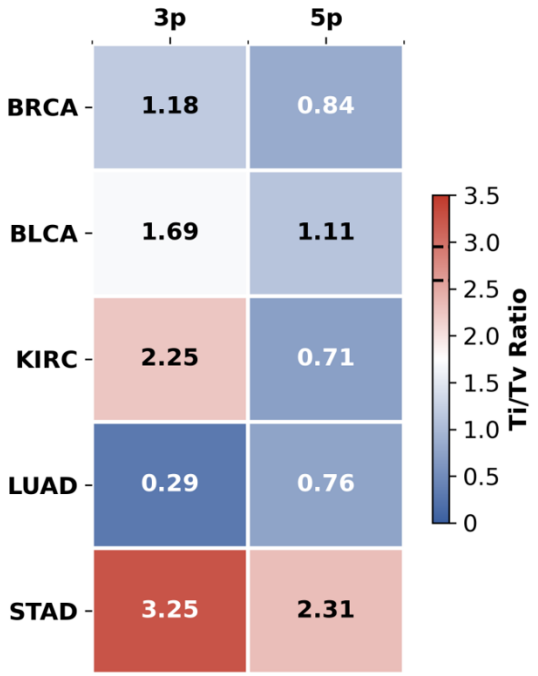


Figure 5. Transition-transversion (Ti/Tv) ratios across five cancer subtypes and miRNA strand orientation. Each cell displays the Ti/Tv ratio for the indicated cancer type (rows) and strand (3p or 5p, columns). Color intensity reflects the magnitude of the ratio: red indicates transition-dominant mutational patterns, blue indicates transversion-dominant patterns, and white approximates the midpoint of the established baseline Ti/Tv range of 2.59-2.95 for healthy human miRNA regions [15]. KIRC results should be interpreted cautiously due to limited mutation counts (3p: n = 13; 5p: n = 12). Ticks on the color scale denote the established baseline Ti/Tv range of 2.59-2.95 for healthy human miRNA regions [15].

natures widely reported in lung adenocarcinoma [18]. Conversely, STAD most closely approached or exceeded expected transition-dominant baselines, potentially reflecting mutational processes associated with microsatellite instability or other gastric-specific signatures [25-29]. These findings suggest that while miRNA loci maintain conserved sequence composition, external mutagenic pressures can override intrinsic nucleotide predispositions, producing cancer-specific mutational landscapes. Such mutation-associated signatures have increasingly been explored as biomarkers for cancer classification, exposure history, and therapeutic stratification [30]. Thus, sequence composition alone does not determine mutational outcome; rather, somatic mutagenesis appears to be the dominant force shaping cancer-associated miRNA mutation patterns.

Our miRNA-gene interaction analysis further contextualizes these sequence-level findings by demonstrating that predicted miRNA targeting remained overwhelmingly enriched in canonical 3'UTR regions across all cancers, consistent with established models of miRNA-mediated post-transcriptional repression [2-6]. Differences in the distribution of oncogenic and tumor-suppressive miRNA-gene interaction categories further support the presence of cancer-specific regulatory landscapes. Variability in the balance between oncogenic and tumor-suppressive interactions may contribute to differences in post-transcriptional dysregulation across tumor types. Although CDS interactions were consistently less frequent, they were detectable across multiple cancer types, suggesting that coding-sequence targeting may represent a secondary but biologically relevant layer of regulatory complexity [10-13]. The relative enrichment of CDS interactions in select cohorts such as KIRC suggests that some cancers may exhibit modestly greater potential for noncanonical regulatory targeting, though 3'UTR dominance remained the overarching pattern.

Importantly, direct comparison of predicted CDS miRNA binding coordinates with somatic missense mutation sites identified only three overlap events (DIS3 in KIRC, FLT3 in LUAD, and PTPRT in STAD). While rare, these loci are potentially significant because they represent candidate sites where coding mutations may simultaneously alter protein structure and disrupt predicted miRNA-mediated regulation [31]. This rarity is itself biologically informative: most somatic missense mutations and most predicted miRNA regulatory interactions appear to operate independently, suggesting that direct

mechanistic convergence is uncommon rather than widespread. However, where such convergence occurs, it may represent disproportionately important regulatory vulnerabilities that warrant functional validation through luciferase assays, CLIP-seq, or transcriptomic analysis.

