SCIENCE CHINA PRESS

SCIENCE CHINA Life Sciences
life.scichina.com link.springer.com
https://doi.org/10.1007/s11427-024-2913-7

SPECIAL TOPIC: RNA modifications

RESEARCH PAPER

Peripheral blood RNA modifications as a novel diagnostic signature for polycystic ovary syndrome

Liwen Zhang[1,2†], Xinxin Liu[2,3†], Yu Zhang[4†], Lang Qin[4†], Shijia Pan[1], Xueqi Yan[4], Sen Dong[1], Zerong Feng[2,3], Song-jia Fan[2], Rusong Zhao[4], Xueying Gao[4], Shigang Zhao[4*], Junchao Shi[2,3*], Han Zhao[4*], Ying Zhang[1*] & Zi-Jiang Chen[4]

[1]The Key Laboratory of Cell Proliferation and Regulation Biology, Ministry of Education, Beijing Key Laboratory of Genetic Engineering Drug and Biotechnology, College of Life Sciences, Beijing Normal University, Beijing 100875, China
[2]China National Center for Bioinformation and Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]State Key Laboratory of Reproductive Medicine and Offspring Health, Center for Reproductive Medicine, Institute of Women, Children and Reproductive Health, Shandong University, Jinan 250012, China

†Contributed equally to this work
*Corresponding authors (Ying Zhang, email: yingzhanglab@gmail.com; Han Zhao, email: hanzh80@sdu.edu.cn; Junchao Shi, email: shijc.lab@gmail.com; Shigang Zhao, email: zsg0108@126.com)

Received 22 November 2024; Accepted 13 February 2025; Published online 3 June 2025

Polycystic ovary syndrome (PCOS) is the most prevalent ovulatory and endocrine disorder affecting reproductive-aged women, yet the absence of a specific, rapid molecular diagnostic marker results in diagnostic delays and inaccuracies. Given the critical role of RNA modifications in disease pathology, this study utilized a high-throughput RNA modification profiling platform to investigate 15 types of peripheral blood RNA modification patterns in individuals with ovulatory disorders, including PCOS and primary ovarian insufficiency (POI), and control subjects. Our results revealed that distinct modification profiles correspond to specific disease states, with significant shifts in RNA modification inter-correlations observed across conditions. Additionally, specific RNA modifications were associated with clinical features, such as serum levels of testosterone and the follicle number per ovary (FNPO). To optimize diagnostic precision, we evaluated various machine learning models, identifying that combining m^6A and m^7G modifications in a light gradient boosting machine model (LightGBM) achieves the highest accuracy in distinguishing PCOS, outperforming traditional diagnostic markers. This highlights the potential of RNA modification profiling as a novel, high-accuracy diagnostic tool for PCOS in clinical settings.

PCOS | RNA modification | follicle number per ovary | machine learning | prediction model | biomarker

INTRODUCTION

Ovulatory disorders account for over 25% of female infertility cases in clinical settings (Carson and Kallen, 2021), with polycystic ovary syndrome (PCOS) affecting 5% to 20% of reproductive-aged women (Azziz et al., 2016), followed by primary ovarian insufficiency (POI), which affects 1% to 3.7% of this population (Stuenkel and Gompel, 2023). PCOS is clinically characterized by irregular menstrual cycles, polycystic ovaries visible on ultrasound, and biochemical signs of excess androgens (Rosenfield and Ehrmann, 2016). Additionally, many PCOS patients present with obesity and are at increased risk of metabolic disorders and cardiovascular diseases (Wild, 2002), emphasizing the importance of early and accurate identification in clinical practice.

The formal diagnostic criteria for PCOS were first proposed in the early 1990s and have been updated over the past 30 years, most notably with the introduction of the Rotterdam Criteria (Teede et al., 2023). Despite these updates, the lack of a specific diagnostic test for PCOS continues to pose significant challenges in its diagnosis and management. One of the primary challenges in current PCOS diagnostics is the assessment of ovarian antral follicle count, which relies on traditional ultrasound imaging. This method is highly operator-dependent, and variations in operator expertise (Sonigo et al., 2018), as well as differences in ultrasound equipment and techniques, can lead to inconsistent follicle counts (Sonigo et al., 2018), complicating the diagnosis of PCOS in clinical settings. Moreover, as a chronic condition, PCOS presents heterogeneous clinical features, and symptom severity can vary across an individual's lifespan. Some women may exhibit minimal or no overt symptoms. Tracking and monitoring clinical symptoms over time is also necessary for PCOS diagnosis. However, the high variability introduced by subjective factors in the diagnostic process further complicates accurate diagnosis (Escobar-Morreale, 2018). These challenges often result in delayed or misdiagnosed PCOS cases in clinical practice (Agapova et al., 2014). Given these issues, there is an urgent need to identify molecular characteristics and specific biomarkers for PCOS. Such advancements would not only enhance diagnostic precision but also facilitate the discovery of potential therapeutic

Citation: Zhang, L., Liu, X., Zhang, Y., Qin, L., Pan, S., Yan, X., Dong, S., Feng, Z., Fan, S.j., Zhao, R., et al. Peripheral blood RNA modifications as a novel diagnostic signature for polycystic ovary syndrome. Sci China Life Sci, https://doi.org/10.1007/s11427-024-2913-7

targets.

Over the past few decades, peripheral blood biomarkers have garnered increasing attention due to their sensitivity, specificity, and convenience. As diseases progress, circulating cells and secreted cellular materials undergo dynamic changes, carrying a wealth of genetic and non-genetic information (Natalia et al., 2023). Consequently, various blood biomarkers, such as extracellular vesicles (Thompson et al., 2016), proteins (Borrebaeck, 2017), and nucleic acids (Schwarzenbach et al., 2011), have been extensively studied for their potential in disease detection and prediction. Among these, RNA modifications have recently emerged as a promising area of research.

To date, more than 170 types of RNA modifications have been identified across the biosphere. With advancements in detection technologies, even RNA modifications on less abundant RNAs are now being detected and characterized, significantly advancing our understanding of this complex field. Initially considered only structural features of RNA, these modifications have since gained attention for their critical roles in regulating RNA function and their involvement in human diseases (Barbieri and Kouzarides, 2020; Delaunay et al., 2024; Sullenger and Gilboa, 2002). RNA modifications have been shown to change sensitively and function in response to disease in both local tissues and the peripheral circulation. For example, blood levels of N^6-methyladenosine (m^6A) were significantly altered following radiation therapy (Chen et al., 2023) and were found to be elevated in the peripheral blood of breast cancer patients compared with those with benign breast conditions (Xiao et al., 2021). Several studies have demonstrated that m^6A modification is critical for hematopoietic stem cell regeneration (Wang et al., 2018; Zhang et al., 2017) and function (Yin et al., 2022), further emphasizing the importance of RNA modification in the circulatory system. These findings suggest that peripheral RNA modifications hold significant potential for disease diagnosis and treatment. Previously, we developed a high-throughput RNA modification profiling platform based on LC-MS/MS (Chen et al., 2016; Zhang et al., 2022; Zhang et al., 2018), which allows for the detection and quantification of RNA modifications from low-input samples.

In this study, we investigated RNA modification changes in the peripheral blood of individuals with ovarian diseases, including PCOS and POI, as well as control subjects. We identified distinct RNA modification signatures across these groups. Notably, we found that specific RNA modifications, such as m^6A and m^7G, could accurately distinguish PCOS cases from both POI patients and control subjects, with high sensitivity and specificity.

RESULTS

Differential expression and variability of RNA modifications in peripheral blood samples among PCOS, POI, and control groups

The baseline characteristics of the PCOS, POI, and control groups are summarized in Table 1 and Table S1 with age distributions consistent across all groups. In alignment with prior studies, the PCOS group exhibited significantly higher serum testosterone (T) levels ($P<0.01$), an increased follicle number per ovary (FNPO) ($P<0.001$), and elevated serum anti-Müllerian hormone (AMH) levels ($P<0.0001$) compared with controls. In contrast, the POI group showed significantly higher serum follicle-stimulating hormone (FSH) levels ($P<0.0001$), lower AMH levels ($P<0.001$), and reduced FNPO ($P<0.001$). Peripheral blood RNAs were then digested into single nucleotides, and RNA modification profiling was performed using a single-blind, high-throughput LC-MS/MS approach. The overall strategy for identifying RNA modifications is depicted in Figure 1A.

