Towards a consensus microRNA signature of primary and metastatic colorectal cancer

Bastian Fromm[1,2,±], Eirik Høye[1,±], Paul Heinrich Michael Böttger[3], Diana Domanska[4], Annette Torgunrud Kristensen[1], Christin Lund-Andersen[1,5], Torveig Weum Abrahamsen[1], Åsmund Avdem Fretland[5,6,7], Vegar Johansen Dagenborg[1,5], Susanne Lorenz[8], Bjørn Edwin[5,6,7], Eivind Hovig[1,9], Kjersti Flatmark [1,5,10*]

[1] Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway
[2] Science for Life Laboratory, Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, Stockholm, Sweden
[3] Independent researcher, Linz, Austria
[4] Department of Pathology, University of Oslo, Oslo, Norway
[5] Institute of Clinical Medicine, Medical Faculty, University of Oslo, Oslo, Norway
[6] The Intervention Centre, Rikshospitalet, Oslo University Hospital, Oslo, Norway
[7] Department of Hepato-Pancreato-Biliary Surgery, Rikshospitalet, Oslo University Hospital, Oslo, Norway
[8] Department of Core Facilities, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway
[9] Center for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway
[10] Department of Gastroenterological Surgery, The Norwegian Radium Hospital, Oslo University Hospital, Nydalen, Oslo, Norway

[±] contributed equally

[*] corresponding author: Kjersti Flatmark: kjersti.flatmark@rr-research.no

Abstract

Background

Although microRNAs (miRNAs) are involved in all hallmarks of cancer, miRNA dysregulation in the metastatic process remains poorly understood. We investigated the role of miRNAs in metastatic evolution of colorectal cancer (CRC) by analyzing smallRNA-seq datasets from primary CRC, metastatic locations (liver, lung and peritoneum), and corresponding adjacent tissues. Addressing main challenges of miRNA analysis, a bioinformatics pipeline was developed that contains *bona fide* miRNA annotations from MirGeneDB, utilizes the quality control software miRTrace, applies physiologically meaningful cutoffs and accounts for contribution of cell-type specific miRNAs and host tissue effects.

Results

Two hundred-and-seventy-five miRNA sequencing datasets were analyzed, and after adjusting for the contribution of heterogeneity in cellular composition, strong signatures for primary and metastatic CRC were identified. The signature for primary CRC contained many previously known miRNAs with known functions. Deregulation of specific miRNAs was associated with individual metastatic sites, but the metastatic signatures contained overlapping miRNAs involved in key elements of the metastatic process, such as epithelial-to-mesenchymal transition and hypoxia. Notably, four of these miRNAs (MIR-8, MIR-10, MIR-375, MIR-210) belong to deeply conserved families present in many other organisms, triggering questions about their evolutionary functions and opportunities for experimental validation.

Conclusion:

Applying a meticulous pipeline for the analysis of smallRNA-seq data, miRNA signatures for primary and metastatic CRC were identified, contributing novel insights into miRNA involvement in CRC metastatic evolution and site-specific metastatic adaptations. New datasets can easily be included in this publicly available pipeline to continuously improve the knowledge in the field.

Keywords: microRNAs, colorectal cancer, metastases, biomarkers, bioinformatics, MIR-8, MIR-10, MIR-375, MIR-210

Background

Colorectal cancer (CRC) is a heterogeneous disease and a leading cause of cancer-related deaths worldwide [1]. CRC evolution is an only partially understood process that starts with the formation of primary tumors from epithelial cells in the colorectum and commonly leads to metastatic progression, which is the major cause of mortality. The main metastatic sites are the liver, lungs and peritoneal surface, emphasizing the substantial changes these cells undergo during their evolution that enable them to travel and establish in distant organs with very different microenvironments [2,3]. Although the molecular landscape of primary CRC (pCRC) has been extensively characterized and several mechanisms for the onset of tumorigenesis are known [4], the genomic and transcriptomic changes in metastatic CRC (mCRC) are less well understood [5–7]. A better understanding of these changes is warranted, because the adaptation to a metastatic phenotype not only represents a pivotal step in cancer evolution, but could be exploited as potential diagnostic, prognostic, and predictive biomarkers of clinical relevance.

MicroRNAs (miRNAs) are evolutionary ancient post-transcriptional gene regulators involved in cell-type specification, tissue identity and development in animals [8]. With more than 11,000 annual publications in 2019 alone they are the most studied RNA molecules to date [9]. Because their dysregulation correlates with all hallmarks of cancer [10], due to their remarkable chemical stability [11], and the availability of sensitive detection methods, miRNAs have repeatedly been suggested as promising cancer biomarkers [12,13]. However, although numerous candidate biomarkers and miRNA signatures have been suggested, there is a striking lack of overlap between published results from different groups, and to date, no clinically implemented miRNA biomarker is available for pCRC [6,14–19]. In mCRC, the absence of consistent data is even more apparent, and there is no consensus regarding which miRNAs are up- and down-regulated in mCRC, representing a major obstacle to understanding the role of miRNAs in metastatic progression as well as to the identification of relevant biomarkers [6].

This lack of consensus is surprising and clearly is not reflecting biology or evolution of CRC but might rather be due to known challenges in miRNA research. First, varying definitions of human miRNA genes have previously resulted in substantially different sets of miRNAs, and non-miRNAs, being profiled between different studies. Second, a uniform bioinformatics workflow tailored for smallRNA-seq analysis of bulk samples from different organs, tissues and pathological states was missing. Third, for differential expression analysis, commonly reported miRNAs did not meet physiologically relevant expression levels in either of the studied samples [20]. Finally, a strategy to account for cellular composition effects when analyzing bulk tissue cancer specimens has usually not been applied [21].

To overcome these challenges, our group previously showed that the majority of predicted human miRNAs in the most commonly used miRNA reference database, miRBase, were not derived from miRNA genes, and that including such sequences in miRNA studies could affect the results [22]. For instance, in a previous study by Neerincx et al [14], the reported "miRNA" signature based on miRBase annotations for mCRC consisted solely of genes that were all not *bona fide* miRNAs (Mir-320 - miRtron [23], Mir-3117 - overlaps with protein coding gene, Mir-1246 - a U2 derived degradation product [24], and Mir-663 - a rRNA fragment [25]. To enable more meaningful and comparative studies, we set up MirGeneDB, a publicly available, curated miRNA gene database [26]. The challenge of ensuring quality and reproducibility when comparing smallRNA-seq data across studies was resolved by Kang et al who developed the miRTrace software as a universal quality control pipeline specifically for smallRNA-seq data [27]. To ensure physiologically meaningful levels of miRNAs in differential expression analysis between two sample groups such as pCRC and mCRC, we applied a strict cutoff of 100 reads per million (RPM) as the minimum expression level, which also implied lower dependence on technical factors such as, for instance, sequencing depth [14].

The final challenge relates to the analysis of datasets generated from bulk tissue samples of varying cellular composition, such as normal tissue and primary tumor tissue biopsies. Although the majority of miRNAs are ubiquitously expressed in

various tissues, many miRNAs are exclusively expressed in particular cell types, tissues, organs, or at specific developmental time points [21,28–30]. Such (cell-) specific miRNAs may confound the interpretation of results when comparing miRNA expression between different tissue types, such as primary and metastatic tumors and tumors derived from different metastatic sites. To address this we took advantage of previous studies on cell-type specific miRNA expression and incorporated the information of cell specificity of miRNAs into our analysis [28–31].

Taking all these aspects into account and addressing each of them in a comprehensive, reproducible and flexible pipeline, we analyzed novel smallRNA-seq data from pCRC and mCRC, combined with normal adjacent tissues and existing smallRNA-seq datasets, altogether totaling 193 datasets. We derived a robust consensus miRNA signature of pCRC and mCRC that partially overlap with previous reports but which also includes novel miRNAs. Specifically, differential expression analysis revealed a set of deregulated miRNAs when comparing pCRC and multiple metastatic sites, including Mir-10-P1a, Mir-375, Mir-8-P1b and the hypoxamiR Mir-210 that was validated in independent samples and point to an evolutionary adaptation of metastatic tumor cells to hypoxic conditions in distant sites.

Results

Next generation sequencing (NGS) data collection and processing

Sequencing samples from three paired primary CRC (pCRC) and normal colorectum (nCR), 9 pairs of CRC liver metastasis (CLM) and normal liver (nLi), 7 pairs of CRC lung metastasis (mLu) and normal lung (nLu), and 9 peritoneal metastasis (PM-CRC) were included in the study. Five previous studies with relevance for the discovery of NGS based miRNA signatures in pCRC and CLM were available for download [14–17,32], while data from four published studies (pCRC and cell lines) were not accessible despite repeated requests [33–36]. Combined datasets consisting of 275 NGS datasets (pCRC=141, nCR=32, CLM=38, nLi=24, mLu=11, nLu=7, PM-CRC=22) were processed. All datasets were run through the miRTrace quality control pipeline [27] and subsequently mapped to MirGeneDB.org [26].

Consensus datasets

Four of the six tissue-based datasets fulfilled the quality control criteria, yielding 193 total NGS datasets (Additional file 1: Supplementary Methods), while two studies were excluded. One had low quality reads in the majority of samples [15] (Additional file 3: Supplementary Methods), and one because of presence of contamination [16] (Additional file 2: Supplementary Methods), resulting in a highly congruent group of datasets ('consensus datasets' N=193 (70.2%), pCRC=94 (66.7%), nCR=24 (75.0%), CLM=18 (47.4%), nLi=19 (79.2%), mLu=10 (90.9%), nLu=7 (100%), PM-CRC=19 (86.4%)) [14,17,32], and these datasets were included in further analyses (Figure 1A).

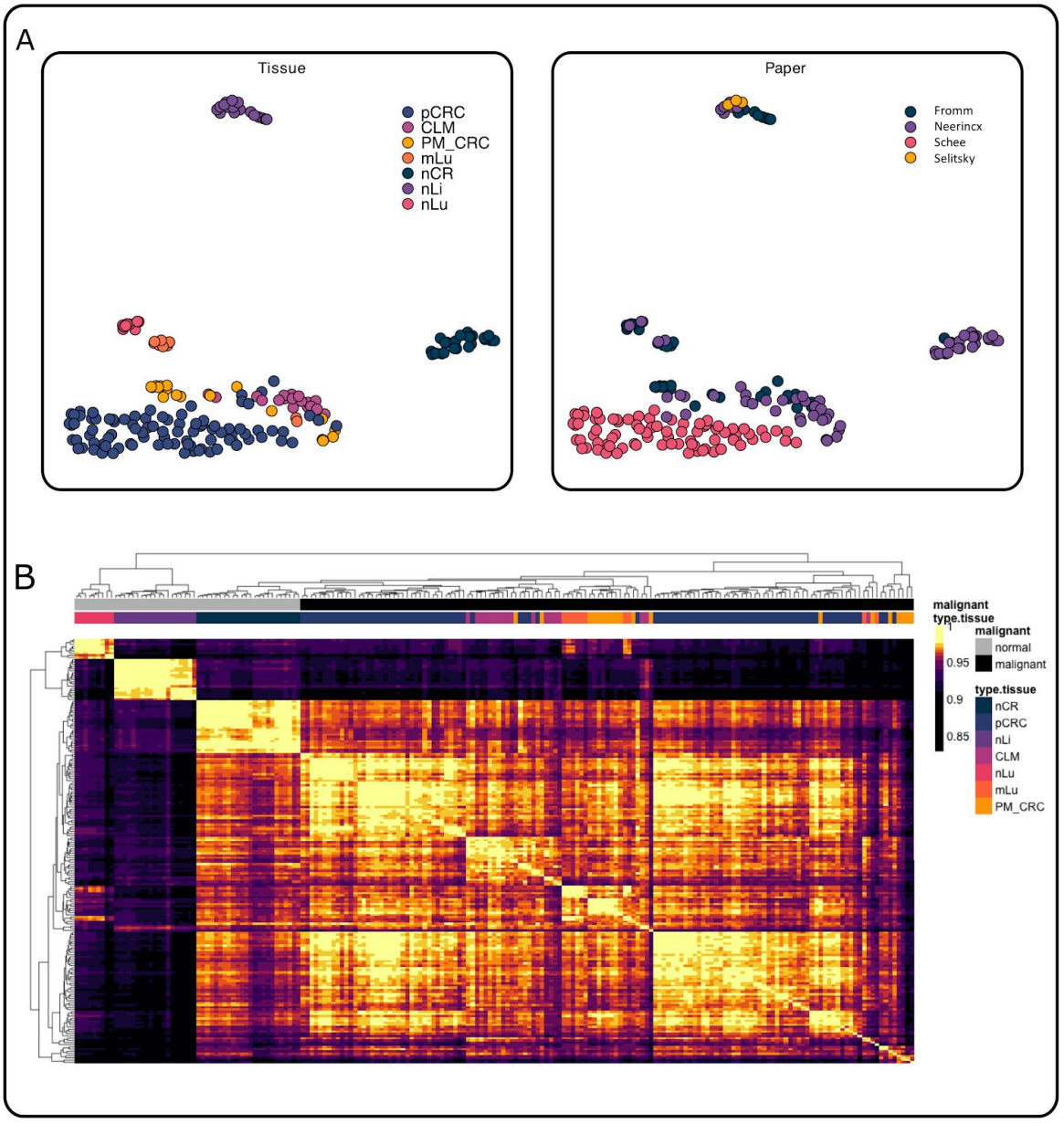


Figure 1: MiRNA expression profiles can distinguish normal adjacent and cancerous tissues at the global level independent of study. **a** Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) cluster plot based on global miRNA expression, normalized by varianceStabilizingTransformation (VST), from the DESeq2 bioconductor package, colored by tissue type (left) and study (right). **b** Correlation plot of global, VST transformed miRNA expression, annotated by normal or malignant tissue, as well as by tissue type of origin. Primary colorectal cancer (pCRC), normal colorectal tissue (nCR), CRC liver metastasis (CLM), normal adjacent liver tissue (nLi), CRC lung metastasis (mLu), normal adjacent lung tissue (nLu), CRC peritoneal metastasis (PM-CRC).

