False discovery rate control: Moving beyond the Benjamini–Hochberg method

Nilanjana Laha[1*], Salil Koner[2], and Navonil De Sarkar[3*]

[1]Department of Statistics, Texas A&M, College Station, Texas, USA

[2]*Department of Bioinformatics, Duke University, Durham, North Carolina, USA*

[3]*Prostate Cancer Center of Excellence, Cancer Center; Medical College of Wisconsin, Milwaukee, Wisconsin, USA*

Correspondence*:
Corresponding Authors
nlaha@tamu.edu (Nilanjana Laha), ndesarkar@mcw.edu (Navonil De Sarkar)

ABSTRACT

Modern bioinformatics studies often involve numerous simultaneous statistical tests, increasing the risk of false discoveries. To control the false discovery rate (FDR), these studies typically employ a statistical method called the Benjamini–Hochberg (BH) method. Often, the BH approach tends to be overly conservative and overlooks valuable biological insights associated with data structures, particularly those of groups. Group structures can manifest when closely located genomic coordinates are functionally active and closely related because of co-regulation. Recent statistical advancements have led to the development of updated BH methods tailored for datasets featuring pre-existing group structures. These methods can improve the statistical power and potentially enhance scientific discoveries. In this study, we elucidated the advantages of contemporary group-aware BH methods using a previously published microRNA (miRNA) dataset. For this dataset, group-aware BH methods identified a larger set of miRNAs with significantly deregulated expression (p-value <0.05) than the traditional BH method. These new findings are supported by existing literature on miRNAs and a related 2017 study. Our results underscore the potential of specialized BH methods for controlling the FDR in high throughput omics studies with pre-defined group structures.

1 INTRODUCTION

High throughput and massively parallel genome-scale molecular profiling technologies transformed the field of genomic and precision medicine by making top-down molecular mapping approaches feasible. For example, RNA-seq or microarray studies routinely analyze several thousands of genes and gene regulatory elements to identify mRNA expression in both normal and aberrant tissues under disease conditions [1, 2]. In genome-wide association studies (GWASs), numerous genetic variants are simultaneously examined at the population level to explore their potential associations with specific traits [3]. Statistical analysis of such data often requires simultaneous evaluation of many hypotheses, each corresponding to one biological element, such as one gene expression in a microarray or RNA-seq study or a single nucleotide polymorphism in a GWAS. The rejection of a hypothesis is generally associated with biologically relevant elements, indicating statistical “discovery.” However, the simultaneous testing of multiple hypotheses using statistical tests leads to an increased likelihood of falsely rejecting a hypothesis, resulting in a high number of false discoveries, necessitating additional statistical methods for their control [4]. Although several strategies have been proposed for addressing the false discovery rate in multiple hypothesis-testing problems, the false discovery rate (FDR) control method introduced by Benjamini and Hochberg [5] has become the most widely used method in high throughput bioinformatics analysis [6]. The FDR is the expected proportion of false rejections among all

rejected hypotheses. In other words, if the expressions of 100 genes are found to be significantly deregulated following the statistical analysis of a microarray study, a 5% FDR indicates that 5/100 identified genes are false discoveries and that their expression is not really deregulated.

In the past decades, various statistical methods have been proposed for controlling the FDR. However, the Benjamini and Hochberg (BH) procedure remains the most prominent tool for this purpose in the realms of bioinformatics [5]. The BH method, initially described for independent hypotheses, preserves the FDR provided the hypotheses are weakly dependent [7]. The latter assumption generally holds for most biological applications [7]. The BH method relies solely on p-values of multiple hypothesis tests. It provides a p-value cutoff, and only tests when lower p-values are rejected. Although the BH method reliably controls the FDR, it is known to be conservative when the number of hypotheses to be tested is much larger than sample size, which is common in many modern biological studies, including high throughput studies [8, 9]. The high dimensionality of the data increases the noise within the data, making it harder to distill a weaker signal from the noise. The conservativeness of the BH method can therefore potentially hinder scientific discoveries based on modern data.

The conservativeness of the BH method stems from its inability to utilize the available biological information in the hypotheses [6]. It treats all hypotheses as exchangeable, ignoring any extra information on the dependence between the hypotheses [10]. However, in some domains, such as bioinformatics, extra information for distinguishing the hypotheses may be available. This study focuses on hypotheses with pre-existing group structures. These groups often arise naturally in biological data. For example, closely located genes are often correlated and may be co-expressed because of their shared transcription programs and regulatory switches, thus forming positional clusters [11]. Several studies have repeatedly shown that grouped FDR control methods that exploit the information on groups exhibit higher statistical power than the BH method with such data [6, 8, 9]. A higher statistical power indicates the ability to detect lower signals, thus potentially increasing scientific discoveries without compromising the FDR. The grouped BH method achieves this by recalibrating the p-values to place more importance on potentially critical groups. However, such group-aware FDR methods are rarely adopted in high throughput data analysis by bioinformaticians.

Briefly, a group's criticality is determined by the proportion of true null hypotheses within it; smaller proportions indicating greater criticality. Grouped BH methods prioritize rejections from the potential critical groups, estimating groupwise null proportions via statistical methods, such as the two-stage method (TST), least-slope method (LSL), and likelihood-based methods. This results in the grouped BH methods of TST-GBH, LSL-GBH, and structure adaptive BH algorithm (SABHA), respectively [8, 9]. The TST-GBH and LSL-GBH methods were previously applied by Zhang et al. [12] to discover differentially methylated regions in the human genome and by Liu et al. [13] to discover differentially expressed unigenes during the blooming process of Asteraceae flowers. SABHA is a more general method that works for a range of structured data and is not limited to grouped hypotheses.

When the p-values are independent or weakly correlated, the grouped BH methods reportedly control the FDR at the desired level if group-wise null proportions are reasonably well-estimated [8, 9]. Apart from the grouped BH methods, an emerging statistical technique called knock-off has also shown promise for FDR control and is flexible enough to incorporate group structures [3]. Notably, the abovementioned methods do not use groups as the unit of rejection. Instead, rejections are conducted at the level of the individual hypotheses, where information on the groups is used to improve statistical power. However, FDR control methods that reject hypotheses at the group level are also available [14, 15].