Across samples from PCOS, POI, and control subjects, we identified and quantified 15 types of modified nucleosides, including m^1A, m^6A, Am, I, Im, m^5U, Um, m^5C, Cm, m^1G, m^2G, m^7G, $m^{2,2}G$, $m^{2,2,7}G$, and Gm. Principal component analysis (PCA) (Figure 1B) revealed distinct RNA modification

Table 1. Clinical characteristics among different groups in the discovery and validation cohorts[a)]

Variables	Discovery cohort				Validation cohort			
	Control	PCOS	POI	*P* value	Control	PCOS	POI	*P* value
Subjects (*n*)	25	27	15		20	21	14	
Age (years)	30.00 [28.00–34.00]	29.00 [27.50–31.00]	31.00 [28.00–34.00]	0.4044	30.00 [26.75–33.00]	29.00 [26.00–29.00]	32.00 [30.50–36.75]	0.0185
BMI (kg m^{-2})	22.79 [20.14–24.42]	23.63 [21.47–27.07]	21.60 [20.15–24.35]	0.2928	21.56 [19.96–29.04]	24.50 [21.63–26.99]	23.00 [21.78–26.83]	0.6897
FSH (mIU mL^{-1})	6.45 [5.79–7.46]	6.11 [5.32–7.38]	20.80 [16.10–28.13]	<0.001	6.34 [5.84–7.33]	5.82 [5.16–6.23]	31.59 [18.70–45.93]	<0.001
LH (mIU mL^{-1})	4.95 [4.13–6.10]	8.50 [5.23–12.07]	6.63 [4.97–14.66]	0.0045	5.48 [4.73–6.24]	9.26 [7.55–12.82]	18.78 [8.43–26.63]	<0.001
Testosterone (ng dL^{-1})	20.62 [15.41–24.59]	30.83 [19.20–46.49]	13.04 [10.24–28.92]	0.0033	15.38 [12.82–20.01]	45.70 [40.85–55.75]	22.42 [10.98–28.18]	<0.001
E_2 (pg mL^{-1})	32.30 [25.60–42.30]	36.40 [27.75–44.35]	23.21 [17.30–33.70]	0.5916	34.25 [23.93–42.5]	39.60 [34.70–58.30]	28.35 [5.68–129.30]	0.0459
P_4 (ng mL^{-1})	0.14 [0.09–0.19]	0.15 [0.09–0.29]	0.28 [0.16–0.39]	0.1620	0.15 [0.11–0.20]	0.14 [0.12–0.24]	0.17 [0.05–0.39]	0.1793
AMH (ng mL^{-1})	3.43 [2.45–5.03]	9.53 [6.58–13.29]	0.12 [0.06–0.55]	<0.001	4.10 [2.87–6.43]	8.32 [6.64–14.07]	0.06 [0.06–0.36]	<0.001
FNPO (*n*)	7.00 [6.00–9.00]	12.00 [11.50–14.00]	1.00 [0.50–2.75]	<0.001	8.25 [5.88–9.13]	13.50 [11.50–14.50]	0.75 [0.00–2.00]	<0.001

a) Values were presented as median [interquartile range]. Statistical significance was determined by two-tailed one-way ANOVA test. BMI, body mass index; FSH, follicle-stimulating hormone; LH, luteinizing hormone; E_2, estradiol; P_4, progesterone.

https://doi.org/10.1007/s11427-024-2913-7

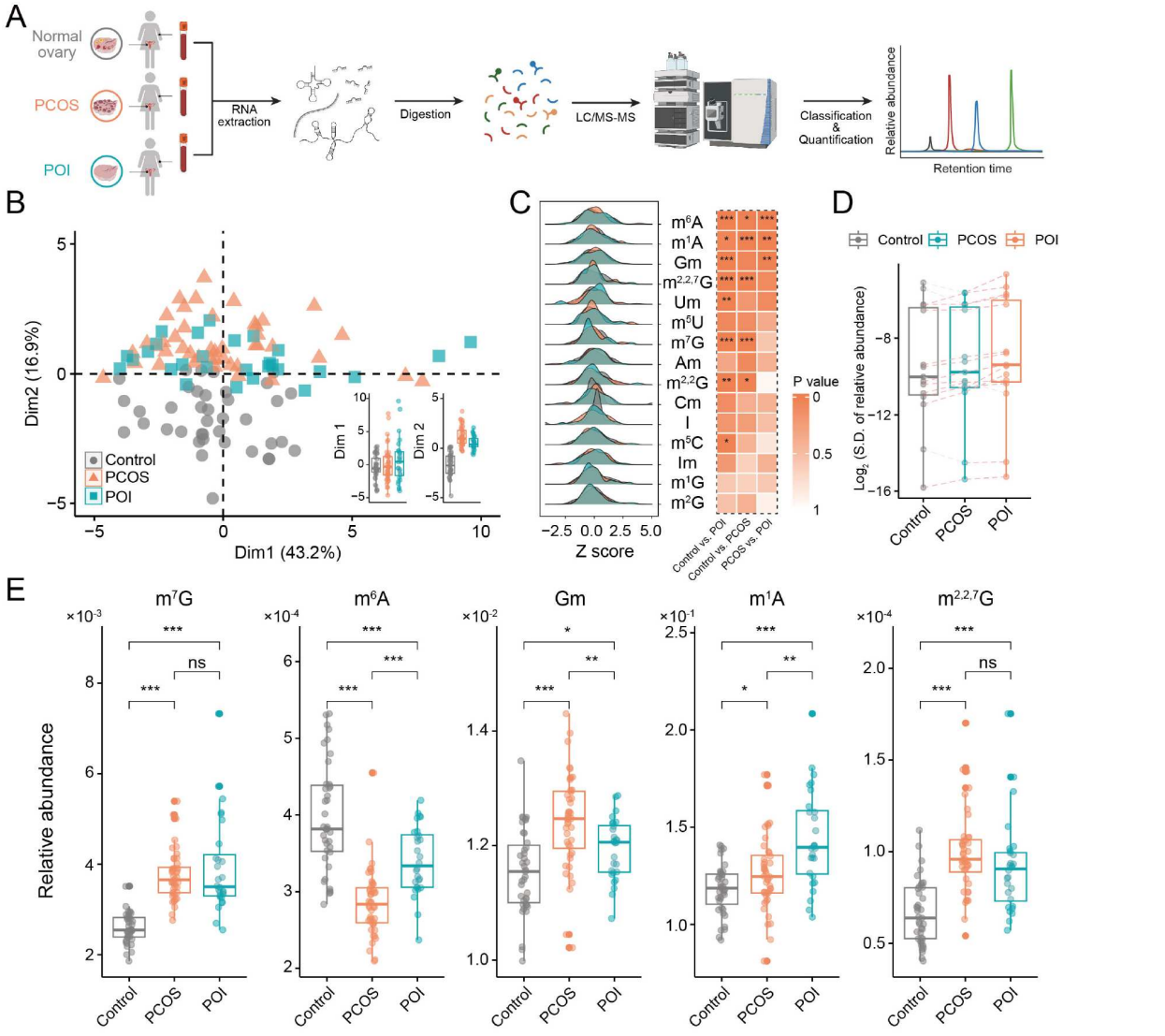

Figure 1. Comprehensive analysis of RNA modification profiles in peripheral blood from control, PCOS, and POI patients. A, Schematic overview of the experimental workflow for identifying RNA modifications. B, PCA of RNA modification abundance across all samples. C, Density plots (left) showing the distribution of Z scores for each RNA modification across the three groups, with a corresponding heatmap (right) displaying the statistical significance of the differences. D, The standard deviation of RNA modification abundance in each group. The pink dashed line indicates modifications with increased variability between groups, while the grey dashed line indicates decreased variability. E, The relative abundance of selected RNA modifications across the three groups. *, $P \leqslant 0.05$; **, $P \leqslant 0.01$, ***, $P \leqslant 0.001$; ns, non-significant differences.

patterns among the three groups, indicating significant alterations in blood RNA modifications associated with disease. Notably, we observed that RNA modification expression levels became more variable with disease onset (Figure 1C and D), suggesting that the stability of RNA modifications is disrupted in PCOS and POI patients. Specific modifications, including m^7G, m^6A, Gm, m^1A, and $m^{2,2,7}G$, are significantly altered in relative abundance across groups (Figure 1E), with Cm, Um, and $m^{2,2}G$ also showing differential expression between the PCOS and control groups (Figure S1).

