

Detection of early-onset severe preeclampsia by cell-free DNA fragmentome

Running Head: Plasma DNA predict preeclampsia

Haiqiang Zhang^{1†}, Longwei Qiao^{2†}, Xintao Hu^{1,3}, Chunhua Zhang², Yu Lin¹, Jingyu Zhao⁴, Xiaojuan Wu², Xiaoyan Song², Hui Tang², Ying Xue², Yang Sun¹, Rijing Ou¹, Xinxin Wang¹, Yan Zhang¹, Xin Jin^{1*}, Ting Wang^{2,6*}

¹BGI Research, Shenzhen 518083, China

²Center for Reproduction and Genetics, School of Gusu, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Nanjing Medical University, Suzhou, China

³State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang, 550025, China

⁴Clin Lab, BGI Genomics, Nanjing 210019, China

*Correspondence: biowt@njmu.edu.cn (T.W.), jinxin@genomics.cn (X.J.)

[†]These authors contributed equally

22 Acknowledgement

23 We would like to thank China National GeneBank (CNGB) for the technical
24 support.

25

26

27 Abstract

28 Early-onset severe preeclampsia (EO-PE) is a distinct and highly consequential form
29 of preeclampsia (PE), presenting significant challenges for early detection. Here, we
30 investigated the fragmentation pattern of plasma cell-free DNA (cfDNA) in EO-PE
31 patients. We uncovered that the nucleotide composition at the 5' end (i.e. ends motif)
32 of plasma cfDNA showed a unique pathological preference in EO-PE pregnancies and
33 gestational-psychology preference in healthy pregnancies. By integrating 91 EO-PE
34 specific motifs into a machine-learning model, we achieved accurate prediction of
35 EO-PE development in early pregnancies. Remarkably, our model demonstrated
36 robust performance in an independent cohort of 74 early pregnancies and 1,241
37 non-invasive prenatal testing (NIPT) samples with ultra-low sequencing depth.
38 Additionally, we revealed that these PE-specific motif signatures lacked tissue
39 specificity, originating extracellularly, and were associated with the abnormal
40 concentration of DNA fragmentation factor subunit beta (DFFB) in EO-PE patients'
41 plasma. These findings establish the plasma DNA fragmentome as a non-invasive and

42 cost-effective biomarker that can be simultaneously captured during NIPT for early EO-PE
43 detection and provide valuable insights into cfDNA production mechanisms in
44 preeclampsia patients.

45

46 **Keywords:** plasma DNA; end motif; non-invasive biomarker; pregnancy
47 complication

48

49

50 **Introduction**

51 Preeclampsia (PE) is one of the most serious pregnancy-specific complications
52 and the leading cause of maternal and perinatal morbidity and mortality worldwide(1,
53 2). This multisystem maternal syndrome is characterized by new-onset hypertension
54 and maternal organ damage that occurs after 20 weeks of gestation(3). Among the
55 various forms of PE, early-onset preeclampsia, occurring before 34 weeks of gestation,
56 represents the most severe manifestation, associated with a higher incidence of
57 maternal complications, perinatal mortality, and morbidity compared to late-onset
58 PE(4, 5). However, the complex etiology, diverse clinical manifestations, and
59 pathophysiology of PE present significant challenges in terms of prediction,
60 management, and prevention(6–8). Despite extensive efforts, clinically satisfactory
61 biomarkers and prediction models for early detection of PE are lacking, primarily due
62 to limited clinical availability and low predictive efficacy(9, 10). Nevertheless, recent

63 advancements in the field of circulating cell-free DNA (cfDNA) provide new
64 opportunities for developing effective PE biomarkers. Promising results have shown
65 that cfDNA concentration is elevated in PE and correlates with disease severity(11,
66 12). Moreover, disparities in cfDNA coverage across promoter regions have been
67 observed between PE and healthy pregnancies(13). However, further in-depth
68 investigations and verifications are required to assess the applicability of cfDNA in
69 early PE detection.

70 Recently, the study of cfDNA fragmentation patterns, known as
71 "fragmentomics," is an emerging field with the potential to advance our
72 understanding of cfDNA's biology and develop new biomarkers for clinical
73 applications(14). In addition to fragment size(15), recent advances in this new "omics"
74 area have identified several crucial characteristics of cfDNA fragments, including
75 nucleosome footprints(16, 17), fragment end motifs(18, 19), preferred ends(20, 21),
76 jagged ends(22), cleavage profile(23) and coverage/end imbalance in regulatory
77 elements(24, 25). These characteristics are believed to preserve valuable information
78 about the tissue of origin of cfDNA(26) For instance, the preferred ends of short and
79 long plasma cfDNA are related to fetal and maternal origin(27). The presence of
80 distinct characteristic end motifs in plasma cfDNA derived from cancers, placenta,
81 and hematopoietic cells enables the monitoring of specific pathophysiological
82 condition(28) Therefore, cfDNA fragmentomics shows great promise in clinical
83 applications, particularly for noninvasive early disease detection(29, 30).

84 In this study, we present a comprehensive analysis and comparison of end motif
85 patterns in plasma DNA from pregnancies with early-onset severe preeclampsia and
86 healthy pregnancies. Our investigation focuses on identifying specific sets of motifs
87 associated with PE pathology and gestational physiology. To validate their potential
88 for early PE detection, we assess the performance of these PE-specific motifs, and
89 notably, our findings demonstrate robust performance even with low-coverage
90 non-invasive prenatal testing (NIPT) data. Additionally, we explore the tissue origin
91 of the PE-specific motif signatures, examining their relationship with cfDNA size,
92 nucleosomal structure, and the concentrations of nucleases involved in cfDNA
93 digestion in both EO-PE and non-EO-PE pregnancies.