This study demonstrates the advantage of grouped BH methods by utilizing the microRNA (miRNA) data from a study by De Sarkar et al. [16] on 18 cancer patients with gingival buccal squamous cell carcinoma (GBSCC). In the study, the authors determined the expression of several miRNAs to investigate their role in this type of oral carcinogenesis. They compared the miRNA expression in histopathologically confirmed malignant and healthy normal (control) tissues using one-sided paired t-tests. As there were 522 miRNA assays in the reverse transcription polymerase chain reaction-based miRNAome panel, 522 t-tests were performed. The BH method was used to control the FDR at 5%, and seven miRNAs whose expression were significantly deregulated in the malignant tissues were identified. However, some critical miRNAs for oral squamous cell carcinoma (OSCC), including hsa-miR-21-5p and hsa-miR-99a-3p, were either borderline or below the BH threshold. Some of the target genes of these miRNAs were significantly deregulated by Singh et al. [17] in a follow-up whole transcriptome analysis study, which analyzed a case series substantially overlapping with the current study case series. This observation raised the question of whether the BH method has enough statistical power to detect miRNAs with deregulated expression in this De Sarkar et al. [16] case series.

Considering our limited sample size (n=18), the number of miRNAs analyzed (n=522), and their genome wide distribution pattern, we opted for a relatively straightforward grouping approach in this proof-of-concept analysis, focusing on the physical location of miRNAs on the chromosome. The broad assumption was proximal miRNAs tend to be co-expressed [11]. Therefore, we divided the chromosomes into positional groups based on their shared chromosome number, arm, and strand. We incorporated this group information into the statistical analysis of our data via the grouped BH methods LSL-GBH, TST-GBH, and SABHA, and compared their performance to that of the BH method. Existing literature and whole transcriptome analysis by Singh et al. [17] were used to support the findings.

2 RESULTS

The study participants were 18 unrelated Indians aged 39-80 years with tobacco habits, with a male:female ratio of 5:4. All patients had histopathologically confirmed GBSCC, which is a type of OSCC prevalent in the tobacco-chewing population in South Asia. The dataset comprised ΔCt values for 522 miRNAs derived from 18 tumor-normal sample pairs of the 18 patients. ΔCt values serve as surrogate estimates for relative miRNA expression levels (relative to geometric mean expression of three endogenous control genes). The data collection method was previously described [16].

2.1 Missingness:

Some miRNAs were not expressed in all 18 patients in the data from De Sarkar et al. [16]. For each miRNA, if one data point in the paired miRNA expression data fell below the detection threshold, De Sarkar et al. [16] excluded both data points to reduce complexity. Consequently, some miRNA pairs were missing in the miRNA expression dataset, with 59.77% of miRNAs having at least one pair of missing observations. Moreover, 30% of miRNAs had more than two pairs of missing observations. Each patient had 12 or more missing miRNA expression pairs, with the majority having fewer than 50 missing pairs. For further details on missingness, refer to Supplementary Section S1.

The missingness in the miRNA expressions could have stemmed from low detection threshold which led to the inability to detect low miRNA expression. The parallel miRNA expression assay was conducted using TLDA-A (V2) and TLDA-B (V3) cards on the 7900HT FAST Real Time PCR system (Applied Biosystems, USA) with the TLDA flat block, which is known to have a fixed lower detection limit [16]. Although this could be one of the reasons for data missingness, various other factors could have also contributed to the data-missingness. Nevertheless, it is evident that the missingness of the miRNAome data was not at random (MNAR) [18]. Although De Sarkar et al. [16] imputed the missing

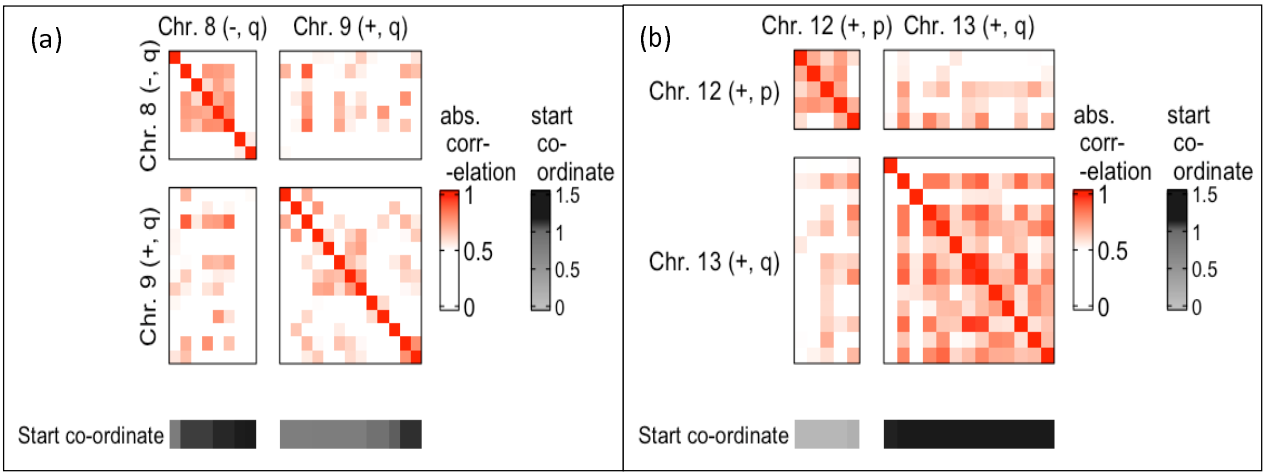
observations using a simple median imputation strategy in their statistical analysis, we observed a sharp increase in the number of discoveries regardless of the FDR control method when using median or mean imputation. This is unsurprising because it is well-known that imputation with the median or mean can artificially decrease the sample variance, especially when the sample size is small, leading to false discoveries [19, 20]. Given our limited sample size and the complexity of measuring all biological factors behind miRNA expression, we chose not to impute missing values. Instead, we conducted our statistical analysis using only complete cases. Hughes et al. [18] showed that complete case analysis can be statistically valid in MNAR data.