Intrinsic correlations of RNA modifications across ovarian diseases and controls

We performed a pairwise correlation analysis of all RNA modifications across samples. Interestingly, we observed strong intrinsic associations among various types of RNA modifications. Guanine modifications (including $m^{2,2}G$, m^2G, m^1G, m^7G, and $m^{2,2,7}G$) exhibited robust positive correlations with each other, while modifications such as Am, Im, and I were negatively correlated with the guanine modifications. These intrinsic associations were consistently observed across PCOS, POI, and control samples (Figure 2A). Notably, RNA modifications in PCOS patients displayed a higher number of significant positive correlations compared with POI and control cases, while the POI group demonstrated an increased number of negative correlation pairs relative to the other groups (Figure 2A).

Furthermore, correlation patterns for specific modifications varied distinctly across the PCOS, POI, and control groups (Figure 2B; Figure S2). For instance, some correlations were stable across groups—for example, m^2G and $m^{2,2}G$ maintained strong positive correlations with comparable coefficients across all groups. However, other correlations differed markedly: m^7G and $m^{2,2}G$, while positively correlated in all groups, showed lower correla-

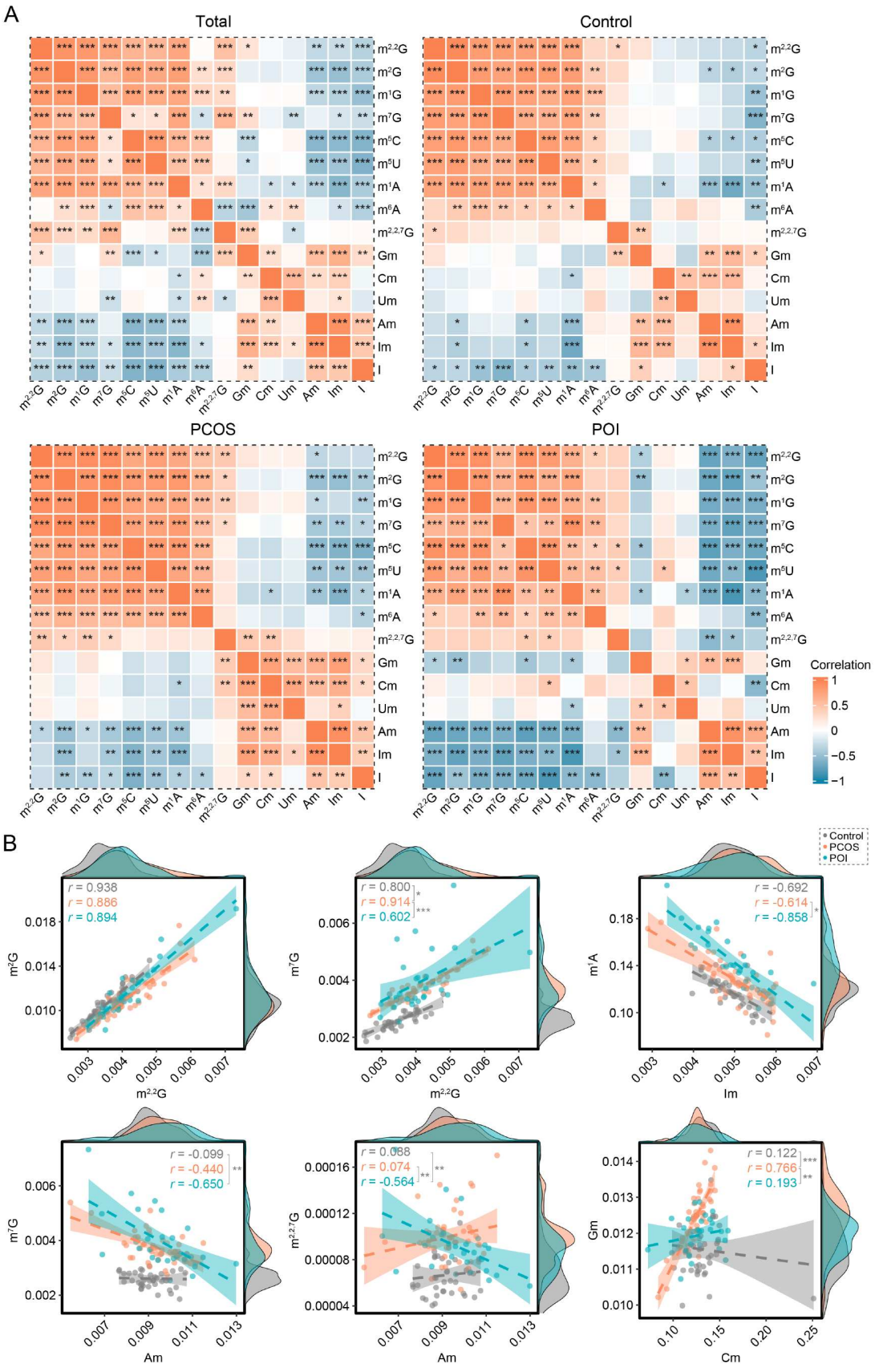

Figure 2. Correlation analysis of RNA modifications. A, Correlation matrices displaying pairwise correlations among RNA modifications for Control, PCOS, and POI groups, as well as for all samples combined. B, The correlations between selected pairs of RNA modifications. *, $P \leqslant 0.05$; **, $P \leqslant 0.01$, ***, $P \leqslant 0.001$.

https://doi.org/10.1007/s11427-024-2913-7

tion coefficients in control (r=0.800) and POI (r=0.602) samples compared with PCOS (r=0.914). Similar shifts were observed between Im and m^1A, which displayed negative correlations across groups (Figure 2B). In some cases, correlation patterns shifted under specific disease conditions. For instance, Am and m^7G showed no correlation in controls but were negatively correlated in both the PCOS (r=−0.440) and POI (r=−0.650) groups. In contrast, Am and $m^{2,2,7}G$ exhibited a positive correlation in PCOS, a relationship that was absent in POI and control groups. The correlation between Gm and Cm also showed significant variability across the three groups (Figure 2B).

While previous studies have primarily focused on individual RNA modifications as potential mechanisms of disease, our results reveal that RNA modifications form intrinsic correlation networks, some of which may represent regulatory interactions specific to ovarian diseases.

Predictive power of RNA modification signatures for clinical features

Given the dynamic changes in blood RNA modifications and their correlations observed across PCOS, POI, and control groups, we investigated whether these RNA modifications were related to specific clinical features associated with these ovarian disorders. We selected key clinical features for analysis, including FNPO, serum testosterone (T), and FSH levels—critical diagnostic factors for both PCOS and POI. Additional variables, such as AMH and body mass index (BMI), known to correlate with PCOS (Álvarez-Blasco et al., 2006; Teede et al., 2019), were also incorporated, alongside age and serum levels of P_4, FSH, and LH. In the PCOS group, RNA modifications m^2G and m^7G exhibited a positive correlation with AMH levels, a proposed diagnostic marker for PCOS (Teede et al., 2019), and a negative correlation with BMI, a recognized PCOS risk factor. In the POI group, modifications Am and I showed a positive association with BMI, while Cm and Um demonstrated a strong negative correlation with P_4 (Figure 3A).

These distinct correlation patterns were further observed among the different groups. For instance, Gm displayed a mild negative correlation with age in the control group, yet a positive correlation with age in the PCOS group. Additionally, correlations between m^1A and FNPO, Am and BMI, m^7G and serum T levels exhibited unique patterns or significant shifts in correlation coefficients across groups (Figure 3B), indicating the sensitivity of RNA modifications to clinical feature fluctuations. These findings suggest that RNA modifications may hold predictive potential for clinical features in ovarian disorders.