94

95 **Results**

96 **Characteristics of pregnancies in the discovery cohort**

97 For the discovery cohort (SZ-1 cohort), we collected a total of 151 plasma
 98 samples from 140 pregnant women with early-onset PE (N=79) and without PE
 99 (N=61) at different gestational ages. Specifically, one set of plasma samples was
 100 obtained before labor during the third trimester: 16 samples were collected at the time
 101 of diagnosis of early-onset severe PE (mean gestational age, 29.9 weeks) (PT3), and 8
 102 samples were collected before labor from healthy control women (mean gestational
 103 age, 39.7 weeks) (NT3). Another set of plasma samples was obtained during the
 104 second trimester: 71 samples were collected before the diagnosis of early-onset severe
 105 PE (mean gestational age, 17.6 weeks) (PT2), and 56 samples were collected from
 106 gestational age-matched healthy control women (mean gestational age, 19.0 weeks)
 107 (NT2). Among these subjects, 8 PE and 3 healthy pregnancies contributed plasma
 108 samples at both the second and third trimesters. (Figure 1A).

109 The demographic and clinical characteristics of all subjects are shown in Table 1.
 110 The maternal age, BMI, times of pregnancies and delivery were comparable between
 111 PE and control subjects in both 2nd and 3rd trimesters. Compared with control subjects
 112 (mean: 39.7 weeks), PE patients exhibited significantly lower gestational age in
 113 groups of 3rd trimester (mean: 29.9 weeks) (P -value<0.0001, Mann-Whitney U test).
 114 The mean gestational age of PE onset in subjects collected at the 2nd and 3rd trimesters

were 30.0 and 29.7, respectively. The maternal plasma cfDNA of group NT2, PT2, NT3, and PT3 was sequenced to a mean of 5.46 (range: 3.06 - 8.85), 5.47 (range: 2.74 - 16.3), 7.94 (range: 5.46 - 10.19) and 10.31 (range: 7.61 - 13.42) haploid human genome coverage, respectively.

Ends motifs of plasma DNA specific to PE pathology and gestational physiology

The distribution of nucleotide string at the end of plasma DNA, referred to as ends motif, was reported to embody the cfDNA fragmentation footprint(18). Therefore, we wonder whether the ends motif could be used to decipher the disease signatures associated with preeclampsia in circulating cell-free DNA. Accordingly, we examined the diversity of 4mer ends motifs in the plasma DNA of both control and PE groups using the motif diversity score (MDS)(18). The results showed that the MDS values between the control (median: 0.946, range: 0.940-0.950) and preeclampsia subjects (median: 0.946, range: 0.934-0.953) were comparable during the second trimester (P -value=0.107; Mann-Whitney U-test). However, during the third trimester, the MDS was significantly reduced in preeclampsia samples (median: 0.943, range: 0.932-0.949) when compared to controls (median: 0.948, range: 0.947-0.950) (P -value=0.003; Mann-Whitney U-test; Figure 1B), implying that preeclampsia-associated ends motif signals might be enhanced as the onset of preeclampsia approaches.

Subsequently, through the unsupervised clustering, we examined the distribution

135 patterns of all 256 types of 4-mer motifs between the healthy and PE pregnancies in
136 the third trimester. As shown in Figure 1C, EO-PE and healthy samples could be
137 readily separated into two distinct clusters, suggesting the profiles of cfDNA ends
138 motifs were highly different between pregnancies with and without EO-PE. Moreover,
139 we compared the motif frequencies between healthy and preeclampsia pregnancies at
140 different gestational ages, as well as between pregnancies during the second and third
141 trimesters in preeclampsia and healthy pregnancies. We identified four sets of motifs
142 ($\text{Log}_{10}(\text{FC})$ ranked top and bottom 25% and Wilcox P -value <0.05): Set A includes 63
143 increased and 64 decreased motifs that exhibited a significant increase and decrease,
144 respectively, in PT3 relative to NT3; Set B includes 58 increased and 50 decreased
145 motifs that exhibited a significant increase and decrease, respectively, in PT2 relative
146 to NT2; Set C includes 63 increased and 62 decreased motifs that exhibited a
147 significant increase and decrease, respectively, in PT3 relative to PT2; and Set D
148 includes 56 increased and 45 decreased motifs that exhibited a significant increase
149 and decrease, respectively, in NT3 relative to NT2 (Figure 1D and Supplemental
150 Table 1). The top 10 increased and decreased motifs of each set are shown in
151 Supplementary Figure 1.

152 Notably, we observed that in both the second and third trimesters, the motifs that
153 were significantly elevated in preeclampsia samples predominantly began with A and
154 C (PT3: 95.24%, PT2: 96.55%), while significantly reduced motifs in preeclampsia
155 samples mainly started with T and G (PT3: 93.75%, PT2: 96%). Such pattern was

156 further pronounced in PT3 pregnancies compared to PT2 pregnancies (A% and C%
157 among increased motifs: 96.83%, T% and G% among decreased motifs: 91.94%),
158 thus demonstrating that the PE-specific motif signals become stronger as the onset of
159 the condition approaches (Figure 1E, F). In contrast, this PE-specific signal was
160 absent in healthy pregnancies between different gestational ages (Figure 1E).
161 Compared with NT2, we found that in the NT3 group, the motifs starting with A and
162 C only accounted for 57.14% of the significantly increased motifs, and the motifs
163 began with G and T only accounted for 48.89% of the significantly decreased motifs
164 (Figure 1F).