Together, these findings support a hierarchical model of miRNA-associated mutational biology in cancer. At the structural level, conserved GC content preserves the biochemical integrity of miRNA loci. At the mutational level, cancer-specific somatic processes reshape Ti/Tv ratios independently of this conserved baseline. At the regulatory level, miRNA interactions remain predominantly concentrated in canonical 3'UTRs, while rare CDS overlaps with missense mutations may define specialized sites of dual functional consequence. Rather than acting as isolated observations, these layers collectively illustrate how stable miRNA biology coexists with divergent tumor-specific mutagenesis to produce distinct regulatory landscapes across cancers.

Several limitations should be considered. First, miRNA binding site predictions were computationally derived using DIANA-microT and were not experimentally validated; therefore, predicted interactions and overlap events do not confirm functional disruption. Second, mutation counts were aggregated across TCGA cohorts without normalization for cohort size or tumor-specific mutation burden, which may influence direct cross-cancer comparisons. Third, Ti/Tv analyses were based on published mutation classifications and do not incorporate broader context-specific mutational frameworks such as trinucleotide signatures. Fourth, limited sample sizes in select cohorts (particularly BLCA and some strand-specific analyses) may reduce statistical power. Finally, while this study identifies associations between mutation patterns and miRNA-associated regions, it does not establish direct mechanistic consequences, which would require molecular or transcriptomic validation.

Future studies should investigate whether transversion-enriched mutational processes alter miRNA hairpin stability, strand selection, or target-binding capacity through RNA secondary-structure modeling and experimental assays. Additionally, the rare CDS-overlap loci identified here may serve as priority candidates for functional validation to determine whether somatic mutations can directly disrupt miRNA-mediated coding-region regulation. Clinically, cancer-specific Ti/Tv deviations in miRNA loci may also warrant exploration as biomarkers of mutational exposure or tumor-specific regulatory disruption [32].

Overall, this study establishes an integrated framework linking conserved miRNA sequence architecture with cancer-specific mutational dynamics and regulatory targeting. By demonstrating that stable structural constraints coexist with divergent mutational pressures and occasional sites of regulatory-coding convergence, these findings expand current understanding of how miRNA biology intersects with somatic mutation processes across cancers.

Conclusion

This study examined how somatic mutational patterns relate to miRNA regulatory regions across five cancer subtypes by integrating GC content, transition-transversion ratios, and miRNA-gene interaction analyses. Predicted miRNA-gene interactions were strongly enriched in 3'UTR regions across all cancer cohorts, with coding sequence interactions detected but consistently less frequent, and three candidate loci were identified where somatic missense mutations directly overlapped predicted CDS miRNA binding sites. Together, these analyses reveal that cancer-associated mutational processes affect miRNA regulatory regions in ways that extend beyond what standard mutation classification captures.

At the sequence level, GC content was largely conserved across cancer types, while Ti/Tv ratios varied substantially and fell below the expected baseline of 2.59-2.95 in all cohorts except STAD. This divergence demonstrates that cancer-specific mutational signatures, such as tobacco smoke-associated G→T transversions in LUAD, reshape miRNA loci independently of underlying sequence composition. The predominance of miRNA interactions in 3'UTR regions suggests that this conserved regulatory architecture remains intact across cancer [3, 5, 13]. However, the rare convergence of somatic missense mutations with predicted CDS miRNA binding sites indicates that coding mutations may occasionally disrupt both protein function and post-transcriptional

regulation simultaneously. This is a dual consequence not captured by standard mutational annotation alone.

