



# A microRNA Transcriptome-wide Association Study of Prostate Cancer Risk

Nicholas B. Larson<sup>1\*</sup>, Shannon K. McDonnell<sup>1</sup>, Zachary Fogarty<sup>1</sup>, Yuanhang Liu<sup>1</sup>, Amy J. French<sup>2</sup>, Lori S. Tillmans<sup>2</sup>, John C. Cheville<sup>2</sup>, Liang Wang<sup>3</sup>, Daniel J. Schaid<sup>1</sup> and Stephen N. Thibodeau<sup>2</sup>

<sup>1</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, United States, <sup>2</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, United States, <sup>3</sup>Department of Tumor Biology, H. Lee Moffitt Cancer Center, Tampa, FL, United States

## OPEN ACCESS

### Edited by:

Robert Klein,  
Icahn School of Medicine at Mount  
Sinai, United States

### Reviewed by:

Benjamin Rybicki,  
Henry Ford Health System,  
United States  
Xia Jiang,  
Karolinska Institutet (KI), Sweden

### \*Correspondence:

Nicholas B. Larson  
Larson.nicholas@mayo.edu

### Specialty section:

This article was submitted to  
Applied Genetic Epidemiology,  
a section of the journal  
Frontiers in Genetics

Received: 16 December 2021

Accepted: 22 February 2022

Published: 30 March 2022

### Citation:

Larson NB, McDonnell SK, Fogarty Z,  
Liu Y, French AJ, Tillmans LS,  
Cheville JC, Wang L, Schaid DJ and  
Thibodeau SN (2022) A microRNA  
Transcriptome-wide Association Study  
of Prostate Cancer Risk.  
Front. Genet. 13:836841.  
doi: 10.3389/fgene.2022.836841

Large genome-wide association studies have identified hundreds of single-nucleotide polymorphisms associated with increased risk of prostate cancer (PrCa), and many of these risk loci is presumed to confer regulatory effects on gene expression. While eQTL studies of long RNAs has yielded many potential risk genes, the relationship between PrCa risk genetics and microRNA expression dysregulation is understudied. We performed a microRNA transcriptome-wide association study of PrCa risk using small RNA sequencing and genome-wide genotyping data from N = 441 normal prostate epithelium tissue samples along with N = 411 prostate adenocarcinoma tumor samples from the Cancer Genome Atlas (TCGA). Genetically regulated expression prediction models were trained for all expressed microRNAs using the FUSION TWAS software. TWAS for PrCa risk was performed with both sets of models using single-SNP summary statistics from the recent PRACTICAL consortium PrCa case-control OncoArray GWAS meta-analysis. A total of 613 and 571 distinct expressed microRNAs were identified in the normal and tumor tissue datasets, respectively (overlap: 480). Among these, 79 (13%) normal tissue microRNAs demonstrated significant cis-heritability (median cis-h<sup>2</sup> = 0.15, range: 0.03–0.79) for model training. Similar results were obtained from TCGA tumor samples, with 48 (9%) microRNA expression models successfully trained (median cis-h<sup>2</sup> = 0.14, range: 0.06–0.60). Using normal tissue models, we identified two significant TWAS microRNA associations with PrCa risk: over-expression of mir-941 family microRNAs (P<sub>TWAS</sub> = 2.9E-04) and reduced expression of miR-3617-5p (P<sub>TWAS</sub> = 1.0E-03). The TCGA tumor TWAS also identified a significant association with miR-941 overexpression (P<sub>TWAS</sub> = 9.7E-04). Subsequent finemapping of the TWAS results using a multi-tissue database indicated limited evidence of causal status for each microRNA with PrCa risk (posterior inclusion probabilities <0.05). Future work will examine downstream regulatory effects of microRNA dysregulation as well as microRNA-mediated risk mechanisms via competing endogenous RNA relationships.

**Keywords:** prostate cancer, eQTL, microRNA, expression, TWAS

## INTRODUCTION

Prostate cancer (PrCa) is a highly heritable complex disease, and large genome-wide association studies (GWAS) and meta-analyses have identified hundreds of single-nucleotide polymorphisms (SNPs) associated with increased risk of PrCa (Gudmundsson et al., 2007; Yeager et al., 2007; Salinas et al., 2008; Schaid et al., 2010; Kote-Jarai et al., 2011; Akamatsu et al., 2012; Cheng et al., 2012; Al Olama et al., 2013; Kote-Jarai et al., 2013; Al Olama et al., 2014; Hazelett et al., 2014; Yang et al., 2014; Amin Al Olama et al., 2015; Hoffmann et al., 2015; Teerlink et al., 2016). However, the majority of these risk loci are intergenic and few causal genes for PrCa have been explicitly identified (Eeles et al., 2014). For many identified loci, the underlying causal variants are presumed to confer regulatory effects on expression of nearby genes. Follow-up expression quantitative trait locus (eQTL) studies have begun to explore putative regulatory targets of PrCa risk SNPs (Thibodeau et al., 2015; DeRycke et al., 2019). Recent methodological advancements in colocalization of disease-associated SNPs and eQTLs have yielded powerful transcriptome-wide association study (TWAS) approaches to disease-gene discovery, which have been highly successful in translating single-SNP GWAS hits into gene-level associations with PrCa risk (Mancuso et al., 2018; Liu et al., 2022).

Most eQTL-based functional studies of PrCa risk loci have focused primarily on protein-coding genes, and classes of non-coding RNA transcripts remain understudied. An important class of small non-coding RNAs is microRNAs (miRNAs), which contribute to transcriptome regulation via post-transcriptional silencing mechanisms of translational repression, and RNA degradation. MiRNAs are initially expressed as long precursor primary miRNA (pri-miRNA) transcripts before multiple post-processing steps yield short mature miRNAs typically 20–22 nucleotides in length. These mature miRNAs in turn bind to miRNA response elements (MREs) located primarily in the 3'-prime untranslated regions (3'UTRs) of target RNA transcripts. There is a large body of evidence demonstrating the roles of miRNAs in cancer (Peng and Croce, 2016). Specific to PrCa, upregulation of miR-124 has been shown to repress PrCa growth and inhibit invasion of PrCa tumor cells (Kang et al., 2014), and circulating miRNA expression profiles demonstrate diagnostic value for aggressive PrCa (Alhasan et al., 2016). Previous multi-omic analyses in the Cancer Genome Atlas (TCGA) prostate adenocarcinoma (PRAD) samples also identified associations of unsupervised miRNA expression clusters and tumor Gleason score (Cancer Genome Atlas Research, 2015).