165 To quantify the signatures of preeclampsia in plasma DNA, motif ratios between
166 the frequencies of increased and decreased motifs of Set A were calculated in PE and
167 control pregnancies in both trimesters (See Methods). As shown in Figure 1G, the
168 motif ratios significantly increased from PT2 (median: 1.45) to PT3 (median: 2.1)
169 (P -value<0.0001, Mann-Whitney U test). Both PT3 and PT2 samples exhibited
170 significantly increased motif ratios relative to the healthy pregnancies in the same
171 trimester (median motif ratio of NT3: 1.31, P -value<0.0001; median motif ratio of
172 NT2: 1.34, P -value<0.0001, Mann-Whitney U test). These results indicated that the
173 PE-specific motifs were capable of reflecting the PE disease signals.

174 We further compared the consistency of differential motifs among different
175 comparison sets. According to Fig 1H-J, 48 increased and 43 decreased motifs were
176 co-present across Sets A, B and C, accounting for 71.65%, 84.26% and 72.8% of the

177 motifs in Sets A, B and C, respectively. In contrast, only 10 increased and 2 decreased
178 motifs co-existed in all sets, accounting for 9.45%, 11.11%, 9.6% and 11.88% of
179 motifs in Set A, B, C and D. This result suggests that the motifs related to
180 preeclampsia pathology were not shared in normal pregnancies. Moreover, 23
181 increased and 8 decreased motifs overlapped between Set C and Set D, which were
182 deemed motifs altered with gestational ages, accounting for 24.8% and 30.69% of
183 motifs in Set C, and D, respectively (Figure 1H-J). These shared motifs were mostly
184 gestational physiology-related signals that coexisted in both preeclampsia and healthy
185 samples.

186 **Correlation between gestational age and PE-specific motif abundance**

187 In the Discovery cohort, 3 pregnant women in the control group and 8 pregnant
188 women in the PE group contributed samples during both 2nd trimester and 3rd
189 trimesters. These samples allowed us to further investigate how the PE-specific motif
190 signals (i.e. frequencies of motifs in set A) changed as the pregnancy progressed. As
191 shown in Figure 2A, most of the PE samples displayed enhanced trends for increased
192 and decreased Set A PE-specific motifs in PT3 compared with PT2. While the
193 PE-specific motif frequencies remained consistent between NT3 and NT2 samples
194 (Figure 2B), indicating the absence of PE signals in healthy pregnancies. In order to
195 quantify the enhancement of motif patterns, we calculated the accumulated deviations
196 between motif frequencies in samples collected from 2nd and 3rd trimesters for each

197 pregnant woman. We termed this value as “motif distance”. According to Fig 2C, the
198 PE samples showed dramatically higher motif distances than the control samples. 75%
199 (6/8) of the PE pregnancies showed enhanced motif patterns (i.e. motif distance >0),
200 but none of the healthy pregnancies exhibited this enhancement.

201 **Correlation between fetal DNA fraction and gestational physiology-related motif** 202 **abundance**

203 To assess the gestation-physiology signals in PE and healthy pregnancies, we
204 analyzed the correlation between set D motif ratios and fetal DNA fraction in
205 maternal plasma in samples of the two groups. The motif ratios were observed to be
206 significantly correlated with the fetal DNA fractions in healthy samples (Pearson $r=$
207 0.49, $P\text{-value}<0.0001$), further consolidating that the motifs in set D were mostly
208 gestation related and carry strong physiology signals (Figure 2D). However, the
209 correlation between Set C motifs ratio and the fetal DNA fraction in PE samples was
210 not significant (Figure 2E). This weak correlation in PE samples might be due to the
211 confounder of abundant PE pathology signals (i.e. PE-specific motifs in set C)
212 carried.

213 **Early identification of EO-PE patients by PE-specific motif ratio**

214 As the cfDNA ends motifs of PE patients showed distinct PE signatures, we
215 wondered whether these PE-specific motifs could be used as diagnostic biomarkers in
216 the early detection of preeclampsia during pregnancy. Hence, we further recruited an

independent pregnancy cohort containing early-onset severe PE (N=26, gestational age: 17.5±2.9 weeks) and gestation-age-matched healthy (N=48, gestational age: 17.8±2.8 weeks) pregnancies in the 2nd trimester as a validation cohort (Supplemental Table 2) in the same hospital (SZ-2 cohort). The plasma samples were processed with the same protocol and platform as the discovery cohort, with mean sequencing depths of 19.03 (range: 9.73 - 87.84) and 22.61 (range: 8.32 - 103.27) for control and EO-PE subjects, respectively. Motif ratios based on PE-specific motifs of Sets A and B from the discovery cohort were analyzed in this validation cohort (See Methods). As shown in Figure 3A and B, the median of motif ratios based on set A and B in EO-PE samples were both significantly higher than healthy samples (median set A motif ratio: 1.45 vs 1.30, P -value<0.0001; median set B motif ratio: 1.28 vs 1.16, P -value<0.0001, Mann-Whitney U test), in line with the observations in the discovery cohort (Figure 1G), which validated the robustness the PE-specific motifs. ROC curve analysis was further employed to assess the classification performance between PE and healthy pregnancies in this validation cohort by motif ratios. The AUC was 0.78 [95% CI:0.65-0.92] for the PE determination based on motifs of set A, which was comparable to the use of motifs of set B (AUC:0.77 [95% CI: 0.63-0.90]) (Figure 3C).