To assess the predictive potential of RNA modifications for clinical features of PCOS, we employed six machine learning (ML) regression models. RNA modifications were incrementally incorporated into the corresponding ML models based on their correlation strength with the respective clinical features (Table S2). As shown in Figure 3C, the combination of five RNA modifications ($m^{2,2,7}G$, m^6A, $m^{2,2}G$, m^5U, and m^7G) yielded the highest predictive performance for serum AMH levels using the linear SVM model, with r=0.587 and P=2.21×10^{-5}. Additionally, Cm alone demonstrated the strongest predictive capability for serum E_2 levels, yielding r=0.548 and P=9.74×10^{-5} in a linear regression model. RNA modification combinations proved effective in predicting FNPO and serum T levels with the random forest model (Figure 3C and D; Table S3). While RNA modifications displayed a predictive capacity for BMI, age, and serum LH, FSH, and P_4 levels (Figure S3), these predictions showed weaker associations, indicating the specificity of RNA modifications in predicting key PCOS-related clinical features.

Establishing an RNA modification signature to distinguish PCOS cases

Given the ability of RNA modifications to predict clinical features related to PCOS, we further investigated whether blood RNA modifications could differentiate PCOS cases from POI and control subjects (Figure 4A), which could have potential clinical applications. We used the 15 RNA modifications from the discovery cohort, which included PCOS, POI, and control samples, to develop a prediction model. First, we evaluated all 15 RNA modification features across nine different ML classification models in the discovery cohort. The importance of each RNA modification for classification was assessed and ranked based on the mean decrease in accuracy (MDA) metric. The modifications m^7G, m^6A, Gm, $m^{2,2,7}G$, m^1A, $m^{2,2}G$, Im, and m^5U were identified as significant contributors to the model's predictive power (Figure 4B). We then tested the ML models in the validation cohort (Figure 4C). We figured out that only two modifications—m^6A and m^7G—are sufficient to classify the PCOS cases and the light gradient boosting machine model (LightGBM) achieved the highest predictive accuracy, with an area under the curve (AUC) of 0.9319 (Figure 4C; Table S4).

We then compared our optimized ML model with current clinical diagnostic criteria for PCOS. For clinical diagnosis, we applied criteria based on androgen excess (T>48.1 ng dL^{-1}), ovulatory dysfunction (menstrual cycle>35 days), and polycystic ovaries (FNPO≥12), similar to the Rotterdam Criteria, where a diagnosis of PCOS requires meeting any two of these three conditions. In differentiating PCOS cases from controls, our model using m^6A and m^7G achieved an AUC of 1.0000, surpassing the AUC of 0.8462 observed for clinical criteria (Figure 4D), indicating superior predictive accuracy for RNA modifications in this context. For distinguishing PCOS from POI, the RNA modification signature yielded an AUC of 0.8529, which was comparable to the clinical criteria's AUC of 0.8824. However, in differentiating PCOS from non-PCOS patients, the model showed robust performance with an AUC of 0.9424, as compared with the clinical criteria's AUC of 0.8824. These results suggest that our optimized ML model, incorporating m^6A and m^7G RNA modifications, offers significant predictive accuracy for PCOS detection and may serve as a valuable biomarker approach for clinical diagnosis.

DISCUSSION

In this study, using high-performance liquid chromatography coupled with a mass spectrometry (HPLC-MS) platform, which provides both qualitative and quantitative information on multiple RNA modifications from a single sample, we identified specific RNA modification expression patterns that exhibited intrinsic correlations and were associated with different ovarian conditions. After determining that these blood RNA modification signatures could predict clinical features of PCOS, we further identified an efficient RNA modification signature in peripheral blood—comprising m^6A and m^7G—that could effectively distinguish PCOS cases from POI patients and controls using a random

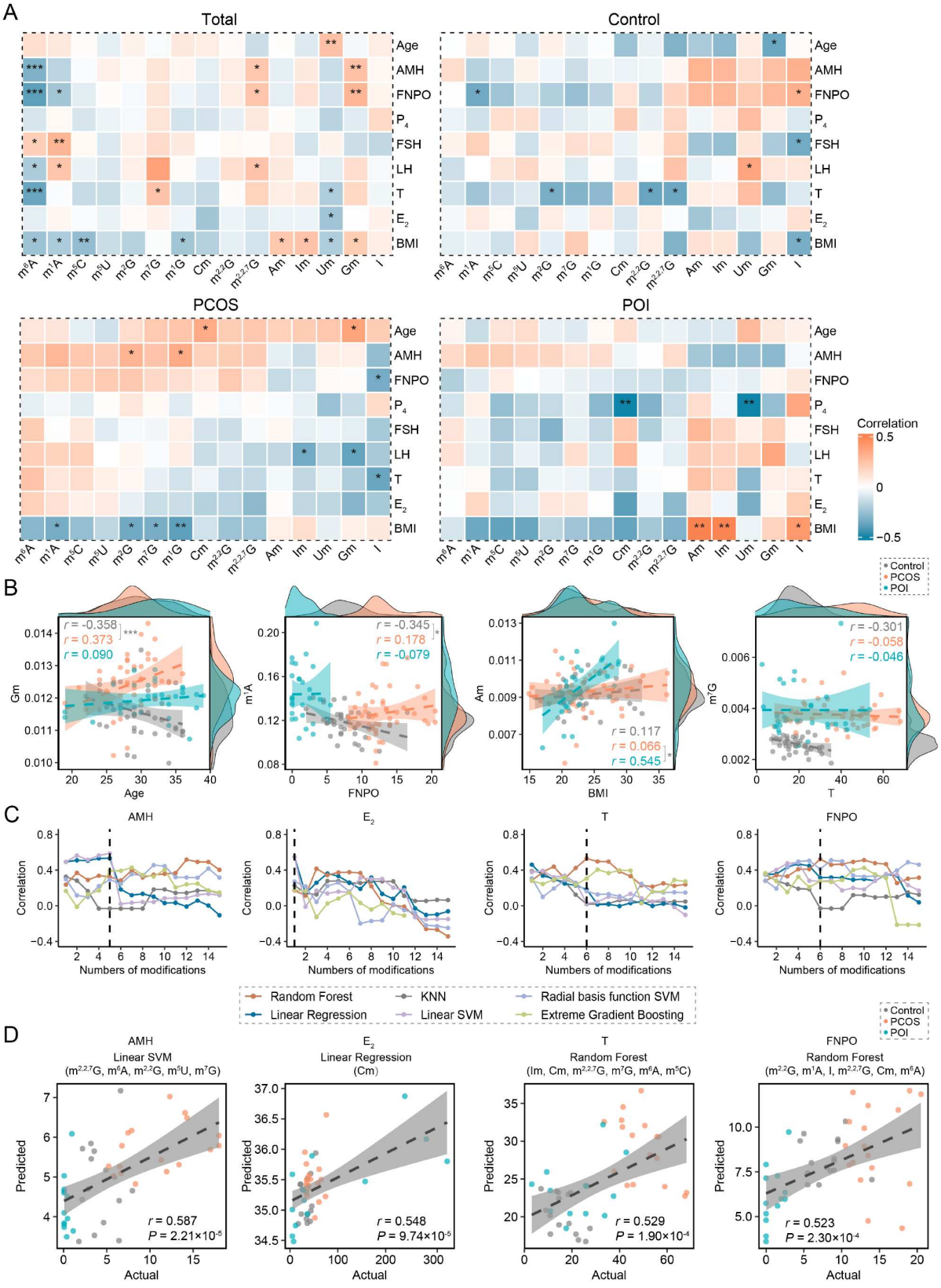

Figure 3. Correlation analysis between RNA modifications and clinical features. A, Correlation matrices displaying the pairwise correlations between 15 RNA modifications and various clinical features for Control, PCOS, and POI groups, as well as for all samples combined. B, The correlations between selected RNA modifications and specific clinical features. C, Correlations between actual and predicted values of AMH, E_2, T, and FNPO using six ML models, with RNA modifications incrementally ranked by feature importance. D, Scatter plots comparing actual and predicted values of AMH, E_2, T, and FNPO in the validation cohort, based on the most predictive RNA modification signatures identified by the best-performing ML models. *, $P \leqslant 0.05$; **, $P \leqslant 0.01$, ***, $P \leqslant 0.001$.