2.2 Grouping:

As mentioned previously, we grouped the miRNAs based on their shared chromosome number, arm, and strand, resulting in 67 groups after excluding groups with no membership among our 522 miRNAs. Supplementary Table S1 shows the number of miRNAs in each group. Some groups were thinly populated with only two to five members, but merging them with the groups on the other arm on the same strand yielded only a nominal difference—only hsa-miR-486-5p ceased to be statistically significant after merging. Most groups were of small to moderate size, but three groups had more than 30 miRNAs each, located on chromosome 14 (positive strand), chromosome 19 (positive strand), and the X chromosome (negative strand); see Supplementary Table S2. Splitting these large groups into adjacent equal halves does not alter the outcome of subsequent statistical analyses. In consideration of these factors, we implemented the grouped BH methods using the original 67 groups.

Groupings are known to be statistically meaningful only when the intra-group association (association between test units within the same group) is stronger than the inter-group association (association between test units in different groups) [21]. Hu et al. [18] observed that the group BH methods demonstrated higher statistical power than the BH method in such scenarios. We heuristically compared the intra-group and inter-group associations between the $\Delta\Delta$Ct values utilizing a widely used statistical approach—a simple random effect model [22]. This model facilitates the quantification and estimation of inter-group and intra-group correlations, which measures the association within and across the groups, respectively [23, 24]. The intra-group correlation corresponds to the correlation between $\Delta\Delta$Ct values of miRNAs within the same group, where the inter-group correlation is the correlation between $\Delta\Delta$Ct values of miRNAs belonging to two different groups. The simple random effect model assumes the inter-group and intra-group correlations to be constant across all groups. Even if these assumptions are violated, the estimated correlations provide meaningful assessments of the overall strength of the inter-group and intra-group association. Details on fitting the simple random effect model can be found in Section *4.3* *4.3* The estimated inter-group correlation was 0.04, indicating a weak inter-group association. In contrast, the intra-group correlation was estimated to be 0.14, noticeably larger than the inter-group correlation. Thus, the simple random effect model suggested a weaker inter-group association compared to intra-group association. However, visual examination of pairwise correlations between $\Delta\Delta$Ct values revealed that miRNAs within the same group tended to exhibit higher correlations compared to those across different groups, supporting the findings of the heuristic analysis. Figure 1 illustrates two cases where overall within-group correlations are higher compared to correlations between miRNA pairs belonging to different groups. In these instances, we also observed that the $\Delta\Delta$Ct values of adjacent miRNAS tended to cluster, aligning with the assertion that adjacent miRNAs are co-expressed.

Figure 1. Heatmap of the absolute correlations between the ΔΔCt values of the miRNAs from two groups. *(a) The first group corresponds to the q arm of the positive strand of chromosome 9 and the second group corresponds to the q arm of the negative strand of chromosome 8. (b) The first group corresponds to the p arm of the positive strand of chromosome 12 and the second group corresponds to the miRNAs on the q arm of the positive strand of chromosome 13. The diagonal blocks in Figures 1a and 1b provide the heatmaps of the absolute correlation between the ΔΔCt values of miRNAs within the same group, whereas the off-diagonal blocks showcase the absolute correlation between ΔΔCt values of miRNAs from different groups. Only absolute correlations above 0.5 are presented in the heatmaps. The ordering of miRNAs is based on their positions on the chromosome, ensuring that adjacent miRNAs are positioned next to each other in the heatmap. The sidebars below the heatmaps provide the start coordinates of the miRNAs.*



2.3 Significantly deregulated miRNAs

Error! Reference source not found. lists the deregulated miRNAs identified by TST-GBH, LSL-GBH, SABHA, and BH methods. **Error! Reference source not found.** shows a heatmap of the ΔΔCt values of these miRNAs. As reported by Sarkar and Zhao [20], the TST-GBH method was the most liberal of the four methods, leading to the highest number of discoveries. **Error! Reference source not found.** includes the missing data frequencies for the identified genes. Notably, 50% of the data for hsa-miR-1293 was missing, rendering the conclusions drawn regarding this miRNA questionable. The LSL-GBH and SABHA methods agreed with each other, with five rejections each. The BH method had the lowest number of discoveries, with only three rejections. Notably, De Sarkar et al. [16] also used the BH method but reported seven discoveries because they used median-imputed data. As mentioned previously, median imputation inflates the number of discoveries for all methods. Supplementary Table S1 lists the miRNAs detected by the four FDR control methods using the median-imputed data.

The grouped BH method identified seven additional miRNAs alongside the seven miRNAs previously identified by De Sarkar et al. (**Error! Reference source not found.**) [16]. Among the newly discovered miRNAs, the expression of hsa-miR-1247-5p, hsa-miR-1, hsa-miR-99a-3p, and hsa-miR-486 3p/5p was downregulated, whereas that of hsa-miR-21-5p and hsa-miR-147b was upregulated in GBSCC cancer