These findings carry potential biological and clinical significance. At the biological level, transversion enrichment in miRNA loci may systematically compromise miRNA regulatory capacity in cancers driven by specific environmental mutagens, providing a mechanistic link between mutagen exposure and post-transcriptional dysregulation. The identification of CDS overlap events further suggests that select somatic mutations may have compounded regulatory consequences beyond their effects on protein structure, representing a potentially underappreciated mechanism of gene regulatory disruption in cancer. At the clinical level, cancer-specific Ti/Tv profiles in miRNA regions may serve as candidate biomarkers for tumor mutational signature characterization, with potential utility in liquid biopsy applications where miRNA mutation patterns could inform tumor identity or treatment response [33-35]. Future experimental validation, including luciferase reporter assays for the identified CDS overlap loci and larger cohort analyses, will be necessary to determine whether these patterns have direct prognostic or therapeutic relevance.

Acknowledgements

We would like to thank Claire Shen for her feedback, which provided detailed comments and structural suggestions that enhanced the clarity and rigor of this manuscript. The results shown here are based in whole or in part on data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Disclosure of conflict of interest

None.

Address correspondence to: Julia Ran, Xiwen Zhang, Yongsheng Bai and Joyce Tang, Eastern Michigan University, Ypsilanti, MI 48197, USA. E-mail: juliawr496@gmail.com (JLR); alinazhxw2@gmail.com (XWZ); bioinformaticsresearchtomorrow@gmail.com (YSB); TangyJoyce@gmail.com (JT)

References

[1] Calin GA, Liu CG, Sevignani C, Ferracin M, Felli N, Dumitru CD, Shimizu M, Cimmino A, Zupo S, Dono M, Dell'Aquila ML, Alder H, Rassenti L, Kipps TJ, Bullrich F, Negrini M and Croce CM. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. Proc Natl Acad Sci U S A 2004; 101: 11755-11760.

[2] de Moor CH, Meijer H and Lissenden S. Mechanisms of translational control by the 3' UTR in development and differentiation. Semin Cell Dev Biol 2005; 16: 49-58.

[3] Lai EC. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet 2002; 30: 363-364.

[4] Robins H and Press WH. Human microRNAs target a functionally distinct population of genes with AT-rich 3' UTRs. Proc Natl Acad Sci U S A 2005; 102: 15557-15562.

[5] Stark A, Brennecke J, Bushati N, Russell RB and Cohen SM. Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell 2005; 123: 1133-1146.

[6] Sun M, Hurst LD, Carmichael GG and Chen JJ. Evidence for a preferential targeting of 3'-UTRs by cis-encoded natural antisense transcripts. Nucleic Acids Res 2005; 33: 5533-5543.

[7] Singh R, Ha SE, Yu TY and Ro S. Dual roles of miR-10a-5p and miR-10b-5p as tumor suppressors and oncogenes in diverse cancers. Int J Mol Sci 2025; 26: 415.

[8] Di Martino MT, Tagliaferri P and Tassone P. MicroRNA in cancer therapy: breakthroughs and challenges in early clinical applications. J Exp Clin Cancer Res 2025; 44: 126.

[9] Koh G, Degasperi A, Zou X, Momen S and Nik-Zainal S. Mutational signatures: emerging concepts, caveats and clinical applications. Nat Rev Cancer 2021; 21: 619-637.

[10] Brummer A and Hausser J. MicroRNA binding sites in the coding region of mRNAs: extending the repertoire of post-transcriptional gene regulation. Bioessays 2014; 36: 617-626.