Early identification of EO-PE patients based on machine learning model

Since a single PE-motif ratio demonstrated the potential for PE identification, we

aimed to develop a machine-learning model based on PE-specific motifs to maximize the PE prediction performance. As shown in Figure 4A, first, we extracted 91 PE-specific motifs shared among Sets A, B and C of the discovery cohort as the model features. As the motifs were mined and validated in the pair-end sequencing data with high sequencing depth, to assess the applicability and robustness of these motif biomarkers in samples with different sequencing platforms, protocols, and ultra-low sequencing depth, we further collected 2,481 sequencing samples from pregnancies (Healthy pregnancy (HP): n=2445, EO-PE pregnancy: n=36) receiving non-invasive pregnant testing (NIPT). From this dataset, we randomly selected 1,222 healthy samples and 18 EO-PE samples (NIPT-1 dataset) and combined them with the discovery cohort as training samples. The remaining NIPT samples (NIPT-2 dataset) were used as testing samples. A support vector machine (SVM) model for PE detection was trained and determined through the 10-fold cross-validation (See Methods). The prediction model was applied to the SZ-2 cohort and NIPT-2 dataset for the EO-PE detection.

As a result, we obtained AUC values of 0.90 ± 0.06 for determining EO-PE patients in the training sample set (Figure 4B). Moreover, we achieved a high AUC of 0.83 [95% CI: 0.72-0.93] for EO-PE detection in the SZ-2 validation cohort (Figure 4C). Notably, even with ultra-low sequencing depth (0.05X) and single-end sequencing mode, we obtained a high AUC of 0.84 [95% CI: 0.74-0.92] for the NIPT-2 sample set (Figure 4D).

258

259 **Tissue-of-origin of cfDNA carrying PE-specific motifs in preeclampsia pregnant** 260 **woman**

261 The cell-free DNA in maternal plasma could be released by the fetus and the
262 mother. To explore whether the PE motif signals were specifically origin from the
263 fetus (i.e. placenta) or the mother (i.e. predominantly the hematopoietic cells), we
264 reanalyzed the plasma DNA of one healthy and one PE pregnant woman with
265 matched sequencing data of maternal and paternal buffy coat from our previously
266 published paper(31) (Figure 5A). Fetal- and maternal-derived cfDNA were
267 determined by the parental specific alleles (See Methods). We categorized the plasma
268 DNA into four origin groups: (1) PE-mother and (2) PE-fetus that differentiated from
269 the PE individual, (3) Control-mother and (4) Control-fetus that differentiated from
270 the healthy individual (Figure 5A).

271 As shown in Figure 5B, the plasma DNA originating from the fetuses was
272 significantly shorter than mother-originated ones. Besides, the PE-mother derived
273 cfDNA was observed to be longer than the cfDNA from the control-mother. To
274 investigate whether the identified PE motif pattern is specific to either the mother or
275 the fetus, we conducted a comparative analysis of motif frequencies between different
276 sources of cfDNA, including (1) fetus-derived cfDNA from PE and control
277 pregnancies, (2) maternal-derived cfDNA from PE and control pregnancies, as well as

278 mother-derived and fetus-derived cfDNA within the (3) PE patient and (4) healthy
279 pregnancy (Figure 5A and C). We calculated the $\log_{10}$ (Fold change) of 256 motifs in
280 terms of frequency within each of the four group comparisons and ranked the motifs
281 based on the $|\log_{10}(\text{Fold change})|$. We took the top and bottom 25% of motifs with
282 positive $\log_{10}(\text{Fold change})$ and top 25% of motifs with negative $\log_{10}(\text{Fold change})$
283 values as the differentially increased and decreased motifs in the latter pregnant
284 woman of the comparison (e.g. PE in the comparison of Control VS PE) (Figure 5C).
285 The motif pattern specific to preeclampsia, characterized by an enrichment of A- and
286 C-started motifs and a reduction of G and T motifs, was consistently observed in both
287 fetal-derived and maternal-derived cfDNA of PE pregnancy compared to those of the
288 healthy pregnant woman (Supplemental Figure 2 and Figure 5D). Notably, the A- and
289 C-started motifs accounted for 78.13% and 70.31% of the differentially increased
290 motifs, while the T- and G-started motifs accounted for 73.44% and 70.31% of the
291 differentially decreased motifs in PE-mother and PE-fetus derived cfDNA,
292 respectively, when compared with those in healthy pregnancy. However, this motif
293 pattern was absent in the comparison between fetus-derived cfDNA and
294 mother-derived cfDNA in either PE or control subjects. Among the increased motifs,
295 the A and C-started motifs only account for 42.19% and 42.19% in the fetus-derived
296 cfDNA compared to the mother derived in PE and control subjects, respectively.
297 Among the decreased motifs, the G and T-started motifs only account for 42.19% and
298 37.50% in the fetus-derived cfDNA compared to the mother derived in PE and control

299 subjects, respectively (Figure 5D).