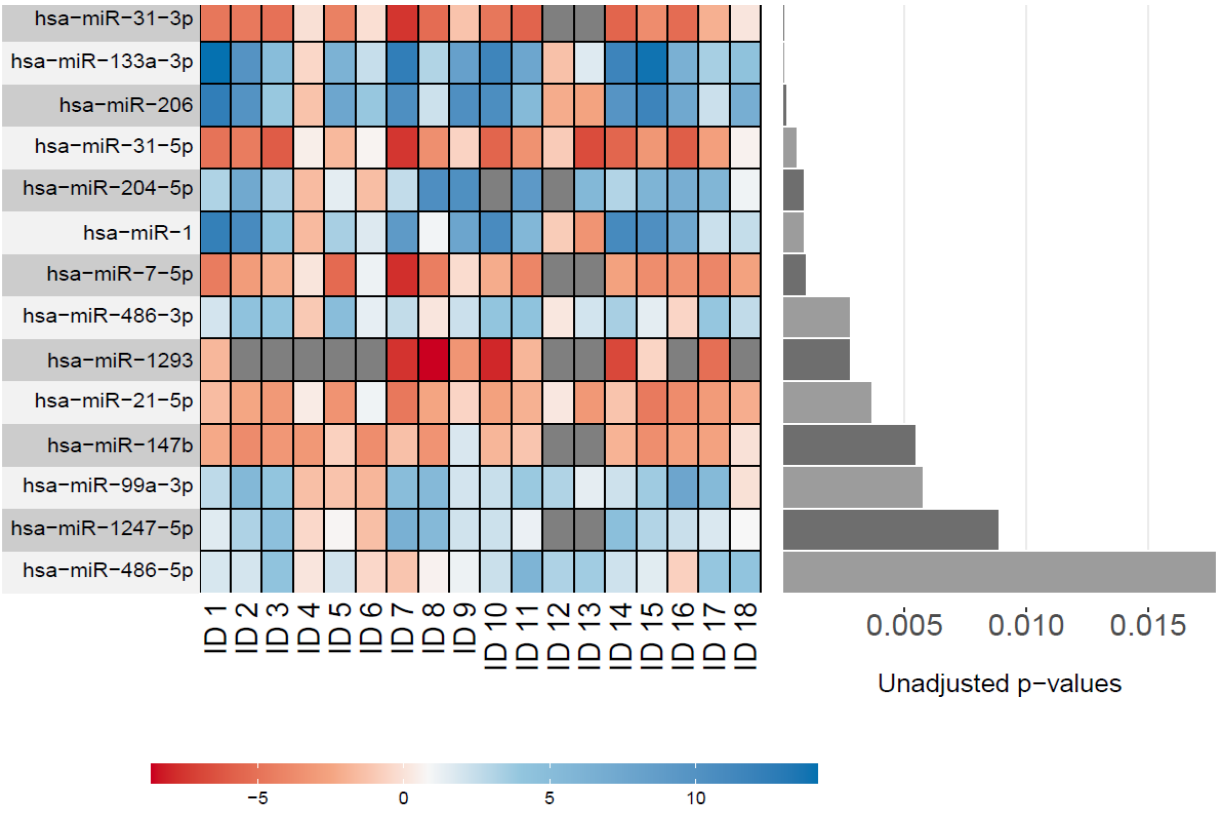
tissues. Although De Sarkar et al. [16] did not identify these miRNAs, the p-value of hsa-miR-1 was just below the BH-corrected level of

significance in their analysis. Moreover, via a cluster analysis, De Sarkar et al. [16] identified a group of 13 patients with 30 deregulated miRNAs in cancer tissues (see Table 6 in De Sarkar et al. [16]). Interestingly, all the newly detected miRNAs, except hsa-miR-486-3p/5p, belonged to this set of 30 miRNAs.

miRNA	Chr.	Strand	Arm	ΔΔCt	p-value	Missing pairs	BH	TST	LSL	SABHA
hsa-miR-133a-3p*	chr18	-	q	6.7	$5.32E^{-05}$↓	0	x	x	x	x
hsa-miR-206*	chr6	+	p	6.0	$1.97E^{-04}$↓	0	x	x		x
hsa-miR-31-3p*	chr9	-	p	-3.8	$1.31E^{-04}$↑	2	x	x	x	x
hsa-miR-1293*	chr12	-	q	-4.8	$2.77E^{-03}$↑	9		x		
hsa-miR-1247-5p	chr14	-	q	2.4	$8.93E^{-03}$↓	2		x		
hsa-miR-147b	chr15	+	q	-2.1	$5.52E^{-03}$↑	2		x		
hsa-miR-7-5p*	chr15	+	q	-3.1	$9.50E^{-04}$↑	2		x	x	
hsa-miR-21-5p	chr17	+	q	-2.2	$3.67E^{-03}$↑	0		x		
hsa-miR-1	chr18	-	q	5.2	$9.38E^{-04}$↓	0		x	x	x
hsa-miR-99a-3p	chr21	+	q	2.8	$5.83E^{-03}$↓	0		x		
hsa-miR-486-3p	chr8	-	p	2.4	$2.81E^{-03}$↓	0		x		
hsa-miR-486-5p^{+}	chr8	-	p	2.0	$1.78E^{-02}$↓	0		x		
hsa-miR-31-5p*	chr9	-	q	-3.4	$6.18E^{-04}$↑	0		x	x	x
hsa-miR-204-5p*	chr9	-	q	4.6	$8.81E^{-04}$↓	2		x		

Table 1. microRNAs (miRNAs) with significantly deregulated expression. The miRNAs listed here were detected using at least one of the four methods tested: BH, TST-GBH (abbreviated as TST), LSL-GBH (abbreviated as LSL), and SABHA. The "Chr." column indicates the chromosome number; the "strand" column indicates the transcribing strand; and the "arm" column indicates the arm of the miRNA gene. The ΔΔCt values are used as surrogate estimates of differential expression for each miRNA. As TST-GBH, LSL-GBH, and SABHA do not yield FDR-corrected p-values, we have reported the raw p-values before FDR correction. The arrow signs after the p-values indicate whether the expression of the miRNAs was upregulated (↑) or downregulated (↓) (per the sign of ΔΔCt). The "Missing pairs" column shows the number of missing observations for each miRNA. The "Significant tests" column indicates the methods that detected a miRNA as one with significantly deregulated expression. The expressions of miRNAs superscripted with an asterisk were reported to be deregulated by De Sarkar et al. [16]. BH, Benjamini–Hochberg; SABHA, structure adaptive Benjamini–Hochberg algorithm; TST-GBH, two-stage step-up–group Benjamini–Hochberg; LSL-GBH, least-squares group Benjamini–Hochberg.

