

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Sievänen, Tero; Korhonen, Tia-Marje; Jokela, Tiina; Ahtiainen, Maarit; Lahtinen, Laura; Kuopio, Teijo; Lepistö, Anna; Sillanpää, Elina; Mecklin, Jukka-Pekka; Seppälä, Toni T.; Laakkonen, Eija K.

**Title:** Systemic circulating microRNA landscape in Lynch syndrome

**Year:** 2023

**Version:** Published version

**Copyright:** © 2022 The Authors. International Journal of Cancer published by John Wiley & Sc




**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Sievänen, T., Korhonen, T., Jokela, T., Ahtiainen, M., Lahtinen, L., Kuopio, T., Lepistö, A., Sillanpää, E., Mecklin, J., Seppälä, T. T., & Laakkonen, E. K. (2023). Systemic circulating microRNA landscape in Lynch syndrome. *International Journal of Cancer*, 152(5), 932-944.  
<https://doi.org/10.1002/ijc.34338>

# Systemic circulating microRNA landscape in Lynch syndrome

Tero Sievänen<sup>1</sup>   | Tia-Marje Korhonen<sup>1</sup> | Tiina Jokela<sup>1</sup> | Maarit Ahtiainen<sup>2</sup> |  
 Laura Lahtinen<sup>3</sup> | Teijo Kuopio<sup>3,4</sup> | Anna Lepistö<sup>5</sup> | Elina Sillanpää<sup>1,6</sup> |  
 Jukka-Pekka Mecklin<sup>7,8</sup> | Toni T. Seppälä<sup>5,9</sup> | Eija K. Laakkonen<sup>1</sup> 

<sup>1</sup>Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

<sup>2</sup>Department of Education and Research, Central Finland Health Care District, Jyväskylä, Finland

<sup>3</sup>Department of Pathology, Central Finland Health Care District, Jyväskylä, Finland

<sup>4</sup>Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

<sup>5</sup>Department of Surgery, Abdominal Center, Helsinki University Hospital, Helsinki, Finland

<sup>6</sup>Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

<sup>7</sup>Department of Surgery, Central Finland Health Care District, Jyväskylä, Finland

<sup>8</sup>Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

<sup>9</sup>Applied Tumor Genomics Research Program, University of Helsinki, Helsinki, Finland

## Correspondence

Tero Sievänen and Eija K. Laakkonen, Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, P.O. Box 35 (VIV), 40014, Jyväskylä, Finland.

Email: [tero.o.sievanen@juu.fi](mailto:tero.o.sievanen@juu.fi) (T. S.) and [eija.k.laakkonen@juu.fi](mailto:eija.k.laakkonen@juu.fi) (E. K. L.)

## Funding information

Emil Aaltosen Säätiö; Finnish Cancer Foundation; HUS State Research Funds, Grant/Award Numbers: TYH2021123, TYH2022323; iCan Digital Precision Medicine Flagship; Jane ja Aatos Erkon Säätiö; KYS State Research Funds; Päivikki ja Sakari Sohlbergin

## Abstract

Circulating microRNAs (c-miRs) are small noncoding RNA molecules that migrate throughout the body and regulate gene expression. Global c-miR expression patterns (c-miRnomes) change with sporadic carcinogenesis and have predictive potential in early detection of cancers. However, there are no studies that have assessed whether c-miRnomes display similar potential in carriers of inherited pathogenic mismatch-repair gene variants (*path\_MMR*), known as Lynch syndrome (LS), who are predisposed to highly increased cancer risk. Using high-throughput sequencing and bioinformatic approaches, we conducted an exploratory analysis to characterize systemic c-miRnomes of *path\_MMR* carriers, sporadic rectal cancer patients and non-LS controls. We showed for the first time that cancer-free *path\_MMR* carriers have a systemic c-miRnome of 40 differentially expressed c-miRs that can distinguish them from non-LS controls. The systemic c-miRnome of cancer-free *path\_MMR* carriers also resembles the systemic c-miRnomes of cancer patients with or without *path\_MMR*. Our pathway analysis linked the found differentially expressed c-miRs to carcinogenesis. A total of 508 putative target genes were identified for 32 out of 40 differentially expressed c-miRs, and 238 of them were enriched in cancer-related pathways. The most enriched c-miR-target genes include well-known oncogenes and tumor suppressor genes such as *BCL2*, *AKT3*, *PIK3CA*, *KRAS*, *NRAS*, *CDKN1A* and *PIK3R1*. Taken together, our findings suggest that LS and sporadic carcinogenesis share common biological pathways and alterations in these pathways can produce a c-miR signature which can track potential oncogenic stress in cancer-free *path\_MMR* carriers. Therefore, c-miRs hold potential in monitoring the LS risk stratification patterns during clinical surveillance or cancer management.

## KEYWORDS

bioinformatics, hereditary cancer, Lynch syndrome, microRNA, next generation sequencing

**Abbreviations:** BP, biological process; c-miR, circulating microRNA; c-miRnome, global circulating miR expression profile; COSMIC-CGC, Cancer Gene Census of the Catalogue of Somatic Mutations in Cancer; DE, differential expression; dMMR, deficient MMR; ERMA, estrogenic regulation of muscle apoptosis; FDR, false discovery rate; GO, Gene Ontology; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LS, Lynch syndrome; miR, microRNA; *path\_MMR*, pathogenic mismatch-repair gene variant; t-SNE, T-distributed stochastic neighbor embedding.

Toni T. Seppälä and Eija K. Laakkonen contributed equally to our study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *International Journal of Cancer* published by John Wiley & Sons Ltd on behalf of UICC.

Säätiö; Relander Foundation; Sigrid Juséliuksen  
 Säätiö; Suomen Lääketieteen Säätiö;  
 Terveystutkimuksen Toimikunta,  
 Grant/Award Numbers: 338657, 341750

### What's new?

Systemic circulating microRNA expression patterns (c-miRNomes) are altered during sporadic carcinogenesis and they have predictive potential in early cancer detection. However, their potential in carriers of inherited pathogenic mismatch-repair gene variants associated with Lynch syndrome remains understudied. Using high-throughput sequencing and bioinformatics, the authors show that Lynch syndrome and sporadic carcinogenesis share common biological pathways. Alterations in these pathways produce a c-miRnome signature that could help track oncogenic stress in cancer-free Lynch syndrome carriers. The findings suggest that systemic c-miRNomes could potentially facilitate the monitoring of Lynch syndrome carriers that require more intensive surveillance or clinical management.

## 1 | INTRODUCTION

Lynch syndrome (LS) is an inherited cancer predisposition syndrome caused by pathogenic gene variants in DNA-mismatch repair (*path\_MMR*) genes *MLH1*, *MSH2*, *MSH6* or *PMS2*.<sup>1</sup> By genetic or epigenetic silencing, deficient MMR (dMMR) significantly increases cellular mutation rates thus predisposing *path\_MMR* carriers to increased cancer risk and excessive cancer occurrence.<sup>1,2</sup> Colorectal cancer is a traditional hallmark cancer of LS that is commonly cured by surveillance, followed by modern surgical and oncological management, with over 90% 10-year overall survival.<sup>2,3</sup> Despite the good recovery rate in first cancers, the persons at risk will develop frequently more lethal cancers still at relatively young age.<sup>4</sup> This highlights the need for an improved molecular assessment and identification of which patients would require more intensive surveillance or clinical management.

MicroRNAs (miRs) are small (18-25 nucleotides) noncoding RNA-molecules that regulate gene expression by translational repression.<sup>5</sup> MiRs play a role in regulation of >30% of the human genes controlling critical biological processes such as cell proliferation, cell differentiation, and apoptosis.<sup>5-7</sup> In cancers, miRs can be regarded as tumor suppressive or oncogenic, thus resulting in down-regulation or upregulation of the affected target genes, respectively.<sup>7</sup> Compared to tissue-based miRs, circulating-miRs (c-miRs) migrate throughout the body within various body fluids and are part of active intertissue crosstalk.<sup>8,9</sup> Nowadays, profiling of the global c-miR expression levels (c-miRnome) has become prevalent and miR expression can be correlated with cancer type, stage, and other clinical variables.<sup>10-13</sup> Therefore, aberrantly expressed miRs could have diagnostic, predictive, and prognostic potential in molecular profiling and early detection of cancers.

LS cohort provides an ideal population for biomarker mining due to well-predicted cancer risk of persons under frequent surveillance. The role of miRs in LS have remained understudied even if various studies have shown that c-miR expression patterns change with carcinogenesis in various sporadic cancers. Balaguer et al have shown that miRs can be used in tumor classification and discrimination of sporadic and hereditary tumors with microsatellite instability,<sup>14</sup> thus highlighting the potential role of miRs as LS biomarkers. In support, Valeri et al, Liccardo et al and Zhou et al postulated that miRs could have functional roles in LS carcinogenesis, for example, by targeting MMR-proteins<sup>15,16</sup> and various tumor-suppressor

genes.<sup>17</sup> However, these studies along with other reports have assessed miR functions in the colorectum and colorectal cancer tissues and cells as well as with microarray data in silico<sup>14-19</sup> but not in circulation.

