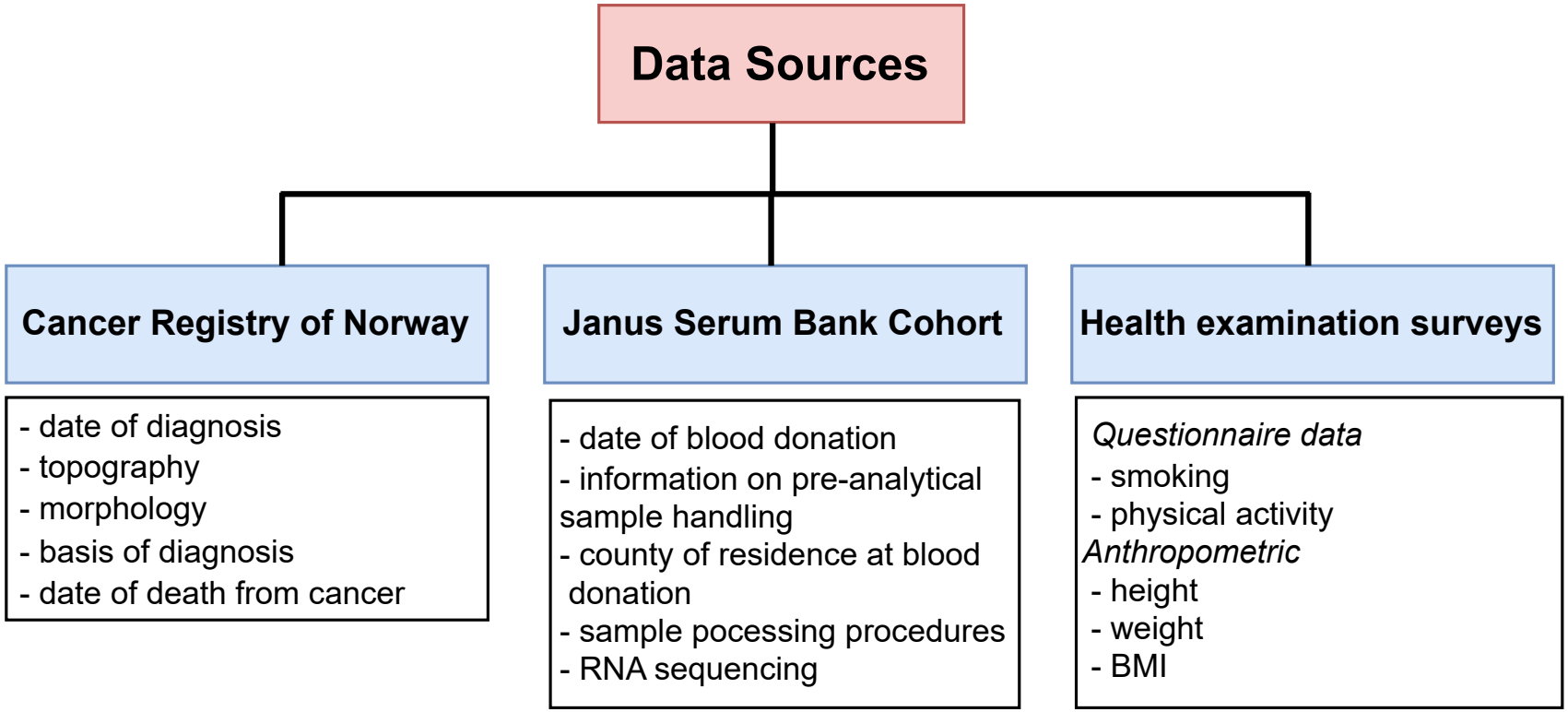
Differential levels of circulating RNAs prior to endometrial cancer diagnosis

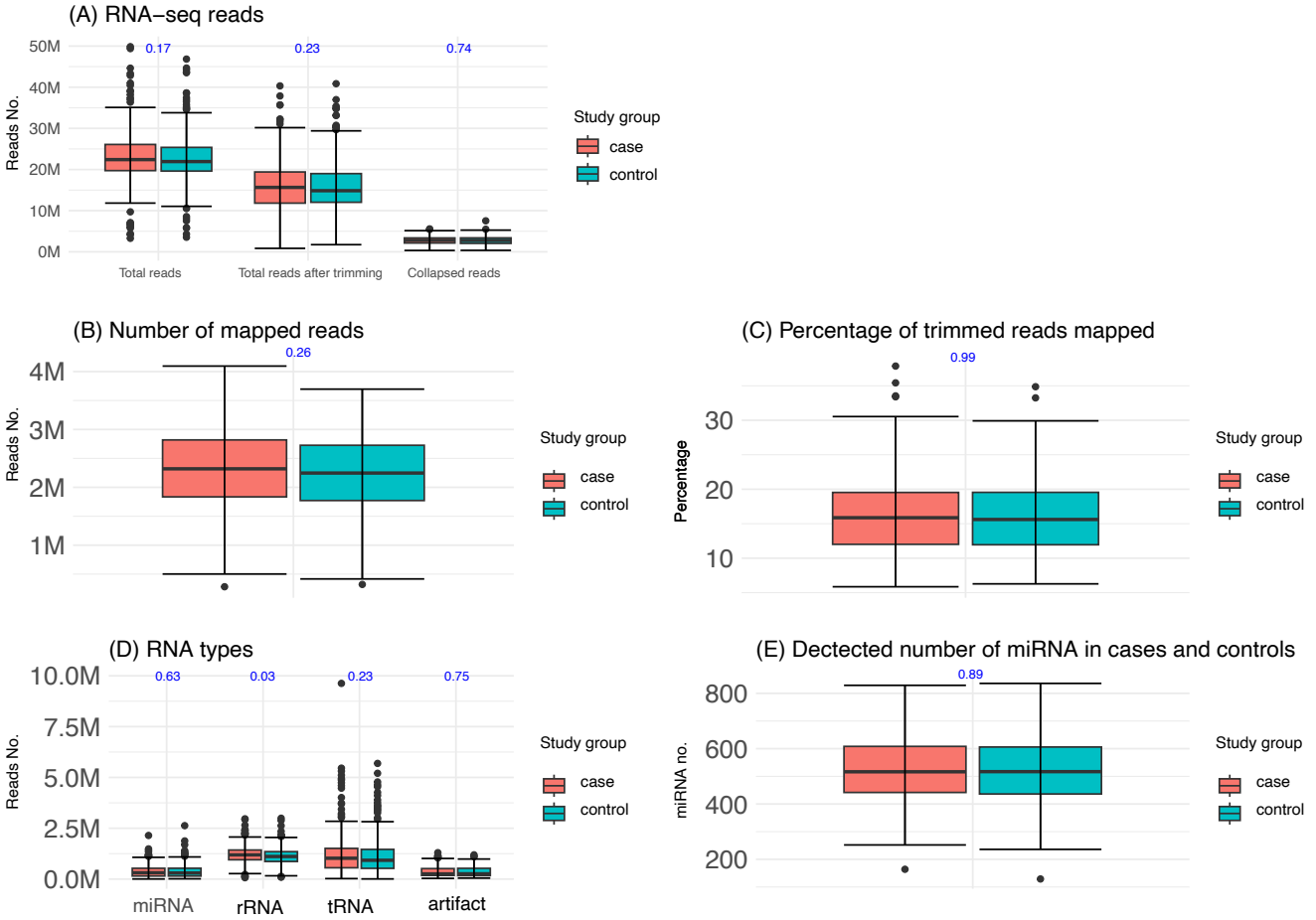
Sina Rostami, Trine B. Rounge, Luca Pestarino, Robert Lyle, Renée Turzanski Fortner, Øystein Ariansen Haaland, Rolv Terje Lie, Fredrik Wiklund, Tone Bjørge, Hilde Langseth