Figure 2. ***Heatmap representing the ΔΔCt values of the deregulated miRNAs in*** *Error! Reference source not found.* ***displaying their patterns of deregulation in cancer tissue samples from 18 gingival buccal squamous cell carcinoma patients****. The miRNAs are ordered as per unadjusted p-values obtained from the paired t-tests.*



2.4 Relevance of the newly discovered miRNAs

Although it would have been desirable to experimentally validate the pathways downstream of the newly detected deregulated miRNAs, such analyses were beyond the scope of the current study because of the exhaustion of biospecimens used by De Sarkar et al [16]. However, several other functional works provided supporting evidence for the role of these newly "discovered" miRNAs in carcinogenesis. In particular, the availability of the whole transcriptome analysis data from Singh et al. (2017) allowed for correlative analysis, and our findings showed concordance with previously published onco-miRNAome work. We utilized these published studies and Singh et al. [17] whole-transcriptome analysis to analyze our findings. Singh et al. assessed 12 cancer tissues, with 10 overlapping with the De Sarkar et al. miRNAome study [16]. Notably, the expression of 5 of the seven newly detected miRNAs was deregulated in the opposite direction of their target genes, as reported in Singh et al. [17] (Table 2). Singh et al. [17] evaluated 12 cancer tissues, 10 of which were included in the miRNAome study of De Sarkar et al. [16]. Expectedly, the expression of five (of the seven) newly detected miRNAs was deregulated in the opposite]direction of their target genes, and reported in Singh et al. [17] (**Error! Reference source not found.**). This observation concordantly supports the validity of our discovery. Below, we summarize the relevance of these five miRNAs. The remaining two miRNAs, hsa-miR-1247-

5p and hsa-miR-147b, also exhibited oncogenic activity. However, there is limited evidence supporting their involvement in OSCC.

miRNA	Deregulated target genes found	References
hsa-miR-21-5p	MEF2C (↓), PDCD4 (↓), TIMP3 (↓), PPARA (↓), RECK (↓), SPRY1 (↓), SPRY2 (↓), THRB (↓)	[28]
hsa-miR-1	FAM101B (↑), TRANK1 (↑), ZNF281 (↑), Runx2 (↑) CXCR4 (↑)	[36], [38], [53], [54]
hsa-miR-486-3p	FLNA (↑)	[43]
hsa-miR-486-5p	KIAA1199 (↑)	[38], [43], [44]
hsa-miR-99a-3p	BCAT1 (↑), MTHFD2 (↑), RRM2 (↑)	[38]

Table 2. Target genes of the newly discovered miRNAs with deregulated expressions, as identified by Singh et al. (2017). Arrows next to the genes indicate the direction of deregulation. Information on the target genes of each miRNA was obtained from the corresponding references.

Among the newly detected miRNAs, hsa-miR-21-5p (upregulated; **Error! Reference source not found.**), and hsa-miR-99a-3p (downregulated; **Error! Reference source not found.**) have been extensively studied in OSCC [25-27]. The miRNA hsa-miR-21 is a ubiquitous oncogene with prognostic potential for OSCC [28-31]. The analysis by Singh et al. [17] identified eight target genes of hsa-miR-21 (MEF2C, PDCD4, TIMP3, PPARA, RECK, SPRY1, SPRY2, and THRB) with downregulated expression, further supporting our new "discovery" (**Error! Reference source not found.**). The downregulation of the expression of these genes has been linked to various hallmarks of cancer [32]. Notably, PDCD4 is one of the main target genes of hsa-miR-21 [33]. The miRNA hsa-miR-99a-3p acts as a tumor suppressor in some squamous cell carcinomas. Three oncogenes regulated by this miRNA (BCAT1, MTHFD2, and RRM2) were significantly upregulated in Singh et al. [17]. These oncogenes are associated with poor cancer prognosis [34, 35].

The list of the seven newly detected miRNA included three additional miRNAs with potential tumor-suppressing activities: hsa-miR-1, hsa-miR-486-3p, and hsa-miR-486-5p [36, 37]. Among them, hsa-miR-1 is a well-established tumor suppressor, and is considered a cancer biomarker [38]. Its target genes *Runx2* and *CXCR4*, which exhibited significantly upregulated expression in the analysis by Singh et al. [17] (**Error! Reference source not found.**). RUNX2 is implicated in breast cancer [39, 40], CXCR4 in thyroid cancer and osteosarcoma [41], and ZNF281 in pancreatic cancer [42]. However, FLNA and KIAA1199, target genes of hsa-miR-486-3p and hsa-miR-486-5p, respectively, which were upregulated in Singh et al. [17] (**Error! Reference source not found.**). FLNA promotes cell migration in laryngeal squamous cell carcinoma [43] and KIAA1199 is a well-known promoter of invasion and metastasis in thyroid cancer [44].

3 DISCUSSION

The importance of adopting an accurate statistical approach is well-recognized in the field of high-throughput biological data analytics. High throughput biological data analysis often requires the adoption of a multitude of statistical tests for a large number of “hypothesis tests” conductd simultaneously, increasing the risk of false discoveries. Controlling the FDR allows investigators to strike a balance between identifying potentially meaningful findings (true positive findings/discoveries) and minimizing the likelihood of spurious results. Researchers have simultaneously examined numerous genes, proteins, and other biomarkers. The BH method is arguably the most widely used approach adopted in high-throughput bioinformatics analysis for FDR control [45]. However, unless the tests are independent or weakly dependent, the BH method compromises statistical power as it remains agnostic to any information on the association between test units [8, 9]. For example, in the study on miRNAs [16], well-known oncogenes, such as hsa-miR-1, failed to meet the FDR cutoff with the stringent BH method. Nevertheless, a subsequent study by Singh et al. [17] revealed that several key hsa-miR-1 target oncogenes were significantly upregulated in GBSCC tumors [16, 17; shared significant sample overlap].

The grouped BH method is a modification of the BH method, which incorporates information on group structures to improve power [8, 9]. We utilized grouped BH methods, incorporating a simple miRNA grouping strategy into the FDR control step, to improve the statistical power and demonstrate a proof of principle. In this proof-of-principle study, we chose a dataset in which miRNAome data was generated independent of the total transcriptome. The sample size was smaller than the number of hypothesis tests, and data generation followed rigorous analysis techniques. The dataset by De Sarkar et al. [16] fulfilled all the required criteria for our study. During analysis, we considered a simplistic group structure in miRNA expressions. We grouped all neighboring miRNAs on the chromosome that could potentially be co-regulated. Although we acknowledge that miRNA associations can arise for different reasons, we opted for this simple approach considering the limited sample size. We also recognize that the adopted positional clustering approach is overly simplistic, but our intent was to illustrate how relevant biological relatedness can improve statistical inference. Despite our simplistic grouping strategy, the grouped BH methods adjusted the q-values to include a few additional miRNAs whose expression was speculated to be deregulated in OSCC (**Error! Reference source not found.**). These new statistical findings were supported by follow-up transcriptome results and independent oral cancer miRNAome study results. The new detections may be attributed to the gain in statistical power from the incorporation of positional information for the miRNAs. However, the sensitivity of the TST-GBH method to missing data, as evidenced by the inclusion of hsa-miR-1293 in **Error! Reference source not found.**, may be of concern when the sample size is small.