300 Moreover, we compared the motif ratio of the top 10 increased and top 10
 301 decreased motifs of set A between fetal cfDNA from both PE and healthy pregnancies,
 302 and maternal cfDNA from both PE and healthy pregnancies. As shown in Figure 5E,
 303 for PE-decreased motifs, 7 out of 10 in PE-fetus derived cfDNA and all of them in
 304 PE-mother derived cfDNA showed consistently decreased patterns compared with
 305 that origin in the healthy individual. Meanwhile, for PE-increased motifs, 9 out of 10
 306 in PE-fetus derived cfDNA and all of them in PE-mother derived cfDNA showed
 307 consistently increased patterns compared with that origin in healthy individuals
 308 (Figure 5F). These results demonstrated that the PE-specific motif pattern co-existed
 309 in both the fetal and maternal released DNA in plasma. Therefore, as the motif
 310 differences observed in both the mother and fetus with preeclampsia exhibit
 311 congruent patterns, we speculated that the fragmentation preference producing the
 312 PE-specific motifs was non-tissue-specific and primarily attributable to extracellular
 313 DNA molecule degradation, which might be influenced by nucleases in the plasma,
 314 and shared the same pattern in both the mother and fetus.

315

316 **Size and nucleosome footprint of cfDNA carrying different ending nucleotide**

317 Ends motifs were generated from the fragmentation of cfDNA in plasma, which
 318 was influenced by both the nucleosome binding patterns(27) and the activity of

nuclease(32). Thus, to explore the potential fragmentation factors leading to the preference of motifs in PE patients, we analyzed the size of DNA molecules and the occurrence of cfDNA ending sites in relation to nucleosome centres. For each DNA fragment, the upstream and downstream 5' ends were labelled as the U-end and D-end, respectively (Figure 6A). We randomly selected 8 samples from each of the groups in the discovery cohort and pooled them to represent that group.

First, we analyzed the content of cfDNA molecules carrying different nucleotides (i.e. A, T, C and G) in the U-end (Figure 6B) and D-end (Supplemental Figure 3A) within the size range of 50-250bp in all groups. As shown in Figure 6B, the PT2 and PT3 groups showed consistently higher contents of cfDNA with A- and C-end and consistently lower contents of cfDNA with G- and T-end compared with control groups. Additionally, cfDNA with different ending nucleotides displayed dramatically different proportion profiles. cfDNA with C-end showed the highest proportion (i.e. >30%) among all groups across all sizes and displayed a predominant peak around 166bp, which was reminiscent of the length of DNA wrapping the nucleosome core, indicating that the nuclease cutting on C-end DNA may primarily occur on the link DNA of nucleosome and contributed the majority of cfDNA. For short cfDNA below 100bp, the proportions of cfDNA with A-end and C-end were initially higher and decreased gradually. Conversely, for cfDNA ending with G and T, the proportions of such cfDNA progressively increased below 150bp and exhibited a strong 10-bp periodicity. These findings suggest that the components of cfDNA with different ends

340 at each size exhibit distinct preferences and fluctuate accordingly.

341 Figure 6C displays the size profiles of cfDNA with different U-end nucleotides
342 between 50bp and 120bp. The cfDNA with A-end exhibited a higher proportion of
343 short DNA (e.g. <120bp) among all groups. The cfDNA with C-end further showed a
344 slightly higher proportion in the PT2 group and a distinctly higher proportion in the
345 PT3 group. Similar patterns were also observed in cfDNA with different D-end
346 nucleotides (Supplemental Figure 3B). These results indicated that the PE-specific
347 cleavage signals existed in cfDNA across various sizes.

348 To explore whether the cfDNA with different ends would have different cutting
349 preferences in relation to nucleosome structure, we mapped the ending sites of cfDNA
350 with A, T, C, and G ends to the upstream and downstream regions of the nucleosome
351 core. Figure 6D illustrates the content of cfDNA with different U-ends in each
352 location relative to the nucleosome core. The cfDNA with different ends showed
353 distinct proportions across nucleosome structures. For example, the cfDNA with A
354 and G ends showed troughs around 75bp upstream to the nucleosome core, while the
355 cfDNA with C and T ends showed peaks at this location. The PE and control groups
356 showed consistent waveforms in the distributions of the cfDNA with the same ending
357 nucleotide. However, compared with control groups, the component of A-end and
358 C-end cfDNA were elevated in PT3 and PT2 groups, the component of G-end and
359 T-end cfDNA were reduced in PT3 and PT2 groups, across the 200bp regions around
360 the nucleosome core. The cfDNA distribution patterns in D-end are shown in

Supplemental Figure 3C. These results show that cfDNA with different ending nucleotides undergoes distinct cutting patterns across nucleosome structures, whereas PE cleavage patterns can occur throughout the entire nucleosome structure rather than being limited to specific regions.

Nuclease levels in plasma DNA of healthy and PE pregnancies

The above findings revealed that the PE cleavage signals were present in all cfDNA sizes and in regions surrounding the nucleosome core. We speculated that the activity of nucleases in plasma might be altered and thus lead to systematically different dynamics of fragmentation patterns. Figure 7A shows the cfDNA concentrations in the plasma of different groups. The patients with preeclampsia were observed to have significantly higher cfDNA levels than the healthy pregnancies at both 2nd (median: 41.2 ng/μl vs 30 ng/μl, *P*-value<0.0001, Mann-Whitney U test) and 3rd (median: 52.4 ng/μl vs 41.8 ng/μl, *P*-value=0.0001, Mann-Whitney U test) trimesters, indicating the nuclease activity and cfDNA generation might be more pronounced in PE patients in comparison to healthy pregnant women.