https://doi.org/10.1007/s11427-024-2913-7

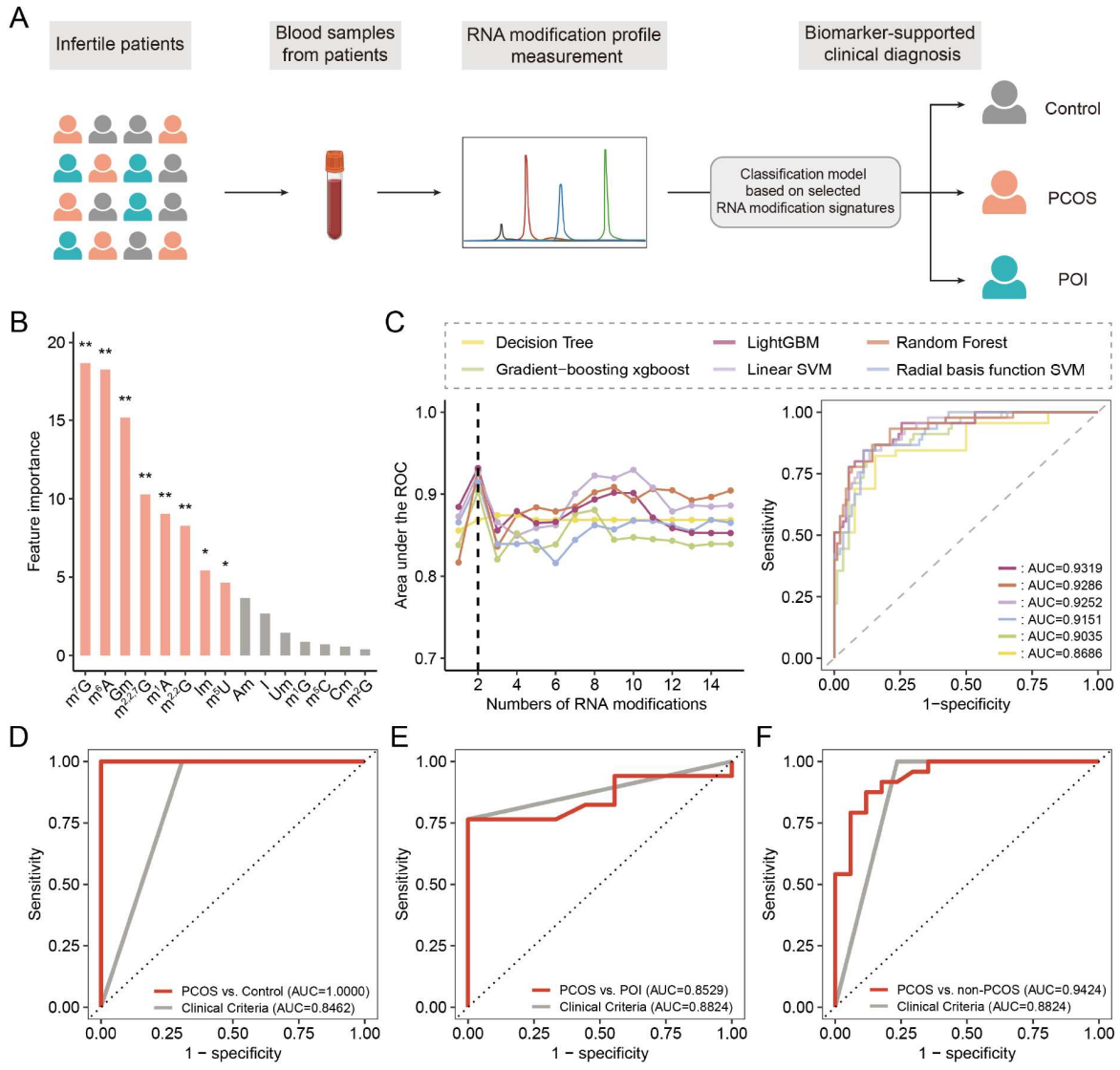

Figure 4. Evaluation of RNA modifications in predicting PCOS. A, Schematic overview illustrating the strategy to identify non-invasive biomarkers for PCOS diagnosis using RNA modification profiles from peripheral blood samples. B, Ranked feature importance of RNA modifications in classifying PCOS within the discovery cohort. C, Left: AUC values of the top six best-performing ML models, organized according to feature importance of RNA modifications; right: ROC curve comparison of the top six best-performing models with RNA modifications m^6A and m^7G. D–F, Comparison of ROC curves of the best-performing model against traditional clinical criteria for identifying PCOS in various comparisons: (D) PCOS *vs.* Control, (E) PCOS *vs.* POI, and (F) PCOS *vs.* non-PCOS samples. *, $P \leqslant 0.05$; **, $P \leqslant 0.01$, ***, $P \leqslant 0.001$.

forest model. These findings suggest the potential for RNA modification signatures to serve as non-invasive biomarkers for the diagnosis of clinical reproductive diseases.

PCOS and POI are the most common ovulatory disorders affecting fertility, and patients with these conditions often present with oligomenorrhea or amenorrhea, necessitating additional clinical tests for proper diagnosis. While serum FSH levels are commonly used for POI diagnosis (European Society for Human et al., 2016), the classification and diagnosis of PCOS remain more challenging. RNA chemical modifications, which alter RNA expression, splicing, stability, structure, and function (Delaunay et al., 2024), have emerged as an additional layer of biological information beyond the RNA sequence itself (Zhang et al., 2018). Since non-coding RNAs play critical roles in both physiological and pathological processes (Esteller, 2011), it is not surprising that RNA modifications have been found to change dynamically in response to various human diseases (Delaunay et al., 2024; Jonkhout et al., 2017; Zhang et al., 2020b). In this study, we identified specific blood RNA modification signatures—particularly m^6A and m^7G—that distinguish PCOS cases from POI and control subjects, thus enhancing our understanding of RNA modifications and their potential role in these disorders.

The origin of these circulating RNA modifications in PCOS raises intriguing questions. Previous studies have shown that circulating non-coding RNAs involved in diseases such as cardiovascular disease (E et al., 2018), kidney disease (van Zonneveld et al., 2023), and cancers (Mugoni et al., 2022; Zen and Zhang, 2012), are predominantly secreted or released from tissues (Barth et al., 2020). This leads us to hypothesize that the observed changes in circulating RNA modifications, especially in PCOS, may partly originate from the ovaries. For example, we found that blood m^6A levels were significantly reduced in PCOS

patients, which is consistent with previous reports linking m^6A to ovarian function and disease regulation. Differential m^6A modification of FOXO3 mRNA has been observed in luteinized granulosa cells (GCs) of non-obese PCOS patients following controlled ovarian hyperstimulation (Zhang et al., 2020a). Furthermore, polymorphisms in the m^6A demethylase *FTO* gene have been associated with PCOS susceptibility and hyperandrogenism (Branavan et al., 2020; Chaudhary et al., 2023; Song et al., 2014), suggesting that ovarian m^6A modifications may differ in PCOS patients. m^6A and its regulatory enzymes undergo dynamic changes during ovarian development (Sun et al., 2020), follicle development (Sun et al., 2020; Xia et al., 2018), oocyte maturation (Sui et al., 2020; Yao et al., 2023) and ovarian cancer (Liu et al., 2020) further supporting their potential involvement in ovarian diseases.