In the present study, the grouping of miRNAs was based on broad spatial information. However, groupings can also occur based on other axis of functional or genetic relatedness information (e.g. ancestral relatedness). Liu et al. [13] used a gene ontology system for group assignments. In addition, Hu et al. [8] suggested using clustering algorithms, such as k-means, to partition the test units into groups when information on the group structure is less apparent. However, recent studies found that reusing the same data for clustering and downstream testing can inflate false discoveries [46, 47].

We remind the readers that in using grouped data, independence between groups is not a requirement. In our motivating example, the miRNAs did not function independently. However, for practitioners to fully benefit from grouped FDR methods, the association between units within groups must be stronger than that between units across groups [21]. Our heuristic analyses suggested that the intra-group association among our groups was stronger than the inter-group association. The grouped BH methods were advantageous over the traditional BH method also when: (a) most hypotheses are null, that is, when the expressions of the majority of miRNAs are not deregulated and (b) the groups are

heterogeneous in terms of the proportion of nulls across the groups [8, 48]. Based on the findings reported by De Sarkar et al. [16] the first scenario applies to this dataset. Verifying the second scenario poses challenges. However, the physical separation of our groups, coupled with stronger intra-group associations than inter-group associations, does not refute the possibility of heterogeneous groups in our case.

Our results underscore the promise of grouped BH methods for managing the FDR in bioinformatics studies characterized by pre-existing group structures. However, the extrapolation of our conclusions to future datasets may be restricted because of our limited sample size. Future experimental designs are required to ascertain the generalizability of our conclusion to different populations and different types of datasets. The replicability of our results in the resulting datasets remains a promising avenue for further investigation. Notably, although we have performed the group BH multiple test correction in the miRNAome dataset, this strategy might be applied to any multiple test scenario with pre-existing groups among the test hypotheses. Controlled future analyses are needed to test such generalizability. Of particular interest are the datasets pertaining to treatment-emerging cancer subtypes, which often have small sample sizes similar to our study [49, 50]. These types of cancers arise as a result of treatment-related factors, potentially as a consequence of the treatment itself or because of the underlying conditions being treated. Given that the genes responsible for such cancer subtypes are still not fully elucidated, future studies aiming to unveil these genetic factors will necessitate FDR control because of multiple hypothesis testing. Additionally, in the advancing era of precision medicine, we increasingly recognize the importance of a systematic approach to ensure durable therapy benefits to patients with minimal side effects. This approach provides a new dimension of information about gene-gene relatedness, making the relevance of GBH more practical and exceedingly important for precision discovery.

4 METHODS

Here we provide a brief overview of the data generation strategies, data preprocessing techniques used by De Sarkar et al. [16], and the implementation of the discussed statistical methods.

4.1 miRNAome data generation using TLDA assay

We downloaded publicly available ΔΔCt data from De Sarkar et al. study [16]. Very briefly, the data generation approach adopted by their study is as follows. TLDA is a high throughput quantitative PCR assay for relative quantification of standard and low expressing miRNAs in biospecimens [51]. The De Sarkar et al. [16] study adopted recommended RNA isolation protocol using mirVana kit (Life Technologies, USA). Samples were QCd for RIN value and ≥6.9 were chosen for this analysis. Ct, ΔCt and ΔΔCt values were derived using SDS and Data Assist (Life Technologies, USA) software packages, (Ct = Cycles at which the PCR product quantity reaches a defined threshold, ΔCt = $Ct_{\text{of a miRNA in cancer tissue}} - Ct_{\text{of geometric mean of expression of the three most stable endogenous control miRNAs in that tissue}}$ and ΔΔCT = $\Delta Ct_{\text{of a miRNA in cancer tissue}} - \Delta Ct_{\text{of that miRNA in control tissue}}$). Assayed miRNA expression was appropriately normalized and trimmed as per recommended TLDA data analysis SOP. De Sarkar et al. [16] pruned the data down to the expression of 531 miRNAs and recommended using 522 high-quality miRNA expression values (ΔΔCT) for future analysis.

4.2 Pairwise t-tests

Upregulation or downregulation of miRNA expression was determined by checking if the change of miRNA expression exceeded 2-fold compared to the control. We performed 522 one-sided pairwise t-tests to determine whether the expression of miRNAs was significantly deregulated, as described in De Sarkar et al [16]. Direction of the one-sided pairwise t-test was determined by checking if the median ΔΔCT value was positive or negative. The null hypothesis of the one-tailed paired t-test was ‘‘expression of a particular miRNA is not greater than 2 fold upregulated (or downregulated)’’.

4.3 Exploratory analysis using the simple random effect model

We used the simple random effect model to quantify and compare the intra-group and inter-group associations. To formalize the setup, we will introduce some new notations. We denote $\{Y_{ij} : j = 1, \ldots, 522\}$ to be the ΔΔCT values of the jth miRNA of the ith patient. Let g(j) be the group indicator of the jth miRNA. For example, if the jth miRNA belongs to the first group, then g(j) = 1. Because there are 67 groups, the group index g(j) takes values between 1 and 67.