Recent eQTL studies have also shown pri-miRNA transcripts are subject to the same *cis*-acting SNP regulatory effects as other long RNAs (Huan et al., 2015), and PrCa risk SNPs may confer systemic impacts on the prostate transcriptome through miRNA expression dysregulation. However, large miRNA expression datasets in relevant tissue types are necessary for eQTL-based disease gene discovery, and such data for normal prostate tissue are lacking. In this study, we present data from a large miRNA sequencing dataset in normal prostate tissue, explore *cis*-based predictive modeling of miRNA expression, and evaluate potential associations between PrCa risk genetics and expression dysregulation of miRNAs by performing a miRNA-based

TWAS (miTWAS) of expressed miRNAs in normal and tumor prostate tissue, which may respectively reveal key risk gene associations relevant to various stages of tumorigenesis.

## METHODS

### Normal Prostate Tissue Samples

Information on study tissue sample selection has been described in greater detail elsewhere (Larson et al., 2015; Thibodeau et al., 2015). Briefly, normal prostate tissue was acquired from an archive collection of fresh frozen material obtained from patients with either radical prostatectomy or cystoprostatectomy. After rigorous evaluation of sample tissue quality, DNA was extracted using the Puregene tissue extraction while RNA was extracted using the QIAGEN miRNeasy Mini Kit and the QIAcube instrument. Informed consent was obtained from all subjects, and the study was approved by the Mayo Clinic Institutional Review Board.

### Normal Prostate eQTL Dataset

Descriptions on genome-wide genotyping and RNA sequencing (RNA-Seq) have been provided elsewhere (Larson et al., 2015; Thibodeau et al., 2015; DeRycke et al., 2019), and these data are publicly available on dbGaP (accession: phs000985.v1.p1). Briefly, genome-wide genotyping was performed using the Illumina Infinium 2.5 M bead arrays (Illumina, San Diego, CA). Quality control excluded variants with low call rate (<95%) and Hardy-Weinberg equilibrium testing ( $P < 1E-05$ ). Untyped and missing autosomal and chrX SNP genotypes were imputed via SHAPEIT (Delaneau et al., 2013) and IMPUTE2 (Howie et al., 2012) using the hg19/GRCh37 1,000 Genomes Phase I reference panel. Imputation quality was assessed using the dosage  $r^2$  metric calculated by BEAGLE v3 utilities (Browning and Browning, 2009), and SNPs with an  $r^2 \leq 0.3$  were excluded from further analysis. After imputation and quality filtering, a total of 9,915,470 variants were available for analysis.

RNA libraries for sequencing were prepared using the TruSeq RNA Sample Prep Kit v2. Paired-end sequencing was performed on an Illumina HiSeq 2000 using TruSeq SBS sequencing kit version 3 and HCS v2.0.12 data collection software. A minimum of 50 million total reads per sample was required for analysis; 234 samples with <50 million total reads were re-sequenced and BAM files were merged if no quality issues were identified. RNA-seq data were processed using the MAP-R-Seq pipeline (Kalari et al., 2014) and gene counts were generated based on ENSEMBL release 78 gene annotation. Conditional quantile normalization (Hansen et al., 2012) was applied to account for GC-bias and sequencing depth. Normalized expression values were then processed using probabilistic estimation of expression residuals (PEER) (Stegle et al., 2012), adjusting for known sample histologic characteristics of percent lymphocytic population and percent epithelium present. PEER residuals were used as the final expression values for all subsequent analyses.

### Normal Prostate Small RNA Sequencing

Among the  $N = 471$  Mayo normal tissue samples with available genome-wide genotyping and RNA-Seq data, we identified

N = 444 total samples with sufficient residual material for small RNA extraction. RNA was extracted using the Qiagen miRNAeasy Mini Kit and the QIAcube instrument. Small RNA sequencing was performed on an Illumina HiSeq 4,000 instrument using NEBNext® Multiplex Small RNA Library Prep Kit, multiplexing up to 48 samples per lane. Sequencing output was processed using the CAP-Mirseq (Sun et al., 2014) bioinformatics pipeline using miRBase v21 (Griffiths-Jones et al., 2006) miRNA reference. An in-house developed RNA-Seq QC (RNASEQQC) pipeline was then applied to identify potential issues based on measures of post-normalization quality (Mahoney et al., 2013).

Mature miRNAs were identified as expressed if  $\geq 5$  aligned reads were identified in  $>25\%$  of samples, and raw expression counts for expressed miRNAs were normalized using trimmed mean of M (TMM) normalization (Robinson and Oshlack, 2010). Leading principal components from the genotyping data did not indicate associations with miRNA expression levels (min  $P > 1e-05$ ). Normalized expression values were then processed using probabilistic estimation of expression residuals (PEER) (Stegle et al., 2012), adjusting for known sample histologic characteristics of percent lymphocytic population, and percent epithelium present. PEER residuals were used as the final expression values for all subsequent analyses.

## TCGA Tumor Data

To additionally explore associations in cancer tissue and provide potential replication of normal tissue discoveries, we also considered miTWAS analyses using miRNA expression data from TCGA prostate adenocarcinoma (PRAD) tumor tissue. Information on the sample preparation and small RNA sequencing are presented in greater detail elsewhere (Chu et al., 2016). To reduce underlying technical differences between our normal tissue miRNA expression dataset and the TCGA small RNA sequencing results, TCGA PRAD small RNA-Seq BAM files were downloaded from the Genomic Data Commons (GDC), of which N = 411 were identified to correspond to tumor tissue. BAM files were reverted to unaligned FASTQ files and reprocessed using the identical miRNA bioinformatics workflow described above. With respect to PEER analyses, sample estimated tumor percentage was used as an adjusting covariate.