Global miRNA expression

Global miRNA expression was analyzed using the dimensionality reducing technique UMAP (Figure 1A) and unsupervised clustering (Figure 1B). UMAP revealed that the consensus datasets clustered according to the tissue of origin (Figure 1A left), and not according to the originating study (Figure 1A right), indicating that the consensus datasets are representative of their respective tissue. All normal tissues (nLi, nLu and nCR) formed distinct clusters distant from each other and from the tumor tissues. Of note, the UMAP plot showed that the CLM and mLu tissues formed distinct clusters. Unsupervised clustering (Figure 1B) showed a clear difference between normal and tumor tissues. The normal tissues (nLi, nLu and nCR) formed distinct clusters (Figure 1B), while nCR and the tumor samples (pCRC, CLM, mLu and PM-CRC), were part of a major cluster separate from nLi and nLu (Figure 1B).

Cell specific miRNA expression

McCall et al 2017 [28] reported a set of 45 miRNAs with cell-type specific expression. These miRNAs were used to analyze the contribution of cell specific miRNAs on expression data derived from bulk tissue (see Additional file 6: Supplementary Methods for all cell specific miRNAs used in this analysis). PCA analysis revealed that when including only cell specific miRNAs in the analysis, datasets still clustered into groups according to the biological origin of the sample, with nLu and nLi being clearly distinct from the tissues of colorectal origin (nCR, pCRC, CLM, mLu and PM-CRC), as shown in Figure 2A (left). The correlation circle (Figure 2A, right) shows to what extent miRNAs contribute to the variance along a dimension, focusing on the top 15 most contributing miRNAs. Notably, the PCA analysis showed that Mir-486_5p, Mir-126_5p, Mir-342_3p and Mir-204-P2_5p correlated with the axis separating nLu and nLi from the malignant tissues, correlating with higher expression in nLu and nLi. Mir-8-P2b_3p, Mir-8-P2a_3p and Mir-17-P3c_5p also correlate with the axis separating malignant tissues from nLi and nLu, although in this case correlating with higher expression in the malignant tissues. Both the mesenchymal specific Mir-145_5p and Mir-143_3p correlated with the axis separating nCR and pCRC and the metastatic tissues, in addition to Mir-129-P1_5p, Mir-133-P1_3p, Mir-133-P2_3p,

Mir-9-P3_5p, Mir-1-P1_3p, and Mir-1-P2_3p.

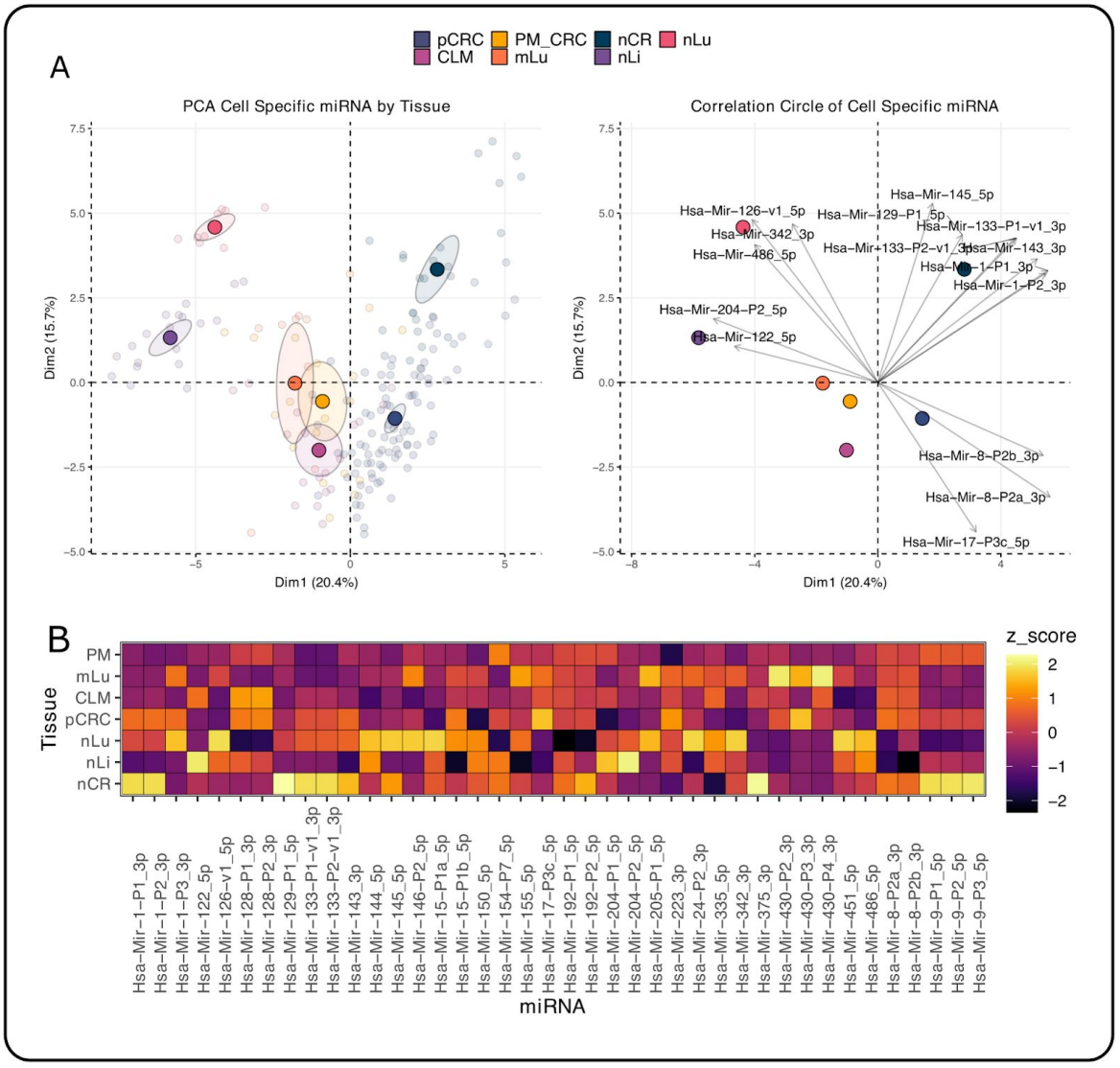


Figure 2: Cell specific miRNAs provide insight into cellular composition of tissue. **a** Left is a Principal Component Analysis (PCA) plot of cell-specific miRNA, VST normalized, colored by tissue. Transparent points are individual samples, solid points represent the mean position of the tissue and shade is 95% confidence of the mean. The correlation circle (Figure 2A, right) shows to what extent miRNAs contribute to the variance along a dimension, focusing on the top 15 most contributing miRNAs. For example, if the direction of an arrow in the correlation plot (right) is the same as the direction that separates two tissues on the PCA plot (left), it means that this particular miRNA contributes to the variance separating the two tissues in dimension 1 and dimension 2. **b** Heatmap of cell-specific miRNAs in each tissue. Color scale is z-score of Reads Per Million (RPM) per miRNA. Only miRNAs with greater than 100 RPM in at least one of the tissues are shown.

Comparing cell specific miRNA expression in the different tissues using a heatmap (Figure 2B), scaled by a z-score of Reads Per Million (RPM) values for each miRNA, cell specific miRNAs tended to be expressed at lower levels in malignant tissues compared to the normal tissues (Figure 2B). Not surprisingly, the hepatocyte specific Mir-122_5p had a much higher expression in nLi compared to nCR. While Mir-145_5p appeared to have a higher expression level in nCR compared to pCRC, Mir-143_3p could not be visually observed to be differentially expressed.

Finally, to assess the influence of cell specific miRNAs on tissue clustering with UMAP, possibly representing the presence of normal adjacent tissue cells in the metastatic tissue samples, cell specific miRNAs were excluded from the analysis (Supplementary Figure 2c, Additional file 6: Supplementary Methods). Even without the cell specific miRNAs, tissues clustered similarly to the datasets with all miRNAs, with nLi, nLu and nCR forming distinct clusters. pCRC, CLM and PM-CRC formed separate clusters, though the CLM and PM-CRC were clearly distinct from the pCRC.

A miRNA signature for pCRC compared to nCR

To test which individual miRNAs are driving the observed global differences between nCR and pCRC, expression levels of all miRNA genes were compared between the two groups, annotating cell specific miRNAs, as well as miRNAs reported to be expressed in plasma of healthy individuals. Thirty-seven miRNAs were up-regulated and 37 miRNAs were down-regulated; and of these, 15 up- and down-regulated miRNAs were cell specific miRNAs and 20 up- and down-regulated miRNAs were reported to be expressed at a high level in peripheral blood cells [37]. Global differential expression results are shown as a volcano plot in Figure 3A, and the signature miRNAs found to be differentially expressed above the defined thresholds are shown in a bar plot in Figure 3B. Of the initial signature, 15 miRNAs were cell-type specific, including Mir-143, Mir-145, Mir-150, Mir-223, and Mir-342. Mir-143 and Mir-145 (lower expression in pCRC) are reported to be exclusively expressed in mesenchymal cells, specifically fibroblasts and smooth muscle cells [38]. Mir-150 and Mir-342 are predominantly expressed by lymphocytes, and we observed a strong reduction in

expression levels in the tumor samples compared to nCR. Mir-223, which is specific to dendritic cells and macrophages [28], was up-regulated in pCRC.

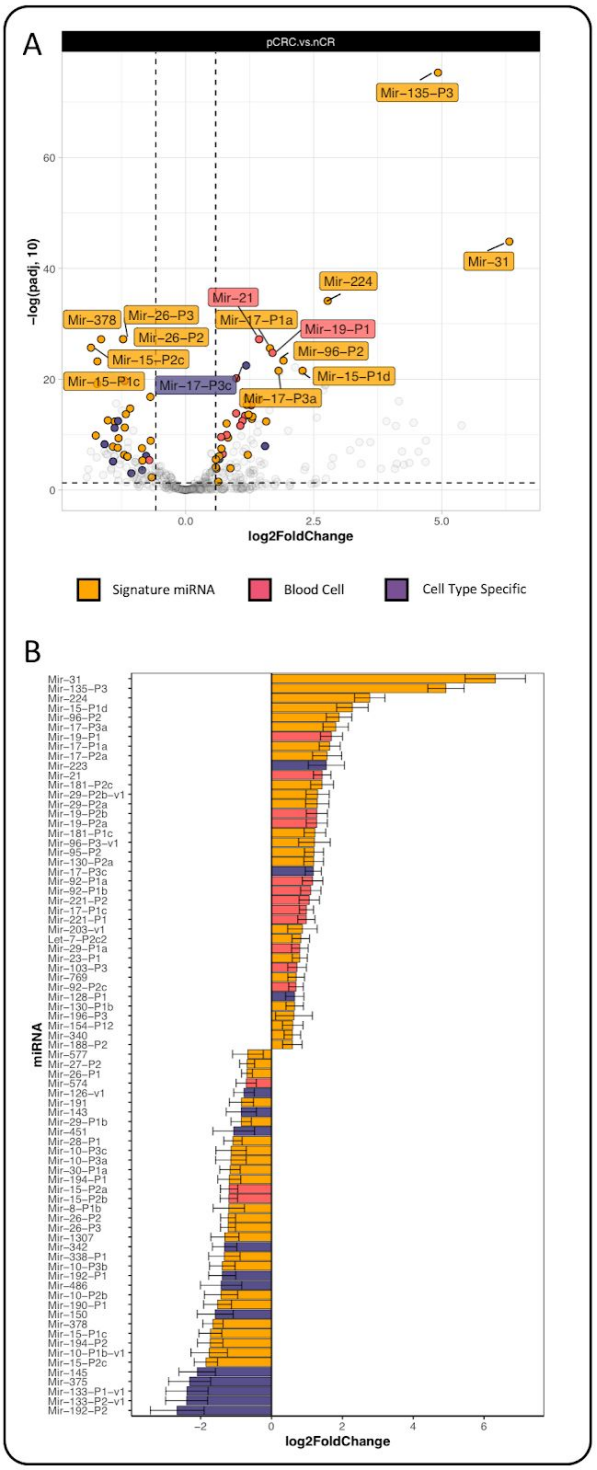


Figure 3: Distinct miRNA signatures in primary colorectal cancer (pCRC). **a** Volcano plot of differentially expressed genes in nCR versus pCRC. The x-axis shows Log 2 Fold Change (LFC) while y-axis shows -Log 10 Benjamini-Hochberg adjusted P-values. Cell-specific miRNAs are known to be exclusively expressed in one or a few cell types. Signature miRNA are miRNA differentially expressed between nCR and pCRC, where expression thresholds are greater than 100 Reads Per Million (RPM) in either nCR or pCRC. Blood cell miRNA are miRNA that have been reported to be expressed in one or more blood cells. **b** Barplot of LFC of differentially expressed miRNA in pCRC compared to nCR.