Instead of using a targeted panel of a priori chosen c-miRs, it is beneficial to characterize the systemic c-miRnome of *path\_MMR* carriers. This “omics-approach” provides a more comprehensive view of how c-miRs could contribute to LS pathogenesis, and plausibly pave way for future use of c-miRs in risk stratification and early detection of LS cancers. Our exploratory study compared the systemic c-miRnome of cancer-free *path\_MMR* carriers with c-miRNomes of non-LS controls (discovery cohort), sporadic rectal cancer patients and *path\_MMR* carriers with cancer (cancer cohort) using high-throughput sequencing and bioinformatic approaches.

## 2 | MATERIALS AND METHODS

### 2.1 | Study subjects

Our study consisted of independent discovery and cancer cohorts. The discovery cohort (n = 118) was composed of 81 currently cancer-free (healthy) Finnish *path\_MMR* carriers and 37 non-LS controls whose c-miRNomes were sequenced. The cancer cohort (n = 37) was composed of 13 *path\_MMR* carriers who currently had cancer and 24 sporadic rectal cancer patients whose c-miRNomes were sequenced.

All *path\_MMR* carriers were enrolled in the study and blood sampling was performed at their regular colonoscopy surveillance appointments at Helsinki University Central Hospital in Helsinki and Central Finland Central Hospital in Jyväskylä, Finland. They were also registered participants in the nationwide Finnish Lynch Syndrome Research Registry (LSRFi, [www.lynchsyndrooma.fi](http://www.lynchsyndrooma.fi), accessed 05/2021). The families and individuals were identified in the registry based on clinical criteria (Amsterdam and Bethesda criteria)<sup>20,21</sup> and subsequently through cascade testing of the families and universal testing of tumors. Adult members of LSRFi with confirmed *path\_MMR* variants (classes 4 and 5 by InSIGHT criteria)<sup>22</sup> were eligible for the study.

Sporadic rectal cancer patients were enrolled, and blood sampling performed at the time of their initial appointment for surgery at surgical clinic at the local tertiary center responsible for management of

rectal cancer in the Southern Finland area (Helsinki University Central Hospital, unit of rectal surgery, Helsinki, Finland).

Non-LS control samples were acquired from Biobank of Eastern Finland, Kuopio, Finland ( $n = 27$ ) in 2020 or were part of the Estrogenic Regulation of Muscle Apoptosis (ERMA) cohort ( $n = 10$ ) consisting of healthy 47–55-years old women.<sup>23</sup> ERMA samples were collected at University of Jyväskylä in Jyväskylä, Finland. Persons with no cancers, blood disorders, acute or chronic infectious diseases, rheumatoid arthritis and known *BRCA* or *MMR*-gene germline mutations were eligible for the non-LS control group. Ethnicity throughout the study population was widely white Caucasian.

## 2.2 | Sample collection

*Path\_MMR* carriers' and sporadic rectal cancer patients' venous blood samples were drawn after surveillance colonoscopy visits at fasted state. All ERMA participants fasted overnight before blood sampling. The duration of fasting is not reported for the samples obtained through biobank ( $n = 27$ ). Samples were taken from antecubital vein to standard serum tubes (455 092, Greiner). To separate serum, the whole blood samples were allowed to clot for 30 minutes at room temperature, centrifuged at 1800g for 10 min and aliquoted.

## 2.3 | Small-RNA isolation and quality evaluation

c-miR isolations from blood serum were carried out using affinity column-based miRNeasy Serum/Plasma Advanced Kit (217204, Qiagen) according to the manufacturer's instructions. Briefly, 0.5 mL of thawed serum was used to isolate miRs. All the required solutions were added in amounts recommended by the manufacturer. Cel-miR-39 miR mimic (MS00019789, Qiagen) was added to each sample to serve as a spike-in control for monitoring the miR purification and amplification. Phase separation centrifugation was executed in 12 000g for 3 min at room temperature (Heraeus, Biofuge Pico and Fresco 17, ThermoFisher) and rest of the centrifugations were performed at 16000g whenever a range of 8000–20000g was recommended. C-miRs were eluted to nuclease-free water. Prior to the library preparation, RNA quality and recovery were checked by RT-qPCR (CFX96-RT-qPCR, Bio-Rad) according to manufacturer's protocol (MiScript Primer assays and II RT kit for cDNA synthesis and MiScript SYBR Green PCR Kit for RT-qPCR, 218 161, Qiagen) from which the recovery of cel-miR-39 spike-in control was confirmed.

## 2.4 | Small-RNA library preparation and sequencing

Small-RNA Library preparations were executed with QIAseq miRNA Library Preparation Kit (1103679, Qiagen) according to the manufacturer's instructions using multiplexing adapters. Briefly, the small RNA

fractions were first ligated to sequencing adapters from both 5' and 3' ends, reverse transcribed into cDNA using UMI-assigning primers and purified using magnetic beads. A universal indexing sequence was also added in the reverse transcription step, thus allowing samples to be distinguished from each other. The samples were then amplified with standard thermocycler (Eppendorf), purified, and eluted into nuclease-free water. Quality assessment of the libraries was completed with TapeStation 4200 (Agilent). The library sample concentrations were measured with Qubit fluorometer (Invitrogen), quantified, diluted, and pooled into a single mixture in equal amounts (1.8 pM per sample) prior to sequencing. Sequencing of the small-RNA libraries were done with NextSeq 500 (Illumina) using NextSeq 500/550 High Output Kit v. 2.5 with 75 cycles (15057934, Illumina) to produce 75-base pair single-end reads with aimed mean sequencing depth of >5 M reads per sample as recommended by the manufacturer (Qiagen).

## 2.5 | Raw data processing and alignment

Sequencing output data was converted to FASTQ-format using bcl2fastq software (v.2.20, Illumina, USA). FastQC was used for quality controls.<sup>24</sup> The QIAseq sequencing adapters were trimmed from the 3' end of the reads with FastX-toolkit<sup>25</sup> using default parameters with minimum alignment length-M 19. Only clipped reads >20 bp in length were selected for downstream analysis. After adapter clipping, the reads were trimmed to 22 bp to enrich miR-sequences and then quality filtered with FastX-toolkit. Only high-quality reads (Phred score >25) were selected for alignment to reference genome. Before alignment, all the four sample lanes were merged to obtain the overall sample read count and to ensure better mapping quality. Samples that had <1 M reads were excluded from the analyses. Subsequently, the preprocessed reads were mapped to human mature miR-genome (miRbase v.22)<sup>26</sup> with Bowtie alignment tool for single end data with v-mode and best strata parameters.<sup>27</sup> Only uniquely mapped miR-reads were selected for differential expression (DE) analysis.

## 2.6 | Differential expression analysis

DE analyses from raw c-miR counts were based on statistical procedures of EdgeR<sup>28</sup> and DESeq2<sup>29</sup> packages and conducted in R-studio (v. 3.6.3)<sup>30</sup> (Supplementary file S3). Briefly, DE analyses were performed on c-miR raw read count matrices after the low expressed genes were filtered out, normalized with the median of ratios method and variance stabilized in DESeq2. C-miRs that had more than 1 count per million in 70% of the samples in a group were selected for DE analyses. Filtered and normalized c-miR counts were used to set up a design matrix in DESeq2 that adjusted for sex and potential batch effect. Benjamini-Hochberg procedure in DESeq2 was used to correct for multiple testing. C-miRs that had a false discovery rate (FDR) <0.05 were considered DE.

## 2.7 | Dimension reduction analysis

Dimension reduction of the DESeq2-normalized data was conducted using the t-distributed stochastic neighbor embedding (t-SNE) method, which is a nonlinear and unsupervised technique to simplify high dimensional data for visualization in low-dimensional space.<sup>31</sup> t-SNE analysis was performed to identify and visualize possible clustering of subpopulations within the dataset. Rtsne package in R-studio was used with output dimensionality set to 2, perplexity set to 35 and theta set to 0.5.

## 2.8 | Target gene prediction and pathway analysis

Putative miR-target gene prediction was performed using mirWalk tool that utilizes a random-forest-based approach, an ensemble learning method based on multiple decision trees, to predict target genes.<sup>32,33</sup> Only the predicted miR-target genes targeting 3' untranslated region with experimental validation from miRTarBase<sup>34</sup> and which were included and verified in mirDB<sup>35</sup> and TargetScan<sup>36</sup> databases were selected for downstream gene set enrichment analysis (GSEA).<sup>37</sup> GSEA of gene ontology biological processes (GO:BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>38</sup> pathways were also conducted with mirWalk. MirWalk provides a standard enrichment analysis based on hypergeometric tests. GO and KEGG terms with FDR-corrected *P*-values of <.05 were considered enriched. Cancer Gene Census of the Catalog of Somatic Mutations in Cancer (COSMIC-CGC)<sup>39</sup> project database were used for target gene investigation.

## 2.9 | Statistical analysis

Data regarding study subjects are presented using means and standard deviations. DE-analyses were based on statistical procedures of DESeq2 package accounting for normalization and exclusion of outliers. Mann-Whitney *U*-test and Kruskal-Wallis-test was used in the validation analysis and cell line experiment (Supplementary file S1, Supplementary materials and methods), respectively. Pearson correlation was used to compare gene fold correlation between the discovery and validation cohorts (Supplementary file S1, Supplementary materials and methods). In all analyses, *P*-value, or FDR <.05 were considered to indicate statistical significance.