Table of contents:

Supplementary Figure 1. Linkage of data sources

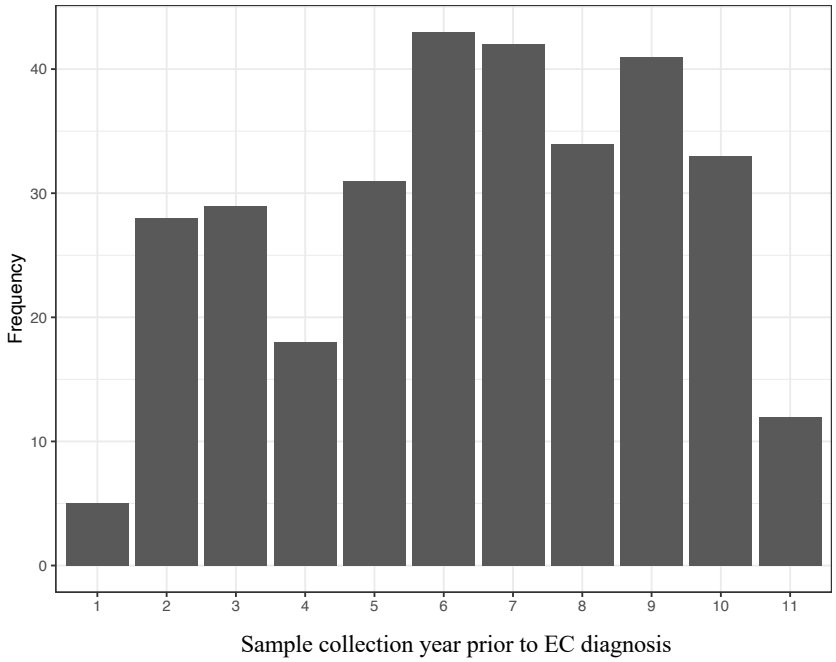


Supplementary Figure 2. Small RNA sequencing coverage and quality statistics. A) Number of reads from small RNA-seq. Total reads after trimming of adapters. Collapsing reads is a relevant step for small RNA-seq to optimize the computational performance of the alignment step. B) Number of reads mapped to the human genome. C) Percentage of trimmed reads mapped to the human genome. D) RNA types based on miRTrace results. E) Number of miRNAs (of the total 2652), with at least 1 read for each sample. Quality statistics represented with box and whisker plots with error bars. Wilcoxon Rank-Sum test was performed between study groups (indicated in blue text above each of the two box plots). Except for rRNA, cases and controls were similar.