Signatures for mCRC compared to pCRC, accounting for normal tissue expression

Common and independent miRNA signatures were identified for CLM, mLu and PM-CRC, compared to pCRC. Altogether, 27 miRNAs were up-regulated and 7 miRNAs were down-regulated in the metastatic lesions when compared to pCRC (Figure 4). To assess miRNAs that are differentially expressed when comparing pCRC and mCRC, an important consideration is the confounding effect of miRNAs derived from the normal tissue surrounding the metastasis, which inevitably will be present in the bulk metastasis tissue samples. This issue was solved by omitting miRNAs that were differentially expressed between nCR and nLi or nLu, as well as between pCRC and nLi or nLu from the signature.

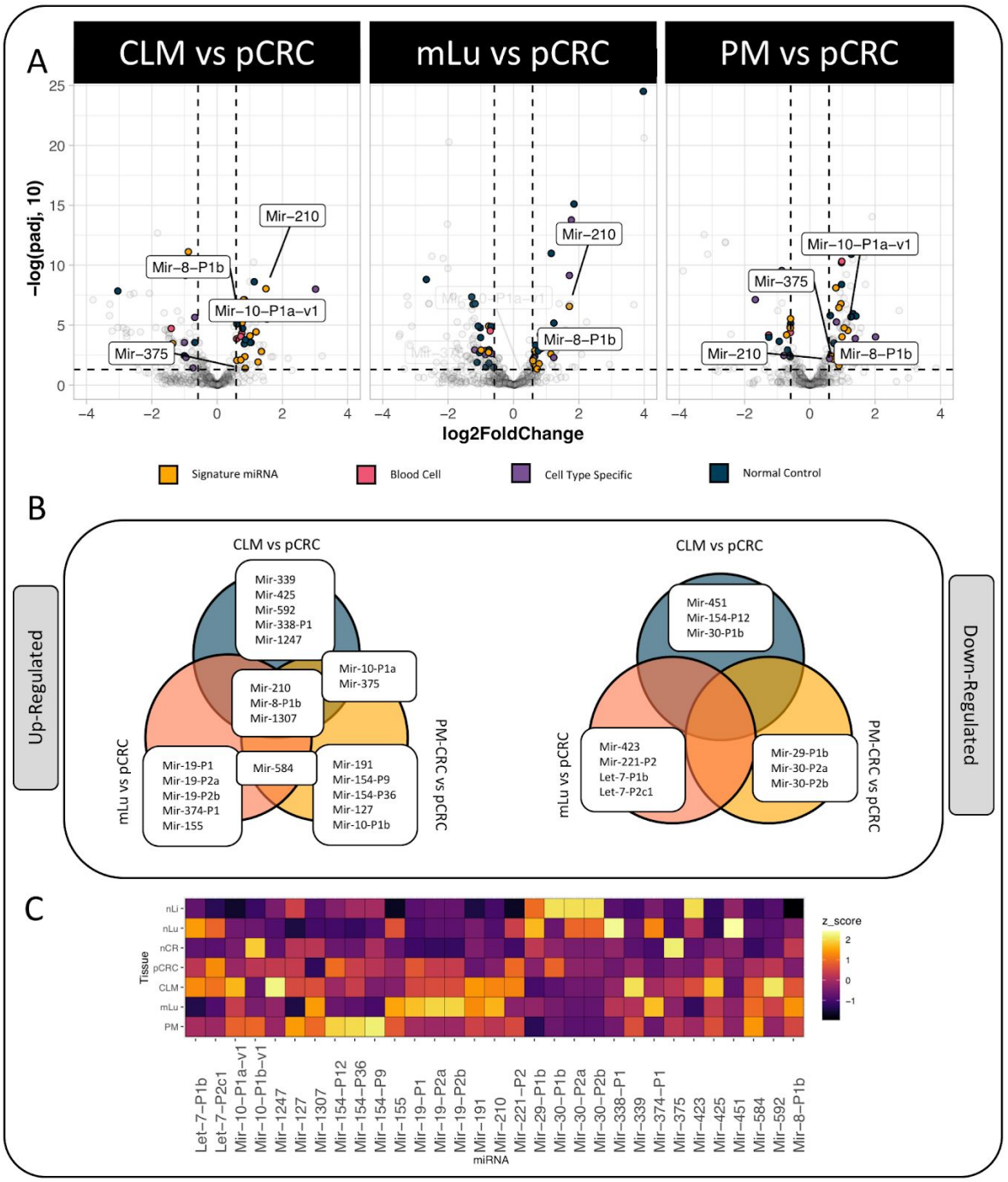


Figure 4: Distinct miRNA signatures in metastatic colorectal cancer (mCRC). **a** Volcano plots show differentially expressed miRNAs in CLM, mLu and PM-CRC compared to pCRC. Cell-specific miRNAs are known to be exclusively expressed in one or a few cell types. Normal control miRNAs are also differentially expressed in normal tissue. Signature miRNA are miRNA differentially expressed between pCRC and mCRC, and where expression levels were greater than 100 Reads Per Million (RPM) in either the metastasis or primary tumor. **b** Venn diagrams show how many of the mCRC signatures overlap between the different metastatic sites. **c** Heatmap compares expression levels of signature miRNA in the tissue types. The color scale is the z-score of RPM values for each individual miRNA.

Five miRNAs were differentially expressed across multiple metastatic sites (Figure 4 and Figure 5). Mir-210_3p showed strong up-regulation in both CLM and mLu, with more than twice as high expression levels compared to pCRC, whereas in PM-CRC, the increase was more modest.

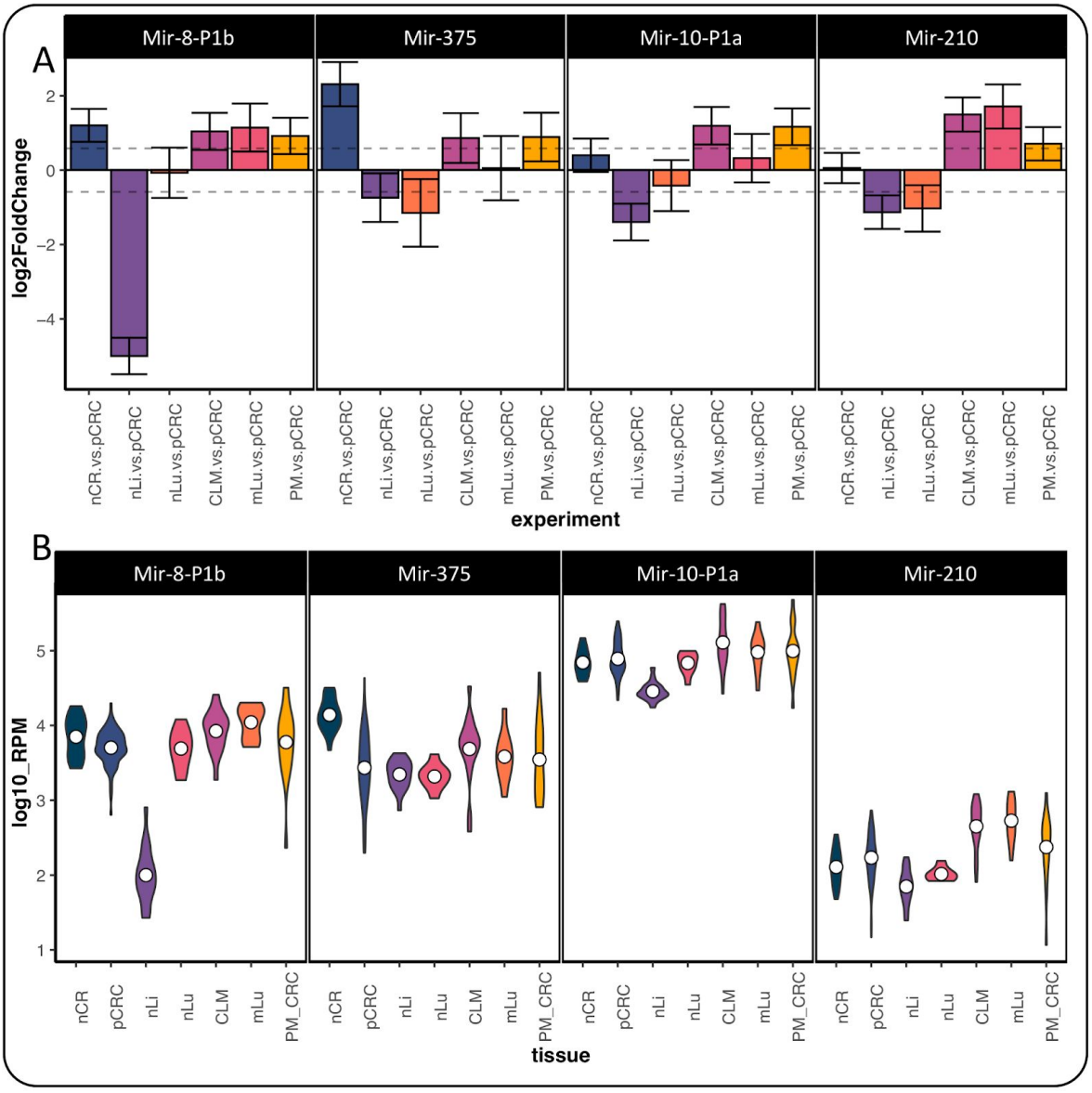


Figure 5: Evolutionarily deeply conserved microRNAs are commonly upregulated at multiple metastatic sites. **a** Bar plots of Log 2 Fold Changes (LFC) of miRNAs differentially

expressed across the normal and metastatic tissue types relative to pCRC. Error bars are 95% Confidence Intervals. **b** Violin plots of Log 10 Reads per Million (Log10 RPM) in the different tissues.

In addition, Mir-1307_5p, Mir-8-P1b_3p (*miR-141*) and Mir-584_5p were up-regulated in all metastatic sites compared to pCRC. Mir-8-P1b_3p was down-regulated in the pCRC (5730 RPM) compared to nCR (8362 RPM), and yet again far lower in the nLi tissue (147 RPM), although it was not differentially expressed between pCRC and nLu. The up-regulation of Mir-8-P1b_3p in CLM tissue (9956 RPM) and mLu (12309 RPM) can therefore not be explained as a normal adjacent tissue effect. Of note, Mir-1307_5p was down-regulated in pCRC (362 RPM) compared to nCR (610 RPM), and again the observed increase in the metastatic sites (CLM: 681 RPM, mLu: 820 RPM, PM-CRC: 737 RPM) cannot be explained by a normal tissue effect (nLi: 423 RPM and nLu: 393 RPM).

Mir-375_3p and Mir-10-P1a_5p were up-regulated in CLM (Mir-375_3p: 7222 RPM, Mir-10-P1a_5p: 164107 RPM) and PM-CRC (Mir-375_3p: 7192 RPM, Mir-10-P1a_5p: 130276 RPM), compared to pCRC (Mir-375_3p: 4500 RPM, Mir-10-P1a_5p: 89188 RPM). Both miRNAs showed lower expression in nLi (Mir-375_3p: 2407 RPM, Mir-10-P1a: 29711 RPM), which means that the detected levels cannot be explained as a normal tissue effect.

CLM vs pCRC specific metastatic miRNA signature

In CLM compared to pCRC, in addition to the six miRNAs differentially expressed in multiple metastatic sites, five miRNAs were up-regulated: Mir-425_5p, Mir-338-P1_3p, Mir-592_5p, Mir-1247_5p and Mir-339_3p, and three miRNAs were down-regulated: Mir-451_5p, Mir-154-P12_3p and Mir-30-P1b_5p. Twenty-five additional miRNAs were differentially expressed between CLM and pCRC, but these were also differentially expressed in the corresponding normal adjacent tissues, and may therefore represent a confounding effect.

mLu vs pCRC specific metastatic miRNA signature

In mLu compared to pCRC, in addition to the miRNAs differentially expressed in multiple metastatic sites, five miRNAs were up-regulated: Mir-374-P1_5p,

Mir-19-P1_3p, Mir-19-P2b_3p, Mir-19-P2a_3p and Mir-155_5p, and four were down-regulated: Mir-423_5p, Let-7-P1b_5p, Let-7-P2c1_5p and Mir-221-P2_3p. Again, 35 additional miRNAs were differentially expressed between mLu and pCRC, but because of similar differential expression in the same direction in the normal adjacent tissues, these were omitted from the signature.