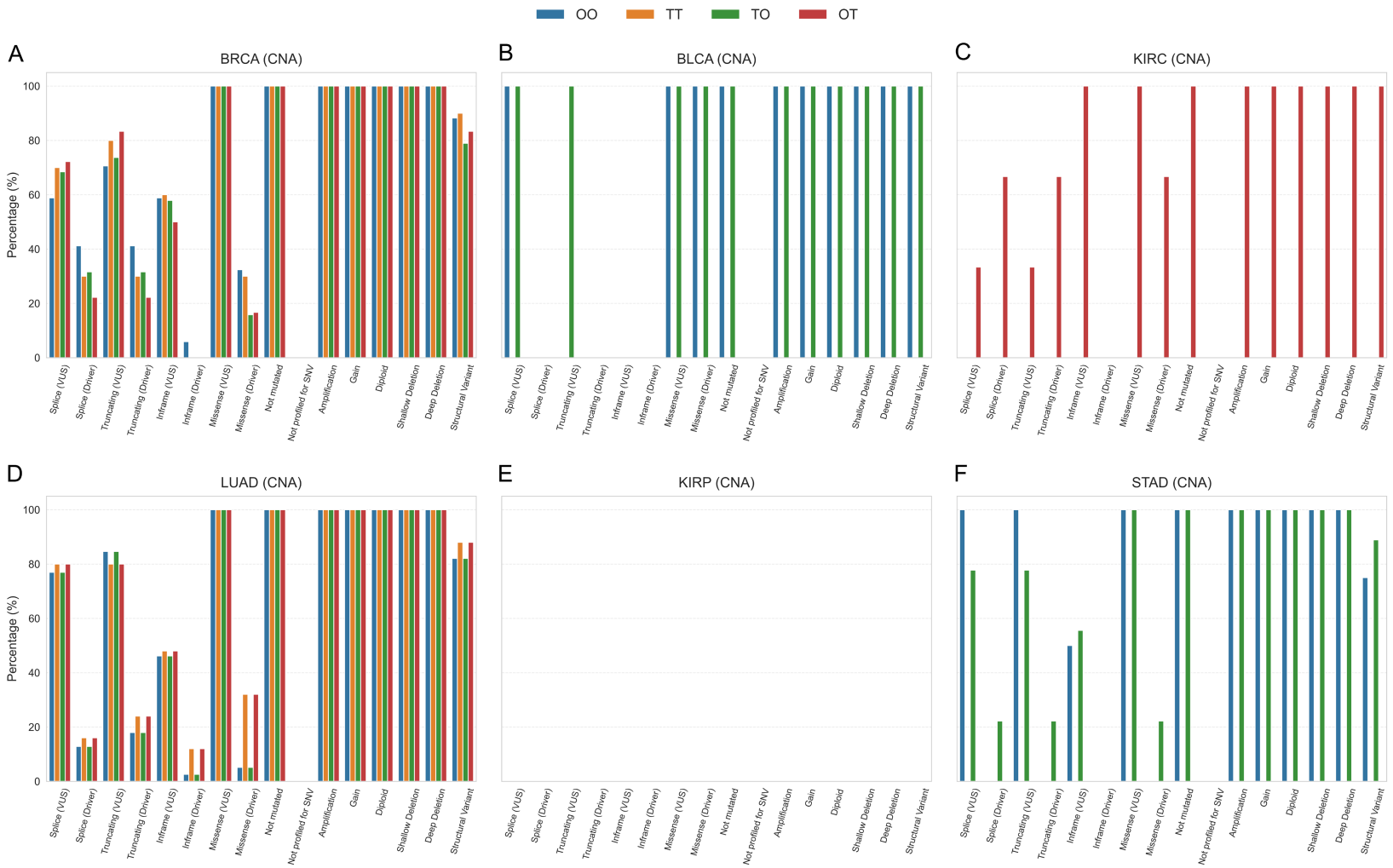
[11] Fang Z and Rajewsky N. The impact of miRNA target sites in coding sequences and in 3'UTRs. PLoS One 2011; 6: e18067.

[12] Anoosha P, Sakthivel R and Gromiha MM. Exploring preferred amino acid mutations in cancer genes: applications to identify potential drug targets. Biochim Biophys Acta 2016; 1862: 155-165.