In addition to m^6A, m^7G emerged as a key modification distinguishing PCOS cases. m^7G methylation is a common modification found in the 5′-cap of mRNAs and in tRNAs, catalyzed by different enzymes. Recent research has shown that METTL1-mediated m^7G modification in tRNAs plays a role in aging regulation (Fu et al., 2024), and METTL1 has also been implicated in prognostic models for ovarian cancer classification (Zheng et al., 2022). This suggests that METTL1-mediated m^7G modification may also influence ovarian function. Our findings highlight m^7G as a characteristic modification in PCOS, indicating that alterations in m^7G and its associated enzymes, such as METTL1, may occur in the ovaries of PCOS patients and be detectable in circulating RNA modifications. If confirmed, these RNA modifications and their regulators could represent novel therapeutic targets for PCOS.

Our findings suggest that circulating RNA modification signatures have the potential to predict key features of PCOS, particularly follicle numbers. Quantitative histological studies of human ovaries have shown that the total number of follicles decreases throughout life, making follicle count an important biomarker for assessing ovarian conditions (Faddy and Gosden, 1995). Currently, ultrasonography is the primary method for detecting follicles in clinical practice. However, ultrasound has limitations; for example, it can only identify sufficiently large follicles, often missing smaller ones. Additionally, variability in operator expertise and differences in the quality of ultrasound machines can introduce challenges and inconsistencies in reporting (Sujata and Swoyam, 2018). Meanwhile, FNPO is also a critical reproductive health indicator, routinely assessed not only for diagnosing PCOS and POI but also for evaluating overall ovarian function. Thus, the lack of simple biological markers linked to follicle numbers limits the ability to achieve accurate and rapid detection. In our study, we found that blood RNA modifications could predict FNPO across all three groups, suggesting a novel avenue for improving clinical assessments. However, the underlying mechanisms connecting RNA modifications to follicle numbers remain unclear and warrant further investigation.

Additionally, we observed a strong intrinsic association among RNA modifications in peripheral blood across our samples, similar to recent findings in mice (Guo et al., 2023) and human sperm (Guo et al., 2022). The origins of these correlations remain unclear, but they may arise from several factors, including shared, sequential, or opposing enzymatic regulation, or the presence of RNA species that carry specific modifications. For instance, under certain conditions, m^1A can convert to m^6A through the Dimroth rearrangement (Macon and Wolfenden, 1968), suggesting a potential negative correlation between these two modifications. Conversely, m^6A reader YTHDF2 can also bind to the m^1A reader HRSP12 (Boo et al., 2022), and YTHDF2 has been shown to interact with m^1A sites as well. This cross-talk between m^6A and m^1A regulators, such as YTHDF2 and ALKBH3, highlights the possibility of a positive correlation between these modifications depending on the context. Therefore, various regulatory mechanisms may influence the correlation patterns of m^1A and m^6A, which could shift under different pathological conditions. In our study, we found that some correlation patterns between RNA modifications were altered in PCOS and POI groups compared with control subjects, indicating that specific regulatory pathways may be affected in these diseases. Although the exact mechanisms underlying these correlations are not fully understood, the inherent relationships among RNA modifications add a layer of complexity and provide valuable information that could be harnessed for diagnostic purposes.

Although the RNA modification detection method used in this study requires only a small blood volume and offers greater accuracy and efficiency compared with traditional PCOS analyses, it has certain limitations. The preprocessing steps, including blood RNA extraction and enzymatic digestion into mononucleotides, are relatively time-consuming and demand specialized molecular biology expertise. Moreover, it is well-established that white blood cells (WBCs), which constitute approximately 1% of blood composition, contribute the majority of RNA content in whole blood (Jiang et al., 2013). Therefore, the presence of immune abnormalities induced by other diseases may alter the proportions of WBCs, potentially affecting RNA modification profiles and influencing diagnostic outcomes. Additionally, PCOS patients have been reported to exhibit increased numbers of total WBCs, lymphocytes, and neutrophils (Agapova et al., 2014). Consequently, the observed changes in m^6A and m^7G modifications may reflect a combination of ovary-derived and immune-related alterations. Thus, further investigation is needed to refine this method and clarify the specific contributions of these changes.

In recent years, advancements in RNA modification detection technologies have uncovered the regulatory roles of RNA modifications in human diseases, marking the beginning of our understanding of this field. The insights gained from our study highlight the relevance of RNA modifications in detecting conditions related to pregnancy and highlight their involvement in the pathophysiology of disorders such as PCOS. In addition to our current omics approach for detecting RNA modifications, recently developed methods, such as Nanopore-based direct RNA sequencing (Lucas et al., 2024) and, advanced mass spectrometry-based method techniques (Yuan et al., 2024), are paving the way for single-molecule level detection and precise mapping of these modifications. Nanopore sequencing, for example, offers the advantage of directly reading RNA modifications, although it currently faces challenges related to read accuracy and requires a high RNA input. On the other hand, advanced mass spectrometry provides high-sensitivity quantification but is limited by low throughput, complex data processing, and the need for specialized instrumentation (Shi et al., 2022). We anticipate that as these technologies evolve, their integration will enhance our ability to pinpoint modification sites and unravel their functional roles in disease processes.

https://doi.org/10.1007/s11427-024-2913-7

MATERIALS AND METHODS

Cohort of study

In this clinical cohort study at the Center for Reproductive Medicine, Shandong University, China, we collected a total of 122 human peripheral blood samples, comprising 67 samples in the discovery cohort and 55 in the validation cohort from August 2019 to September 2019. Informed consent was obtained from all individual participants included in the study. According to Rotterdam Criteria (Teede et al., 2023), PCOS was diagnosed meeting two of the following three criteria: polycystic ovaries; clinical and/or biochemical hyperandrogenism; oligo- or anovulation, the menstrual cycle length over 35 days, excluding other causes of oligomenorrhea or hyperandrogenism, such as Cushing's syndrome. POI patients showed oligomenorrhea or menopause before 40 years of age, with elevated FSH (>25 IU L^{-1}). Women with ovarian surgery, chemo/radiotherapy, or known autoimmune disease were excluded. Control participants were selected based on their history of regular menstrual cycles and absence of PCOS features, such as clinical or biochemical hyperandrogenism.

Ribonucleotides used as LC-MS/MS standards

2′-O-methylguanosine (Gm), *2′-O*-methyluridine (Um), and inosine (I) were purchased from Berry & Associates (USA). 1-methylguanosine (m^1G), N^2-methylguanosine (m^2G), $N^{2,2,7}$-trimethylguanosine ($m^{2,2,7}G$), N^2,N^2-dimethylguanosine ($m^{2,2}G$) and *2′-O*-methylinosine (Im) were purchased from Carbosynth. Cytidine (C), adenosine (A), guanosine (G), uridine (U), *2′-O*-methylcytidine (Cm), *2′-O*-methyladenosine (Am), 5-methyladenosine (m^5C), N^1-methyladenosine (m^1A), 5-methyluridine (m^5U) and N^7-methylguanosine (m^7G) were purchased from Sigma-Aldrich (USA). N^6-methyladenosine (m^6A) was purchased from SelleckChem (USA).

Data collection

Peripheral blood was collected from the veins of PCOS cases, POI cases, and controls. After adding Trizol LS, the samples were stored at −80°C. RNA extraction and a single-blind RNA modification detection in a high-throughput LC-MS/MS approach were performed as previously described (Chen et al., 2016; Zhang et al., 2022; Zhang et al., 2018), with the amount of blood used for each individual corresponding to 0.5 mL. In short, purified RNAs were digested into single ribonucleotides with 1 U benzonase nuclease (Sigma-Aldrich), 0.05 U phosphodiesterase I (Sigma-Aldrich) and 0.5 U alkaline phosphatase (Sigma-Aldrich) at 37°C for 3 h, and redundant enzymes in the digestion mixture were removed by centrifugation with a Nanosep 3K device with Omega membrane (Sigma-Aldrich). Mass spectrometry analysis was performed on a ThermoFisher TSQ Vantage Quadrupole mass spectrometer, which was coupled with an Agilent 1200 HPLC system and equipped with an electrospray ionization source. The MS system was set to run in positive ion mode, utilizing a multiple reaction monitoring (MRM) scan model. Data from LC-MS/MS were acquired and subsequently processed with the Thermo Xcalibur™ mass spectrometry data system for the quantification of modified ribonucleotide concentrations. To minimize or eliminate the impact of variation in sample loading, the calculated percentages of each modified ribonucleotide were standardized against the total quantified ribonucleotides that share an identical nucleobase. For instance, the percentage of m^5C was computed by dividing its molar concentration by the total molar concentration of m^5C, Cm, and C.