PM-CRC vs pCRC specific metastatic miRNA signature

As no normal peritoneal tissue was available, the union of nCR vs nLi and nCR vs nLu, and the union of pCRC vs nLi and pCRC vs nLu, was used to control for a normal adjacent tissue effect in the analysis of PM-CRC compared to pCRC. In addition to the differentially expressed miRNAs in multiple metastatic sites, five miRNAs were up-regulated in PM-CRC, including Mir-154-P36_3p, Mir-191_5p, Mir-127_3p, Mir-154-P9_3p and Mir-10-P1b_5p, and three miRNAs were down-regulated: Mir-29-P1b_3p, Mir-30-P2a_5p and Mir-30-P2b_5p. Again, 31 differentially expressed miRNAs when comparing PM-CRC and pCRC were excluded based on expression in the normal adjacent tissues.

qPCR validation

qPCR analysis of randomly selected CLM (n=12) and pCRC (n=12) samples not included in the previous NGS analysis replicated the findings from the NGS data, validating up-regulation of Mir-210_3p in CLM compared to pCRC, exhibiting a 2.4-fold increase (Welch t-test p-value = 0.027, alternative hypothesis: greater in CLM). The result is close to the NGS data, where Mir-210_3p was 2.6-fold higher in CLM compared to pCRC. Data analysis is shown in (Additional file 8: Supplementary Methods).

Discussion

We here present a rigorous pipeline for the analysis of bulk tissue smallRNA-seq data. The pipeline was developed to address prevailing challenges in the miRNA field to obtain robust insights into the evolution of CRC and in particular to understand the role of miRNAs in the evolution of mCRC. Our analysis of clinical smallRNA-seq datasets produced in this study, as well as publicly available clinical smallRNA-seq data sets represents the most comprehensive analysis of miRNAs in CRC to date. Combining meticulous quality control with open science practices allows for increased power to detect meaningful miRNAs, along with high confidence that the results will be reproducible. In addition, this allows for the inclusion of additional datasets to the analysis as they become available in the future. Importantly, our analytical pipeline is not restricted to CRC and could easily be applied to other cancer types.

pCRC signature miRNAs

More than 40 deregulated miRNAs were identified in the signature for pCRC. This strong signal reflects major and genome-wide changes of malignant transformation on the transcriptional profile of these tissues (Figure 3). Among the up-regulated miRNAs, a subset represents well known "oncoMirs" such as Mir-21, and members of the three Mir-17-92 clusters (8 genes), Mir-31 and Mir-221. These miRNAs have validated and well-studied targets in CRC such as PTEN, TGFBR, SMAD [18,39,40]. The majority of the remaining previously reported genes in pCRC are the MIR-15 family member, Mir-29-P2a/b (*miR-29b*), two MIR-96 family members, Mir-135-P3, Mir-130-P2a (*miR-301a*) [41,42], Mir-181-P1c [43], and Mir-224 [18,39]. Mir-95-P2 (*miR-421*) was also significantly up-regulated when comparing nCR and pCRC in our data. These miRNAs have not been reported for CRC previously, but their deregulation has been demonstrated in other cancers [44–46]. The remaining down-regulated miRNAs were all previously reported to be associated with CRC; for instance the tumor suppressor MIR-10 family (4 genes), MIR-15, MIR-192 and MIR-194 (2 genes each) [18,39] have been implicated in CRC progression. The fact

that the pCRC vs nCR miRNA signature consists of miRNAs previously reported in the cancer literature provides confidence in the subsequent analysis of metastatic tumor samples.

mCRC signature miRNAs

Multiple miRNAs were found to be differentially expressed between pCRC and mCRC, and of these, Mir-8-P1b (miR-141), Mir-375, Mir-10-P1a, Mir-210, and were up-regulated in multiple metastatic sites, representing key starting points for future studies in metastasis biology. A number of miRNAs were differentially expressed at single metastatic sites compared to pCRC and may represent evolutionary adaptations to survival at the specific site. Little is known about these miRNAs in mCRC and the role in organ specific metastatic adaptation, and they may represent new players in metastasis and novel starting points in future research.

Several of the identified miRNAs have previously been implicated in parts of the metastatic process, mechanistically linking them to important evolutionary adaptations of metastasizing tumor cells such as the epithelial-mesenchymal transition (EMT) and, its reversal, the mesenchymal-epithelial transition (MET) [47].

MIR-8 and MIR-375

Mir-8-P1b_3p (*miR-141*), which is a member of the MIR-8 family[1], was down-regulated in pCRC compared to nCR, yet up-regulated at all metastatic sites compared to pCRC (Figure 5). MIR-8 family members have been reported to suppress metastasis by targeting transcription factors ZEB1 and ZEB2, suppressing E-cadherin expression [48,49]. The finding therefore makes sense in terms of contributing to MET in the established metastasis at the distant site and is in line with previous reports in CRC [50,51]. Increased expression of Mir-8-P1b_3p suppresses ZEB1 and ZEB2 expression, and thus reactivates E-cadherin expression and the epithelial phenotype. Further experimental evidence from an *in vivo* study of mouse breast cancer cell lines suggested that MIR-8 family miRNA expression was necessary for

[1] https://mirgenedb.org/browse/hsa?sort=pos&seed=&family=MIR-8

liver and lung colony formation [52], supporting the significance of the MIR-8 family in the evolution of metastases not only for mCRC, but metastases in general.

Two previous studies have also suggested circulating Mir-8-P1b_3p (*miR-141*) to be a predictive biomarker of metastasis in CRC. In a study of 182 pCRC cases (stage I-IV) and 76 healthy controls, plasma Mir-8-P1b_3p (*miR_141*) levels were associated with stage IV CRC and poor survival outcomes [53], while no differences in expression levels were detected in the corresponding pCRC samples. In a more recent study, Meltzer et al 2019 studied plasma exosomal miRNA levels in 83 locally advanced rectal cancer cases collected at the time of diagnosis [54]. Exosomal Mir-8-P1b_3p (*miR-141*) was elevated in patients with synchronous CLM compared to patients with no metastases, and Mir-8-P1b_3p (*miR-141*) was associated with higher risk of CLM development. The study also reported that high Mir-375_3p levels were associated with synchronous CLM. This is interesting in light of the tissue expression levels detected in our study, as Mir-375_3p was expressed at higher levels in both CLM and PM-CRC compared to pCRC. Mir-375_3p is the only member of the MIR-375 family[2] .

Both miRNAs have previously been reported to be present in circulation in other cancer types, including breast cancer [55], suggesting a possible association with the metastatic process. It is worthwhile mentioning that both the MIR-8 and the MIR-375 family are evolutionarily deeply conserved miRNA families, ranging back more than 650 million years to the last common ancestor of all bilaterian animals. Consequently, family members are also found in most other animals [26], and could be studied, for instance, in emerging cancer model systems such as the fruit fly *Drosophila* [56].

MIR-10

Another miRNA family that has been associated with EMT is the MIR-10 family [57] and in particular its Mir-10-P1 paralogues[3]. Levels of Mir-10-P1b_5p were elevated in breast cancer cell lines with metastatic capability, compared to both human mammary epithelial cells, or other human breast cancer cell lines [58]. Furthermore,

[2] https://mirgenedb.org/browse/ALL?seed=&family=MIR-375
[3] https://mirgenedb.org/browse/hsa?family=MIR-10

in vivo silencing of Mir-10-P1b_5p using antagomirs suppressed lung metastasis, but did not slow primary tumor growth [57]. Mir-10-P1b_5p expression is induced by *Twist*, a hypoxia inducible transcription factor involved in EMT, and it targets Hoxd10, which is reported to be a suppressor of metastasis [57]. These findings were further expanded to CRC showing that elevated Mir-10-P1 levels increase invasion and migration in CRC tumors [59]. Although Mir-10-P1b_5p was not up-regulated in mCRC in this study, its paralogue Mir-10-P1a_5p, was strongly up-regulated. With an expression of around 90 000 RPM in pCRC and 165 000 RPM in CLM, its up-regulation in CLM was strong, both in relative and absolute terms. It was also up-regulated in PM-CRC, but more modestly at 130 000 RPM. Although encoded on two separate loci, Mir-10-P1a (chromosome 17) and Mir-10-P1b (chromosome 2), the seed sequences are identical (ACCCUGU), and thus have the same mRNA target range and similar functional roles [60]. The high total expression levels suggest a functional role, possibly representing adaptations to survive at the distant site. Given that the MIR-10 family is the deepest conserved miRNA family known, found even in early multicellular animals such as jellyfish [61], and with high conservation in sequence and position in the HOX-cluster between the human and the cancer model the zebrafish *Danio rerio* [62] and the emerging model nematode *Caenorhabditis elegans* [63], a range of interesting follow-up experiments are possible.

MIR-210

Hypoxia is an important feature of metastatic evolution, influencing processes such as metabolism, angiogenesis, cell proliferation and differentiation [64]. In this context, our finding of up-regulation of the "hypoxamir" Mir-210_3p at all mCRC sites is of particular interest and we further validated the up-regulation of this miRNA by qPCR. The Mir-210_3p promoter has a binding site for HIF-1 α and HIF-2 α, transcription factors involved in maintaining cellular oxygen homeostasis, and Mir-210_3p is therefore up-regulated under hypoxic conditions [65]. However, the impact of Mir-210_3p on the metastatic process is controversial, with studies providing contradictory statements about its role in different cancer types, some reporting functions as an oncoMir and inducer of metastatic progression, while others supports

a role as a tumor suppressor [66]. In pCRC Mir-210_3p was shown to be up-regulated in a study of 193 cases comparing paired pCRC and nCR tissue samples, and high Mir-210_3p levels were associated with inferior prognosis [67]. Higher Mir-210 levels were also reported to enhance invasion and metastasis in cell lines (Mudduluru et al. 2015). We did not observe up-regulation of Mir-210_3p in pCRC compared to nCR; rather, our results show up-regulation at the metastatic sites, which is in accordance with several non-NGS studies in CRC [68–70]. Our finding of increased Mir-210_3p expression at all mCRC sites, suggest that hypoxic stress is a common occurrence at the metastatic sites, and that up-regulation of Mir-210_3p may be a consequence of this hypoxic microenvironment. Curiously, MIR-210 also belongs to a group of ancient miRNA, present in most animals living today, and therefore could be tested in a range of previously mentioned models such as the fruit fly or zebrafish.

the normal tissue effect and cell-type specific miRNAs

When considering miRNA expression at the global level, our results confirmed the importance of analyzing mCRC sites as separate entities. The UMAP analysis (Figure 1A) showed that the metastatic sites, in terms of global miRNA expression, were clearly distinct. In particular, the mLu datasets formed more distinct clusters from the pCRC, compared to the other mCRC samples. This may reflect the presence of host tissue cells in the bulk metastasis tissue, but could also represent adaptive expressional responses in tumor cells. Both this, and the unsupervised clustering analysis (Figure 1B), showed that at the global miRNA expression level of the adjacent tissues, nCR, nLi and nLu were clearly distinct, stressing the importance of controlling for adjacent tissues in the differential expression analysis to moderate the initial apparent large differences between nCR and pCRC, or pCRC and mCRC.

Important in this perspective, exclusion of cell-type specific miRNAs (Figure 2A), from the analysis did not appear to influence the clustering analyses. In particular, mLu samples still clustered close to nLu, and not to the other malignant tissues, pointing to the presence of additional global miRNA expression differences in these organs and tissues. Nevertheless, when comparing tissues, it is important to account

for miRNAs that are exclusively expressed in specific cell types. For instance, Mir-143/145 was previously reported to be down-regulated in pCRC compared to nCR, and was therefore suggested as a biomarker of CRC [71]. However, it has since been conclusively shown that Mir-143/145 are expressed in mesenchymal cells only, and not in epithelial cells [72,73]. Therefore, the observed "down-regulation" of Mir-143/145 must be caused by differences in cellular composition of the tissues, and not in expression changes in epithelial cancer cells. It is therefore necessary to account for the cellular composition of tissues when analyzing bulk tissue samples. Similarly, the down-regulation of platelet and red blood cell specific Mir-486 was previously associated with tumor stage in pCRC [74], but is probably not directly related to tumor progression. Another example is the hepatocyte specific Mir-122 [75], where a very strong up-regulation in CLM likely reflects presence of hepatocytes in the bulk CLM tissue. Not surprisingly, therefore, these cell-type specific miRNAs appear to have lower expression in malignant tissues compared to the normal tissues (Figure 2B). This is an important consideration for future research as the biomarker potential of a miRNA abnormally expressed in diseased tissue must be tightly controlled.

Conclusion

In conclusion, using a consistent pipeline for analysis of bulk tissue smallRNA-seq data across datasets from multiple studies, a set of miRNAs implicated in mCRC evolution was revealed. Interestingly, three of the miRNAs up-regulated in mCRC are among the evolutionary most deeply conserved miRNAs which could be interpreted as a reactivation of early gene regulatory programs and as a support of the previously suggested devolution hypothesis of cancer [76–78].

A limitation of the study is that the datasets originated from heterogeneous studies, trials and time points and included few paired pCRC and mCRC samples. This may have resulted in under reporting of some relevant miRNAs, but the consensus signatures still represent a solid foundation for future research. By setting up a publicly available bioinformatics pipeline for processing and analyzing smallRNA-seq data, we have ensured that it will be straightforward to add new datasets when they become available.

Accounting for known cell-type specific miRNAs is a novel effort in the field, which has contributed to remove false signals and identify a robust signature. However, the effort was limited to available knowledge from cell type profiling *in vitro* and in bulk [28–30,37], and does not account for all cell-type specific expression differences, particularly where the differences in miRNA expression levels are more subtle or unknown. Cell deconvolution of bulk tissue sequencing data based on gene expression is in itself a large and rapidly developing field [79–81]. As of yet, development of rigorous miRNA expression profiles on the cellular level is still in its infancy [82]. The use of bulk tissue data does not allow for direct analysis of tumor heterogeneity, and tumors are not homogenous entities, but made up of distinct sub clonotypes and cell types [83]. Therefore, single cell approaches, which have led to a number of encouraging findings in CRC [84,85], hold promise for future progress in the miRNA field [86,87].