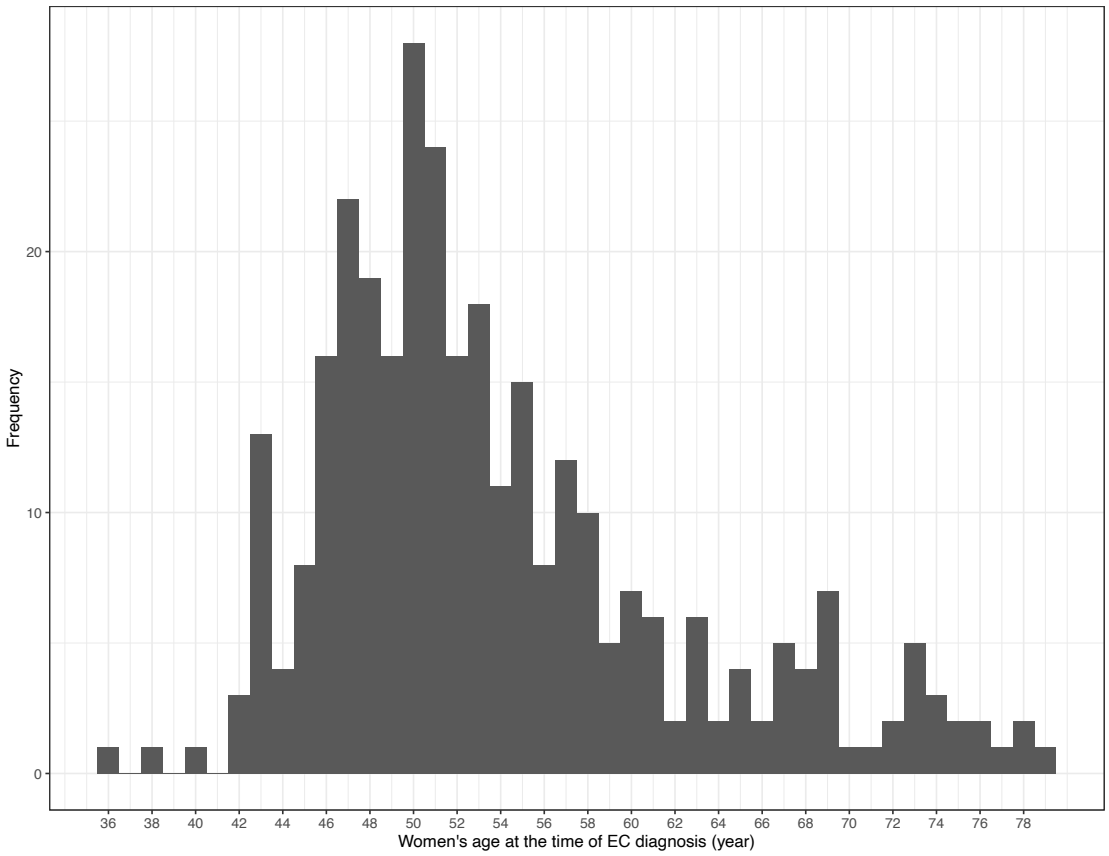


Supplementary Figure 3. Sample batch groups for serum samples based on sample collection periods and serum processing methods.

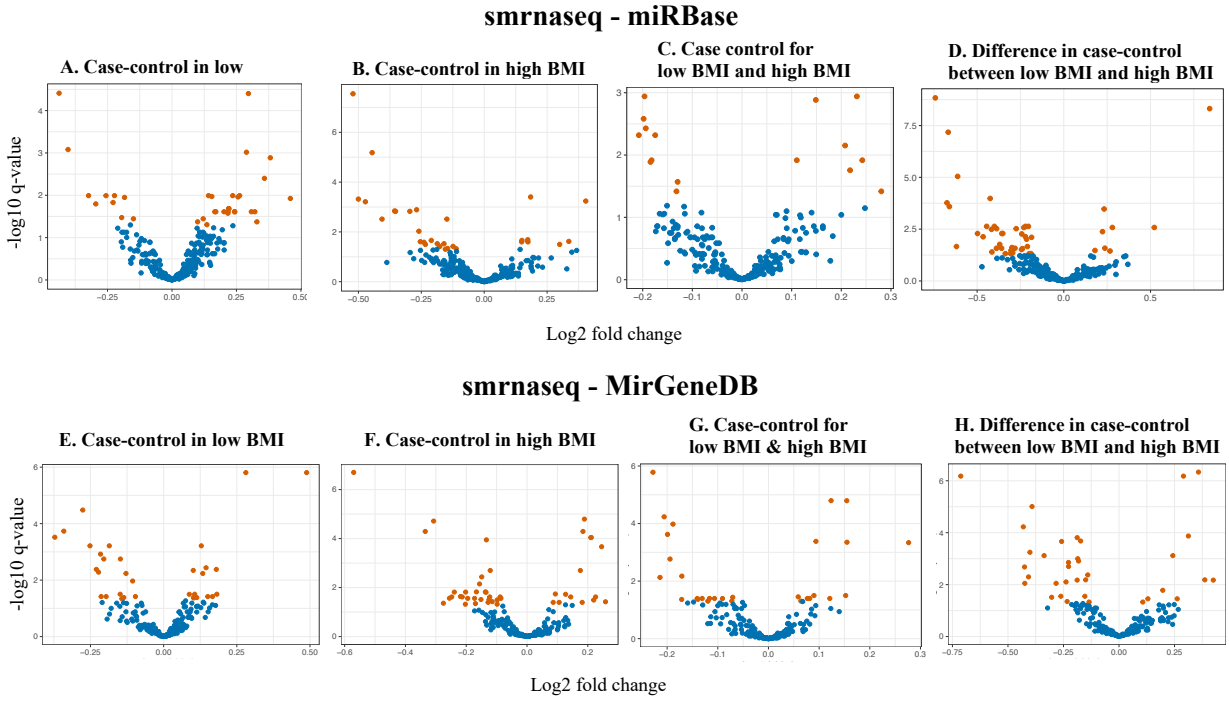
Supplementary Figure 4. Distribution of EC cases with regards to sample collection year prior to diagnosis (1-11 years).