Establishment of predictive models for clinical features using peripheral blood RNA modifications

A total of 15 peripheral blood RNA modifications were considered to develop the prediction models. The importance of each RNA modification was assessed and ranked using the correlation of the actual clinical features and predicted clinical features that were predicted by each RNA modification. Six different ML regression models, including linear regression, linear support vector machine (linear SVM), radial basis function support vector machine (radial basis function SVM), random forest, extreme gradient boosting (XGBoost), and *k*-nearest neighbors (KNN) were utilized to develop the RNA modification signature for clinical features prediction. The clinical features include age, BMI, T, luteinizing hormone (LH), progesterone (P_4), FSH, estradiol (E_2), AMH, and FNPO by using the data from the discovery cohort.

Establishment of predictive models for PCOS classification using peripheral blood RNA modifications

A total of 15 peripheral blood RNA modifications were considered to develop the prediction models. The feature importance of each RNA modification for PCOS classification was ranked using the MDA metric. Nine different ML classification models, including logistic regression, naive Bayes, linear support vector machine (linear SVM), radial basis function support vector machine (radial basis function SVM), decision tree, random forest, light gradient boosting machine (LightGBM), gradient-boosting xgboost, and *k*-nearest neighbors (KNN) were utilized to develop the RNA modification signature for predicting PCOS using data from the discovery cohort.

Statistical analysis

All the statistical analyses were conducted using R (version 4.3.3). The correlation test, *t*-test, and Wilcoxon test were performed using the "cor.test", "t.test", and "wilcox.test" functions, respectively. The PCA was performed using the "FactoMineR" and "factoextra" packages. The permutation test for feature importance evaluation was performed using the "rfPermute" package. A Kolmogorov-Smirnov (KS) test was applied to assess differences in RNA modification distribution. The ML models were established using the "caret", "tidymodels", "neuralnet", and "keras" packages. The predictive power of the models was evaluated by ROC curves with the "multipleROC" package. Hierarchical clustering was performed using the "complete" method with "Euclidean" distance using the "ComplexHeatmap" package. The correlation analyses between actual and predicted clinical features were conducted using Pearson's correlation test to determine the correlation coefficient (r). Statistical significance was set at a two-tailed P value$\leqslant 0.05$.

Compliance and ethics

The authors declare that they have no conflict of interest. This study was approved by the Ethics Committee of the Center for Reproductive Medicine, Shandong University (No. 2019-19).

Acknowledgement

This work was supported by the National Natural Science Foundation of China (92357306, 82122027, 32171110, 32370596, 82071606, 82421004, 32588201, 32370916), the National Key Research and Development Program of China (2019YFA0802600, 2021YFC2700400), Strategic Priority Research Program of the Chinese Academy of Sciences (XDA0460302), and Shandong Provincial Key Research and Development Program (2024CXPT087). Figures 1A and 4A were created with BioRender.com.

Supporting information

The supporting information is available online at https://doi.org/10.1007/s11427-024-2913-7. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

Agapova, S., Cameo, T., Sopher, A., and Oberfield, S. (2014). Diagnosis and challenges of polycystic ovary syndrome in adolescence. Semin Reprod Med 32, 194–201.

Álvarez-Blasco, F., Botella-Carretero, J.I., San Millán, J.L., and Escobar-Morreale, H.F. (2006). Prevalence and characteristics of the polycystic ovary syndrome in overweight and obese women. Arch Intern Med 166, 2081–2086.

Azziz, R., Carmina, E., Chen, Z.J., Dunaif, A., Laven, J.S.E., Legro, R.S., Lizneva, D., Natterson-Horowtiz, B., Teede, H.J., and Yildiz, B.O. (2016). Polycystic ovary syndrome. Nat Rev Dis Primers 2, 16057.

Barbieri, I., and Kouzarides, T. (2020). Role of RNA modifications in cancer. Nat Rev Cancer 20, 303–322.

Barth, D.A., Drula, R., Ott, L., Fabris, L., Slaby, O., Calin, G.A., and Pichler, M. (2020). Circulating non-coding RNAs in renal cell carcinoma—pathogenesis and potential implications as clinical biomarkers. Front Cell Dev Biol 8, 828.

Boo, S.H., Ha, H., and Kim, Y.K. (2022). m1A and m6A modifications function cooperatively to facilitate rapid mRNA degradation. Cell Rep 40, 111317.

Borrebaeck, C.A.K. (2017). Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. Nat Rev Cancer 17, 199–204.

Branavan, U., Wijesundera, S., Chandrasekaran, V., Arambepola, C., and Wijeyaratne, C. (2020). In depth analysis of the association of FTO SNP (rs9939609) with the expression of classical phenotype of PCOS: a Sri Lankan study. BMC Med Genet 21, 30.

Carson, S.A., and Kallen, A.N. (2021). Diagnosis and management of infertility: a review. JAMA 326, 65–76.

Chaudhary, H., Patel, J., Jain, N.K., Panchal, S., Laddha, N., and Joshi, R. (2023). Association of *FTO* gene variant rs9939609 with polycystic ovary syndrome from Gujarat, India. BMC Med Genomics 16, 216.

Chen, H., Zhao, X., Yang, W., Zhang, Q., Hao, R., Jiang, S., Han, H., Yu, Z., Xing, S., Feng, C., et al. (2023). RNA N^6-methyladenosine modification-based biomarkers for absorbed ionizing radiation dose estimation. Nat Commun 14, 6912.

Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., Feng, G., Peng, H., Zhang, X., Zhang, Y., et al. (2016). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science 351, 397–400.

Delaunay, S., Helm, M., and Frye, M. (2024). RNA modifications in physiology and disease: towards clinical applications. Nat Rev Genet 25, 104–122.

E, S., Costa, M.C., Kurc, S., Drożdż, A., Cortez-Dias, N., and Enguita, F.J. (2018). The circulating non-coding RNA landscape for biomarker research: lessons and prospects from cardiovascular diseases. Acta Pharmacol Sin 39, 1085–1099.

Escobar-Morreale, H.F. (2018). Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. Nat Rev Endocrinol 14, 270–284.

Esteller, M. (2011). Non-coding RNAs in human disease. Nat Rev Genet 12, 861–874.

European Society for Human, R., Embryology Guideline Group on, P.O.I., Webber, L., Davies, M., Anderson, R., Bartlett, J., Braat, D., Cartwright, B., Cifkova, R., de Muinck Keizer-Schrama, S., et al. (2016). ESHRE Guideline: management of women with premature ovarian insufficiency. Hum Reprod 31, 926–937.

Faddy, M.J., and Gosden, R.G. (1995). Physiology: a mathematical model of follicle dynamics in the human ovary. Hum Reprod 10, 770–775.

Fu, Y., Jiang, F., Zhang, X., Pan, Y., Xu, R., Liang, X., Wu, X., Li, X., Lin, K., Shi, R., et al. (2024). Perturbation of METTL1-mediated tRNA N^7-methylguanosine modification induces senescence and aging. Nat Commun 15, 5713.

Guo, H., Shen, X., Hu, H., Zhou, P., He, T., Xia, L., Tan, D., Zhang, X., and Zhang, Y. (2022). Alteration of RNA modification signature in human sperm correlates with sperm motility. Mol Hum Reprod 28, gaac031.

Guo, H., Xia, L., Wang, W., Xu, W., Shen, X., Wu, X., He, T., Jiang, X., Xu, Y., Zhao, P., et al. (2023). Hypoxia induces alterations in tRNA modifications involved in translational control. BMC Biol 21, 39.

Jiang, Z., Uboh, C.E., Chen, J., and Soma, L.R. (2013). Isolation of RNA from equine peripheral blood cells: comparison of methods. SpringerPlus 2, 478.

Jonkhout, N., Tran, J., Smith, M.A., Schonrock, N., Mattick, J.S., and Novoa, E.M. (2017). The RNA modification landscape in human disease. RNA 23, 1754–1769.