Materials and Methods

Patient samples

Tumor tissue and corresponding adjacent colorectal tissue samples (nCR) of patients with pCRC were obtained from the tumor biobank of the LARC-EX trial (ClinicalTrials.gov Identifier: NCT02113384). CLM and corresponding adjacent tissue samples were retrieved from the tumor biobank of the OSLO-COMET trial (ClinicalTrials.gov Identifier: NCT01516710) and prepared and processed as described in [88]. mLu and corresponding adjacent tissue samples and PM-CRC samples were retrieved from our lung metastasis biobank (S-06402b) and peritoneal metastasis biobank (2010/2390), respectively. The studies were approved by the Regional Ethics Committee of Southh-East Norway, and patients were included in the respective studies following written informed consent. Patient samples were collected at the time of surgery and were snap frozen in liquid nitrogen at the time of collection and stored at -80°C. Cryo sections were haematoxylin and eosin stained and a pathologist confirmed the presence of >60% tumor content in the tumor samples.

RNA extraction and next generation sequencing

RNA was extracted using Qiagen Allprep DNA/RNA/miRNA universal kit, which simultaneously isolates genomic DNA and total RNA. RNA concentration was evaluated using ThermoFisher NanoDrop spectrophotometer and RNA integrity was evaluated by Agilent Technologies Bioanalyzer RNA 6000 Nano kit. Samples with sufficient quality were then used to prepare small RNA NGS libraries (RIN > 6), using TruSeq Small RNA Library preparation protocol. Successfully prepared libraries were sequenced using Illumina HiSeq 2500 High Throughput Sequencer using single end sequencing (50bp).

Collection of available NGS datasets

A literature search was conducted for the terms “microRNA + CRC + next generation sequencing” in different variations, and reviews were studied [18,39]. Publicly

available data were downloaded from European Genome-phenome Archive (EGA), the Sequence Read Archive (SRA) and the Gene Expression Omnibus (GEO). For download, the latest version of sRNAbench was used [89]. Corresponding SRA data files were automatically downloaded and converted into FASTQ files.

Data processing

All datasets were processed using the same pipeline. miRTrace [27] was used for preprocessing and quality control of raw data (FASTQ files). In brief, low quality reads, defined as reads where less than 50% of nucleotides had a Phread quality score greater than 20 were discarded. 3p adapter sequences were trimmed, and reads made up of repetitive elements and reads that are shorter than 18nt were removed. The output of miRTrace quality control statistics included read length distribution, percentage sequences where adapter was detected, reads < 18 nt, low complexity or low Phred score, percentage of reads miRTrace determined to be miRNA, rRNA, tRNA, artifacts or unknown, number of unique miRNA genes detected in each sample, and degree of contamination of miRNAs derived from other species than human. Samples where either less than 25% of reads were between 20 and 25 nt, or where more than 75% of reads were discarded, or where less than 10% of reads were determined to be miRNA, were excluded. Studies where more than half of datasets failed the QC criteria, or where significant contamination was detected, were excluded from the study.

Read alignment

Reads were mapped to MirGeneDB2.0 (mirgenedb.org) precursor sequences consisting of the pre-miRNA sequence as well as 30 nt upstream and downstream of the Drosha cut site [26] using bowtie1.2 [90], requiring an 18 nucleotide seed sequence of zero mismatches to avoid cross-mapping, and up to five mismatches outside the seed. Mapped reads were counted using the featureCounts 'summarizeOverlaps()' function from the 'GenomicAlignments' Bioconductor package [91,92].

Data exploration

For data exploration, read counts were normalized using the varianceStabilizingTransformation() (VST) function from the DESeq2 package [93]. Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP), and hierarchical clustering, was used to visualize the similarity of datasets on the global miRNA expression level. UMAP is a dimensionality-reducing algorithm used to reduce the numbers of dimensions in each dataset, here 537 miRNA genes, onto a more easily interpretable two dimensions, while retaining the relationship between datasets [94]. The umap R package was used, and datasets were annotated by tissue and by study (Figure 1A, left and right panels, respectively). Hierarchical clustering was performed with the hclust() R function using the complete linkage clustering method on a distance matrix of all datasets computed with the dist() R function using maximum distance measure. Datasets were annotated according to tissue origin.

Analysis of influence of cell specific miRNA

Principal Component Analysis (PCA) was used to analyze the influence of the 45 cell-type specific miRNAs upon global miRNA expression. PCA plots were made with the FactoMineR PCA() function, on the VST transformed datasets, subset to include the 45 miRNAs reported by McCall et al [28] to be cell-type specific. Two plots were made, one visualizing the relationship between datasets in PCA dimension 1 and 2 with the cell specific miRNAs (Figure 2A left). Datasets were annotated by tissue. Plots of each individual dataset are transparent, while the mean values for each tissue were plotted in full color. 95% confidence intervals for the mean are also shown. To analyze the influence of individual cell specific miRNAs, a correlation circle was plotted to illustrate the contribution of each of the top 15 miRNAs to the variance between datasets in each dimension (Figure 2A, right panel). In addition, a heatmap of the 45 cell specific miRNAs, scaled by z-score of RPM values for each individual miNRA, was made to visualize their relative expression in each tissue. Furthermore, another UMAP analysis was done, discarding the 45 cell specific miRNAs from the

analysis, in order to assess clustering when these miRNAs are not present in the datasets.

Differential expression analysis

The differential expression analysis can be viewed in the supplementary R-markdown file, (Additional file 7: Supplementary Methods)

Differential expression analysis was performed using DESeq2 [93]. Multiple testing was corrected using Benjamin-Hochberg, with a false discovery rate (FDR), threshold of 0.05. The Log 2 Fold Change (LFC) shrinkage function in DESeq2, lfcShrink(), was enabled. This shrinks fold changes for miRNAs with higher variance. In addition, miRNAs were filtered if they had a LFC > 0.58 or LFC < -0.58, and where at least one tissue had mean expression >100 Reads Per Million (RPM) a level previously suggested as cutoff for physiological activity [20], and MiRNAs known to be cell-type specific [28] were labelled in each signature. Also, miRNAs that have been reported to be present at high levels in one or more blood cell types were annotated as such, [37]. When comparing metastatic tissue samples, normal adjacent tissues were used to control for the confounding effect of surrounding normal tissue.

qPCR validation

To validate the NGS finding of increased expression of Mir-210_3p in CLM compared to pCRC, 12 new randomly chosen CLM and pCRC tissue samples were selected for qPCR validation. RNA was isolated with Qiagen Allprep DNA/RNA/miRNA universal kit (Qiagen Cat. No. 80224). Synthetic RNA Spike-Ins UniSp2, UniSp4 and UniSp5 (Qiagen Cat. No. 339390) were added pre-isolation. RNA integrity, RIN, was assessed with Agilent Technologies Bioanalyzer RNA 6000 Nano. RNA was diluted to 5ng/uL. RT-PCR was done with miRCURY LNA RT Kit (Qiagen Cat. No. 339340), adding UniSp6 and cel-miR-39-3p RNA Spike-Ins (Qiagen Cat. No 339390). qPCR was done using Qiagen miRCURY SYBR Green Kit (Qiagen Cat. No. 339345) and miRCURY LNA miRNA PCR Assays (Qiagen Cat. No. 339306). Two miRNAs were used for reference, Mir-191 and Mir-103 (Qiagen primer Cat. No. YP00204063 and Cat. No.

YP00204306), while the Mir-210_3p primer was Qiagen Cat. No. YP00204333. Two PCR replicates were run per primer assay. Mir-210_3p Cq values were normalized to the mean of the reference miRNAs, to obtain the dCq value. Welch t-test was used to test if the expression levels were different between CLM and pCRC, alternative being greater in CLM. LFC between CLM and pCRC was obtained with the 2^(-ddCq) method. Calculations are shown in (Additional file 8: Supplementary Methods).

Declarations

Ethics approval and consent to participate

All patients whose material has been used in this study have provided consent to participate in the study. The Norwegian Regional Committees for Medical and Health Research Ethics (REK) has approved all research on patient material developed in this study (CLM and nLi: Ref. nr.: 2011/1285, mLu and nLu: REK S-06402b, PM-CRC: REK:2010/2390).

Consent for publication

Not Applicable

Availability of data and materials

The datasets generated and/or analysed during the current study will be made available in the European Genome-phenome Archive repository. Bioinformatics pipeline can be found at (https://github.com/eirikhoye/mirna_pipeline)

Funding

BF was supported by the South-Eastern Norway Regional Health Authority (grant number 2014041 to KF). CLA was supported by the Norwegian Cancer Society (grant number 4499184 to KF). Part of this work was supported by the Research Council of Norway (MetAction grant number 218325). EH was supported by the South-Eastern Norway Regional Health Authority (grant number 2018014 to KF).

Authors' contributions

KF conceived the study with input from BF. BF, EIRH, PHMB, DD and EH performed all computational analyses. BF, EIRH, ATH, CLA, TWA and SL performed all experiments. AAF, VJD, BE and KF collected all specimens. BF, EIRH, PHMB and EH contributed to the design of the study. All authors contributed to writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Michael Hackenberg for help with data acquisition and Marc Halushka for discussions.

Competing Interests

The authors declare that they have no competing interests.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136:E359–86.

2. Nguyen DX, Bos PD, Massagué J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009;9:274–84.

3. Riihimäki M, Hemminki A, Sundquist J, Hemminki K. Patterns of metastasis in colon and rectal cancer. Sci Rep. 2016;6:29765.

4. Angelova M, Mlecnik B, Vasaturo A, Bindea G, Fredriksen T, Lafontaine L, et al. Evolution of Metastases in Space and Time under Immune Selection. Cell. 2018;175:751–65.e16.

5. Massague J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529:298–306.

6. Flatmark K, Høye E, Fromm B. microRNAs as cancer biomarkers. Scand J Clin Lab Invest [Internet]. 2016;76. Available from: http://dx.doi.org/10.1080/00365513.2016.1210330

7. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017;66:683–91.

8. Bartel DP. Metazoan MicroRNAs. Cell. 2018;173:20–51.

9. Fromm B, Keller A, Yang X, Friedlander MR, Peterson KJ, Griffiths-Jones S. Quo vadis microRNAs? Trends Genet [Internet]. 2020; Available from:

http://www.sciencedirect.com/science/article/pii/S0168952520300718

10. Dalmay T, Edwards DR. MicroRNAs and the hallmarks of cancer. Oncogene. 2006;25:6170–5.

11. Jung M, Schaefer A, Steiner I, Kempkensteffen C, Stephan C, Erbersdobler A, et al. Robust microRNA stability in degraded RNA preparations from human tissue and cell samples. Clin Chem. 2010;56:998–1006.

12. Volinia S, Calin GA, Liu C-G, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A. 2006;103:2257–61.

13. Baffa R, Fassan M, Volinia S, O'Hara B, Liu C-G, Palazzo JP, et al. MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. J Pathol. 2009;219:214–21.

14. Neerincx M, Sie DL, van de Wiel MA, van Grieken NC, Burggraaf JD, Dekker H, et al. MiR expression profiles of paired primary colorectal cancer and metastases by next-generation sequencing. Oncogenesis. 2015;4:e170.

15. Rohr C, Kerick M, Fischer A, Kuhn A, Kashofer K, Timmermann B, et al. High-throughput miRNA and mRNA sequencing of paired colorectal normal, tumor and metastasis tissues and bioinformatic modeling of miRNA-1 therapeutic applications. PLoS One. 2013;8:e67461.

16. Goossens-Beumer IJ, Derr RS, Buermans HP, Goeman JJ, Bohringer S, Morreau H, et al. MicroRNA classifier and nomogram for metastasis prediction in colon cancer. Cancer Epidemiol Biomarkers Prev. 2015;24:187–97.

17. Schee K, Lorenz S, Worren MM, Gunther CC, Holden M, Hovig E, et al. Deep Sequencing the MicroRNA Transcriptome in Colorectal Cancer. PLoS One. 2013;8:e66165.

18. Cekaite L, Eide PW, Lind GE, Skotheim RI, Lothe RA. MicroRNAs as growth regulators, their function and biomarker status in colorectal cancer. Oncotarget. 2016;7:6476–505.

19. To KK, Tong CW, Wu M, Cho WC. MicroRNAs in the prognosis and therapy of colorectal cancer: From bench to bedside. World J Gastroenterol. 2018;24:2949–73.

20. Mullokandov G, Baccarini A, Ruzo A, Jayaprakash AD, Tung N, Israelow B, et al. High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. Nat Methods. 2012;9:840–6.

21. Witwer KW, Halushka MK. Toward the promise of microRNAs – Enhancing reproducibility and rigor in microRNA research. RNA Biol. Taylor & Francis; 2016;13:1103–16.

22. Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, et al. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu Rev Genet. 2015;49:213–42.

23. Xie M, Li M, Vilborg A, Lee N, Shu M-D, Yartseva V, et al. Mammalian 5'-capped microRNA precursors that generate a single microRNA. Cell. 2013;155:1568–80.