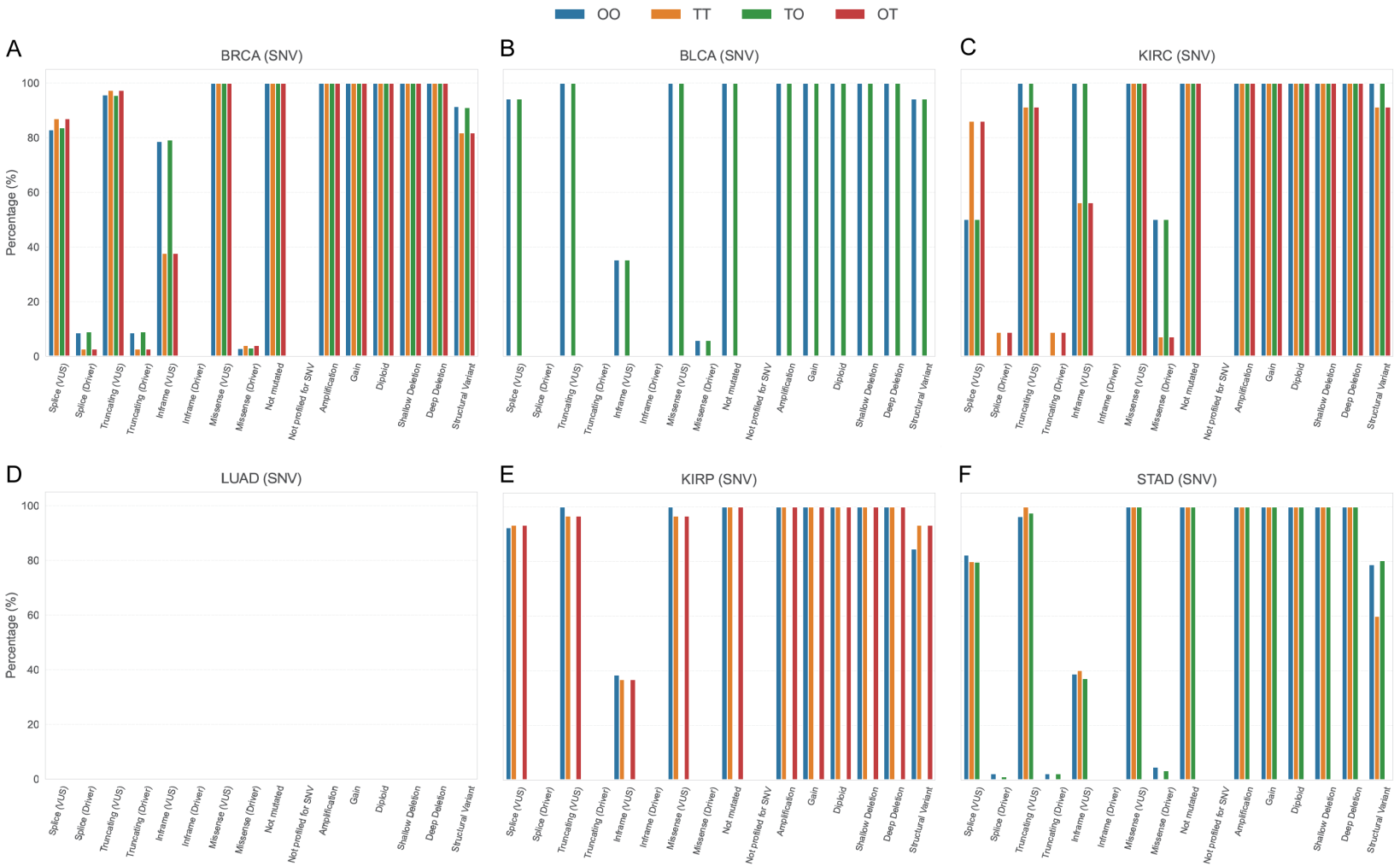
[13] Hausser J and Zavolan M. Identification and consequences of miRNA-target interactions: beyond repression of gene expression. Nat Rev Genet 2014; 15: 599-612.