Raw genome-wide genotyping Birdseed files from the Affymetrix 6.0 genotyping panel were similarly retrieved from the GDC online portal for all TCGA PRAD samples. We identified one unique non-tumor file per subject (either whole blood or normal tissue), with preferential selection of whole blood data where available. This yielded N = 496 samples (439 blood, 57 normal tissue), which were processed for genotype calling using a genotype confidence threshold of 0.5. Genotype data were then further processed for SNP- and sample-level QC as similarly described for the normal tissue genotyping data. Based on output from STRUCTURE, N = 415 samples were identified that corresponded to subjects of European descent (Caucasian probability  $>0.8$ ). The resulting genome-wide genotyping dataset was then imputed using the Michigan Imputation Server (Das et al., 2016) based on the hg19/GRCh37 Haplotype Reference Consortium (McCarthy et al., 2016) hrc.

r1.1.2016 reference panel, with phasing performed using Eagle v2.4.

## miRNA Transcriptome-wide Association Study

Our miTWAS was based on PrCa case-control GWAS meta-analysis summary statistics from the PRACTICAL Consortium (Schumacher et al., 2018), downloaded from the PRACTICAL website (<http://practical.icr.ac.uk>). For the respective Mayo and TCGA miRNA eQTL datasets, genetically-regulated expression (GrEx) prediction models were trained using the TWAS software FUSION (Gusev et al., 2016). Eligibility for GrEx model training was determined based on a cis-heritability (cis- $h^2$ ) test  $p$ -value  $<0.01$  and corresponding cis- $h^2$  estimate  $>0$ , and training was performed using elastic-net with a 1 Mb buffer for cis-variant inclusion, with all other settings left at default values. Similar methods were used to train GrEx models for all expressed long RNAs (coding and non-coding) from the previously mentioned Mayo normal tissue RNA-Seq data for purposes of comparison and downstream TWAS finemapping at significant miTWAS loci. Expressed miRNAs overlapping the major histocompatibility complex (MHC) region, defined as hg19/GRCh37 positions chr6:28,477,797–33,448,354, were excluded from analysis due to the complex LD structure in this region.

FUSION was then used in conjunction with the PRACTICAL PrCa risk GWAS meta-analysis summary statistics to generate miTWAS results for all expressed miRNAs that had successfully trained GrEx models. The corresponding LD reference was derived from the normal tissue genotyping dataset previously described. Results were declared statistically significant separately for normal and tumor tissue analyses based on a Benjamini-Hochberg false discovery rate (FDR) criterion of  $FDR < 0.05$ .

## TWAS Finemapping

Significant non-causal TWAS results may cluster at a given risk region due to LD and/or shared regulatory regions. TWAS finemapping methods can help resolve which gene(s) may be causal by simultaneously evaluating multiple genes in a given region. For significant normal tissue miTWAS associations, we applied the TWAS finemapping software FOCUS (Mancuso et al., 2019) using a customized version of the FOCUS multi-tissue GrEx database (focus.db, downloaded November 2020) that was augmented by additionally including all miRNA and mRNA GrEx models derived from the Mayo normal tissue dataset. FOCUS was otherwise run under default settings per chromosome, which included output for marginal posterior inclusion probabilities and inclusion in the 90% credible set.

## RESULTS

Small RNA sequencing yielded a median total throughput of 8.0 million reads per sample, with a range of 1.2–22.8 million reads. Of the N = 444 normal tissue samples processed for small RNA sequencing, 441 (99.3%) passed sample-level quality control

thresholds for analysis. There was a median of 4.1 million mature miRNA-aligned reads per sample (range: 0.3–15.5 million; **Supplementary Figure S1**). A total of 613 distinct mature miRNAs were identified as expressed per our criterion of  $\geq 5$  aligned reads in  $>25\%$  of samples. Among the 411 TCGA PRAD tumor samples with reprocessed miRNA expression data, no QC sample exclusions were made, and there was a median 7.5 (range 1.9–34.6 million; **Supplementary Figure S2**) million miRNA-aligned reads. A total of 571 distinct expressed miRNAs were identified (overlap with normal prostate expression data: 480). A total of 3 and 2 expressed miRNAs overlapped the MHC region in the normal and tumor tissue datasets, respectively, and were excluded from further analysis.

## MiRNA GReX Models

Among the remaining 610 non-MHC miRNAs expressed in normal prostate tissue,  $cis-h^2$  testing via FUSION was successfully performed for 596 (97.7%) and 79 miRNAs passed our training eligibility criteria (**Table 1**). The median estimated  $cis-h^2$  among this subset was 0.151 and interquartile range (IQR) was [0.096, 0.233]. Similarly, for the TCGA tumor miRNA expression data, 560/569 (98.4%) had  $cis-h^2$  testing results and 48 met GReX training criteria (overlap with normal: 22). The median estimated  $cis-h^2$  among this subset was 0.143 (IQR = [0.106, 0.267]). **Supplementary Figure S3** compares the  $cis-h^2$  estimates and corresponding Wald 95% confidence intervals for all miRNAs expressed in both datasets and where GReX models were trained in at least one dataset. A total of 77/79 normal tissue GReX models and 47/48 tumor tissue GReX models corresponded to at least one non-zero SNP weight and were considered for further miTWAS analysis, resulting in successful GReX model training proportions of 12.6 and 8.3% for the normal and tumor tissue data, respectively.

To provide relative context on miRNA training, we compared these GReX training results to those for protein-coding genes in our comparably sized normal prostate tissue RNA-Seq expression dataset. Among 16,592 expressed protein-coding genes, GReX models were successfully trained for 7,088 genes (43%), nearly 3-4X that of the respective miRNA data. Among genes with trained models, the distribution of  $cis-h^2$  estimates was comparable (median = 0.171; IQR = [0.112, 0.288]).