Supplementary Figure 5. Distribution of women's age at the time of EC diagnosis

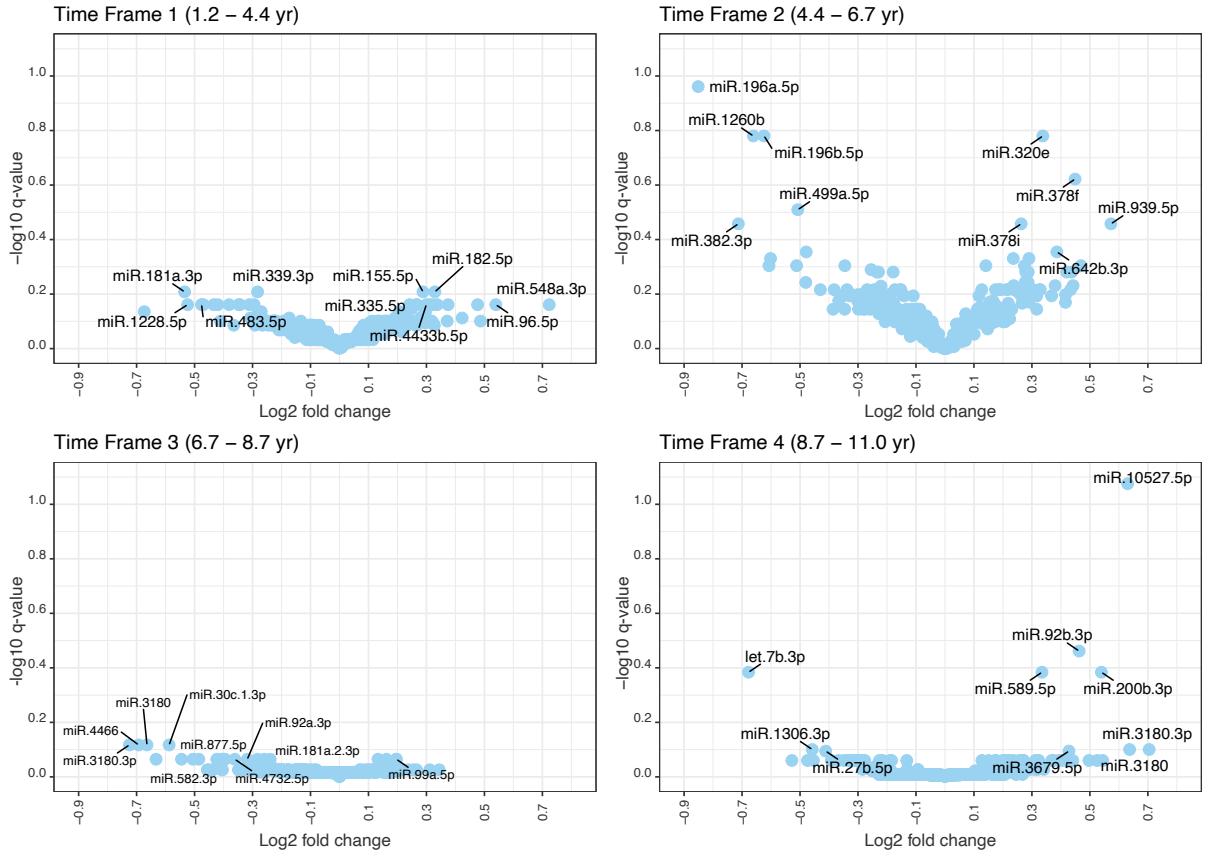


Supplementary Figure 6. Differential abundance of miRNAs for endometrial cancer cases with samples collected prior to diagnosis and controls, and its interaction in a model with BMI levels. RNA counts were based on smrnaseq pre-processing pipeline and miRNA annotation databases miRBase and MirGeneDB. (A and E) case-control status in low BMI, (B and F) case-control status in high BMI, (C and G) case-control status for low BMI and high BMI, (D and H) Interaction term: difference in case-control status between low BMI and high BMI. Differential abundance models were based on mixed model with sample batch groups as random effect, and included age at sample collection. WHO BMI levels were recoded to represent 'low BMI' (underweight and normal weight) and 'high BMI' (overweight and obese) groups.



Supplementary Figure 7. Differential abundance of miRNAs for endometrial cancer based on Model 1 in different Time Frames prior to diagnosis. Top 10 (by the lowest q-value) differentially abundant (DA) miRNAs are labeled in each Time Frame. There were 79 case-control paired samples in each Time Frame. Volcano plots are based on results in Model 1 in Supplementary Table 5.

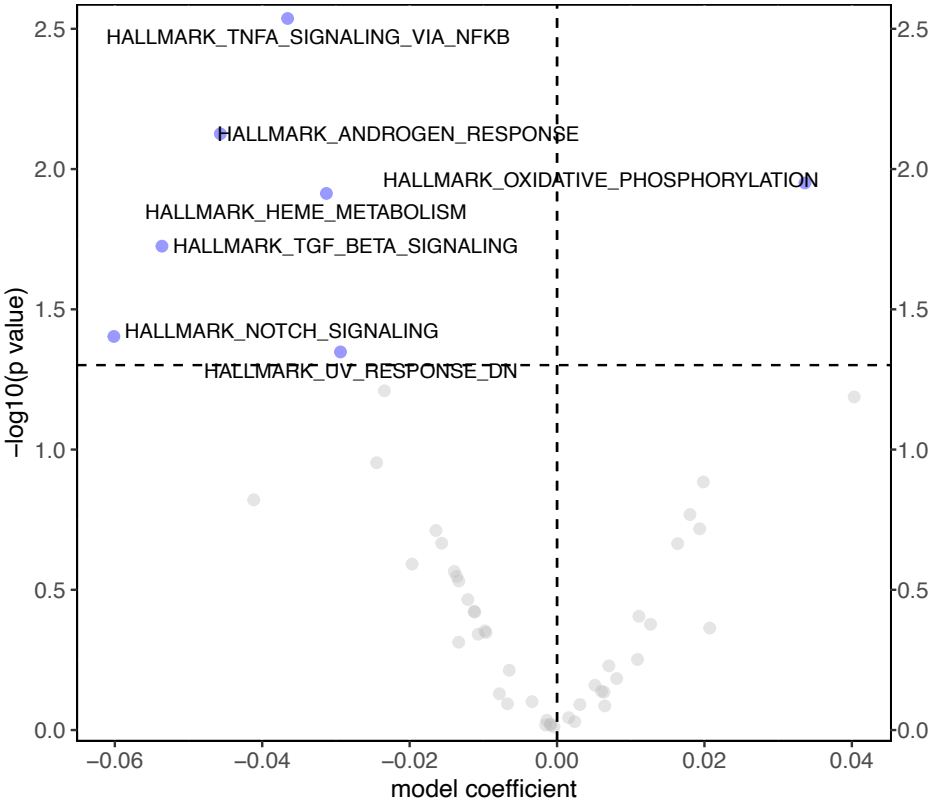
We did another sensitivity analyses, and included all controls in all Time Frames. Using this approach, we saw more DA RNAs (listed in Supplementary Table 6). Note that, however, this made the analyses less comparable to the analysis with all samples, or analysis of matched cases and controls in the 4 Time Frames. For instance, we see that the age at sample collection, which was a criterion for matching cases and controls, is now different between cases in Time Frame 2, and Time Frame 4 and all controls.