## 3 | RESULTS

### 3.1 | A pool of 228 c-miRs is shared between the discovery and cancer cohorts

Descriptive characteristics of study subjects in the discovery cohort and cancer cohort are presented in Table 1.

Human genome encodes approximately 2600 mature miRs (miR-base, v.22).<sup>26</sup> To inspect the systemic c-miR content in the discovery and cancer cohorts, we performed small-RNA sequencing experiment

to characterize the serum c-miRNomes. We identified a total of 1349 distinct c-miRs in three separate sequencing runs with an average sequencing depth of 3.2 M reads per sample (Supplementary file S1, Supplementary materials and methods and Supplementary file S1, Table S1 and Supplementary file S2, Table S1). After processing of raw data and filtering of low expressed c-miRs, 228 c-miRs common to both cohorts were identified (Supplementary file S1, Figure S1 and Supplementary file S2, Table S2).

The most highly expressed c-miRs among *path\_MMR* carriers with or without cancer were hsa-let-7a-5p, hsa-let-7b-5p, hsa-miR-122-5p, hsa-miR-16-5p and hsa-miR-223-3p (Supplementary file S1, Figure S2). The most abundant c-miRs in non-LS control group were the same as in *path\_MMR* carriers with or without cancer (Supplementary file S1, Figure S3). Among sporadic rectal cancer group, the top c-miRs were otherwise the same except hsa-miR-451a replaced hsa-miR-122-5p (Supplementary file S1, Figure S4). All these top c-miRs in total accounted for approximately 50% of all c-miR counts in all cohorts, thus displaying major overrepresentation that could have possibly affected the c-miR pool size. In summary, our sequencing analysis provided moderate coverage of c-miRNomes in LS.

### 3.2 | Healthy *path\_MMR* carriers have a c-miRnome that differs from non-LS controls but resembles the c-miRNomes of patients with sporadic or hereditary cancer

The phenotype and cancer risk spectrum vary within LS cohort, for example, due to *path\_MMR* variant and sex.<sup>1</sup> As our discovery cohort consisted of males and females with all *path\_MMR* variants included, we first explored whether these traits influenced c-miR expression in healthy *path\_MMR* carriers. We used the pool of identified 228 c-miRs to form the count matrix for all DE-analyses (Supplementary file S3). Hsa-miR-206 and hsa-miR-223-5p were observed downregulated in males compared to females and thus sex was added as a covariate to further analyses (Supplementary file S3). We did not find DE c-miRs when *path\_MMR* variants were compared to each other or when *path\_MLH1* carriers were compared to all other *path\_MMR* variants combined (Supplementary file S3). These results show that different *path\_MMR* variants do not cause heterogeneity that would generate a recognizable c-miR profile, thus suggesting a shared systemic response common to all *path\_MMR* variants. Furthermore, we also tested if the c-miR expression profile is altered in persons who had had cancer or multiple cancers previously, but we did not find significant differences (Supplementary file S3).

Alterations in the immune cell abundance of normal colorectal mucosa in cancer-free *path\_MMR* carriers separate them from those with cancer.<sup>40</sup> To see whether we can identify a LS-specific c-miR signature, our primary objective was to characterize systemic c-miRnome of healthy *path\_MMR* carriers, which has not been done previously. We thus performed DE-analysis within the discovery cohort and RT-

**TABLE 1** Descriptive characteristics of study subjects in the discovery cohort and cancer cohort

Variable	Discovery cohort		Cancer cohort	
	<i>Path_MMR</i> , healthy	non-LS, healthy	<i>Path_MMR</i> , cancer	Sporadic rectal cancer patients
N	81	37	13	24
Sex (N [%])				
Male	40 (49.4)	18 (48.6)	10 (76.9)	10 (41.6)
Female	41 (50.6)	19 (51.4)	3 (23.1)	14 (58.4)
Age, years (mean ± SD)	59.5 (10.7)	54.9 (10.7)	60.7 (15.3)	69.8 (9.9)
Body mass index, kg/m <sup>2</sup> (mean ± SD) <sup>a</sup>	27.3 (5.7)	28.0 (6.2)	28.2 (3.4)	27.6 (6.3)
<i>Path_MMR</i> (N [%])				
<i>MLH1</i>	50 (61.7)	-	8 (61.5)	-
<i>MSH2</i>	17 (21.0)	-	2 (15.4)	-
<i>MSH6</i>	12 (14.8)	-	3 (23.1)	-
<i>PMS2</i>	2 (2.5)	-	0 (0.0)	-
Previous cancers (N [%])				
Yes	42 (51.9)	-	10 (76.9)	-
No	39 (48.1)	-	3 (23.1)	-
Cancer type (N [%])				
Colorectal cancer	-	-	5 (38.5)	-
Prostate cancer	-	-	3 (23.0)	-
Other cancer <sup>b</sup>	-	-	5 (38.5)	-
Rectal cancer	-	-	-	24 (100.0)

<sup>a</sup>Missing data: Discovery cohort, n = 12 in *path\_MMR* carriers; Cancer cohort, n = 3 in *path\_MMR* carriers.

<sup>b</sup>Other cancer include esophageal cancer, n = 1; spinocellular cancer, n = 1; glioblastoma, n = 1; gastric cancer, n = 1 and thymic cancer, n = 1.

qPCR validation analysis within similar but independent validation cohort (Supplementary file S1, Supplementary materials and methods) to compare healthy *path\_MMR* carriers to healthy non-LS controls (Supplementary file S1, Figure S5). In DE-analysis, we found 40 out of 228 c-miRs to display aberrant expression in healthy *path\_MMR* carriers (Table 2). Of them, 15 were upregulated and 25 downregulated in *path\_MMR* carriers compared to non-LS controls, but the fold changes remained low varying from minimum of -0.88 to maximum of 1.25 (Figure 1A). Hsa-miR-155-5p, hsa-let-7c-5p and -let-7 e-5p and -122b-3p had the most significant upregulation within healthy *path\_MMR* carriers (Table 2). Of the downregulated c-miRs, hsa-miR-15a-5p was the most significantly downregulated followed by hsa-miR-185-5p, -320a-3p and -186-5p, respectively (Table 2). Overall, aberrant expression of multiple c-miRs in healthy *path\_MMR* carriers might indicate that some systemic alterations in c-miR-mediated regulation of biological pathways associated with dMMR may be ongoing even at cancer-free state in *path\_MMR* carriers.

To understand this phenomenon further, we explored whether the *path\_MMR* carriers who currently have cancer also display unique c-miR expression. By using tumor samples, Balaguer et al have shown that miR expression can distinguish LS tumors from sporadic tumors with microsatellite instability.<sup>14</sup> To test if we can similarly reveal differences

in c-miRs, we first inspected c-miRnomes within the cancer cohort but did not find any differences (Figure 1B and Supplementary file S1, Table S2), thus suggesting a mutual c-miR response among the cancer types. Furthermore, our second analysis scheme comparing healthy *path\_MMR* carriers to sporadic rectal cancer patients (Figure 1C and Table 2), our third analysis scheme comparing healthy *path\_MMR* carriers to *path\_MMR* carriers with cancer (Figure 1D and Table 2) and our fourth analysis scheme comparing *path\_MMR* carriers with cancer to healthy non-LS controls (Figure 1E and Supplementary file S1, Table S2) were also unable to detect DE c-miRs. These observations imply that c-miRnomes within our dataset cannot discern healthy *path\_MMR* carriers from cancer patients with or without dMMR.

Several DE c-miRs have been implicated to sporadic cancer progression.<sup>41,42</sup> To study this in our dataset, we compared sporadic rectal cancer patients to non-LS controls. We found that hsa-miR-200a-3p, -10a-5p, -196a-5p and -200c-3p were significantly upregulated in sporadic rectal cancer patients differentiating them from non-LS controls (Figure 1F and Table 2). All of these c-miRs have earlier been shown to associate with colorectal cancer, and of them, hsa-miR-200a-3p was also significantly upregulated in healthy *path\_MMR* carriers compared to non-LS controls with fold change of 0.88. In this analysis scheme, the fold change in hsa-miR-