## MiRNA TWAS Analyses

TWAS analyses were performed on all unique miRNAs with successfully trained GReX models (77 Mayo, 47 TCGA), of which 22 were shared across datasets. Initial review of these results revealed a number of instances of miRNAs from the same miRNA cluster and identical TWAS output (e.g., hsa-miR-3910 and hsa-miR-3910.1). To reduce redundancy, results were reviewed for members of the same family and, if identical (i.e., same GReX model characteristics, same TWAS statistics), a representative finding was carried forward to the final set of results. This yielded 71 and 39 miTWAS results in the normal and tumor tissue datasets, respectively, with 18 expressed miRNAs shared. Comparison of the TWAS Z-statistics among the 18 miRNAs analyzed revealed fairly consistent findings across datasets (**Supplementary Figure S4**).

Using the  $FDR < 0.05$  significance criterion, we identified two significant miTWAS results on chr20 in at least one expression dataset (**Table 2**): hsa-miR-941 ( $P_{\text{normal}} = 2.9e-04$  in normal;  $P_{\text{tumor}} = 1.0e-03$ ) and hsa-miR-3617-5p ( $P_{\text{normal}} = 1.5e-04$ ; not expressed in tumor). A Hudson plot of the complete miTWAS results across the Mayo normal and TCGA tumor prostate samples is presented in **Figure 1**, while complete TWAS results across both datasets are presented in **Supplementary Table S1**.

Top miTWAS result hsa-miR-3617-5p was evaluated solely in normal tissue (not considered expressed in tumor data), with the TWAS Z-score =  $-3.80$  indicating reduced miR-3617-5p expression associated with increased PrCa risk. The best GWAS PrCa risk SNP result reported by FUSION was rs432448 (chr20:44332298.G.T), located approximately 1.5 kb from the *MIR3617* gene stem-loop sequence.

The miRNA hsa-miR-941 corresponded to a notably high expression level heritability in the Mayo normal tissue ( $cis-h^2 = 0.79$ ), which was similarly well-captured by the trained GReX model (cross-validated  $R^2 = 0.71$ ). The top GWAS PrCa risk SNP was consistent across miRNA expression datasets (rs1058319; chr20:62374389.C.T). TWAS Z-scores were also consistent in direction across datasets (TWAS  $Z_{\text{normal}} = 3.62$ , TWAS  $Z_{\text{tumor}} = 3.28$ ), indicating higher predicted miR-941 expression associated with increased PrCa risk. The mature miRNA originates from one of five tightly clustered stem-loop sequences within a  $\sim 500$  bp region, located at the 62.5 Mb region of chr20, approximately 177 kb from rs1053819.

## TWAS Finemapping Results

We first examined previously reported multi-tissue PrCa risk TWAS results by Mancuso et al. (Mancuso et al., 2018) for protein-coding mRNAs to determine if associations had been previously identified in proximity to (i.e., within 1 Mb) hsa-miR-941 and hsa-miR-3617-5p. For hsa-miR-941, multiple results were reported for a variety of genes and source tissues about the 62.3 Mb region of chr20, with the best GWAS SNP also identified for the majority of results to be rs1058319. The top overall result was *ARFRP1* in CMC. BRAIN.RNASEQ (TWAS  $p = 2.03E-20$ ), while the top prostate-specific result was *ZGPAT* in the TCGA PRAD tumor dataset (TWAS  $p = 2.5E-11$ ). No results were reported in proximity to miR-3617-5p.

Finemapping results from FOCUS for chr20 encompassed the two significant miRNA discoveries from our miTWAS analyses. The corresponding fine-mapping regions were chr20:42680754–44838826 (hsa-miR-3617-5p) and 20:62190180–62964897 (hsa-miR-941). The null model  $p$ -value for the hsa-miR-3617-5p region was the leading result, with a posterior inclusion probability (PIP) of 0.078. The PIP for hsa-miR-3617-5p was 0.0117, with no strong result for any given gene (leading gene result: *L3MBTL1* from brain in GTEx with PIP = 0.054). The 90% credible set was also relatively large (345 GReX models), which included hsa-miR-3617-5p. For the hsa-miR-941 region, a total of 10 genes comprised the 90% credible set, with the leading result corresponding to *DIDO1* from the brain\_cortex in GTEx (PIP = 0.658). No prostate tissue GReX models were included in the credible set. Complete FOCUS output for the two relevant loci is presented in **Supplementary Table S2**.

**TABLE 1** | Summary of miRNA GRex model training for tumor and normal expression datasets.

MiRNA source	Number of mature miRNAs			GRex trained models		
	Expressed <sup>a</sup>	Failed	Eligible for GRex <sup>b</sup>	Trained <sup>c</sup>	cis-h <sup>2</sup>	# SNPs
			N (%)			
Mayo (Normal)	610	14	79 (13%)	77 (12.6%)	0.154 (0.032, 0.793)	27 (1, 82)
TCGA (Tumor)	569	9	48 (8.4%)	47 (8.3%)	0.144 (0.061, 0.598)	30 (4, 94)