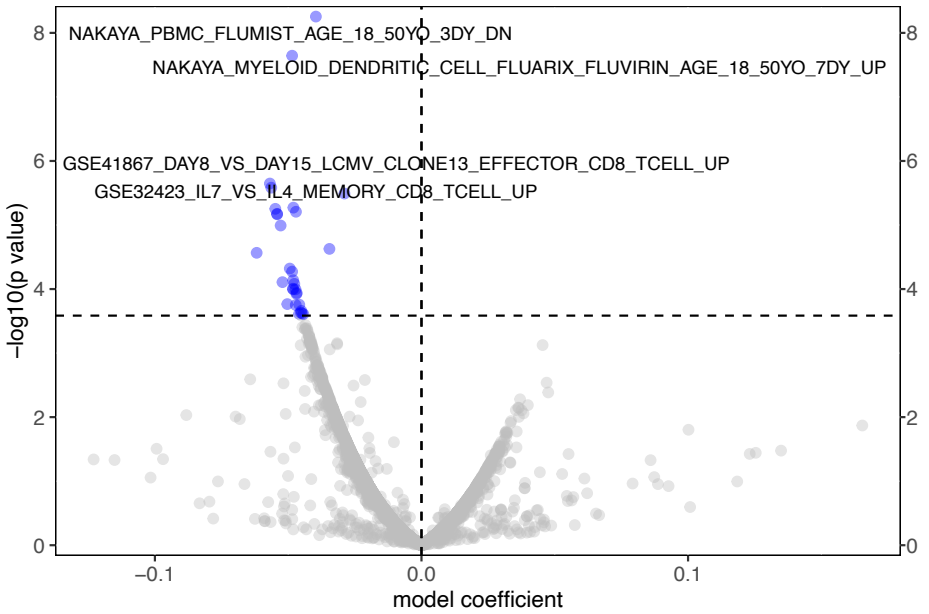
Supplementary Figure 8. miRNA to category heatmap for EC case-control status (53 miRNAs)

Supplementary Figure 9. miRNA gene set analysis with RBiomirGS for selected gene sets from GSEA. Based on miRNAs for EC in Model 1 (Table 2).

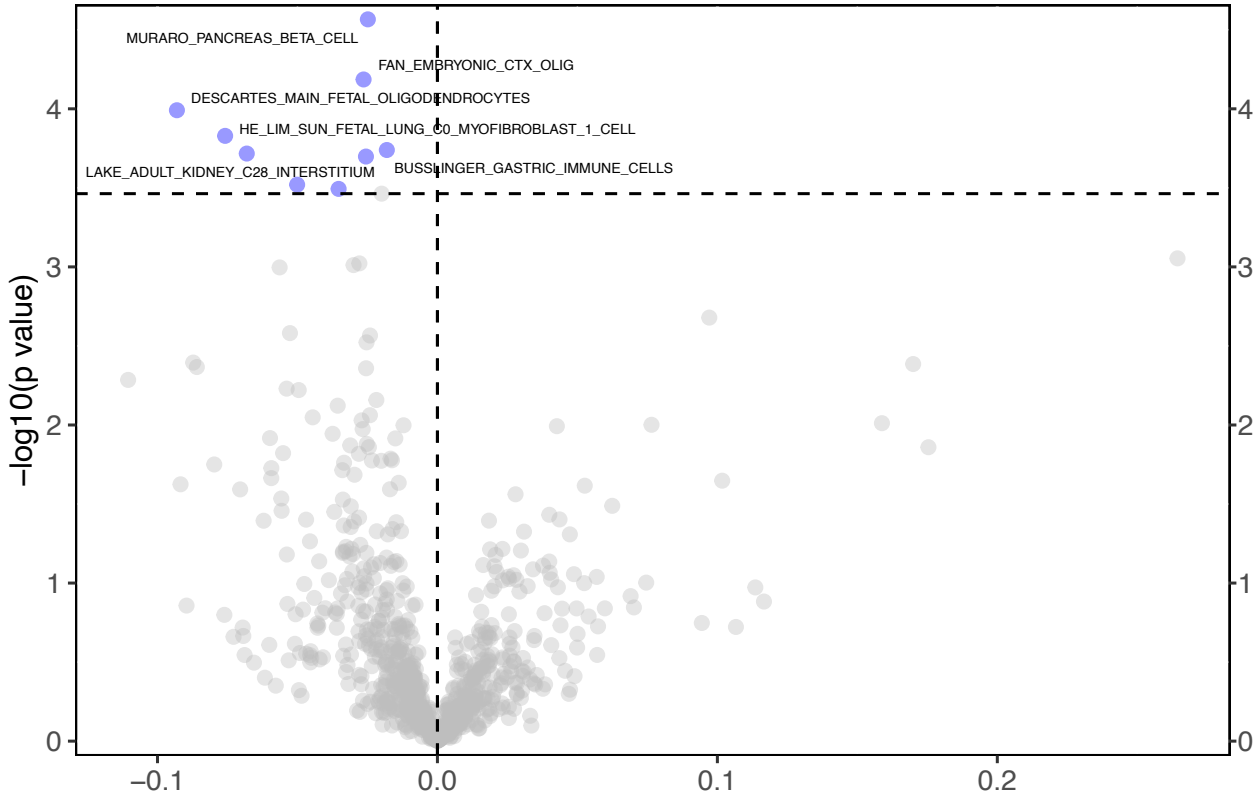
A) for "H: Hallmark gene sets"