Previous studies(32) have implicated the involvement of nuclease DFFB and Deoxyribonuclease I (DNASE1) in cfDNA fragmentation in plasma. Specifically, DFFB has been shown to preferentially generate cfDNA with A ends, while DNASE1 is associated with the production of cfDNA with T-ends. Thus, we further quantified

the concentration of nuclease DFFB and DNASE1 in the plasma of 21 healthy and 15 EO-PE pregnancies during the 3rd trimester. The median levels of DFFB were approximately two-fold higher in PE subjects (median: 0.73 ng/ml, range: 0.25 ng/ml -1.15 ng/ml) compared to healthy subjects (median: 0.36 ng/ml, range: 0.20 ng/ml -1.66 ng/ml) (P -value=0.011, Mann-Whitney U test) (Figure 6F). This increase in DFFB levels might contribute to the higher proportion of cfDNA with A ends observed in PE pregnancies. Besides, we also observed significantly increased levels of DNASE1 in PE patients (median: 4.00 ng/ml, range: 1.71 ng/ml -8.45 ng/ml) compared with healthy individuals (median: 1.98 ng/ml, range: 0.86 ng/ml -9.22 ng/ml) (P -value=0.0110, Mann-Whitney U test) (Supplemental Figure 4).

Discussion

In this study, we investigated the end motif patterns in the plasma DNA of pregnancies with and without early-onset severe preeclampsia in different trimesters. Our findings revealed that these end motifs exhibit distinct pathological signatures in preeclampsia patients, while displaying physiological signatures in healthy pregnant women.

By leveraging the PE-specific motifs identified in the discovery cohort, we successfully validated their utility as a highly effective biomarker for the early detection of EO-PE in another independent cohort processed with a similar protocol. Remarkably, incorporating the PE-specific motifs into a predictive model enabled us

401 to accurately identify PE patients even in samples from the NTPT group, which
402 underwent disparate protocols, including variations in library preparation, sequencing
403 platforms, ultra-low sequencing depth (~0.05X), and sequencing mode (35bp
404 single-end read). These findings reinforce the robustness and versatility of the
405 PE-specific motifs in early PE detection and highlight their potential for integration
406 into the current non-invasive prenatal testing protocol.

407 By deciphering the motifs in maternal- and fetal-derived cfDNA in maternal
408 plasma, we revealed that the PE-specific motif preference was likely generated within
409 the plasma rather than in the somatic tissue (e.g. placental). Furthermore, through
410 dissecting the component of cfDNA with different ending nucleotides at varying
411 lengths and nucleosome structures, we uncovered that the fragmentation process
412 leading to the formation of PE-specific motifs was enhanced across various sizes and
413 the entirety of nucleosome-wrapped cfDNA. Specifically, the
414 fragmentation-generating ends with nucleotides A and C showed systematic
415 enhancement. This led us to hypothesize that the DFFB nuclease(32), known for its
416 preference in generating cfDNA with A ends, might be involved in generating the
417 specific motif pattern associated with PE. Supporting this hypothesis, we have
418 observed a significant elevation in the concentration of DFFB in PE patients,
419 implying the potential role of DFFB nuclease in the etiopathology of preeclampsia.
420 Additionally, we observed a significant increase of DNASE1 in the plasma of PE
421 patients compared to healthy pregnancies. Previous studies(32) have suggested that

422 DNASE1 is prone to generating cfDNA molecules ending with T. However, in our
423 study, we observed a reduced frequency of cfDNA ending with T in the plasma of PE
424 patients, leading to a contradiction. One possible explanation for this discrepancy is
425 that a published research by Cheng et al(33) has indicated that DNASE1 may not be a
426 major nuclease responsible for DNA digestion in plasma. Therefore, while the
427 concentration of DNASE1 increases in plasma, its impact on DNA fragmentation may
428 be negligible or counterbalanced by the activity of DFFB. Instead, DNASE1 has been
429 shown to play a major role in cfDNA fragmentation in urine(34). Exploring the motifs
430 present at cfDNA ends and DNASE1 activity in urine, along with its relationship with
431 the onset of preeclampsia, would be a worthwhile avenue for further research.

432 Recently, using single-molecule sequencing, Yu et al(35). also reported a
433 preference of end motifs in the plasma DNA (a significant amount of long cfDNA,
434 e.g., >500bp) of PE patients in the third trimester. Their findings, however, differ from
435 our study in terms of the specific motif preferences observed. In our study, we
436 observed that the majority of increased motifs in PE patients were characterized by
437 ends with A and C (e.g., >95%), while the majority of decreased motifs in PE patients
438 were characterized by ends with T and G (e.g., >91%). In contrast, Yu et al. found that
439 motifs with reduced frequencies in preeclampsia predominantly ended with G or A,
440 while those with increased frequencies all ended with T or C. This inconsistency
441 suggests that the long DNA fragments captured by single-molecule sequencing may
442 originate from different fragmentation patterns compared to the short cfDNA

443 fragments sequenced by second-generation sequencing. These variations in
444 fragmentation patterns could potentially contribute to the observed differences in
445 motif preferences.

446 Despite the valuable insights gained from our study, there are certain limitations
447 that need to be addressed in future research. Firstly, our discovery cohort consisted
448 only of early-onset preeclampsia patients. Thus, a larger cohort containing multiple
449 types of preeclampsia collected from multiple centers is required to explore the
450 applicability of our findings to other types of preeclampsia. Another limitation of our
451 study is the lack of multiple sampling for the majority of pregnancies. Only 24
452 pregnant women were sampled twice in our study, which limits the exploration of the
453 earliest time for preeclampsia detection based on motif biomarkers. Including
454 multi-point sampling in future studies would also be highly valuable in monitoring the
455 progression of PE by capturing dynamic changes in the abundance of PE-specific
456 motifs.