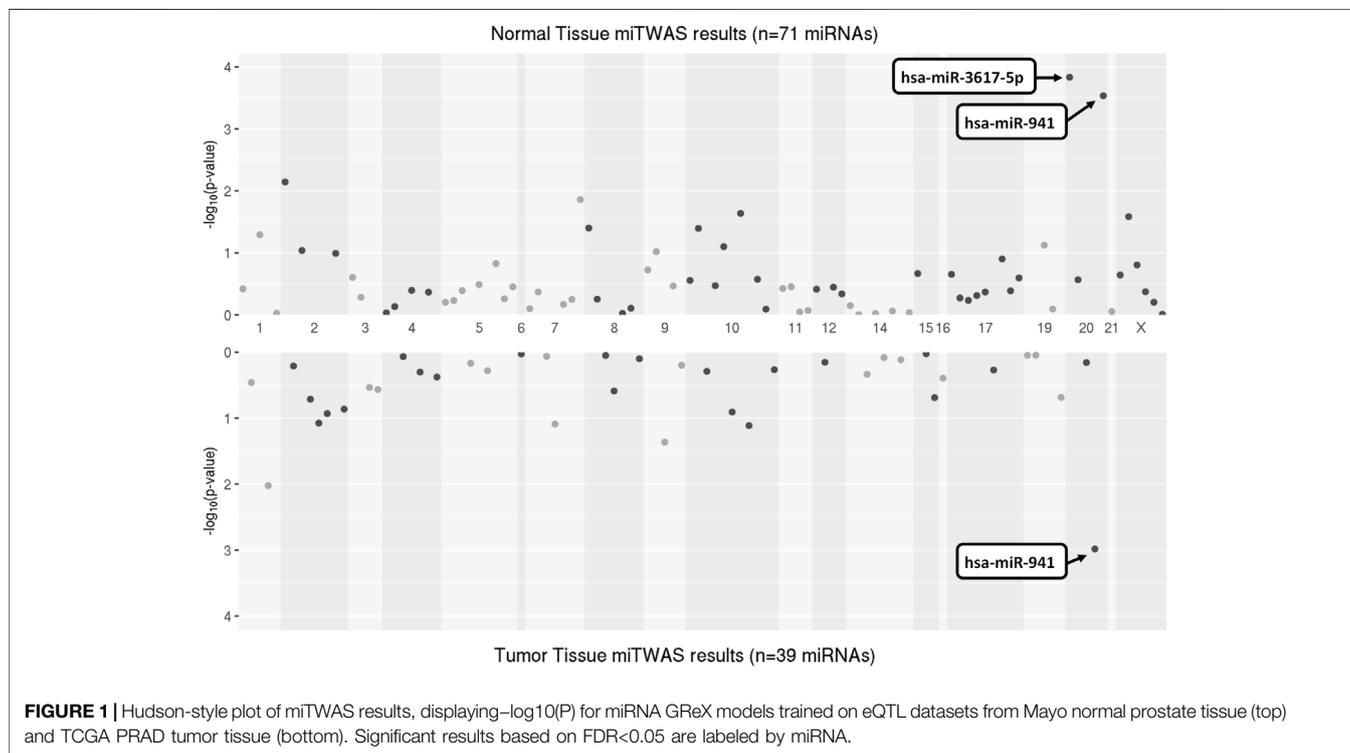
<sup>a</sup>Expressed mature miRNAs were identified as  $\geq 5$  reads in  $>25\%$  of samples. Excludes miRNAs in the MHC region.

<sup>b</sup>MiRNAs with estimated cis-heritability  $> 0$  and heritability test p-value  $< 0.01$  were considered eligible for GRex training.

<sup>c</sup>MiRNAs with at least one non-zero SNP weight in the final GRex model.

**TABLE 2** | MiTWAS results for miRNAs identified as significant (FDR $<0.05$ ) in at least one miRNA expression dataset.

Dataset	TWAS results								
	Mature miRNA	MiRNA Gene	Chr	Start (hg19)	cis-h <sup>2</sup>	Best GWAS SNP	Z	p-value	Q-value
Normal	hsa-miR-3617-5p	MIR3617	20	44333789	0.323	rs432448 chr20:44332298.G.T	-3.80	1.47E-04	0.0103
	hsa-miR-941	MIR941	20	62551155	0.793	rs1058319 chr20:62374389.C.T	3.62	2.92E-04	0.0103
Tumor	hsa-miR-941	MIR941	20	62551155	0.267	rs1058319 chr20:62374389.C.T	3.28	1.04E-03	0.0406



## DISCUSSION

In this study, we leveraged large prostate normal and tumor small RNA sequencing datasets to explore miRNA-based TWAS associations with genetic risk of PrCa and identified two significant associations: miR-941 and miR-3617-5p. Through

the GRex training process, we observed a substantially lower overall proportion of miRNAs with successfully trained expression prediction models relative to protein-coding genes. This may be attributable to both biological and technical factors. First, the absolute expression levels were generally much lower in our normal tissue miRNA

expression dataset than our much deeper sequenced long RNA-Seq data, which may reduce heritability testing power and predictive modeling performance. Moreover, there is evidence that the *cis*-regulatory contributions to overall heritability for miRNAs may be relatively modest compared to mRNAs. In the large multi-generational miRNA expression analysis of N = 5,239 Framingham Heart Study participants by Huan et al. (Huan et al., 2015), only 9/247 (3.6%) of expressed miRNAs corresponded to a narrow-sense  $h^2$  estimate  $>0.3$ , of which the mean single *cis*-eQTL contribution to explained variation was 0.08. As miRNAs are also regulatory non-coding RNAs, they are subject to sponging effects of target mRNAs and may be more susceptible to overall transcriptomic variability. Finally, as similarly speculated in Huan et al., it is likely that evolutionary pressures may constrain *cis*-regulatory effects on pri-miRNAs, given the roles of miRNAs on critical biological processes.

We also noted a sizable difference in number of trained models by dataset (77 for normal vs. 47 for tumor). We believe this may be again driven by both biological and technical differences between datasets. Firstly, the microRNA transcriptome can be heavily dysregulated in the tumor environment, which substantially increases the variance of expression and reduces the relative contribution of *cis*-acting genetic effects. Second, the relative density of the respective genotyping platforms could yield differential imputation efficiency for key eQTL SNPs, potentially yielding more accurate GREx models derived from the normal tissue samples. Despite these caveats, we did observe consistent results across the independent tumor and normal tissue miRNA expression datasets, both with respect to identified expressed miRNAs and miTWAS test statistics.

Our top miTWAS finding was miR-3617-5p, encoded by *MIR3617*, identified in the normal prostate tissue analysis. This miRNA was only identified as expressed in normal prostate tissue, with reduced levels corresponding to increased PrCa risk. Upon literature review, little is currently known regarding the biology of miR-3617-5p and its potential involvement to tumorigenic processes. A previous study of small-cell lung cancer identified downregulation of miR-3617-5p to be associated with chemoresistance (Kuang et al., 2020).