24. Xu Y-F, Hannafon BN, Khatri U, Gin A, Ding W-Q. The origin of exosomal miR-1246 in human cancer cells. RNA Biol. 2019;16:770–84.

25. Yoshikawa M, Fujii YR. Human Ribosomal RNA-Derived Resident MicroRNAs as the Transmitter of Information upon the Cytoplasmic Cancer Stress. Biomed Res Int. 2016;2016:7562085.

26. Fromm B, Domanska D, Høye E, Ovchinnikov V, Kang W, Aparicio-Puerta E, et al. MirGeneDB 2.0: the metazoan microRNA complement. Nucleic Acids Res [Internet]. 2020; Available from: https://academic.oup.com/nar/advance-article-abstract/doi/10.1093/nar/gkz885/5584683

27. Kang W, Eldfjell Y, Fromm B, Estivill X, Biryukova I, Friedländer MR. miRTrace reveals the organismal origins of microRNA sequencing data. Genome Biol. 2018;19:213.

28. McCall MN, Kim M-S, Adil M, Patil AH, Lu Y, Mitchell CJ, et al. Toward the human cellular microRNAome. Genome Res. 2017;27:1769–81.

29. Rie D de, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, et al. An integrated expression atlas of miRNAs and their promoters in human and mouse. Nat Biotechnol [Internet]. 2017; Available from: http://www.nature.com.ezproxyhost.library.tmc.edu/nbt/journal/vaop/ncurrent/full/nbt.3947.html?foxtrotcallback=true

30. Juzenas S, Venkatesh G, Hübenthal M, Hoeppner MP, Du ZG, Paulsen M, et al. A comprehensive, cell specific microRNA catalogue of human peripheral blood [Internet]. Nucleic Acids Research. 2017. p. 9290–301. Available from: http://dx.doi.org/10.1093/nar/gkx706

31. Halushka MK, Fromm B, Peterson KJ, McCall MN. Big Strides in Cellular MicroRNA Expression. Trends Genet. 2018;34:165–7.

32. Selitsky SR, Dinh TA, Toth CL, Lisa Kurtz C, Honda M, Struck BR, et al. Transcriptomic Analysis of Chronic Hepatitis B and C and Liver Cancer Reveals MicroRNA-Mediated Control of Cholesterol Synthesis Programs [Internet]. mBio. 2015. Available from: http://dx.doi.org/10.1128/mbio.01500-15

33. Sun Y, Wang L, Guo S-C, Wu X-B, Xu X-H. High-throughput sequencing to identify miRNA biomarkers in colorectal cancer patients [Internet]. Oncology Letters. 2014. p. 711–3. Available from: http://dx.doi.org/10.3892/ol.2014.2215

34. Sun G, Cheng Y-W, Lai L, Huang T-C, Wang J, Wu X, et al. Signature miRNAs in colorectal cancers were revealed using a bias reduction small RNA deep sequencing protocol. Oncotarget. 2016;7:3857–72.

35. Liang G, Li J, Sun B, Li S, Lu L, Wang Y, et al. Deep sequencing reveals complex mechanisms of microRNA deregulation in colorectal cancer. Int J Oncol. 2014;45:603–10.

36. Hamfjord J, Stangeland AM, Hughes T, Skrede ML, Tveit KM, Ikdahl T, et al. Differential expression of miRNAs in colorectal cancer: comparison of paired tumor tissue and adjacent

normal mucosa using high-throughput sequencing. PLoS One. 2012;7:e34150.

37. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, et al. Blood Cell Origin of Circulating MicroRNAs: A Cautionary Note for Cancer Biomarker Studies [Internet]. Cancer Prevention Research. 2012. p. 492–7. Available from: http://dx.doi.org/10.1158/1940-6207.capr-11-0370

38. Kent OA, McCall MN, Cornish TC, Halushka MK. Lessons from miR-143/145: The importance of cell-type localization of miRNAs. Nucleic Acids Res. 2014;42:7528–38.

39. Strubberg AM, Madison BB. MicroRNAs in the etiology of colorectal cancer: pathways and clinical implications. Dis Model Mech. 2017;10:197–214.

40. Schee K, Boye K, Abrahamsen TW, Fodstad O, Flatmark K. Clinical relevance of microRNA miR-21, miR-31, miR-92a, miR-101, miR-106a and miR-145 in colorectal cancer. BMC Cancer. 2012;12:505.

41. Zhang W, Zhang T, Jin R, Zhao H, Hu J, Feng B, et al. MicroRNA-301a promotes migration and invasion by targeting TGFBR2 in human colorectal cancer. J Exp Clin Cancer Res. 2014;33:113.

42. Ma X, Yan F, Deng Q, Li F, Lu Z, Liu M, et al. Modulation of tumorigenesis by the pro-inflammatory microRNA miR-301a in mouse models of lung cancer and colorectal cancer. Cell Discov. 2015;1:15005.

43. Guo X, Zhu Y, Hong X, Zhang M, Qiu X, Wang Z, et al. miR-181d and c-myc-mediated inhibition of CRY2 and FBXL3 reprograms metabolism in colorectal cancer. Cell Death Dis. 2017;8:e2958.

44. Scaravilli M, Porkka KP, Brofeldt A, Annala M, Tammela TLJ, Jenster GW, et al. MiR-1247-5p is overexpressed in castration resistant prostate cancer and targets MYCBP2 [Internet]. The Prostate. 2015. p. 798–805. Available from: http://dx.doi.org/10.1002/pros.22961

45. Chen L, Tang Y, Wang J, Yan Z, Xu R. miR-421 induces cell proliferation and apoptosis resistance in human nasopharyngeal carcinoma via downregulation of FOXO4. Biochem Biophys Res Commun. 2013;435:745–50.

46. Zhao J, Tao Y, Zhou Y, Qin N, Chen C, Tian D, et al. MicroRNA-7: a promising new target in cancer therapy. Cancer Cell Int. 2015;15:103.

47. Gupta GP, Massague J. Cancer metastasis: building a framework. Cell. 2006;127:679–95.

48. Pencheva N, Tavazoie SF. Control of metastatic progression by microRNA regulatory networks. Nat Cell Biol. 2013;15:546–54.

49. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 2008;10:593–601.

50. Hur K, Toiyama Y, Takahashi M, Balaguer F, Nagasaka T, Koike J, et al. MicroRNA-200c modulates epithelial-to-mesenchymal transition (EMT) in human colorectal cancer

metastasis. Gut. 2013;62:1315–26.

51. O'Brien SJ, Carter JV, Burton JF, Oxford BG, Schmidt MN, Hallion JC, et al. The role of the miR-200 family in epithelial--mesenchymal transition in colorectal cancer: a systematic review. International journal of cancer. Wiley Online Library; 2018;142:2501–11.

52. Dykxhoorn DM, Wu Y, Xie H, Yu F, Lal A, Petrocca F, et al. miR-200 enhances mouse breast cancer cell colonization to form distant metastases. PLoS One. 2009;4:e7181.

53. Cheng H, Zhang L, Cogdell DE, Zheng H, Schetter AJ, Nykter M, et al. Circulating plasma MiR-141 is a novel biomarker for metastatic colon cancer and predicts poor prognosis. PLoS One. 2011;6:e17745.

54. Meltzer S, Bjørnetrø T, Lyckander LG, Flatmark K, Dueland S, Samiappan R, et al. Circulating Exosomal miR-141-3p and miR-375 in Metastatic Progression of Rectal Cancer | Elsevier Enhanced Reader [Internet]. 2019 [cited 2019 Aug 24]. Available from: https://www.sciencedirect.com/science/article/pii/S1936523319300919/pdfft?isDTMRedir=true&download=true

55. Madhavan D, Zucknick M, Wallwiener M, Cuk K, Modugno C, Scharpff M, et al. Circulating miRNAs as surrogate markers for circulating tumor cells and prognostic markers in metastatic breast cancer. Clin Cancer Res. 2012;18:5972–82.

56. Gonzalez C. Drosophila melanogaster: a model and a tool to investigate malignancy and identify new therapeutics. Nat Rev Cancer. 2013;13:172–83.

57. Ma L, Reinhardt F, Pan E, Soutschek J, Bhat B, Marcusson EG, et al. Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. Nat Biotechnol. 2010;28:341–7.

58. Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature. 2007;449:682–8.

59. Wang Y, Li Z, Zhao X, Zuo X, Peng Z. miR-10b promotes invasion by targeting HOXD10 in colorectal cancer. Oncol Lett. 2016;12:488–94.

60. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife [Internet]. 2015;4. Available from: http://dx.doi.org/10.7554/eLife.05005

61. Grimson A. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature. 2008;455:1193–7.

62. Feitsma H, Cuppen E. Zebrafish as a cancer model. Mol Cancer Res. 2008;6:685–94.

63. Kirienko NV, Mani K, Fay DS. Cancer models in Caenorhabditis elegans. Dev Dyn. 2010;239:1413–48.

64. Rankin EB, Giaccia AJ. Hypoxic control of metastasis. Science. 2016;352:175–80.

65. Chan YC, Banerjee J, Choi SY, Sen CK. miR-210: the master hypoxamir. Microcirculation. 2012;19:215–23.

66. Qin Q, Furong W, Baosheng L. Multiple functions of hypoxia-regulated miR-210 in cancer. J Exp Clin Cancer Res. 2014;33:50.

67. Qu A, Du L, Yang Y, Liu H, Li J, Wang L, et al. Hypoxia-inducible MiR-210 is an independent prognostic factor and contributes to metastasis in colorectal cancer. PLoS One. 2014;9:e90952.

68. Mudduluru G, Abba M, Batliner J, Patil N, Scharp M, Lunavat TR, et al. A Systematic Approach to Defining the microRNA Landscape in Metastasis. Cancer Res. 2015;75:3010–9.

69. Bigagli E, Luceri C, Guasti D, Cinci L. Exosomes secreted from human colon cancer cells influence the adhesion of neighboring metastatic cells: Role of microRNA-210. Cancer Biol Ther. 2016;17:1062–9.

70. Ellermeier C, Vang S, Cleveland K, Durand W, Resnick MB, Brodsky AS. Prognostic microRNA expression signature from examination of colorectal primary and metastatic tumors. Anticancer Res. 2014;34:3957–67.

71. Michael MZ, O' Connor SM, van Holst Pellekaan NG, Young GP, James RJ. Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res. 2003;1:882–91.

72. Boettger T, Beetz N, Kostin S, Schneider J, Krüger M, Hein L, et al. Acquisition of the contractile phenotype by murine arterial smooth muscle cells depends on the Mir143/145 gene cluster. J Clin Invest. 2009;119:2634–47.

73. Chivukula RR, Shi G, Acharya A, Mills EW, Zeitels LR, Anandam JL, et al. An essential mesenchymal function for miR-143/145 in intestinal epithelial regeneration. Cell. 2014;157:1104–16.

74. Lian H, Liu C, Li M, Hu Y, Shi N, Yu H, et al. miR-486-5p attenuates tumor growth and lymphangiogenesis by targeting neuropilin-2 in colorectal carcinoma [Internet]. OncoTargets and Therapy. 2016. p. 2865. Available from: http://dx.doi.org/10.2147/ott.s103460

75. Vychytilova-Faltejskova P, Pesta M, Radova L, Liska V, Daum O, Kala Z, et al. Genome-wide microRNA Expression Profiling in Primary Tumors and Matched Liver Metastasis of Patients with Colorectal Cancer. Cancer Genomics Proteomics. 2016;13:311–6.

76. Chen H, Lin F, Xing K, He X. The reverse evolution from multicellularity to unicellularity during carcinogenesis. Nat Commun. 2015;6:6367.

77. Davies PCW, Lineweaver CH. Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. Phys Biol. 2011;8:015001.

78. Gibson G. Decanalization and the origin of complex disease. Nat Rev Genet. 2009;10:134–40.

79. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7.

80. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. bioRxiv. 2018;223180.

81. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics. Narnia; 2019;35:i436–45.

82. Zhang Y, Cuerdo J, Halushka MK, McCall MN. The effect of tissue composition on gene co-expression. Brief Bioinform [Internet]. 2019; Available from: http://dx.doi.org/10.1093/bib/bbz135

83. Lawson DA, Kessenbrock K, Davis RT, Pervolarakis N, Werb Z. Tumour heterogeneity and metastasis at single-cell resolution. Nat Cell Biol. 2018;20:1349–60.

84. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet [Internet]. 2017; Available from: http://www.nature.com/ng/journal/vaop/ncurrent/full/ng.3818.html

85. Hu Z, Li Z, Ma Z, Curtis C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. Nat Genet [Internet]. 2020; Available from: http://dx.doi.org/10.1038/s41588-020-0628-z

86. Faridani OR, Abdullayev I, Hagemann-Jensen M, Schell JP, Lanner F, Sandberg R. Single-cell sequencing of the small-RNA transcriptome. Nat Biotechnol. 2016;34:1264–6.

87. Hagemann-Jensen M, Abdullayev I, Sandberg R, Faridani OR. Small-seq for single-cell small-RNA sequencing. Nat Protoc. 2018;13:2407–24.