**TABLE 2** DE and non-DE c-miRs within and between the discovery and cancer cohorts

Healthy <i>path_MMR</i> vs non-LS control			Sporadic rectal cancer patients vs healthy <i>path_MMR</i>			Healthy <i>path_MMR</i> vs <i>path_MMR</i> with cancer			Sporadic rectal cancer patients vs non-LS control		
c-miR	log2FC	FDR	c-miR	log2FC	FDR	c-miR	log2FC	FDR	c-miR	log2FC	FDR
hsa-miR-155-5p	0.905	<b>&lt;0.001</b>	hsa-miR-10a-5p	0.700	0.088	hsa-miR-127-3p	1.548	0.277	hsa-miR-200a-3p	1.755	<b>&lt;0.001</b>
hsa-let-7c-5p	0.729	<b>&lt;0.001</b>	hsa-miR-1180-3p	1.147	0.155	hsa-let-7b-5p	-0.250	0.991	hsa-miR-10a-5p	0.981	<b>0.003</b>
hsa-let-7 e-5p	0.955	<b>&lt;0.001</b>	hsa-miR-126-3p	-0.395	0.155	hsa-let-7c-5p	0.191	0.991	hsa-miR-196a-5p	1.813	<b>0.028</b>
hsa-miR-122b-3p	1.252	<b>0.001</b>	hsa-miR-148b-3p	-0.336	0.155	hsa-let-7d-3p	-0.146	0.991	hsa-miR-200c-3p	1.133	<b>0.028</b>
hsa-miR-15a-5p	-0.677	<b>0.001</b>	hsa-miR-196a-5p	1.414	0.155	hsa-let-7d-5p	-0.325	0.991			
hsa-miR-185-5p	-0.483	<b>0.001</b>	hsa-miR-320a-3p	0.557	0.155	hsa-let-7 e-5p	0.331	0.991			
hsa-miR-320a-3p	-0.709	<b>0.001</b>	hsa-miR-320b	0.845	0.155	hsa-let-7f-5p	0.174	0.991			
hsa-miR-186-5p	-0.548	<b>0.002</b>	hsa-miR-486-5p	0.542	0.243	hsa-let-7i-5p	0.067	0.991			
hsa-let-7a-5p	0.535	<b>0.003</b>	hsa-miR-320c	1.097	0.262	hsa-miR-100-5p	-0.453	0.991			
hsa-miR-10b-5p	0.500	<b>0.003</b>	hsa-miR-185-5p	0.344	0.288	hsa-miR-101-3p	0.171	0.991			
hsa-miR-3613-5p	-0.880	<b>0.003</b>	hsa-miR-223-3p	0.413	0.288	hsa-miR-103a-3p	0.246	0.991			
hsa-miR-22-3p	-0.522	<b>0.004</b>	hsa-miR-483-5p	0.774	0.319	hsa-miR-103b	0.239	0.991			
hsa-miR-19b-3p	-0.490	<b>0.005</b>	hsa-miR-2110	1.198	0.342	hsa-miR-106b-3p	-0.216	0.991			
hsa-miR-125a-5p	0.490	<b>0.007</b>	hsa-miR-222-3p	0.750	0.342	hsa-miR-106b-5p	0.535	0.991			
hsa-miR-451a	-0.714	<b>0.007</b>	hsa-miR-486-3p	0.475	0.462	hsa-miR-107	0.243	0.991			
hsa-miR-125b-5p	0.600	<b>0.009</b>	hsa-let-7d-3p	0.462	0.462	hsa-miR-10a-5p	-0.194	0.991			
hsa-miR-15b-5p	-0.525	<b>0.009</b>	hsa-miR-11 400	-0.824	0.462	hsa-miR-10b-5p	-0.170	0.991			
hsa-miR-32-5p	-0.564	<b>0.009</b>	hsa-miR-134-5p	-0.670	0.462	hsa-miR-11 400	0.549	0.991			
hsa-miR-339-5p	-0.806	<b>0.009</b>	hsa-miR-193a-5p	0.532	0.462	hsa-miR-1180-3p	-0.864	0.991			
hsa-miR-107	-0.464	<b>0.012</b>	hsa-miR-196b-5p	0.447	0.462	hsa-miR-1255b-5p	0.373	0.991			
hsa-miR-484	-0.748	<b>0.012</b>									
hsa-let-7f-5p	0.328	<b>0.015</b>									
hsa-miR-206	0.994	<b>0.015</b>									
hsa-miR-25-3p	-0.375	<b>0.015</b>									
hsa-miR-27a-3p	-0.373	<b>0.015</b>									
hsa-miR-486-3p	-0.565	<b>0.015</b>									
hsa-miR-141-3p	0.874	<b>0.016</b>									
hsa-miR-3074-5p	-0.537	<b>0.020</b>									
hsa-miR-126-3p	0.328	<b>0.021</b>									
hsa-miR-200a-3p	0.884	<b>0.021</b>									
hsa-miR-221-3p	-0.312	<b>0.033</b>									
hsa-miR-424-5p	-0.662	<b>0.034</b>									
hsa-let-7i-5p	0.275	<b>0.040</b>									
hsa-miR-23a-3p	-0.437	<b>0.040</b>									
hsa-miR-27b-3p	0.420	<b>0.040</b>									
hsa-miR-486-5p	-0.447	<b>0.040</b>									
hsa-miR-19a-3p	-0.441	<b>0.046</b>									
hsa-miR-222-3p	-0.647	<b>0.046</b>									
hsa-miR-363-3p	-0.537	<b>0.049</b>									
hsa-miR-92a-3p	-0.370	<b>0.050</b>									

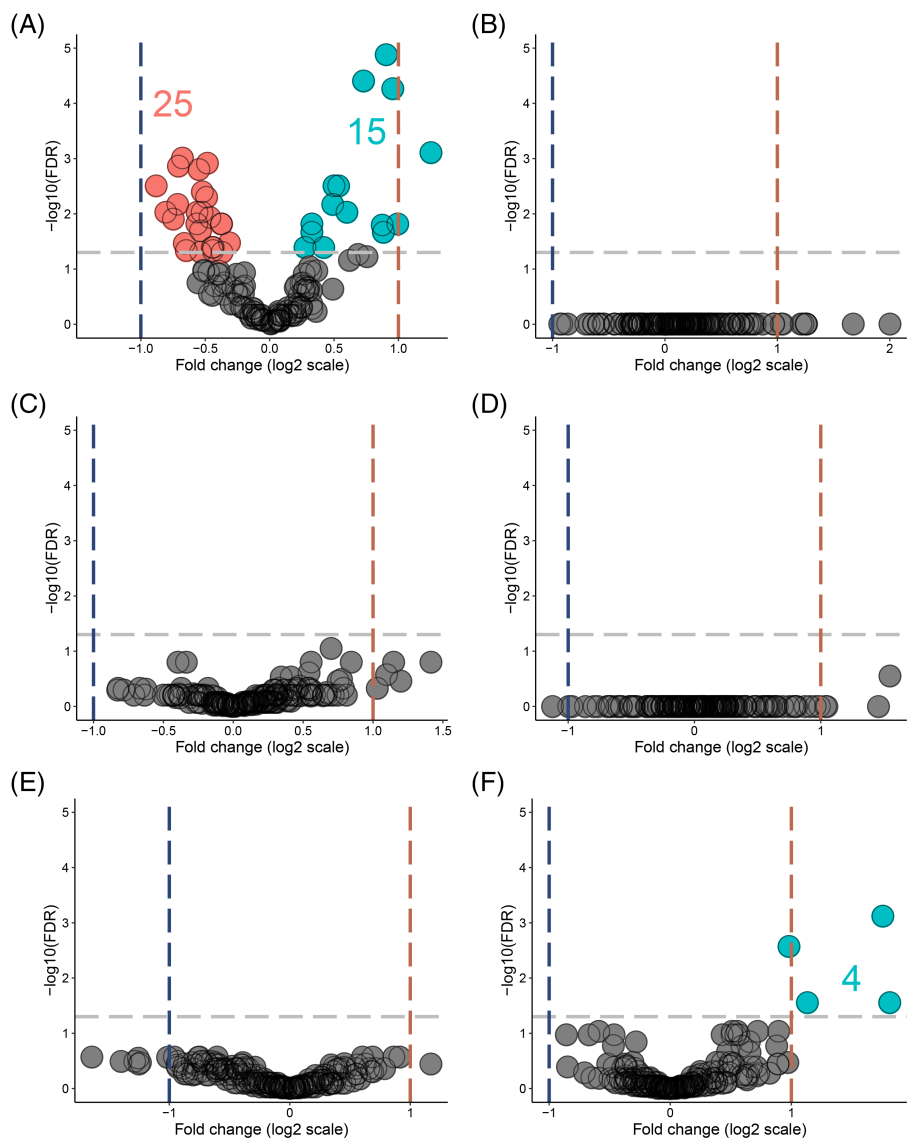
Note: N, healthy *path\_MMR* = 81; N, *path\_MMR* with cancer = 13; N, sporadic rectal cancer patients = 24; N, non-LS controls = 37. FDR <0.05 highlighted with bold. Abbreviations: c-miR, circulating microRNA; FDR, false discovery rate; log2FC, logarithmic2 fold change.

200a-3p was 1.76, indicating significantly higher expression compared to the healthy non-LS controls (Table 2).

Taken together, our findings imply that healthy *path\_MMR* carriers have a systemic c-miRnome that separates them from healthy non-LS persons but resemble the c-miRnome of cancer

patients with or without dMMR. Thus, these findings suggest that sporadic and dMMR-directed carcinogenesis share common miR-targeted biological pathways where potential alterations may produce a detectable c-miR signature in the healthy *path\_MMR* carriers.