Liu, T., Wei, Q., Jin, J., Luo, Q., Liu, Y., Yang, Y., Cheng, C., Li, L., Pi, J., Si, Y., et al. (2020). The m^6A reader YTHDF1 promotes ovarian cancer progression via augmenting EIF3C translation. Nucleic Acids Res 48, 3816–3831.

Lucas, M.C., Pryszcz, L.P., Medina, R., Milenkovic, I., Camacho, N., Marchand, V., Motorin, Y., Ribas de Pouplana, L., and Novoa, E.M. (2024). Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. Nat Biotechnol 42, 72–86.

Macon, J.B., and Wolfenden, R. (1968). 1-Methyladenosine. Dimroth rearrangement and reversible reduction. Biochemistry 7, 3453–3458.

Mugoni, V., Ciani, Y., Nardella, C., and Demichelis, F. (2022). Circulating RNAs in prostate cancer patients. Cancer Lett 524, 57–69.

Natalia, A., Zhang, L., Sundah, N.R., Zhang, Y., and Shao, H. (2023). Analytical device miniaturization for the detection of circulating biomarkers. Nat Rev Bioeng 1, 481–498.

Rosenfield, R.L., and Ehrmann, D.A. (2016). The pathogenesis of polycystic ovary syndrome (PCOS): the hypothesis of PCOS as functional ovarian hyperandrogenism revisited. Endocrine Rev 37, 467–520.

Schwarzenbach, H., Hoon, D.S.B., and Pantel, K. (2011). Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer 11, 426–437.

Shi, J., Zhou, T., and Chen, Q. (2022). Exploring the expanding universe of small RNAs. Nat Cell Biol 24, 415–423.

Song, D.K., Lee, H., Oh, J.Y., Hong, Y.S., and Sung, Y.A. (2014). *FTO* gene variants are associated with PCOS susceptibility and hyperandrogenemia in young Korean women. Diabetes Metab J 38, 302–310.

Sonigo, C., Jankowski, S., Yoo, O., Trassard, O., Bousquet, N., Grynberg, M., Beau, I., and Binart, N. (2018). High-throughput ovarian follicle counting by an innovative deep learning approach. Sci Rep 8, 13499.

Stuenkel, C.A., and Gompel, A. (2023). Primary ovarian insufficiency. N Engl J Med 388, 154–163.

Sui, X., Hu, Y., Ren, C., Cao, Q., Zhou, S., Cao, Y., Li, M., Shu, W., and Huo, R. (2020). METTL3-mediated m^6A is required for murine oocyte maturation and maternal-to-zygotic transition. Cell Cycle 19, 391–404.

Sujata, K., and Swoyam, S. (2018). 2D and 3D trans-vaginal sonography to determine cut-offs for ovarian volume and follicle number per ovary for diagnosis of polycystic ovary syndrome in Indian women. J Reprod Infertil 19, 146–151.

Sullenger, B.A., and Gilboa, E. (2002). Emerging clinical applications of RNA. Nature 418, 252–258.

Sun, X., Zhang, J., Jia, Y., Shen, W., and Cao, H. (2020). Characterization of m^6A in mouse ovary and testis. Clin Transl Med 10, e141.

Teede, H., Misso, M., Tassone, E.C., Dewailly, D., Ng, E.H., Azziz, R., Norman, R.J., Andersen, M., Franks, S., Hoeger, K., et al. (2019). Anti-Müllerian hormone in PCOS: a review informing international guidelines. Trends Endocrinol Metab 30, 467–478.

Teede, H.J., Tay, C.T., Laven, J.J.E., Dokras, A., Moran, L.J., Piltonen, T.T., Costello, M. F., Boivin, J., Redman, L.M., Boyle, J.A., et al. (2023). Recommendations from the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome. J Clin Endocrinol Metab 108, 2447–2469.

Thompson, A.G., Gray, E., Heman-Ackah, S.M., Mäger, I., Talbot, K., Andaloussi, S.E., Wood, M.J., and Turner, M.R. (2016). Extracellular vesicles in neurodegenerative disease—pathogenesis to biomarkers. Nat Rev Neurol 12, 346–357.

van Zonneveld, A.J., Zhao, Q., Rotmans, J.I., and Bijkerk, R. (2023). Circulating non-coding RNAs in chronic kidney disease and its complications. Nat Rev Nephrol 19, 573–586.

Wang, H., Zuo, H., Liu, J., Wen, F., Gao, Y., Zhu, X., Liu, B., Xiao, F., Wang, W., Huang, G., et al. (2018). Loss of YTHDF2-mediated m^6A-dependent mRNA clearance facilitates hematopoietic stem cell regeneration. Cell Res 28, 1035–1038.

Wild, R.A. (2002). Long-term health consequences of PCOS. Hum Reprod Update 8, 231–241.

Xia, H., Zhong, C., Wu, X., Chen, J., Tao, B., Xia, X., Shi, M., Zhu, Z., Trudeau, V.L., and Hu, W. (2018). *Mettl3* mutation disrupts gamete maturation and reduces fertility in zebrafish. Genetics 208, 729–743.

Xiao, H., Fan, X., Zhang, R., and Wu, G. (2021). Upregulated N^6-methyladenosine RNA in peripheral blood: potential diagnostic biomarker for breast cancer. Cancer Res Treat 53, 399–408.

Yao, H., Gao, C.C., Zhang, D., Xu, J., Song, G., Fan, X., Liang, D.B., Chen, Y.S., Li, Q., Guo, Y., et al. (2023). scm^{6}A-seq reveals single-cell landscapes of the dynamic m^6A during oocyte maturation and early embryonic development. Nat Commun 14, 315.

Yin, R., Chang, J., Li, Y., Gao, Z., Qiu, Q., Wang, Q., Han, G., Chai, J., Feng, M., Wang, P., et al. (2022). Differential m^6A RNA landscapes across hematopoiesis reveal a role for IGF2BP2 in preserving hematopoietic stem cell function. Cell Stem Cell 29,

149–159.e7.
Yuan, X., Su, Y., Johnson, B., Kirchner, M., Zhang, X., Xu, S., Jiang, S., Wu, J., Shi, S., Russo, J.J., et al. (2024). Mass spectrometry-based direct sequencing of tRNAs *de novo* and quantitative mapping of multiple RNA modifications. J Am Chem Soc 146, 25600–25613.
Zen, K., and Zhang, C. (2012). Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Med Res Rev 32, 326–348.
Zhang, C., Chen, Y., Sun, B., Wang, L., Yang, Y., Ma, D., Lv, J., Heng, J., Ding, Y., Xue, Y., et al. (2017). m^{6}A modulates haematopoietic stem and progenitor cell specification. Nature 549, 273–276.
Zhang, L., Li, Y., Wang, J., Hong, C., Lu, W., Qu, Y., Black, S.M., Chen, Q., Zhou, T., Jin, W., et al. (2022). RNA modification signature of peripheral blood as a potential diagnostic marker for pulmonary hypertension. Hypertension 79, e67–e69.
Zhang, S., Deng, W., Liu, Q., Wang, P., Yang, W., and Ni, W. (2020a). Altered m^{6}A modification is involved in up-regulated expression of *FOXO3* in luteinized granulosa cells of non-obese polycystic ovary syndrome patients. J Cell Mol Med 24, 11874–11882.
Zhang, Y., Zhang, X., Shi, J., Tuorto, F., Li, X., Liu, Y., Liebers, R., Zhang, L., Qu, Y., Qian, J., et al. (2018). Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. Nat Cell Biol 20, 535–540.
Zhang, Z., Luo, K., Zou, Z., Qiu, M., Tian, J., Sieh, L., Shi, H., Zou, Y., Wang, G., Morrison, J., et al. (2020b). Genetic analyses support the contribution of mRNA N^6-methyladenosine (m^{6}A) modification to human disease heritability. Nat Genet 52, 939–949.
Zheng, P., Li, N., and Zhan, X. (2022). Ovarian cancer subtypes based on the regulatory genes of RNA modifications: novel prediction model of prognosis. Front Endocrinol 13, 972341.