[14] Diederichs S, Bartsch L, Berkmann JC, Fröse K, Heitmann J, Hoppe C, Iggena D, Jazmati D, Karschnia P, Linsenmeier M, Maulhardt T, Möhrmann L, Morstein J, Paffenholz SV, Röpenack P, Rückert T, Sandig L, Schell M, Steinmann A, Voss G, Wasmuth J, Weinberger

ME and Wullenkord R. The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. EMBO Mol Med 2016; 8: 442-457.
[15] Sved J and Bird A. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proc Natl Acad Sci U S A 1990; 87: 4692-4696.
[16] Urbanek-Trzeciak MO, Galka-Marciniak P, Nawrocka PM, Kowal E, Szwec S, Giefing M and Kozlowski P. Pan-cancer analysis of somatic mutations in miRNA genes. EBioMedicine 2020; 61: 103051.
[17] Shen C, Geng R, Zhu S, Huang M, Liang J, Li B and Bai Y. Characterization of tumor suppressors and oncogenes evaluated from TCGA cancers. Am J Clin Exp Immunol 2024; 13: 187-194.
[18] Wang J, Raskin L, Samuels DC, Shyr Y and Guo Y. Genome measures used for quality control are dependent on gene function and ancestry. Bioinformatics 2015; 31: 318-323.
[19] Van Hoeck A, Tjoonk NH, van Boxtel R and Cuppen E. Portrait of a cancer: mutational signature analyses for cancer diagnostics. BMC Cancer 2019; 19: 457.
[20] Zeng Y and Cullen BR. Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. J Biol Chem 2005; 280: 27595-27603.
[21] Zhang X and Zeng Y. The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res 2010; 38: 7689-7697.
[22] Kim B, Jeong K and Kim VN. Genome-wide mapping of DROSHA cleavage sites on primary microRNAs and noncanonical substrates. Mol Cell 2017; 66: 258-269, e5.
[23] Fang W and Bartel DP. The menu of features that define primary microRNAs and enable de novo design of microRNA genes. Mol Cell 2015; 60: 131-145.
[24] Duchene S, Ho SY and Holmes EC. Declining transition/transversion ratios through time reveal limitations to the accuracy of nucleotide substitution models. BMC Evol Biol 2015; 15: 36.
[25] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014; 513: 202-209.
[26] Cortes-Ciriano I, Lee S, Park WY, Kim TM and Park PJ. A molecular portrait of microsatellite instability across multiple cancers. Nat Commun 2017; 8: 15180.
[27] Zhang Z, Liu Z, Gu Y, Xu Y, Zhu L and Zheng Z. Microsatellite instability in gastric cancer: molecular features and clinical implications. Clin Cancer Bull 2024; 3: 12.
[28] Puliga E, Corso S, Pietrantonio F and Giordano S. Microsatellite instability in gastric cancer: between lights and shadows. Cancer Treat Rev 2021; 95: 102175.
[29] Farmanbar A, Kneller R and Firouzi S. Mutational signatures reveal mutual exclusivity of homologous recombination and mismatch repair deficiencies in colorectal and stomach tumors. Sci Data 2023; 10: 430.
[30] Nana-Sinkam SP and Croce CM. MicroRNA regulation of tumorigenesis, cancer progression and interpatient heterogeneity: towards clinical use. Genome Biol 2014; 15: 445.
[31] Kim SY, Na MJ, Yoon S, Shin E, Ha JW, Jeon S and Nam SW. The roles and mechanisms of coding and noncoding RNA variations in cancer. Exp Mol Med 2024; 56: 1909-1920.
[32] Chen B, Dragomir MP, Yang C, Li Q, Horst D and Calin GA. Targeting non-coding RNAs to overcome cancer therapy resistance. Signal Transduct Target Ther 2022; 7: 121.
[33] Calin G and Croce C. MicroRNA signatures in human cancers. Nat Rev Cancer 2006; 6: 857-866.
[34] Schwab S and Nonaka T. Circulating miRNAs as liquid biopsy biomarkers for diagnosis in patients with colorectal cancer: a systematic review and meta-analysis. Front Genet 2025; 16: 1574586.
[35] Cheng H, Shi X, Wang L, Li J and Sun Y. A liquid biopsy signature predicts treatment response to fluoropyrimidine plus platinum therapy in patients with metastatic or unresectable gastric cancer. NPJ Precis Oncol 2022; 6: 2.



Supplementary Figure 1. Significant mutation analysis of CNA.



Supplementary Figure 2. Significant mutation analysis of SNV.