We posit a random effect model [23] for the $\Delta\Delta CT$ values as

$$Y_{ij} = \mu + S_i + G_{ig(j)} + E_{ij},$$

where μ is the overall mean, S_i is the random effect of the ith patient, $G_{ig(j)}$ is the random effect of the $g(j)$th group for the ith patient, and E_{ij} is the random error for the ith patient and the jth miRNA. The random error can represent independent measurement errors or other underlying biological factors. We let the collections of random variables $\{S_i : i = 1, \ldots, 18\}$, $\{G_{ig(j)} : 1 \le i \le 18,\ 1 \le j \le 522\}$, and $\{E_{ij} : 1 \le i, j \le 522\}$ be independent of each other. We also assumed that $S_1, \ldots, S_{18}$ are independent and identically distributed centered Gaussian random variables with variance σ_s^2, $G_{ig(j)}$'s identically distributed independent centered Gaussian random variables with variance σ_g^2, and the E_{ij}'s are also identically distributed independent centered Gaussian random variables with variance σ_e^2. Under the above model, the correlation between

the $\Delta\Delta Ct$ values of any two miRNAs belonging to any two different groups within the same patient remained constant. The above correlation will thus be a natural candidate for the inter-group correlation. Straight-forward algebra shows that the inter-group correlation equals $\sigma^2_s / (\sigma^2_s + \sigma^2_g + \sigma^2_e)$. Likewise, under our model, the $\Delta\Delta Ct$ values of two miRNAs belonging to the same group within the same patient have identical correlations across the patients and the miRNAs – this correlation will be a natural candidate for the intra-group correlation. The intra-group correlation is given by $(\sigma^2_s + \sigma^2_g) / (\sigma^2_s + \sigma^2_g + \sigma^2_e)$. We fit the simple random-effect model using SAS proc mixed [52].

4.4 Implementation of BH and grouped BH methods

The BH method was implemented using the package *Stats in the software R.* The TST-GBH and LSL-GBH methods were implemented following the algorithms provided by Hu et al. [8]. To implement SABHA, we used the R codes provided by the authors of the corresponding paper by Li and Barber [9]. SABHA requires two tuning parameters: (1) □, a lower bound on the proportion of the true null hypotheses, and (2) a threshold τ, used for calibrating the individual p-value thresholds. Similar to Li and Barber [9], we used □ = 0.1 and $\tau = 0.5$. We found that the discoveries were not sensitive to small perturbations of either of these tuning parameters.

REFERENCES

1. Dudoit, S., Shaffer, J. P. & Boldrick, J. C. Multiple hypothesis testing in microarray experiments. *Stat. Sci.* **18**, 71–103 (2003).
2. Goeman, J. J. & Solari, A. Multiple hypothesis testing in genomics. *Stat. Med.* **33**, 1946–1978 (2014).
3. Sesia, M., Bates, S., Candes, E., Marchini, J. & Sabatti, C. False discovery rate control in` genome-wide association studies with population structure. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2105841118 (2021).
4. Menyhart, O., Weltz, B. & Gyorffy, B. Multipletesting.com: A tool for life science researchers˝ for multiple hypothesis testing correction. *PLoS One* **16**, e0245824 (2021).
5. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Methodol.* **57**, 289–300 (1995).
6. Korthauer, K., et al. A practical guide to methods controlling false discoveries in computational biology. *Genome Biol.* **118**, 1–21 (2019).
7. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat*. **29**, 1165–1188 (2001).
8. Hu, J. X., Zhao, H. & Zhou, H. H. False discovery rate control with groups. *J. Am. Stat. Assoc.* **105**, 1215–1227 (2010).
9. Li, A. & Barber, R. F. Multiple testing with the structure-adaptive Benjamini–Hochberg algorithm. *J. R. Stat. Soc. Methodol.* **81**, 45–74 (2019).
10. Genovese, C. R., Roeder, K. & Wasserman, L. False discovery control with p-value weighting. *Biometrika* **93**, 509–524 (2006).

11. Koutna, I., et al. New insights into´ gene positional clustering and its properties supported by large-scale analysis of various differentiation pathways. *Genomics* **89**, 81–88 (2007).
12. Zhang, B., et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res.* **23** 1522–1540 (2013).
13. Liu, H., et al. Whole-transcriptome analysis of differentially expressed genes in the vegetative buds, floral buds and buds of chrysanthemum morifolium. *PLoS One* **10**, e0128009 (2015).
14. Efron, B. Simultaneous inference: When should hypothesis testing problems be combined? *Ann. Appl. Stat.* **2**, 197–223 (2008).
15. Benjamini, Y. & Heller, R. False discovery rates for spatial signals. *J. Am. Stat. Assoc.* **102**, 1272–1281 (2007).
16. De Sarkar, N., et al. A quest for mirna bio-marker: a track back approach from gingivo buccal cancer to two different types of precancers. *PloS One* **9**, e104839 (2014).
17. Singh, R., et al. Analysis of the whole transcriptome from gingivo-buccal squamous cell carcinoma reveals deregulated immune landscape and suggests targets for immunotherapy. *PloS One* **12**, e0183606 (2017).
18. Hughes, R. A., Heron, J., Sterne, J. A. C. & Tilling, K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int. J. Epidemiol.* **48**, 1294–1304 (2019).
19. Zhang, Z. Missing data imputation: focusing on single imputation. *Ann. Transl. Med.* **4** (2016).
20. Sarkar, S. K. & Zhao, Z. Local false discovery rate based methods for multiple testing of one-way classified hypotheses. *Electron. J. Stat.* **16**, 6043–6085 (2022).
21. Chu, Y., et al. mir-1247-5p functions as a tumor suppressor in human hepatocellular carcinoma by targeting wnt3. *Oncol. Rep.* **38**, 343–351 (2017).
22. Nakagawa, S., Johnson, P. C. & Schielzeth, H. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface* **14**, 20170213 (2017).
23. Searle, S. R., Casella, G. & McCulloch, C. E. *Variance Components*. (John Wiley & Sons, 2009).
24. Montgomery, D. C., Peck, E. A. & Vining, G. G. *Introduction to Linear Regression Analysis.* (John Wiley & Sons, 2021).
25. Troiano, G., et al. Circulating miRNAs from blood, plasma or serum as promising clinical biomarkers in oral squamous cell carcinoma: A systematic review of current findings. *Oral Oncology* 63, 30–37 (2016).26. Setién-Olarra, A., et al. Genomewide miRNA profiling of oral lichenoid disorders and oral squamous cell carcinoma. *Oral Dis.* **22**, 754–760 (2016).
27. Chamorro Petronacci, C. M., et al. miRNAs expression of oral squamous cell carcinoma patients: Validation of two putative biomarkers. *Medicine* **98** (2019).
28. Buscaglia, L. E. B. & Li, Y. Apoptosis and the target genes of microrna-21. *Chin. J. Cancer* **30**, 371–380 (2011).
29. Dioguardi, M., et al. Microrna-21 expression as a prognostic biomarker in oral cancer: Systematic review and meta-analysis. *Int. J. Environ. Res. Public Health* **19**, 3396 (2022).