The miR-941 gene family consists of a cluster of five miRNA sequences spanning approximately 500 bp on chr20 (bp: 62550778–62551292), and hsa-miR-941 upregulation was identified to be associated with increased PrCa risk in both tumor and normal datasets. Hsa-miR-941 is a human-specific miRNA that is expressed in a wide variety of tissue types, and has demonstrated higher expression levels in cancer-derived cell lines (Hu et al., 2012). Hu et al. identified tumor suppressor lncRNA *TP73-AS1* as a sponge transcript for hsa-miR-941, showing that over-expression of *TP73-AS1* attenuated cell migration and led to increased expression of hsa-miR-941 targets (Hu et al., 2018). Zhao et al. (2020) investigated serum exosomal miR-941 levels in laryngeal squamous cell carcinoma, identifying increased expression as a diagnostic biomarker. The authors further demonstrated miR-941 overexpression promoted cell

proliferation and invasion via cell-line studies. These findings are commensurate with the observed positive association with PrCa risk in our study.

The relatively modest miTWAS associations along with our finemapping results raise questions regarding the potential causal relationship of PrCa genetic risk mediated by direct expression dysregulation of these two miRNAs and, to a larger extent, the prostate miRNA transcriptome in general. Similarly, while both identified miRNAs have previously reported associations with cancer in the literature, this is true of many miRNAs and thus such connections should be interpreted with caution. Consequently, follow-up functional studies that can study the direct effects of up/down-regulation of these miRNAs in prostate cell-lines are warranted.

There are limitations to our study that warrant mention. Firstly, no true replication dataset was available to verify discoveries, as the two datasets we analyzed were from normal, and tumor prostate tissue, respectively. Thus, further validation of these findings is warranted in independent and comparable datasets. Additionally, the high degree of sample multi-plexing in the normal tissue small RNA sequencing may have reduced sensitivity for association analysis with lower expressed mature miRNAs. However, we observed a high degree of overlap in identified miRNAs, and the individual miTWAS results across datasets were comparable. Our analyses were also restricted to subjects of European ancestry, which limits the generalizability of our results to other racial/ethnic groups. Finally, TWAS methods focus solely on *cis*-regulatory effects of trait-associated SNPs on expression. As miRNAs are themselves regulatory in nature, *trans*-effects of PrCa risk variants may be more relevant via competing endogenous RNA (ceRNA) networks.

In conclusion, our exploration of prostate miRNA expression and PrCa risk genetics revealed two moderate miTWAS associations, including an association in one region with no previously reported TWAS associations. Future efforts studying the role of miRNAs in the genetic risk of PrCa will focus on more sophisticated examination of ceRNA-based *trans*-effects and multi-tissue analyses.

## DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the dbGaP repository, accession number phs000985.v2.p1. Additional data from TCGA are publicly accessible and were downloaded from the GDC.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Mayo Clinic Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

NL conceived of the study, contributed to the data analyses, and drafted the manuscript. SM and ZF contributed to data analyses and draft manuscript preparation. YL contributed to data analyses. All authors reviewed the results, provided critical feedback, and approved the final version of the manuscript.

## FUNDING

This work was supported by the Eagles 5th District Cancer Telethon Funds for Cancer Research, Pilot Program. Generation of data used for this project was in part funded by grants from the Department of Defense (W81XWH-11-1-0261) and National Cancer Institute (R01CA151254, CA151254, CA157881, and CA015083). The Prostate cancer genome-wide association analyses are supported by the Canadian Institutes of Health Research, European Commission's Seventh Framework Programme grant agreement n 223175 (HEALTH-F2-2009-223175), Cancer Research United Kingdom Grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, and The National Institutes of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative). Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) (U19 CA 148537 for ELucidating Loci Involved in Prostate cancer Susceptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN2682012000081) and by Cancer Research United Kingdom grant A8197/A16565. Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher).

## REFERENCES

- Akamatsu, S., Takata, R., Haiman, C. A., Takahashi, A., Inoue, T., Kubo, M., et al. (2012). Common Variants at 11q12, 10q26 and 3p11.2 Are Associated with Prostate Cancer Susceptibility in Japanese. *Nat. Genet.* 44, 426S421–429. doi:10.1038/ng.1104
- Al Olama, A. A., Kote-Jarai, Z., Kote-Jarai, Z., Berndt, S. I., Conti, D. V., Schumacher, F., et al. (2014). A Meta-Analysis of 87,040 Individuals Identifies 23 New Susceptibility Loci for Prostate Cancer. *Nat. Genet.* 46, 1103–1109. doi:10.1038/ng.3094
- Alhasan, A. H., Scott, A. W., Wu, J. J., Feng, G., Meeks, J. J., Thaxton, C. S., et al. (2016). Circulating microRNA Signature for the Diagnosis of Very High-Risk Prostate Cancer. *Proc. Natl. Acad. Sci. USA* 113, 10655–10660. doi:10.1073/pnas.1611596113
- Amin Al Olama, A., Dadaev, T., Hazelett, D. J., Li, Q., Leongamornlert, D., Saunders, E. J., et al. (2015). Multiple Novel Prostate Cancer Susceptibility Signals Identified by fine-mapping of Known Risk Loci Among Europeans. *Hum. Mol. Genet.* 24, 5589–5602. doi:10.1093/hmg/ddv203
- Amin Al Olama, A., Kote-Jarai, Z., Schumacher, F. R., Wiklund, F., Berndt, S. I., Benlloch, S., et al. (2013). A Meta-Analysis of Genome-wide Association Studies to Identify Prostate Cancer Susceptibility Loci Associated with Aggressive and Non-aggressive Disease. *Hum. Mol. Genet.* 22, 408–415. doi:10.1093/hmg/ddt425
- Browning, B. L., and Browning, S. R. (2009). A Unified Approach to Genotype Imputation and Haplotype-phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am. J. Hum. Genet.* 84, 210–223. doi:10.1016/j.ajhg.2009.01.005

## ACKNOWLEDGMENTS

In accordance with usage of the PRACTICAL meta-analysis summary statistics, we would like to acknowledge the PRACTICAL consortium, CRUK, BPC3, CAPS, PEGASUS and include the following statement: We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign United Kingdom (now PCUK), The Orchid Cancer Appeal, Rosetrees Trust, The National Cancer Research Network United Kingdom, The National Cancer Research Institute (NCRI) United Kingdom. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. The Prostate Cancer Program of Cancer Council Victoria also acknowledge grant support from The National Health and Medical Research Council, Australia (126402, 209057, 251533, 396414, 450104, 504700, 504702, 504715, 623204, 940394, 614296), VicHealth, Cancer Council Victoria, The Prostate Cancer Foundation of Australia, The Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. EAO, DMK, and EMK acknowledge the Intramural Program of the National Human Genome Research Institute for their support.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.836841/full#supplementary-material>