457 In summary, our study highlights the potential of cfDNA end motif as a
458 promising non-invasive biomarker for early detection of early-onset severe
459 preeclampsia in pregnancy. Additionally, we have provided significant insights into
460 the mechanisms involved in the generation and release of cfDNA from preeclampsia
461 patients. By extending the clinical applications of cfDNA fragmentomics, our findings
462 offer a valuable approach for the fast and cost-effective detection of pregnancy
463 complications.

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478 **Methods**

479 **Statistics**

480 We utilized Mann-Whitney U-test (Wilcoxon Rank Sum Tests) and the

481 Kruskal-Wallis test to evaluate differences in continuous variables among different

482 groups. For the comparison of categorical variables, we employed the Chi-square test

483 or Fisher's exact test. Pearson's test was used to assess correlation. P-values below

0.05 were considered statistically significant. P-values less than 0.05, 0.01, 0.001, and 0.0001 were denoted by *, **, ***, and ****, respectively. We conducted the statistical analysis using R software (version 3.6.1).

Study approval

Our study was approved by the Institutional Review Board of BGI (BGI-IRB 20040-T1) and the Institutional Review Board of Suzhou Municipal Hospital (K-2019-019-K01).

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The key raw data have been deposited onto the China National GeneBank DataBase (CNGBdb) with accession number CNP0003056.

Patients and samples

For the discovery cohort (SZ-1 cohort), we recruited 140 singleton pregnancies including healthy pregnancies and those diagnosed with preeclampsia, during the 2nd and 3rd trimesters at the Suzhou Municipal Hospital, Jiang Su, China.

For the validation cohort (SZ-2 cohort), 26 PE and 48 healthy pregnancies, all in the 2nd trimester, were further recruited at the same hospital.

For the NIPT cohort, the sequencing data of NIPT from 36 pregnancies who

503 developed early-onset severe PE and 2,445 healthy pregnancies accepted NIPT
504 between January 2016 and March 2021 were obtained from the Suzhou Municipal
505 Hospital, Jiang Su, China. The plasma cfDNA was sequenced by Illumina Hiseq
506 sequencing platform in a 35bp single-end format, resulting in a median of 3,497,067
507 sequencing reads (range: 1,060,392-11,201,854). From these samples, we randomly
508 selected 1,222 healthy samples and 18 EO-PE samples to be included in the training
509 dataset for the PE prediction model. The remaining samples were allocated for testing
510 the EO-PE prediction.

511 To quantify nuclease concentration in maternal plasma, we collected plasma
512 samples from 21 healthy pregnancies (8 samples were sequenced in the discovery
513 cohort) and 15 EO-PE pregnancies (12 samples were sequenced in the discovery
514 cohort) at the same hospital.

515 All participants involved in the study provided written informed consent. This
516 study was approved by the Ethics Committee of Suzhou Municipal Hospital
517 (K20190019H01) and the Institute Review Board of BGI (BGI-IRB 20040).

518 **Sample collection and processing**

519 Peripheral blood samples were collected from each subject using EDTA tubes. The
520 blood samples were centrifuged within 4 hours of collection at $1,600 \times g$ for 10
521 minutes at 4°C . Subsequently, the supernatant was isolated and centrifuged again at
522 $16,000 \times g$ for 10 minutes at 4°C . This process yielded 5ml of plasma.

523 **Sequencing Library Preparation**

524 CfDNA was extracted from the plasma samples using the MagPure Circulating DNA
525 Mini KF Kit (Magen, MD5432-02). DNA libraries were prepared using MGIEasy
526 Cell free DNA library preparation reagent kit (MGI, 1000007037) according to the
527 manufacturer's instructions. cfDNA concentration was measured by Qubit dsDNA HS
528 Assay Kit (Q32854, Invitrogen).

529 **Quantification of Nuclease levels in plasma**

530 For the quantification of human DFFB and DNASE1 levels in plasma, we used
531 commercial enzyme-linked immunosorbent assay (ELISA) kits purchased from
532 Jianglai Industrial (catalog number JL10215) and CUSABIO (catalog number
533 CSB-E09068h). The assay values ranged from 0.15 ng/mL to 10 g/mL and 1.56
534 ng/mL to 100 ng/mL, respectively. The principle of the kit is based on the indirect
535 sandwich ELISA technique, which utilizes a capture antibody and a biotin-labelled
536 detection antibody for capture and detection purposes, respectively.

537 **Sequencing and Alignment**

538 The DNA libraries were sequenced using the DNBSEQ platform (MGI) in a 100bp ×
539 2 paired-end format. Fastp (v0.20.1)(36) was employed to remove adaptor sequences
540 and eliminate low-quality reads (i.e., quality score < 20) from the raw sequencing data.
541 Subsequently, Minimap2 (v2.2.11) was used to align the clean reads to the human
542 reference genome (GRCh38/hg38). PCR duplicates were filtered out using

543 Biobambam2 (v2-2.0.87)(37). Reads with more than three mismatches or aligning to
544 multiple locations were filtered. only paired-end reads that aligned to the same
545 chromosome with the correct orientation and had an insert size of ≤ 600 bp were
546 retained for downstream analysis.

547 **Motif analysis**

548 The 4-nucleotide sequence located at the 5' end of plasma DNA fragments is
549 previously defined as the ends motif(18). Motif frequency was computed by
550 normalizing the count of this specific motif against the total count of all 256 types of
551 4mer motifs. The motif diversity score based on the entropy principle, as described
552 before(18), was employed to gauge the uniformity of motif distribution. A low MDS
553 value indicated a skewed distribution, while a high MDS value indicated a more
554 uniform distribution of ends motifs in terms of their frequencies. Additionally, the
555 motif ratio was determined as the ratio between the accumulated frequencies of
556 increased motifs and decreased motifs.