**FIGURE 1** Healthy *path\_MMR* carriers have a c-miRnome that differ from non-LS controls but resembles the c-miRnoms of patients with sporadic or hereditary cancer. (A) DE c-miRs in healthy *path\_MMR* carriers vs non-LS controls. (B) DE c-miRs in sporadic rectal cancer patients vs *path\_MMR* carriers with cancer. (C) DE c-miRs in healthy *path\_MMR* carriers vs sporadic rectal cancer patients. (D) DE c-miRs in healthy *path\_MMR* carriers vs *path\_MMR* carriers with cancer. (E) DE c-miRs in *path\_MMR* carriers with cancer vs non-LS controls. (F) DE c-miRs in sporadic rectal cancers patients vs non-LS controls. Blue dash lines indicate negative fold change of expression, red dash line indicate positive fold change of expression and gray dash line indicate FDR <0.05. Downregulated c-miRs are highlighted in red, upregulated c-miRs are highlighted in cyan and nonsignificantly expressed c-miRs are highlighted in gray. Dots represents c-miRs. c-miR, circulating microRNA; FDR, false discovery rate [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



### 3.3 | Dimension reduction analysis of multiple traits was unable to discern *path\_MMR* carriers from sporadic rectal cancer patients

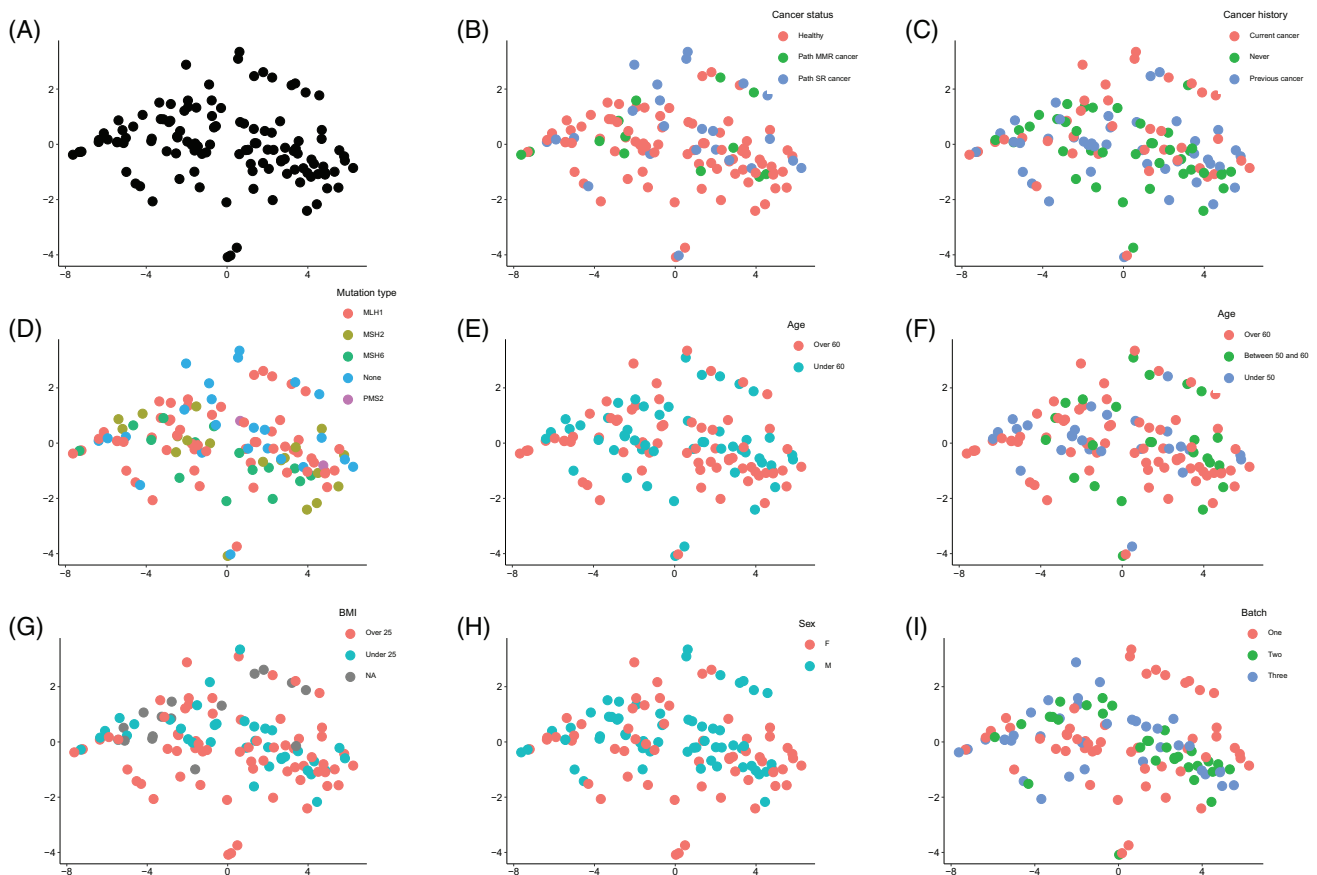
We did not identify DE c-miRs between *path\_MMR* carriers and sporadic rectal cancer patients. Therefore, by using the expression data of all 228 c-miRs shared between the discovery and cancer cohorts, we performed a dimension reduction analysis with t-SNE to identify possible subpopulations within *path\_MMR* carriers and sporadic rectal cancer patients. First, we investigated if phenotypic traits such as being *path\_MMR* carrier, current cancer status, cancer history or *path\_MMR* variant type, would reveal clustering of samples, but did not find any clear patterns (Figure 2A-D). We also investigated if age or BMI would be the discerning traits, but they also failed to reveal any clustering (Figure 2E-G). Finally, sex and the sequencing batch did not form clusters within our dataset (Figure 2H,I). Taken together, the t-SNE analysis supported the DE c-miR findings and was not able to differentiate *path\_MMR* carriers from sporadic cancer patients, which

may be an indicative of shared c-miR-mediated regulation as seen in the DE-analyses.

### 3.4 | Pathway analysis revealed putative c-miR-target genes that are linked to biological processes and pathways associated with cancer

To further evaluate our hypothesis that healthy *path\_MMR* carriers might have a c-miRnome that resembles the c-miRnome of cancer cohort due to shared miR-targeted biological pathways, we next investigated what are the target genes of the observed DE c-miRs. We also inspected what biological processes and pathways these target genes associate with. With mirWalk, we used random-forest-based approach to predict the target genes using databases with experimental validation and high confidence of reported miR-target gene interactions. MirWalk identified a total of 1731 miR-target gene interactions with 508 distinct putative target genes for 32 out of





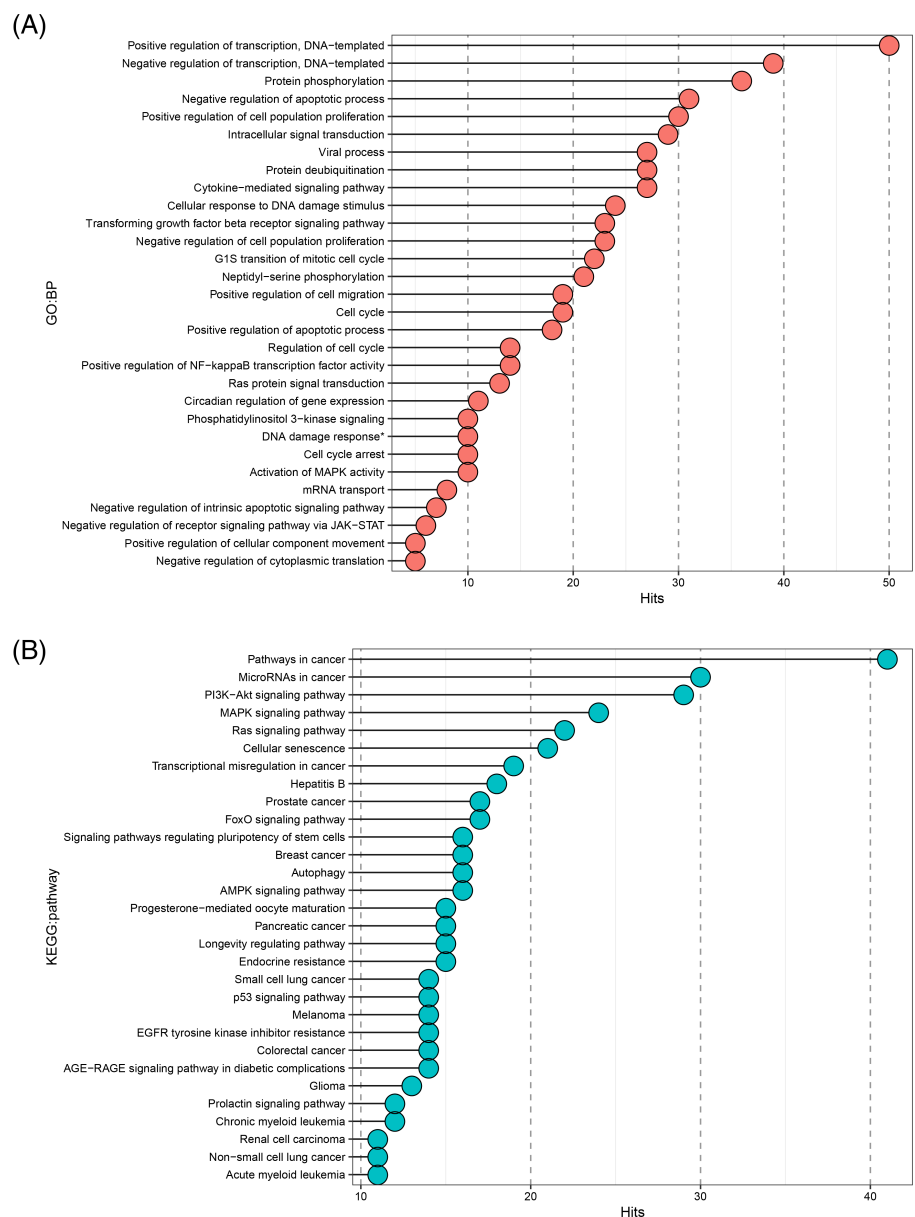
**FIGURE 2** Dimension reduction analysis of multiple traits was unable to discern *path\_MMR* carriers from sporadic rectal cancer patients. (A) *Path\_MMR* carriers and sporadic rectal cancer patients. (B) Cancer status. Healthy, cancer-free *path\_MMR* carriers; *path\_MMR* cancer, *path\_MMR* carriers with cancer; path SR cancer, sporadic rectal cancer patients. (C) Cancer history. Current cancer, has cancer currently; Never, currently healthy, never had cancers; Previous cancer, currently healthy, had had cancer or multiple cancers; (D) *path\_MMR* variant. (E) Dichotomous age. Over 60, persons >60-years of age; Under 60, persons <60-years of age. (F) Nondichotomous age. Over 60, persons >60-years of age; Between 50 and 60, persons between 50 and 60-years of age; Under 50, persons <50-years of age. (G) BMI. Over 25, persons with BMI > 25; Under 25, persons with BMI < 25; NA, no reported BMI. (H) Sex. M, males; F, females. (I) Batch effect of three separate sequencing runs in running order. All t-SNE plots are 2D constructions. Dots represent study subjects. BMI, body mass index