88. Østrup O, Dagenborg VJ, Rødland EA, Skarpeteig V, Silwal-Pandit L, Grzyb K, et al. Molecular signatures reflecting microenvironmental metabolism and chemotherapy-induced immunogenic cell death in colorectal liver metastases [Internet]. Oncotarget. 2017. Available from: http://dx.doi.org/10.18632/oncotarget.19350

89. Rueda A, Barturen G, Lebrón R, Gómez-Martín C, Alganza Á, Oliver JL, et al. sRNAtoolbox: an integrated collection of small RNA research tools [Internet]. Nucleic Acids Research. 2015. p. W467–73. Available from: http://dx.doi.org/10.1093/nar/gkv555

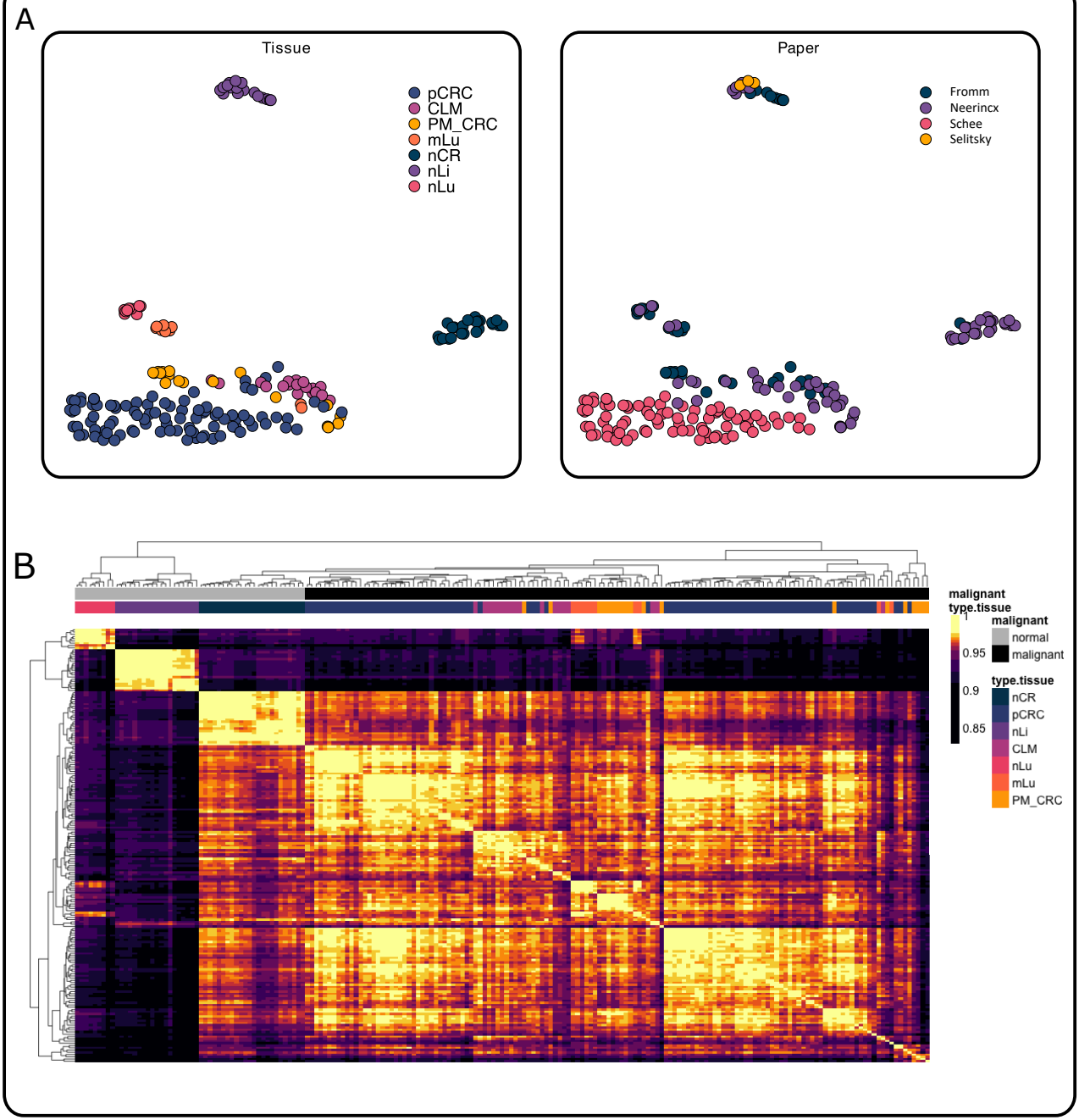
90. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

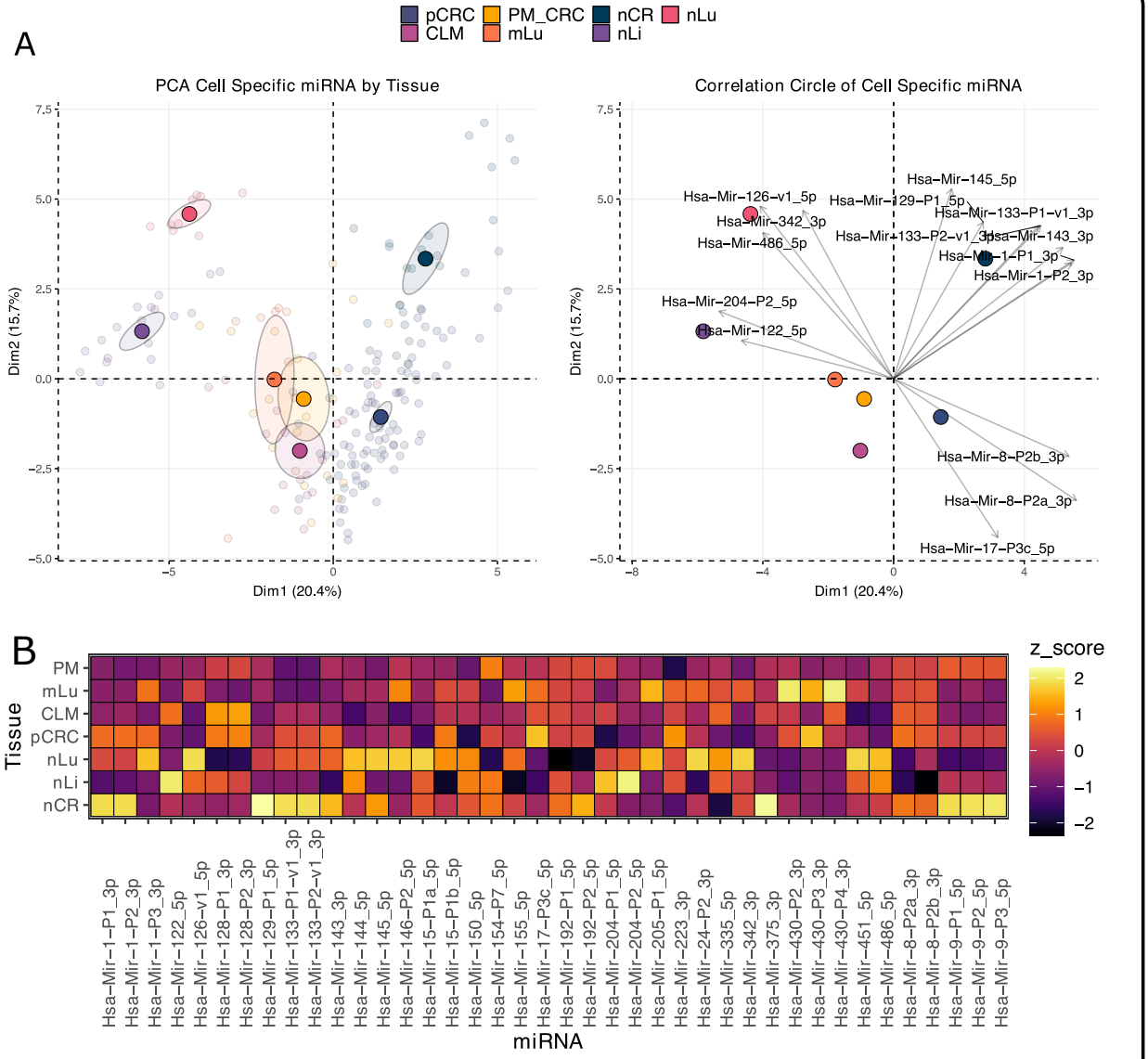
91. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

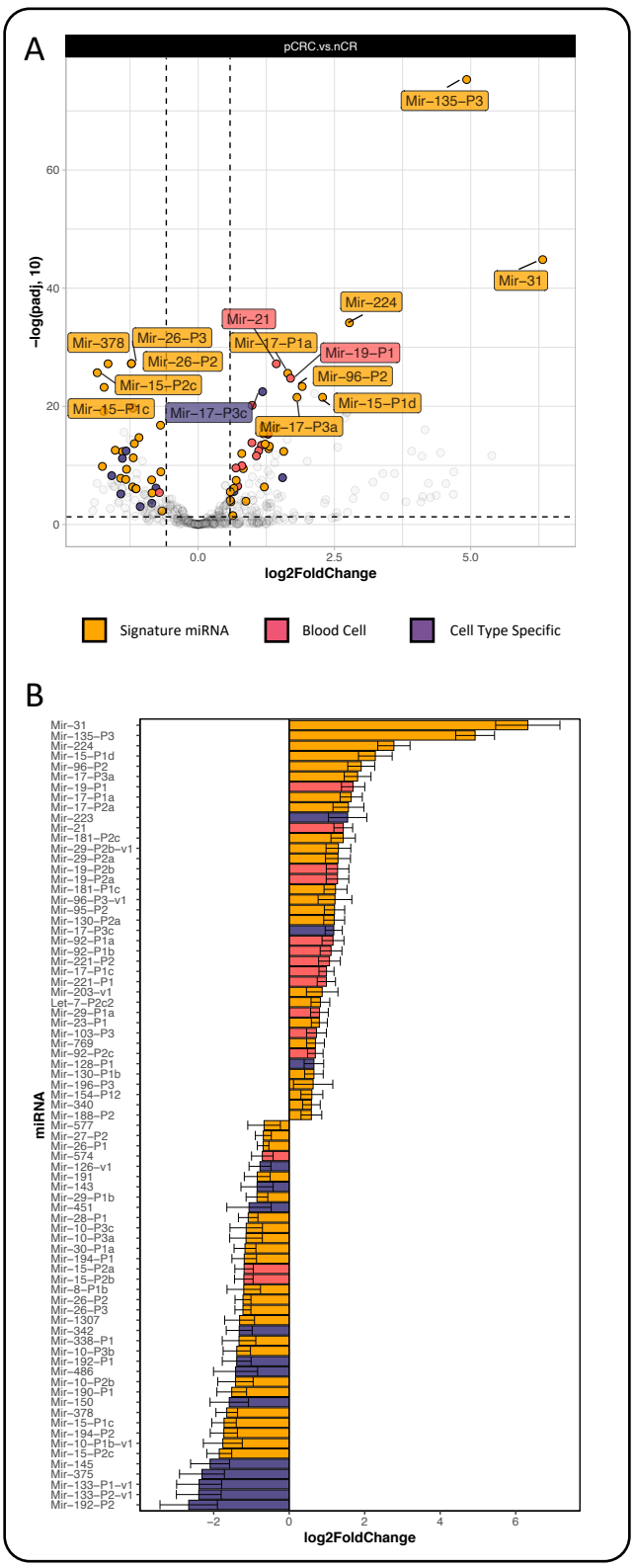
92. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges [Internet]. PLoS Computational Biology. 2013. p. e1003118. Available from: http://dx.doi.org/10.1371/journal.pcbi.1003118