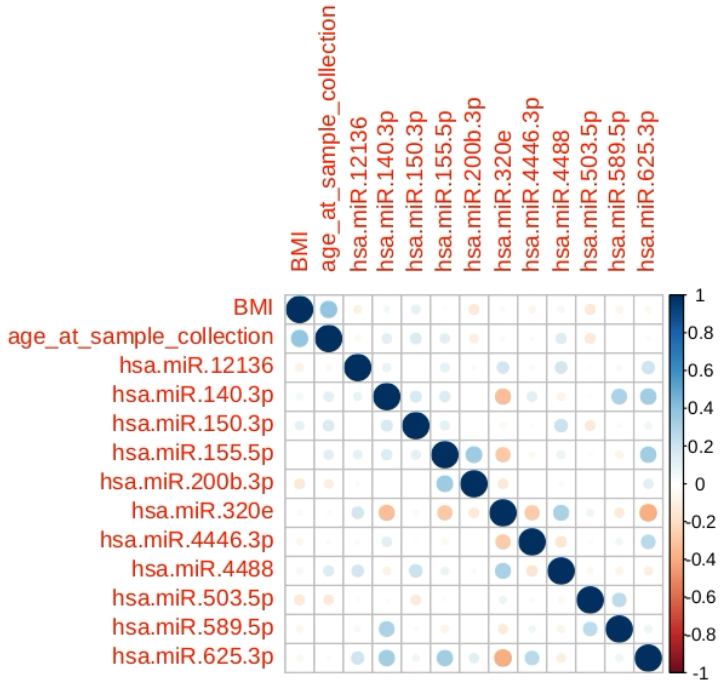
B) for "C7: immunologic signature gene sets"



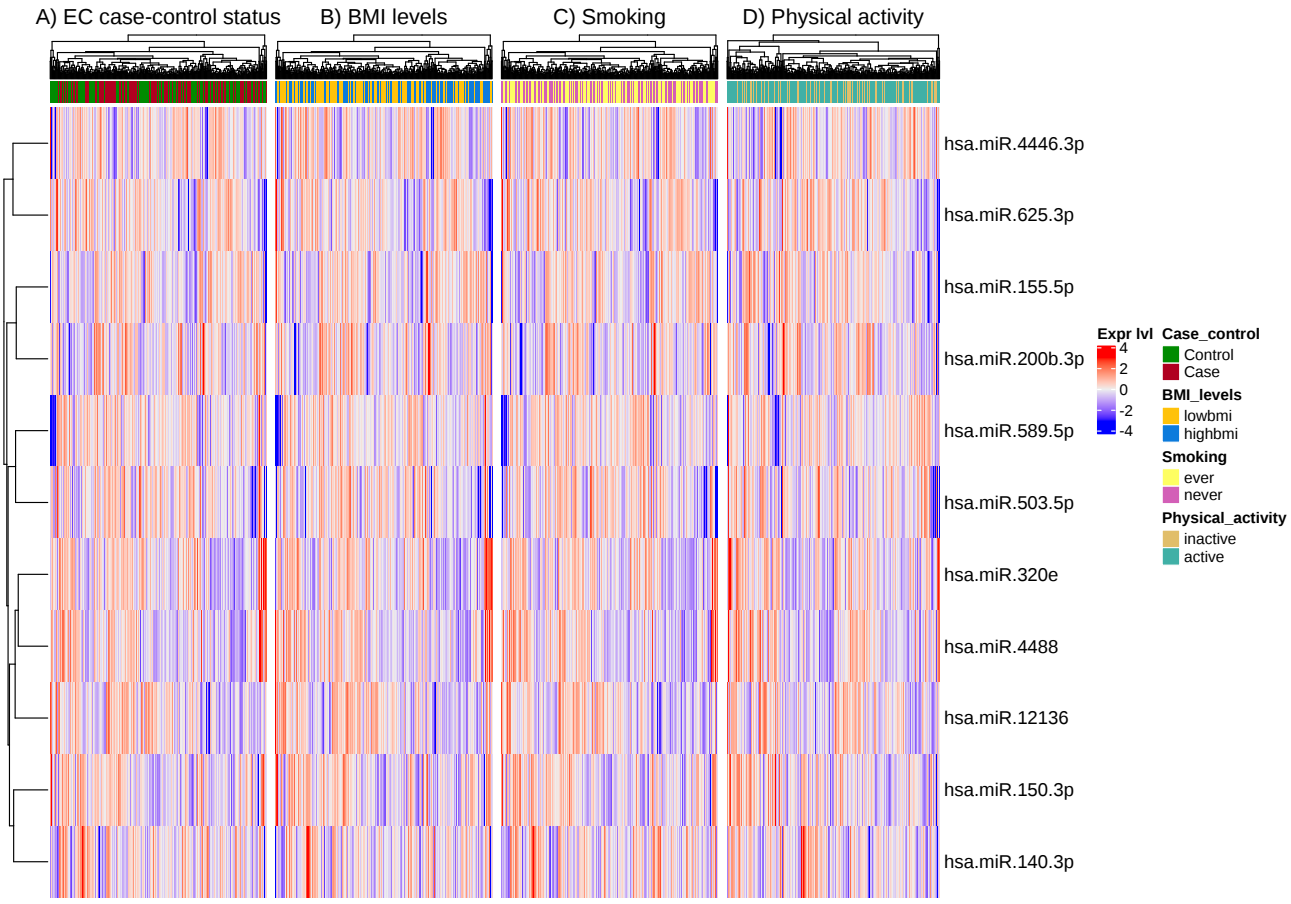
C) for “C8: cell type signature gene sets”