40 observed DE c-miRs from discovery cohort analysis (Supplementary file S2, Tables S3 and S4).

We then performed mirWalk-GSEA analysis on the 508 predicted target genes to explore what functional roles the DE c-miRs might possess. The GSEA analysis revealed 195 distinct significantly enriched biological processes (Supplementary file S2, Tables S5 and S6). To identify the key biological processes, we then narrowed the given output list based on FDR and the number of involved target genes to the top 30 most significantly enriched biological processes (Supplementary file S2, Table S7). Most of the discovered biological processes were linked to apoptosis, regulation of transcription, cell cycle, cell proliferation, DNA damage and signal transduction (Figure 3A). We then conducted a small-scale cell line experiment to investigate how c-miR over- and underexpression affect Human colorectal cell line (HCT116) viability (Supplementary file S1, Supplementary materials and methods). We chose hsa-miR-122b and -451a as representatives of over- and

underexpressed miRs found in healthy *path\_MMR* carriers vs non-LS control comparisons. HCT116 cell line was chosen to mimic LS colorectal cancer. The cell line experiment hinted that overexpression of hsa-miR-122b could reduce cell viability via increased apoptosis whereas underexpression of hsa-miR-451a also resulted in reduced viability but did not induce apoptosis of HCT116 cells (Supplementary file S1, Figure S6). We observed considerable overlap between the identified pathways since 208 out of 508 identified distinct c-miR-target genes contributed to the top biological processes (Supplementary file S2, Table S8). *TGFBR1*, *CDKN1A*, *IGF1*, *TRAF6* and *BCL2* genes were present in most of the observed biological processes along with several other genes (Supplementary file S2, Table S8). The performed in silico target analysis showed that *TGFBR1* was targeted by hsa-miR-27b-3p, *CDKN1A* and *IGF1* were targets of hsa-let-7 e-5p, *TRAF6* was targeted by hsa-miR-125a-5p and *BCL2* was targeted by hsa-miR-125b-5p and hsa-miR-15b-5p (Supplementary file S2, Table S3). Of these c-miRs, all

**FIGURE 3** Pathway analysis revealed putative c-miR-target genes that are linked to biological processes and pathways associated with cancer. (A) Top 30 most enriched biological processes annotated to the identified target genes of 32 out of 40 DE c-miRs found in healthy *path\_MMR* carriers. FDR, false discovery rate; GO:BP, Gene Ontology: biological process; Hits, number of target genes annotated to the biological process. \*Signal transduction by p53 class mediator resulting in cell cycle arrest. (B) Top 30 most enriched KEGG pathways annotated to the identified target genes of 32 out of 40 DE c-miRs found in healthy *path\_MMR* carriers. c-miR, circulating microRNA; FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes pathway; Hits, number of target genes annotated to the pathway [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



except hsa-miR-15b-5p were upregulated in *path\_MMR* compared to controls (Table 3).

Next, we explored how the c-miR-target genes interact with KEGG pathways. GSEA analysis of the same gene set discovered 88 significantly enriched KEGG biological pathways (Supplementary file S2, Tables S9 and S10). Again, to focus on the possible key pathways, we narrowed the output list to the top 30 of the most significant pathways based on similar parameters than in the previous analysis (Supplementary file S2, Table S11). A great majority of the discovered pathways linked to cancer, cancer signaling and cell aging (Figure 3B). Of the 508 predicted target genes, 113 were involved in the discovered top KEGG pathways (Supplementary file S2, Table S12). *AKT3*, *PIK3R1* and *PIK3CA* genes were involved in 27 out of 30 KEGG pathways, whereas *KRAS* had 26 and *NRAS* had 24 hits, respectively (Supplementary file S2, Table S12). *AKT3* was targeted by hsa-miR-15b-5p, *PIK3R1* was targeted by hsa-miR-107

and hsa-miR-486-5p, *PIK3CA* was targeted by hsa-miR-19a-3p, *KRAS* was targeted by hsa-miR-27a-3p and *NRAS* was a target of hsa-let-7a and -7c-5p (Supplementary file S2, Table S3). Of these c-miRs, all except hsa-let7a and -7c, were downregulated in *path\_MMR* compared to controls (Table 3).

As these key target genes were interacting in the majority of the identified cancer-associated biological processes and pathways, we then explored and validated their potential carcinogenic roles. We submitted the gene set to COSMIC-CGC database and found that *BCL2*, *AKT3*, *PIK3CA*, *KRAS* and *NRAS* possess oncogenic functions, whereas *CDKN1A* is a potential oncogene or tumor suppressor gene and *PIK3R1* functions as a tumor suppressor gene (Table 3). All these genes have well-documented roles in multiple tumor types, including colorectal cancer, and with most having functions in hallmarks of cancer.<sup>43</sup> Of the target gene set, *TGFBR1*, *IGF1* nor *TRAF6* were not included in COSMIC-CGC database. These results support our

hypothesis that the observed resemblance of the c-miRNomes between *path\_MMR* carriers and sporadic rectal cancer patients can be due to shared biological processes and pathways that include well-known oncogenes and tumor-suppressor genes.

Taken together, our in silico analysis shows that the c-miRs in hsa-let-7 family, as well as hsa-miR-15b-5p, hsa-miR-19a-3p, hsa-miR-27a-3p and -27b-3p, hsa-miR-107, hsa-miR-125b-5p and hsa-miR-486-5p could target genes that are ubiquitous in cancer-associated biological processes and pathways. These findings imply that the altered c-miRnome expression pattern of cancer-free *path\_MMR* carriers may hold predictive value by tracking potential oncogenic stress caused by dMMR-driven distortions.

## 4 | DISCUSSION

Our study pioneered in characterizing the systemic c-miRNomes of *path\_MMR* carriers. By utilizing high throughput sequencing, a total of 228 distinct c-miRs common to all study subjects were detected. Of these, we showed healthy *path\_MMR* carriers to have an exclusive c-miRnome of 40 DE c-miRs that differs from non-LS-controls, but that does not differ from the c-miRnome of cancer patients with or without dMMR. Our c-miR expression analysis combined with in silico tools revealed that the observed resemblance in the c-miRNomes is possibly caused by distortions in several biological networks that are governed by well-known oncogenes and tumor suppressor genes, thus suggesting that c-miRnome could be used to track potential oncogenic stress at cancer-free state.

There is a growing interest in exploiting miRs as cancer biomarkers. Balaguer et al studied miRs that were extracted from tumors of *path\_MMR* carriers and sporadic colorectal cancer patients with verified microsatellite instability and normal tissue samples.<sup>14,18</sup> They used a set of >700 miR-probes with microarray analysis and detected hundreds of DE miRs among the tissue samples, showing that LS tumors can be separated from sporadic tumors with microsatellite instability, as well as that suspected LS samples discern from confirmed LS samples. Aligned with their study, we also showed that different *path\_MMR* variants do not display unique c-miR expression thus implying a shared systemic response. However, we could not pinpoint DE c-miRs that would distinguish *path\_MMR* carriers from sporadic cancer patients although we did, as well as in numerous other studies, detect a c-miR signature unique to sporadic cancer patients when compared to healthy non-LS controls. The observation that *path\_MMR* carriers do not differ from sporadic cancer patients in their c-miRnome was also supported by our t-SNE analysis that did not reveal any clustering within our dataset based on several variables. The reason behind the substantial difference in DE c-miR numbers between our and the study by Balaguer et al is likely explained by the study setting, used specimen type and methodology. In our study, the DE c-miRs were sequenced from the circulation of cancer-free persons where such a robust c-miR signature is not presumably detected when compared to miRs at the site of pathology.