30. Troiano, G., et al. Predictive prognostic value of tissue-based microRNA expression in oral squamous cell carcinoma: a systematic review and meta-analysis. *J. Dent. Res*. **97**, 759–766 (2018).
31. McQueen, C. *Comprehensive Toxicology* (Elsevier, 2017).
32. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
33. Jenike, A. E. & Halushka, M. K. mir-21: a non-specific biomarker of all maladies. *Biomarker Res.* **9**, 1–7. (2021).
34. Okada, R., et al. Regulation of oncogenic targets by mir-99a-3p (passenger strand of mir-99a-duplex) in head and neck squamous cell carcinoma. *Cells* **8**, 1535 (2019).
35. Osako, Y., et al. Potential tumor-suppressive role of microrna-99a-3p in sunitinib-resistant renal cell carcinoma cells through the regulation of rrm2. *Int. J. Oncol.* **54**, 1759–1770 (2019).
36. Khan, P., et al. Microrna-1: Diverse role of a small player in multiple cancers. *Semin. Cell Dev. Biol.* **124**, 114–126 (2022).
37. Yang, H., et al. Mir-486-3p inhibits the proliferation, migration and invasion of retinoblastoma cells by targeting ecm1. *Biosci. Rep*. **40** (2020).
38. Safa, A., et al. mir-1: A comprehensive review of its role in normal development and diverse disorders. *Biomed. Pharmacother*. **132**, 110903 (2020).
39. Pratap, J., et al. Regulatory roles of runx2 in metastatic tumor and cancer cell interactions with bone. *Cancer Metastasis Rev.* **25**, 589–600 (2006).
40. Wysokinski, D., Blasiak, J. & Pawlowska, E. Role of runx2 in breast carcinogenesis. *Int. J. Mol. Sci.* **16**, 20969–20993 (2015).
41. Li, B., et al. Epigenetic regulation of CXCL12 plays a critical role in mediating tumor progression and the immune response in osteosarcomaos fate determined by epigenetic regulation of cxcl12. *Cancer Res.* **78**, 3938–3953 (2018).
42. Qian, Y., Li, J. & Xia, S. Znf281 promotes growth and invasion of pancreatic cancer cells by activating wnt/β-catenin signaling. *Dig. Dis. Sci.* **62**, 2011–2020 (2017).
43. ElKhouly, A. M., Youness, R. & Gad, M. Microrna-486-5p and microrna-486-3p: Multifaceted pleiotropic mediators in oncological and non-oncological conditions. *Non-coding RNA Res.* **5**, 11–21 (2020).
44. Jiao, X., et al. Kiaa1199, a target of micorna-486-5p, promotes papillary thyroid cancer invasion by influencing epithelial-mesenchymal transition (emt). *Med. Sci. Monit. Basic Res.* **25**, 6788–6796 (2019).
45. Haynes, W. *Benjamini–Hochberg Method* (Springer New York, 2013).
46. Lähnemann, D., et al. Eleven grand challenges in single-cell data science. *Genome Biol.* **21**, 1–35 (2020).
47. Gao, L. L., Bien, J. & Witten, D. Selective inference for hierarchical clustering. *arXiv Preprint arXiv:2012.02936* (2020).
48. Francois, O., Martins, H., Caye, K., and Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Mol Ecol.* **25**, 454–469

49. Aggarwal, R. R., et al. Whole-genome and transcriptional analysis of treatment-emergent small-cell neuroendocrine prostate cancer demonstrates intraclass heterogeneity. *Mol. Cancer Res*. **17**, 1235–1240 (2019).
50. Clermont, P. L., Ci, X., Pandha, H., Wang, Y. & Crea, F. Treatment-emergent neuroendocrine prostate cancer: molecularly driven clinical guidelines. *Int. J. Endocrinol.* **6**, IJE20. (2019).
51. Wang B, et al. Systematic evaluation of three microRNA profiling platforms: Microarray, beads array, and quantitative real-time PCR array. *PLoS One* **6** e17167 (2011).
52. SAS Institute Inc. *SAS/STAT® 9.2 User's Guide.* (SAS Institute Inc., 2008).
53. Leone, V., et al. MiR-1 Is a tumor suppressor in thyroid carcinogenesis targeting CCND2, CXCR4, and SDF-1Œ±. *J. Clin. Endocrinol. Metab.* **96**, E1388–E1398 (2011). doi:10.1210/jc.2011-0345
54. Nohata, N., et al. mir-1 as a tumor suppressive microRNA targeting tagln2 in head and neck squamous cell carcinoma. *Oncotarget* **2**, 29 (2011).

Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870–878 (2002).
Benjamini, Y., Krieger, A. M. & Yekutieli, D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507 (2006).

ACKNOWLEDGEMENTS

Nilanjana Laha's research was partially supported by the NSF-DMS grant DMS-2311098. The authors are thankful to Editage for providing editing support.

AUTHOR CONTRIBUTIONS

Salil Koner contributed to the statistical analyses. Navonil De Sarkar contributed to the paper's writing and conceived the idea. Nilanjana Laha supervised the project, secured funding for editing-related services, wrote the paper, and conceptualized the statistical part of the project.

COMPETING INTERESTS

The authors declare that there are no competing interests associated with this manuscript.

DATA AVAILABILITY

The authors affirm the unrestricted availability of all data supporting the findings. Relevant data can be accessed in the Supplementary files of De Sarkar et al. on the journal's website: (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4134240/).