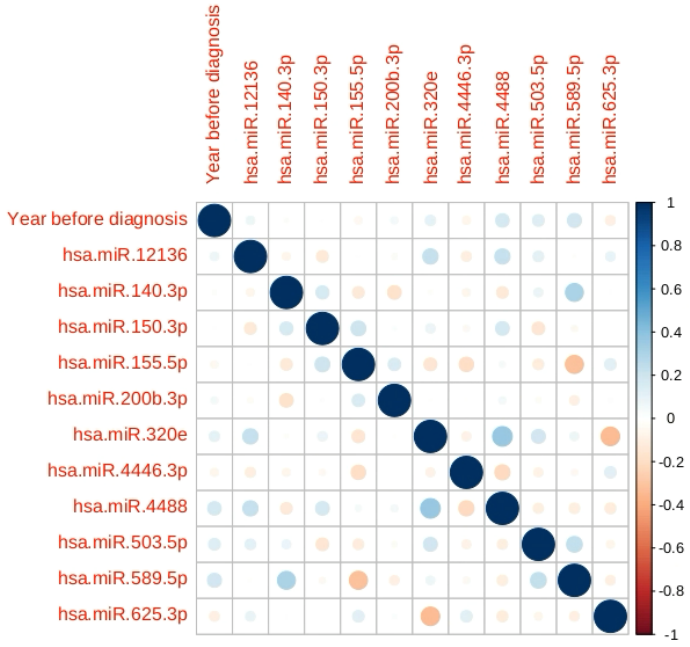
Supplementary Figure 10. Spearman correlation heatmap for 11 DA miRNAs shared for EC in Models 1, 2 and 3 (Figure 2P, and Supplementary Tables 1-3), BMI, and age at sample collection. There was no strong correlation between BMI and age at sample collection with any of the 11 miRNAs. Given the sample batch effect, we only considered samples in group 3 (largest sample group with 398 samples). miRNAs were normalized with TMM and CPM normalization in edgeR.



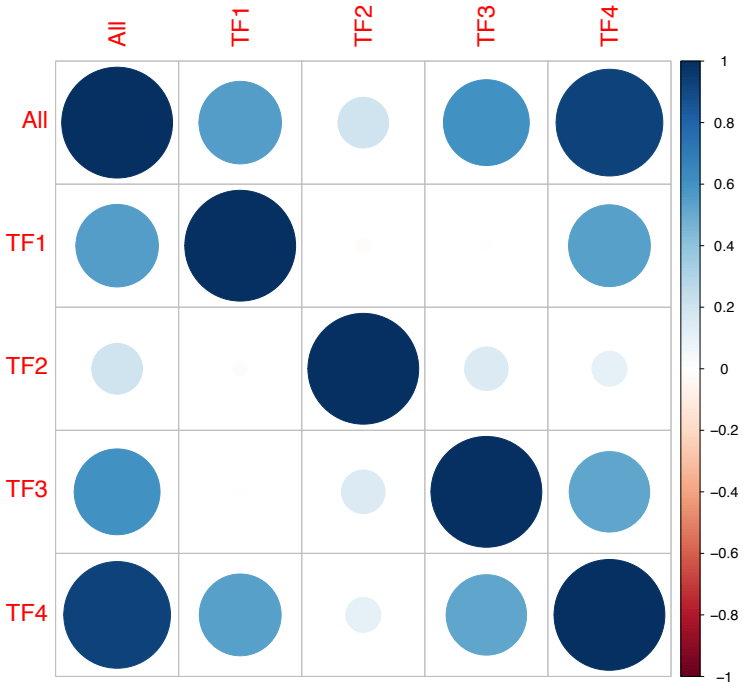
Supplementary Figure 11. Hierarchical clustering heatmap with samples labelled for EC case-control status (A), BMI levels (B), smoking levels (C), and physical activity levels (D). To ensure that the heatmaps were not reflecting sample batch groups, we restricted the analysis to sample batch group 3 only (398 samples). miRNA counts were normalized with TMM and CPM using edgeR and transformed to Z scores. Here, we present heatmaps for 11 DA miRNAs shared for EC in Models 1, 2 and 3 (Figure 2P, and Supplementary Tables 1-3). Overall, there was no specific pattern for sample clustering.