Furthermore, Balaguer et al detected several DE miRs with diagnostic potential in LS, including hsa-miR-125b-5p, -137, -622, -192 and -1238, whereas Zhou et al displayed that hsa-miR-137, -520 e and -590-3p are indicative of LS by using a subset of *path\_MMR* cancer tumor samples and normal tissue samples from the study by Balaguer et al.<sup>17</sup> We did not find significant overlapping of DE miR content between our c-miRs and tumor-miRs from those studies, except for hsa-miR-125b-5p, that was also identified by Balaguer et al. Aberrant expression of hsa-miR-125b-5p has been reported for multitude of cancer types and it has been implied to serve as a circulating cancer biomarker by targeting apoptosis-regulating oncogene *BCL2*.<sup>44</sup>

The most significant DE c-miR in our setting was hsa-miR-155-5p, followed by hsa-let-7c-5p and -7 e-5p, -122b-3p and 15a-5p, which all except hsa-miR-15a-5p were upregulated in healthy *path\_MMR* carriers. Valeri et al demonstrated that hsa-miR-155-5p targets several MMR-genes and that overexpression of hsa-miR155-5p downregulates *MLH1* and *MSH2* in colorectal cancer cell lines.<sup>15</sup> Within this concept, our DE findings also support the role of hsa-miR-155p modulation in LS pathogenesis even though the performed in silico analysis could not identify MMR-genes as targets of hsa-miR-155-5p. miRs in hsa-let-7 family have been suggested to increase colorectal cancer risk in *path\_MMR* carriers with proficient MMR by lowering the expression of *TGFBR1* haplotype.<sup>45</sup> We found hsa-let-7 family to target *TGFBR3* and hsa-miR-27b-3p to target *TGFBR1*. We did not find experimentally verified target genes for hsa-miR-122b-3p. However, we could see that overexpression of hsa-miR-122b might result in reduced cell viability, plausibly due to increased apoptosis. Previous studies have linked hsa-miR-15a-5p to sporadic endometrial cancer<sup>46</sup> and colorectal cancer,<sup>47</sup> both being hallmark cancers of LS. In our analysis, hsa-miR-15a-5p was seen to target several genes, including known oncogenes and tumor suppressor genes such as *CCND1*, *CDK6* and *DICER1*, thus suggesting biomarker potential also in LS.

MiRs have critical functions across various biological processes and pathways involved in carcinogenesis. We found 508 putative target genes for 32 out of 40 observed DE c-miRs that associate with several pathways common to cancer. In addition to above mentioned c-miRs, we also identified several other c-miRs that could be key regulators in dMMR-driven carcinogenesis. The performed in silico analysis indicated that all these c-miRs target several well-known oncogenes and tumor suppressor genes such as *KRAS*, *NRAS*, *PIK3RI*, and *PIK3CA*, that were significantly enriched in our pathway analysis. Supported by our DE-analysis, the observation that these identified DE c-miRs target known oncogenes and tumor suppressors, could indicate upregulation of the oncogenes and consequently downregulation of the tumor suppressor genes. However, since we studied cell-free c-miRs without possibility to investigate expression levels of their putative target genes, this suggestion remains hypothetical. Unfortunately, c-miRs are not easily tracked where tracking of c-miRs would provide us clues to what tissues they will be affecting and where to seek further signs of cancer development. Matching pairwise tissue samples to observed c-miRs could help elucidate these issues but we

**TABLE 3** Key target genes of DE c-miRs in healthy *path\_MMR* carriers compared to non-LS controls

Key target gene	Gene name	Hits	COSMIC-CGC Role in cancer	c-miR
GO:BP				
<i>TGFBR1</i>	Transforming growth factor-beta receptor type 1	10	NA	hsa-miR-27b-3p ↑
<i>CDKN1A</i>	Cyclin dependent kinase inhibitor 1A	8	Oncogene, tumor suppressor gene	hsa-let-7 e-5p ↑
<i>IGF1</i>	Insulin growth factor 1	7	NA	hsa-let-7 e-5p ↑
<i>TRAF6</i>	TNF receptor-associated factor 6	7	NA	hsa-miR-125a-5p ↑
<i>BCL2</i>	B-cell CLL/lymphoma 2	6	Oncogene, fusion	hsa-miR-125b-5p ↑ hsa-miR-15b-5p ↓
KEGG				
<i>AKT3</i>	V-akt murine thymoma viral oncogene homolog 3	27	Oncogene	hsa-miR-15b-5p ↓
<i>PIK3R1</i>	Phosphoinositide-3-kinase, Regulatory subunit 1 (alpha)	27	Tumor suppressor gene	hsa-miR-107 ↓ hsa-miR-486-5p ↓
<i>PIK3CA</i>	Phosphoinositide-3-kinase, Catalytic, alpha polypeptide	27	Oncogene	hsa-miR-19a-3p ↓
<i>KRAS</i>	KRAS Proto-Oncogene, GTPase	26	Oncogene	hsa-miR-27a-3p ↓
<i>NRAS</i>	NRAS Proto-Oncogene, GTPase	24	Oncogene	hsa-let-7a-5p ↑ hsa-let-7c-5p ↑

Note: Arrows indicate up- (↑) or downregulation (↓) of c-miR in DE-analysis. Hits indicate the number of top GO:BP or KEGG-pathways where the gene is present.

Abbreviations: c-miR, circulating microRNA; COSMIC-CGC, The Catalogue of Somatic Mutations in Cancer and Cancer Gene Census database; GO:BP, Gene Ontology:biological process; KEGG, Kyoto Encyclopedia of Genes and Genomes pathway.

had no possibility to do so. Nevertheless, our exploratory findings indicate that *path\_MMR* carriers display oncogenic stress even when they are cancer-free, but more studies are needed to verify our results and to show if they have true power as a biomarkers of early cancer development. A future goal is to determine whether the longitudinal change or development of c-miRomes appears in conjunction with cancer incidence and treatment. The biological basis for aberrant c-miR expression between *path\_MMR* carriers and non-LS controls remains a clinical question to be elucidated also in the future work.

A major strength of our study is that the study subjects had undergone comprehensive screenings of LS-predisposing mutations, with ascertainment utilizing Amsterdam and Bethesda clinical criteria and cascade testing. Also, instead of a priori chosen gene panel, we conducted a systemic level investigation of c-miRnome, which provides a more comprehensive view of how already identified c-miRs and putative target genes contribute to distorted biological networks in sporadic and hereditary cancer. For example, our findings allow construction of c-miRnome-target gene collection to be explored for potentially distorted biological networks associated with dMMR. Also, it can be used for establishing candidate hypotheses to drive further research and for further exploratory c-miR analyses of potential contributing gene clusters not previously discovered. Finally, the bioinformatic analyses in our study were performed in precise detail according to the latest knowledge using state-of-the-art tools and algorithms.

Our study has potential pitfalls. Although largest to date, the study sample was relatively small especially in the cancer cohort, which could have reduced the statistical power of DE-analyses. Regarding the methodology, there is no conclusive rule which

sequencing depth should be aimed at when assessing DE of c-miRs. In our study, the aimed mean sequencing depth was 5 M reads per sample, but the achieved mean sequencing depth was 3.2 M reads due to underclustering issues in sequencing. The underclustering might have affected c-miR detection by favoring highly expressed c-miRs and thus resulting in overrepresentation of these c-miRs and underrepresentation or masking of c-miRs with low expression and potential cancer- or dMMR-relevant functions. A common issue with c-miRs is the identification of their primary and target locations, and alike in many other studies, we did not track the observed c-miRs to certain locations, which introduce a certain degree of uncertainty over the interpretations of the observations. Unfortunately, our efforts to validate DE findings with RT-qPCR were not completely successful when using an independent validation cohort, although we observed a trend of parallel expression in both cohorts in eight out of nine validation c-miRs. Overall considerable variation in c-miR expression levels were detected with both methods and cohorts, which could explain why significant differences between groups in the smaller validation cohort were not detected. Furthermore, we cannot completely exclude the possibility that varying ascertainment site for sample collection may have increased between sample variation and could thereby have affected our analyses.

To conclude, our exploratory study was the first to characterize the systemic c-miRomes of *path\_MMR* carriers. We showed that systemic c-miRnome can be used to track potential oncogenic stress in cancer-free *path\_MMR* carriers thus paving way for the future investigation of c-miRs in monitoring the risk stratification patterns during the risk-reducing clinical surveillance and possible cancer

management. Our study also produced novel insight that allows construction of a c-miRnome-target gene collection to be explored for potentially distorted biological networks and c-miRnome-target gene interactions in LS.