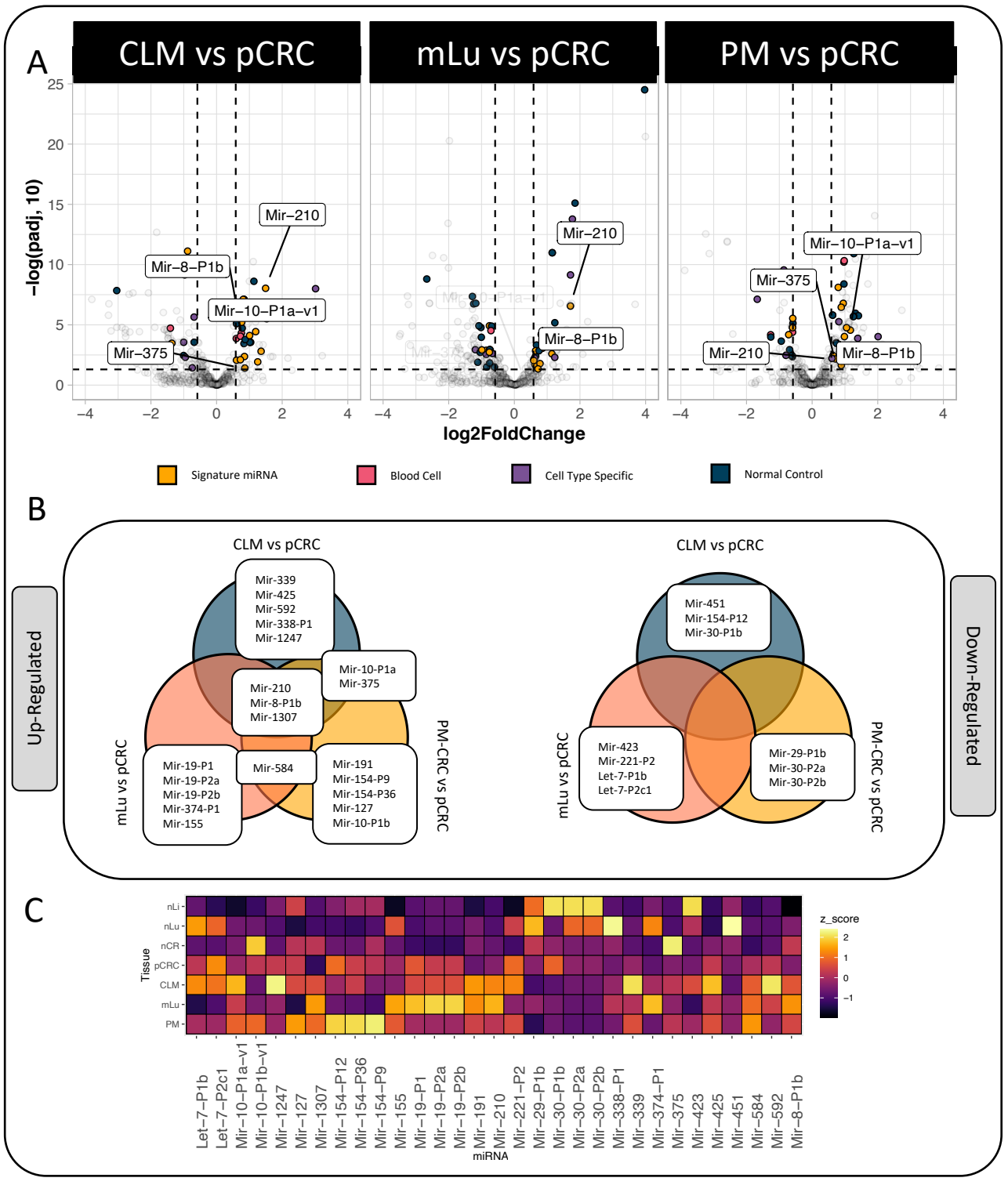
93. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 [Internet]. Genome Biology. 2014. Available from: http://dx.doi.org/10.1186/s13059-014-0550-8

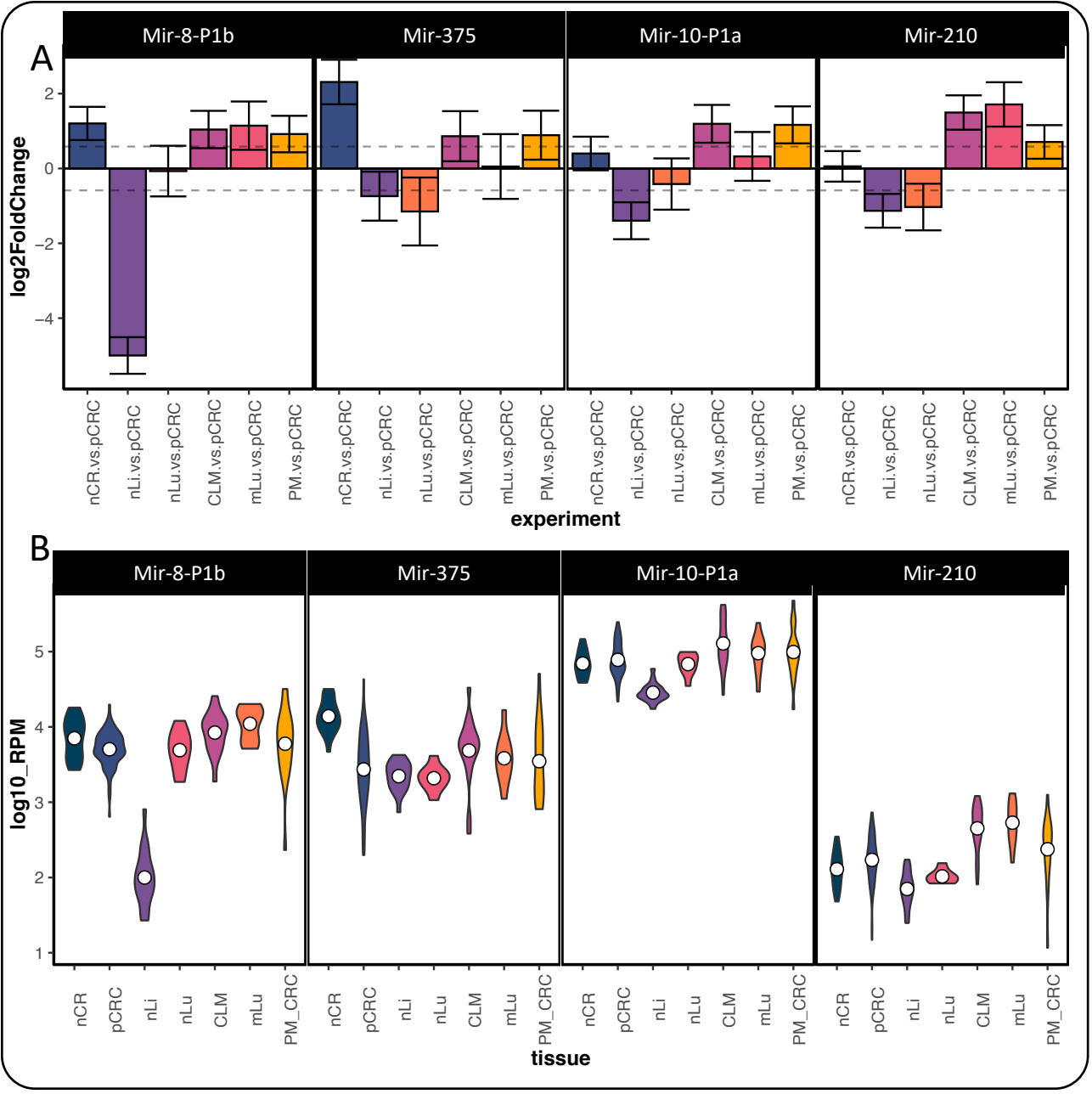
94. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Internet]. arXiv [stat.ML]. 2018. Available from: http://arxiv.org/abs/1802.03426


A
Tissue
pCRC
CLM
PM_CRC
mLu
nCR
nLi
nLu
Paper
Fromm
Neerincx
Schee
Selitsky
B
malignant
type.tissue
1
0.95
0.9
0.85
malignant
normal
malignant
type.tissue
nCR
pCRC
nLi
CLM
nLu
mLu
PM_CRC


A
pCRC
CLM
PM_CRC
mLu
nCR
nLi
nLu
PCA Cell Specific miRNA by Tissue
Dim2 (15.7%)
Dim1 (20.4%)
Correlation Circle of Cell Specific miRNA
Hsa−Mir−145_5p
Hsa−Mir−126−v1_5p
Hsa−Mir−129−P1_5p
Hsa−Mir−133−P1−v1_3p
Hsa−Mir−342_3p
Hsa−Mir−133−P2−v1_3p
Hsa−Mir−143_3p
Hsa−Mir−486_5p
Hsa−Mir−1−P1_3p
Hsa−Mir−1−P2_3p
Hsa−Mir−204−P2_5p
Hsa−Mir−122_5p
Hsa−Mir−8−P2b_3p
Hsa−Mir−8−P2a_3p
Hsa−Mir−17−P3c_5p
B
Tissue
PM
mLu
CLM
pCRC
nLu
nLi
nCR
z_score
2
1
0
−1
−2
miRNA
Hsa−Mir−1−P1_3p
Hsa−Mir−1−P2_3p
Hsa−Mir−1−P3_3p
Hsa−Mir−122_5p
Hsa−Mir−126−v1_5p
Hsa−Mir−128−P1_3p
Hsa−Mir−128−P2_3p
Hsa−Mir−129−P1_5p
Hsa−Mir−133−P1−v1_3p
Hsa−Mir−133−P2−v1_3p
Hsa−Mir−143_3p
Hsa−Mir−144_5p
Hsa−Mir−145_5p
Hsa−Mir−146−P2_5p
Hsa−Mir−15−P1a_5p
Hsa−Mir−15−P1b_5p
Hsa−Mir−150_5p
Hsa−Mir−154−P7_5p
Hsa−Mir−155_5p
Hsa−Mir−17−P3c_5p
Hsa−Mir−192−P1_5p
Hsa−Mir−192−P2_5p
Hsa−Mir−204−P1_5p
Hsa−Mir−204−P2_5p
Hsa−Mir−205−P1_5p
Hsa−Mir−223_3p
Hsa−Mir−24−P2_3p
Hsa−Mir−335_5p
Hsa−Mir−342_3p
Hsa−Mir−375_3p
Hsa−Mir−430−P2_3p
Hsa−Mir−430−P3_3p
Hsa−Mir−430−P4_3p
Hsa−Mir−451_5p
Hsa−Mir−486_5p
Hsa−Mir−8−P2a_3p
Hsa−Mir−8−P2b_3p
Hsa−Mir−9−P1_5p
Hsa−Mir−9−P2_5p
Hsa−Mir−9−P3_5p


A
pCRC.vs.nCR
−log(padj, 10)
log2FoldChange
Mir−135−P3
Mir−31
Mir−224
Mir−21
Mir−378
Mir−26−P3
Mir−17−P1a
Mir−19−P1
Mir−26−P2
Mir−96−P2
Mir−15−P2c
Mir−15−P1d
Mir−15−P1c
Mir−17−P3c
Mir−17−P3a
Signature miRNA
Blood Cell
Cell Type Specific
B
miRNA
log2FoldChange
Mir−31
Mir−135−P3
Mir−224
Mir−15−P1d
Mir−96−P2
Mir−17−P3a
Mir−17−P1a
Mir−17−P2a
Mir−223
Mir−21
Mir−181−P2c
Mir−29−P2b−v1
Mir−29−P2a
Mir−19−P2b
Mir−19−P2a
Mir−181−P1c
Mir−96−P3−v1
Mir−95−P2
Mir−130−P2a
Mir−17−P3c
Mir−92−P1a
Mir−92−P1b
Mir−221−P2
Mir−17−P1c
Mir−221−P1
Mir−203−v1
Let−7−P2c2
Mir−29−P1a
Mir−23−P1
Mir−103−P3
Mir−769
Mir−92−P2c
Mir−128−P1
Mir−130−P1b
Mir−196−P3
Mir−154−P12
Mir−340
Mir−188−P2
Mir−577
Mir−27−P2
Mir−26−P1
Mir−574
Mir−126−v1
Mir−191
Mir−143
Mir−29−P1b
Mir−451
Mir−28−P1
Mir−10−P3c
Mir−10−P3a
Mir−30−P1a
Mir−194−P1
Mir−15−P2a
Mir−15−P2b
Mir−8−P1b
Mir−26−P2
Mir−26−P3
Mir−1307
Mir−342
Mir−338−P1
Mir−10−P3b
Mir−192−P1
Mir−486
Mir−10−P2b
Mir−190−P1
Mir−150
Mir−378
Mir−15−P1c
Mir−194−P2
Mir−10−P1b−v1
Mir−15−P2c
Mir−145
Mir−375
Mir−133−P1−v1
Mir−133−P2−v1
Mir−192−P2


A
CLM vs pCRC
mLu vs pCRC
PM vs pCRC
−log(padj, 10)
log2FoldChange
Mir−210
Mir−8−P1b
Mir−10−P1a−v1
Mir−375
Signature miRNA
Blood Cell
Cell Type Specific
Normal Control
B
Up-Regulated
Down-Regulated
CLM vs pCRC
mLu vs pCRC
PM-CRC vs pCRC
Mir-339
Mir-425
Mir-592
Mir-338-P1
Mir-1247
Mir-10-P1a
Mir-375
Mir-210
Mir-8-P1b
Mir-1307
Mir-19-P1
Mir-19-P2a
Mir-19-P2b
Mir-374-P1
Mir-155
Mir-584
Mir-191
Mir-154-P9
Mir-154-P36
Mir-127
Mir-10-P1b
Mir-451
Mir-154-P12
Mir-30-P1b
Mir-423
Mir-221-P2
Let-7-P1b
Let-7-P2c1
Mir-29-P1b
Mir-30-P2a
Mir-30-P2b
C
Tissue
nLi
nLu
nCR
pCRC
CLM
mLu
PM
z_score
miRNA
Let−7−P1b
Let−7−P2c1
Mir−10−P1a−v1
Mir−10−P1b−v1
Mir−1247
Mir−127
Mir−1307
Mir−154−P12
Mir−154−P36
Mir−154−P9
Mir−155
Mir−19−P1
Mir−19−P2a
Mir−19−P2b
Mir−191
Mir−210
Mir−221−P2
Mir−29−P1b
Mir−30−P1b
Mir−30−P2a
Mir−30−P2b
Mir−338−P1
Mir−339
Mir−374−P1
Mir−375
Mir−423
Mir−425
Mir−451
Mir−584
Mir−592
Mir−8−P1b


A
Mir-8-P1b
Mir-375
Mir-10-P1a
Mir-210
log2FoldChange
2
0
−2
−4
nCR.vs.pCRC
nLi.vs.pCRC
nLu.vs.pCRC
CLM.vs.pCRC
mLu.vs.pCRC
PM.vs.pCRC
experiment
B
Mir-8-P1b
Mir-375
Mir-10-P1a
Mir-210
log10_RPM
5
4
3
2
1
nCR
pCRC
nLi
nLu
CLM
mLu
PM_CRC
tissue