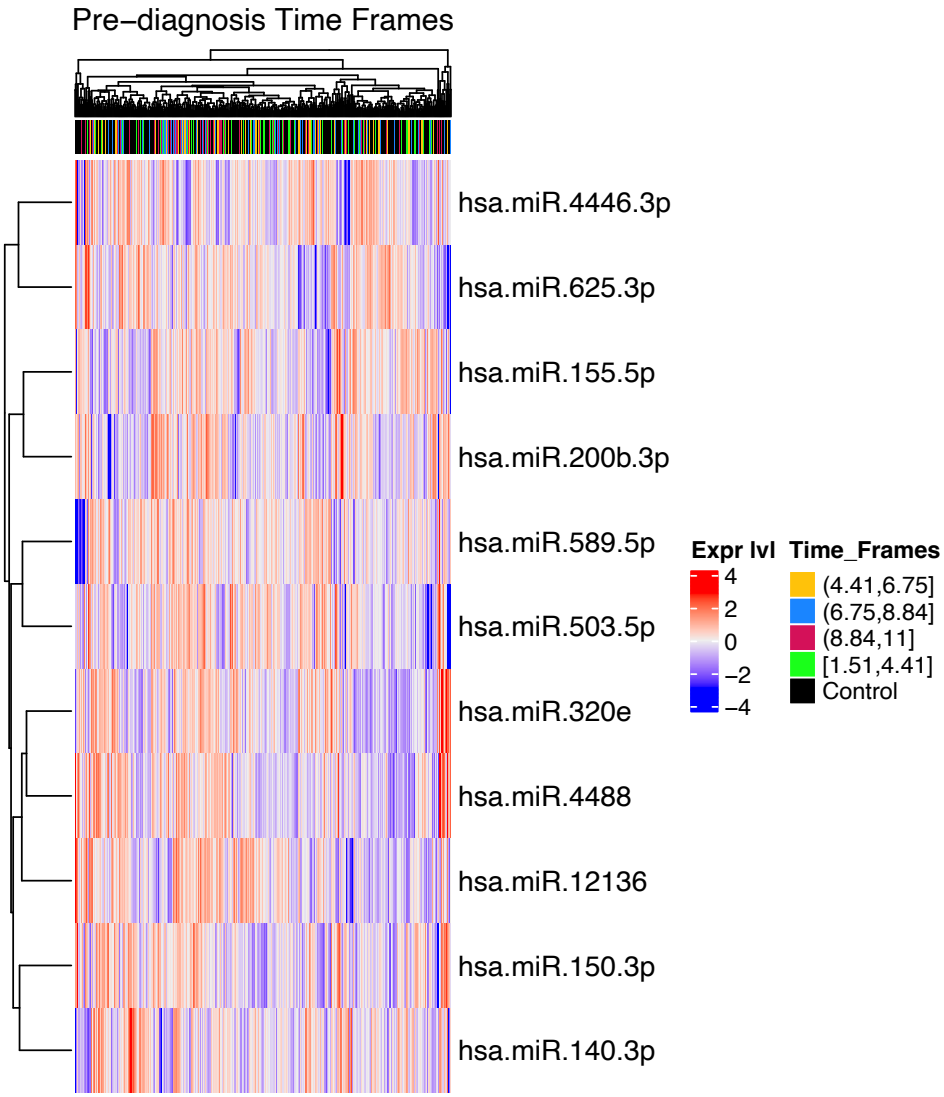
Supplementary Figure 12. Spearman correlation heatmap for 11 DA miRNAs shared for EC in Models 1, 2 and 3 (Figure 2P, and Supplementary Tables 1-3), and sample collection year prior to diagnosis for EC cases. There was no strong correlation between year before diagnosis and any of the 11 miRNAs. Given the sample batch effect, we only considered EC cases in group 3 (largest sample group with 398 samples). miRNAs were normalized with TMM and CPM normalization in edgeR.



Supplementary Figure 13. Spearman's correlation heatmap for selected miRNAs (11 DA miRNAs shared for EC in Models 1-3 (Figure 2P, and Supplementary Tables 1-3)), and their logFC between the analysis with all samples, and those in Time Frames 1- 4. There was a strong correlation between the analysis with all samples and Time Frame 4 (0.93). TF2 showed the least correlation with the analysis with all samples and all other Time Frames.



Supplementary Figure 14. Hierarchical clustering heatmap with samples labelled for 4 Time Frames prior to diagnosis and controls. To ensure that the heatmaps were not reflecting sample batch groups, we restricted the analysis to sample in batch group 3 only (398 samples). miRNA counts were normalized with TMM and CPM using edgeR and transformed to Z scores. Here, we present heatmaps for 11 DA miRNAs shared for EC in Models 1, 2 and 3 (Figure 2P, and Supplementary Tables 1-3). Overall, there was no specific pattern for sample clustering. Note that the Time Frames are not exactly corresponding to Time Frames in Supplementary Figure 7. This is because for this analysis we have included samples in batch group 3 only. Number of samples in Time Frames [1.51-4.41], (4.41-6.75], (6.75-8.84], (8.84-11], and controls were 50, 49, 49, 49, and 201, respectively.



Description of Supplementary Tables:

Table_s1: List of differentially abundant RNAs (with q-value < 0.05) for endometrial cancer cases with samples collected prior to diagnosis and controls based on Model 1 in Table 2. Raw count tables were produced by sncRNA pipeline and miRNAs were annotated by miRbase.

Table_s2: List of differentially abundant RNAs (with q-value < 0.05) for endometrial cancer cases with samples collected prior to diagnosis and controls based on Model 2 in Table 2. Raw count tables were produced by sncRNA pipeline and miRNAs were annotated by miRbase.

Table_s3: List of differentially abundant RNAs (with q-value < 0.05) for endometrial cancer cases with samples collected prior to diagnosis and controls based on Model 3 in Table 2. Raw count tables were produced by sncRNA pipeline and miRNAs were annotated by miRbase.

Table_s4: List of differentially abundant miRNAs (with q-value < 0.05) for endometrial cancer cases with samples collected prior to diagnosis and controls based on Models 1-3 in Table 2. Raw count tables were produced by smrnaseq pipeline and miRNAs were annotated by miRbase and MirGeneDB.

Table_s5: List of top 10 (by the lowest q-value) miRNAs for endometrial cancer based on Models 1-3 (Table 2) in Time Frames prior to diagnosis (1.16-4.39], (4.39-6.66], (6.66-8.67], and (8.67-11] years. There were 79 case-control paired samples in each Time Frame.

Table_s6: List of top 10 (by the lowest q-value) miRNAs for endometrial cancer based on Models 1-3 (Table 2) in Time Frames prior to diagnosis (1.16-4.39], (4.39-6.66], (6.66-8.67], and (8.67-11] years. There were 79 cases in each Time Frame and all controls (316) were used in all Time Frames.

Table_s7: List of top (by the lowest q-value) miRNAs for endometrial cancer based on Models 1-3 (Table 2), after removal of samples in group 1 sample batch group which showed the most noticeable batch effect. Highlighted in red are miRNAs shared with DA miRNAs in Models 1-3 in Figures 2C, 2G, and 2K, respectively.

Table_s8: Comparison of differentially abundant miRNAs for endometrial cancer based on Models 1-3 in the present study to previous studies on circulating miRNAs for endometrial cancer (Bloomfield et. al (doi: 10.3390/cells11111836), Fan et. al (doi: 10.1042/BSR20210111), and Zhou et. al (doi: 10.1186/s12943-021-01352-4)).