## AUTHOR CONTRIBUTIONS

**Tero Sievänen:** Formal analysis, investigation, methodology, software, validation, visualization, writing original draft, writing-review and -editing. **Tia-Marje Korhonen:** Formal analysis, methodology, software, visualization, writing-review and -editing. **Tiina Jokela:** Methodology, software, validation, visualization. **Maarit Ahtiainen:** Methodology, writing-review and -editing. **Laura Lahtinen:** Methodology, writing-review and -editing. **Teijo Kuopio:** Methodology, writing-review and -editing. **Anna Lepistö:** Writing-review and -editing. **E. Sillanpää:** Conceptualization, supervision, writing-review and -editing. **Jukka-Pekka Mecklin:** Conceptualization, resources, writing-review and -editing. **Toni T. Seppälä:** Conceptualization, funding acquisition, resources, supervision, writing-review and -editing. **Eija K. Laakkonen:** Conceptualization, data curation, funding acquisition, project administration, resources, supervision, writing-review and -editing. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

## ACKNOWLEDGEMENTS

We acknowledge the support from Biosciences team of IT Center for Science Finland (CSC) for providing HPC-resources for our data analytics. We would also like to thank research assistant Mintä Kärkkäinen for her valuable help with the cell line experiments.

## FUNDING INFORMATION

Eija K. Laakkonen was supported by the Päivikki and Sakari Sohlberg Foundation. Elina Sillanpää was supported by the Academy of Finland research fellowship (grant number: 341750). Toni T. Seppälä was supported by Finnish Medical Foundation, Emil Aaltonen Foundation, Jane and Aatos Erkko Foundation, Sigrid Juselius Foundation, Finnish Cancer Foundation, Relander Foundation, Academy of Finland (grant number: 338657), HUS State Research Funds (TYH2021123 and TYH2022323) and iCAN Digital Precision Medicine Flagship. Jukka-Pekka Mecklin was supported by Jane and Aatos Erkko Foundation, Finnish Cancer Foundation and KYS State Research Funds.

## CONFLICT OF INTEREST

Toni T. Seppälä declares being CEO and co-owner of HealthFund Finland and Consultation fees from Boehringer Ingelheim and Amgen Finland. The other authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The datasets supporting the conclusions of this article are available in the Gene Expression Omnibus (GSE198834). Other data that support the findings of our study are available from the corresponding author upon request. Step-by-step analysis protocols can be accessed via GitHub repository (<https://zenodo.org/badge/latestdoi/467491700>).

## ETHICS STATEMENT

Informed consent was obtained from all participants, and the Helsinki and Uusimaa Health Care District (HUS/155/2021) and Central Finland Health Care District Ethics Committee (KSSHP D# 1U/2018 and 1/2019 and KSSHP 3/2016) approved the study protocol. The study was conducted according to the guidelines of the Declaration of Helsinki.

## ORCID

Tero Sievänen  <https://orcid.org/0000-0002-9660-4559>

## TWITTER

Tero Sievänen  @SievanenTero

Eija K. Laakkonen  @eija\_laakkonen

## REFERENCES

- Dominguez-Valentin M, Sampson JR, Seppälä TT, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the prospective Lynch syndrome database. *Genet Med*. 2020;22:15-25.
- Møller P, Seppälä T, Bernstein I, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. *Gut*. 2017;66:464-472.
- Renkonen-Sinisalo L, Seppälä TT, Järvinen HJ, Mecklin JP. Subtotal colectomy for colon cancer reduces the need for subsequent surgery in Lynch syndrome. *Dis Colon Rectum*. 2017;60:792-799.
- Møller P, Seppälä T, Bernstein I, et al. Incidence of and survival after subsequent cancers in carriers of pathogenic MMR variants with previous cancer: a report from the prospective Lynch syndrome database. *Gut*. 2017;66:1657-1664.
- Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;6:857-866.
- Jung G, Hernández-Illán E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat Rev Gastroenterol Hepatol*. 2020;17:111-130.
- Goodall GJ, Wickramasinghe VO. RNA in cancer. *Nat Rev Cancer*. 2021;21:22-36.
- Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res*. 2008;18:997-1006.
- Mori MA, Ludwig RG, Garcia-Martin R, Brandão BB, Kahn CR. Extracellular miRNAs: from biomarkers to mediators of physiology and disease. *Cell Metab*. 2019;30:656-673.
- FrancaVilla A, Turoczi S, Tarallo S, Vodicka P, Pardini B, Naccarati A. Exosomal microRNAs and other non-coding RNAs as colorectal cancer biomarkers: a review. *Mutagenesis*. 2020;35:243-260.
- Muñelo-Romay L, Casas-Arozamena C, Abal M. Liquid biopsy in endometrial cancer: new opportunities for personalized oncology. *Int J Mol Sci*. 2018;19:19.
- He Y, Deng F, Yang S, et al. Exosomal microRNA: a novel biomarker for breast cancer. *Biomark Med*. 2018;12:177-188.
- Hu C, Meiners S, Lukas C, Stathopoulos GT, Chen J. Role of exosomal microRNAs in lung cancer biology and clinical applications. *Cell Prolif*. 2020;53:e12828.
- Balaguer F, Moreira L, Lozano JJ, et al. Colorectal cancers with microsatellite instability display unique miRNA profiles. *Clin Cancer Res*. 2011;17:6239-6249.
- Valeri N, Gasparini P, Fabbri M, et al. Modulation of mismatch repair and genomic stability by miR-155. *Proc Natl Acad Sci*. 2010;107:6982-6987.

16. Liccardo R, Sessa R, Trombetti S, et al. Mir-137 targets the 3' untranslated region of msh2: potential implications in lynch syndrome-related colorectal cancer. *Cancer*. 2021;13:1-12.
17. Zhou C, Li J, Li J, et al. Hsa-miR-137, hsa-miR-520 e and hsa-miR-590-3p perform crucial roles in lynch syndrome. *Oncol Lett*. 2016;12:2011-2017.
18. Balaguer F, Link A, Lozano JJ, et al. Epigenetic silencing of miR-137 is an early event in colorectal carcinogenesis. *Cancer Res*. 2010;70:6609-6618.
19. Pavicic W, Perkiö E, Kaur S, Peltomäki P. Altered methylation at microRNA-associated CpG islands in hereditary and sporadic carcinomas: a methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA)-based approach. *Mol Med*. 2011;17:726-735.
20. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004;96:261-268.
21. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the international collaborative group on HNPCC. *Gastroenterology*. 1999;116:1453-1456.
22. Thompson BA, Spurdle AB, Plazzer J-P, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSIGHT locus-specific database. *Nat Genet*. 2014;46:107-115.
23. Kovanen V, Auque P, Kokko K, et al. Design and protocol of estrogenic regulation of muscle apoptosis (ERMA) study with 47 to 55-year-old women's cohort: novel results show menopause-related differences in blood count. *Menopause*. 2018;25:1020-1032.
24. Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data* [Online]. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed April 26, 2022.
25. Hannon GJ. *FASTX-Toolkit*. 2010. [http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit). Accessed April 26, 2022.
26. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;34:140-144.
27. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009;26:139-140.
29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1-21.
30. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Version 4.0.5. Vienna, Austria: R Foundation for Statistical Computing; 2021. <https://www.R-project.org/>. Accessed April 26, 2022.
31. van der Maaten L, Hinton G. Visualizing data using t-SNE. *Laurens*. *J Mach Learn Res*. 2008;9:2579-2605.
32. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: an online resource for prediction of microRNA binding sites. *PLoS One*. 2018;13:e0206239.
33. Ding J, Li X, Hu H. TarPmiR: a new approach for microRNA target site prediction. *Bioinformatics*. 2016;32:2768-2775.
34. Huang HY, Lin YCD, Li J, et al. MiRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res*. 2020;48:D148-D154.
35. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48:D127-D131.
36. McGeary SE, Lin KS, Shi CY, et al. The biochemical basis of microRNA targeting efficacy. *Science*. 2019;366:6472.
37. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102:15545-15550.
38. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28:27-30.
39. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18:696-705.
40. Bohaumilitzky L, Kluck K, Hüneburg R, et al. The different immune profiles of normal colonic mucosa in cancer-free Lynch syndrome carriers and Lynch syndrome colorectal cancer patients. *Gastroenterology*. 2022;162:907-919.e10.
41. Saberinia A, Alinezhad A, Jafari F, Soltany S, Akhavan SR. Oncogenic miRNAs and target therapies in colorectal cancer. *Clin Chim Acta*. 2020;508:77-91.
42. Carter JV, Galbraith NJ, Yang D, Burton JF, Walker SP, Galandiuk S. Blood-based microRNAs as biomarkers for the diagnosis of colorectal cancer: a systematic review and meta-analysis. *Br J Cancer*. 2017;116:762-774.
43. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646-674.
44. Wang Y, Zeng G, Jiang Y. The emerging roles of miR-125b in cancers. *Cancer Manag Res*. 2020;12:1079-1088.
45. Xicola RM, Bontu S, Doyle BJ, et al. Association of a let-7 miRNA binding region of TGFBR1 with hereditary mismatch repair proficient colorectal cancer (MSS HNPCC). *Carcinogenesis*. 2016;37:751-758.
46. Zhou L, Wang W, Wang F, et al. Plasma-derived exosomal miR-15a-5p as a promising diagnostic biomarker for early detection of endometrial carcinoma. *Mol Cancer*. 2021;20:57.
47. Li Z, Zhu Z, Wang Y, et al. hsa-miR-15a-5p inhibits colon cell carcinoma via targeting CCND1. *Mol Med Rep*. 2021;24:735.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Sievänen T, Korhonen T-M, Jokela T, et al. Systemic circulating microRNA landscape in Lynch syndrome. *Int J Cancer*. 2022;1-13. doi:10.1002/ijc.34338