- Cancer Genome Atlas Research (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 163, 1011–1025. doi:10.1016/j.cell.2015.10.025
- Cheng, I., Chen, G. K., Nakagawa, H., He, J., Wan, P., Laurie, C. C., et al. (2012). Evaluating Genetic Risk for Prostate Cancer Among Japanese and Latinos. *Cancer Epidemiol. Biomarkers Prev.* 21, 2048–2058. doi:10.1158/1055-9965.epi-12-0598
- Chu, A., Robertson, G., Brooks, D., Mungall, A. J., Birol, I., Coope, R., et al. (2016). Large-scale Profiling of microRNAs for the Cancer Genome Atlas. *Nucleic Acids Res.* 44, e3. doi:10.1093/nar/gkv808
- Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation Genotype Imputation Service and Methods. *Nat. Genet.* 48, 1284–1287. doi:10.1038/ng.3656
- Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved Whole-Chromosome Phasing for Disease and Population Genetic Studies. *Nat. Methods* 10, 5–6. doi:10.1038/nmeth.2307
- Derycke, M. S., Larson, M. C., Nair, A. A., McDonnell, S. K., French, A. J., Tillmans, L. S., et al. (2019). An Expanded Variant List and Assembly Annotation Identifies Multiple Novel Coding and Noncoding Genes for Prostate Cancer Risk Using a normal Prostate Tissue eQTL Data Set. *PLoS One* 14, e0214588. doi:10.1371/journal.pone.0214588
- Eeles, R., Goh, C., Castro, E., Bancroft, E., Guy, M., Olama, A. A. A., et al. (2014). The Genetic Epidemiology of Prostate Cancer and its Clinical Implications. *Nat. Rev. Urol.* 11, 18–31. doi:10.1038/nrurol.2013.266
- Griffiths-Jones, S., Grocock, R. J., Van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA Sequences, Targets and Gene Nomenclature. *Nucleic Acids Res.* 34, D140–D144. doi:10.1093/nar/gkj112
- Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L. T., Gudbjartsson, D., Helgason, A., et al. (2007). Genome-wide Association Study Identifies a

- Second Prostate Cancer Susceptibility Variant at 8q24. *Nat. Genet.* 39, 631–637. doi:10.1038/ng1999
- Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., et al. (2016). Integrative Approaches for Large-Scale Transcriptome-wide Association Studies. *Nat. Genet.* 48, 245–252. doi:10.1038/ng.3506
- Hansen, K. D., Irizarry, R. A., and Wu, Z. (2012). Removing Technical Variability in RNA-Seq Data Using Conditional Quantile Normalization. *Biostatistics* 13, 204–216. doi:10.1093/biostatistics/kxr054
- Hazelett, D. J., Rhie, S. K., Gaddis, M., Yan, C., Lakeland, D. L., Coetzee, S. G., et al. (2014). Comprehensive Functional Annotation of 77 Prostate Cancer Risk Loci. *Plos Genet.* 10, e1004102. doi:10.1371/journal.pgen.1004102
- Hoffmann, T. J., Van Den Eeden, S. K., Sakoda, L. C., Jorgenson, E., Habel, L. A., Graff, R. E., et al. (2015). A Large Multiethnic Genome-wide Association Study of Prostate Cancer Identifies Novel Risk Variants and Substantial Ethnic Differences. *Cancer Discov.* 5, 878–891. doi:10.1158/2159-8290.cd-15-0315
- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G. R. (2012). Fast and Accurate Genotype Imputation in Genome-wide Association Studies through Pre-phasing. *Nat. Genet.* 44, 955–959. doi:10.1038/ng.2354
- Hu, H., Liu, J.-M., Hu, Z., Jiang, X., Yang, X., Li, J., et al. (2018). Recently Evolved Tumor Suppressor Transcript TP73-AS1 Functions as Sponge of Human-specific miR-941. *Mol. Biol. Evol.* 35, 1063–1077. doi:10.1093/molbev/msy022
- Hu, H. Y., He, L., Fominykh, K., Yan, Z., Guo, S., Zhang, X., et al. (2012). Evolution of the Human-specific microRNA miR-941. *Nat. Commun.* 3, 1145. doi:10.1038/ncomms2146
- Huan, T., Rong, J., Liu, C., Zhang, X., Tanriverdi, K., Joehanes, R., et al. (2015). Genome-wide Identification of microRNA Expression Quantitative Trait Loci. *Nat. Commun.* 6, 6601. doi:10.1038/ncomms7601
- Kalari, K. R., Nair, A. A., Bhavsar, J. D., O'Brien, D. R., Davila, J. I., Bockol, M. A., et al. (2014). MAP-RSeq: Mayo Analysis Pipeline for RNA Sequencing. *BMC bioinformatics* 15, 224. doi:10.1186/1471-2105-15-224
- Kang, S., Zhao, Y., Hu, K., Xu, C., Wang, L., Liu, J., et al. (2014). miR-124 Exhibits Antiproliferative and Antiaggressive Effects on Prostate Cancer Cells through PACE4 Pathway. *Prostate* 74, 1095–1106. doi:10.1002/pros.22822
- Kote-Jarai, Z., Olama, A. A., Olama, A. A., Giles, G. G., Severi, G., Schleutker, J., et al. (2011). Seven Prostate Cancer Susceptibility Loci Identified by a Multi-Stage Genome-wide Association Study. *Nat. Genet.* 43, 785–791. doi:10.1038/ng.882
- Kote-Jarai, Z., Saunders, E. J., Leongamornlert, D. A., Tymrakiewicz, M., Dadaev, T., Jugurnauth-Little, S., et al. (2013). Fine-mapping Identifies Multiple Prostate Cancer Risk Loci at 5p15, One of Which Associates with TERT Expression. *Hum. Mol. Genet.* 22, 2520–2528. doi:10.1093/hmg/ddt086
- Kuang, P., Chen, P., Wang, L., Li, W., Chen, B., Liu, Y., et al. (2020). RNA Sequencing Analysis of Small Cell Lung Cancer Reveals Candidate Chemotherapy Insensitivity Long Noncoding RNAs and microRNAs. *Ann. Transl. Med.* 8, 121. doi:10.21037/atm.2020.01.75
- Larson, N. B., McDonnell, S., French, A. J., Fogarty, Z., Cheville, J., Middha, S., et al. (2015). Comprehensively Evaluating Cis-Regulatory Variation in the Human Prostate Transcriptome by Using Gene-Level Allele-specific Expression. *Am. J. Hum. Genet.* 96, 869–882. doi:10.1016/j.ajhg.2015.04.015
- Liu, D., Zhu, J., Zhou, D., Nikas, E. G., Mitanis, N. T., Sun, Y., et al. (2022). A Transcriptome-wide Association Study Identifies Novel Candidate Susceptibility Genes for Prostate Cancer Risk. *Int. J. Cancer* 150, 80–90. doi:10.1002/ijc.33808
- Mahoney, D. W., Therneau, T. M., Anderson, S. K., Jen, J., Kocher, J.-P. A., Reinholz, M. M., et al. (2013). Quality Assessment Metrics for Whole Genome Gene Expression Profiling of Paraffin Embedded Samples. *BMC Res. Notes* 6, 33. doi:10.1186/1756-0500-6-33
- Mancuso, N., Freund, M. K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., et al. (2019). Probabilistic fine-mapping of Transcriptome-wide Association Studies. *Nat. Genet.* 51, 675–682. doi:10.1038/s41588-019-0367-1
- Mancuso, N., Gayther, S., Gayther, S., Gusev, A., Zheng, W., Penney, K. L., et al. (2018). Large-scale Transcriptome-wide Association Study Identifies New Prostate Cancer Risk Regions. *Nat. Commun.* 9, 4079. doi:10.1038/s41467-018-06302-1
- Mccarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A Reference Panel of 64,976 Haplotypes for Genotype Imputation. *Nat. Genet.* 48, 1279–1283. doi:10.1038/ng.3643
- Peng, Y., and Croce, C. M. (2016). The Role of MicroRNAs in Human Cancer. *Sig Transduct Target. Ther.* 1, 15004. doi:10.1038/sigtrans.2015.4
- Robinson, M. D., and Oshlack, A. (2010). A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data. *Genome Biol.* 11, R25. doi:10.1186/gb-2010-11-3-r25
- Salinas, C. A., Kwon, E., Carlson, C. S., Koopmeiners, J. S., Feng, Z., Karyadi, D. M., et al. (2008). Multiple Independent Genetic Variants in the 8q24 Region Are Associated with Prostate Cancer Risk. *Cancer Epidemiol. Biomarkers Prev.* 17, 1203–1213. doi:10.1158/1055-9965.epi-07-2811
- Schaid, D. J., McDonnell, S. K., Riska, S. M., Carlson, E. E., and Thibodeau, S. N. (2010). Estimation of Genotype Relative Risks from Pedigree Data by Retrospective Likelihoods. *Genet. Epidemiol.* 34, 287–298. doi:10.1002/gepi.20460
- Schumacher, F. R., Al Olama, A. A., Al Olama, A. A., Berndt, S. I., Benlloch, S., Ahmed, M., et al. (2018). Association Analyses of More Than 140,000 Men Identify 63 New Prostate Cancer Susceptibility Loci. *Nat. Genet.* 50, 928–936. doi:10.1038/s41588-018-0142-8
- Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using Probabilistic Estimation of Expression Residuals (PEER) to Obtain Increased Power and Interpretability of Gene Expression Analyses. *Nat. Protoc.* 7, 500–507. doi:10.1038/nprot.2011.457
- Sun, Z., Evans, J., Bhagwate, A., Middha, S., Bockol, M., Yan, H., et al. (2014). CAP-miRSeq: a Comprehensive Analysis Pipeline for microRNA Sequencing Data. *BMC Genomics* 15, 423. doi:10.1186/1471-2164-15-423
- Teerlink, C. C., Leongamornlert, D., Leongamornlert, D., Dadaev, T., Thomas, A., Farnham, J., et al. (2016). Genome-wide Association of Familial Prostate Cancer Cases Identifies Evidence for a Rare Segregating Haplotype at 8q24.21. *Hum. Genet.* 135, 923–938. doi:10.1007/s00439-016-1690-6
- Thibodeau, S. N., French, A. J., McDonnell, S. K., Cheville, J., Middha, S., Tillmans, L., et al. (2015). Identification of Candidate Genes for Prostate Cancer-Risk SNPs Utilizing a normal Prostate Tissue eQTL Data Set. *Nat. Commun.* 6, 8653. doi:10.1038/ncomms9653
- Yang, Z., Hu, S., Cheng, J., Xu, J., Shi, W., Zhu, B., et al. (2014). Prevalence and Risk of Cancer of Incidental Uptake in Prostate Identified by Fluorine-18 Fluorodeoxyglucose Positron Emission Tomography/computed Tomography. *Clin. Imaging* 38, 470–474. doi:10.1016/j.clinimag.2014.01.019
- Yeager, M., Orr, N., Hayes, R. B., Jacobs, K. B., Kraft, P., Wacholder, S., et al. (2007). Genome-wide Association Study of Prostate Cancer Identifies a Second Risk Locus at 8q24. *Nat. Genet.* 39, 645–649. doi:10.1038/ng2022
- Zhao, Q., Zheng, X., Guo, H., Xue, X., Zhang, Y., Niu, M., et al. (2020). Serum Exosomal miR-941 as a Promising Oncogenic Biomarker for Laryngeal Squamous Cell Carcinoma. *J. Cancer* 11, 5329–5344. doi:10.7150/jca.45394

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Larson, McDonnell, Fogarty, Liu, French, Tillmans, Cheville, Wang, Schaid and Thibodeau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.