557 **Ends analysis**

558 cfDNA molecules possess an upstream (U) and a downstream (D) 5' endpoint,
559 determined by their coordinates on the human reference genome. We refer to these
560 endpoints as the U-end and D-end, respectively. The U-end nucleotide can be one of
561 the A/T/C/G bases, leading to the classification of cfDNA into UA/UT/UC/UG
562 fragments. Similarly, the D-end nucleotide categorizes cfDNA into DA/DT/DC/DG

563 fragments. To achieve a high sequencing depth of cfDNA for deep investigation, we
564 randomly selected 8 samples from each group in the discovery cohort and pooled
565 them to represent the group. We analyzed and compared the size profiles of cfDNA
566 belonging to different categories.

567 Meanwhile, the occurrence of the ending sites of cfDNA from different categories
568 was interrogated across the 200bp region surrounding each of the nucleosome cores
569 determined in the plasma DNA from Synder et al's paper (Figure 6A)(17). By
570 accumulating and normalizing the cfDNA ends relative to the nucleosome cores
571 within this 200bp region, we determined the composition of cfDNA with different end
572 types at each location of this region, as well as the relative frequencies of cfDNA with
573 a specific end type across the region.

574 **Analysis of fetal- and maternal-derived DNA**

575 The sequencing data of plasma DNA and paired parental genomic DNA of a healthy
576 and a preeclampsia pregnancy were obtained from our previous study(31). The
577 parental informative single-nucleotide polymorphisms (SNPs) where the mother was
578 homozygous (AA) and the fetus was heterozygous (i.e. fetus: AB, deduced from
579 father: AB/BB) were used for the differentiation of fetal- and maternal-derived
580 cfDNA in maternal plasma. We determined fetal-derived cfDNA based on those
581 molecules carrying fetal-specific alleles (B). Meanwhile, as the majority of cfDNA in
582 maternal plasma is mother released, the molecules carrying shared alleles (A) were
583 deemed as predominately maternal-derived.

584 **Development of Machine-learning model for EO-PE detection**

585 We identified 91 differential motifs that were shared among Sets A, B, and C in the
586 discovery cohort and utilized them as model features. The training dataset consisted
587 of all samples from the discovery cohort, along with a random selection of 1,222
588 control samples and 18 EO-PE samples. To train our EO-PE prediction model, we
589 employed the Support Vector Machine (SVM) algorithm and performed 10-fold
590 cross-validation. The hyperparameters and model that yielded the highest AUC value
591 in PE classification within the training sample set were selected as the final EO-PE
592 prediction model. We then evaluated the performance of the developed EO-PE
593 prediction model by applying it to the validation sets of the SZ-2 cohort and the
594 NIPT-2 dataset.

595 **Fundings:**

596 This work was supported by the National Natural Science Foundation of China (Grant
597 No. 82001576), Primary Research & Development Plan of Jiangsu Province
598 (BE2022736), Jiangsu Maternal and Children health care key discipline
599 (FXK202142), National Natural Science Foundation of China (32171441, 32000398),
600 National Key R&D Program of China (2022YFC2502402).

601 **Competing interests**

602 The authors declare that there is no competing interest in this study.

603

604 Author contributions

605 Conception and design, HQ.Z., X.J., T.W.; sample collection and processing,
606 LW.Q., CH.Z., XJ.W., XY.S., H.T., Y.X., RJ.O., XX.W., Y.Z.; analysis and
607 interpretation of data, HQ.Z., LW.Q., XT.H., Y.L., JY.Z., Y.S., X.J., T.W.; writing –
608 original draft, HQ.Z., LW.Q., XT.H., Y.L., JY.Z.; writing – review & editing: HQ.Z.,
609 LW.Q., XT.H., X.J., T.W.; All authors reviewed the manuscript and approved the final
610 revision.

611

612

613 Reference

614

- 615 1. N. J. Kassebaum, R. M. Barber, Z. A. Bhutta, L. Dandona, P. W. Gething, S. I.
616 Hay, Y. Kinfu, H. J. Larson, X. Liang, S. S. Lim, Global, regional, and national
617 levels of maternal mortality, 1990–2015: a systematic analysis for the Global
618 Burden of Disease Study 2015. *The lancet* **388**, 1775–1812 (2016).
- 619 2. N. J. Kassebaum, A. Bertozzi-Villa, M. S. Coggeshall, K. A. Shackelford, C.
620 Steiner, K. R. Heuton, D. Gonzalez-Medina, R. Barber, C. Huynh, D. Dicker,
621 Global, regional, and national levels and causes of maternal mortality during
622 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013.
623 *The Lancet* **384**, 980–1004 (2014).

- 624 3. J. M. Roberts, P. A. August, G. Bakris, J. R. Barton, I. M. Bernstein, M. Druzin,
625 R. R. Gaiser, J. P. Granger, A. Jeyabalan, D. D. Johnson, Hypertension in
626 pregnancy: executive summary. *Obstetrics and Gynecology* **122**, 1122–1131
627 (2013).
- 628 4. D. Raymond, E. Peterson, A critical review of early-onset and late-onset
629 preeclampsia. *Obstet Gynecol Surv* **66**, 497–506 (2011).
- 630 5. A. L. Tranquilli, M. A. Brown, G. G. Zeeman, G. Dekker, B. M. Sibai, The
631 definition of severe and early-onset preeclampsia. Statements from the
632 International Society for the Study of Hypertension in Pregnancy (ISSHP).
633 *Pregnancy Hypertension: An International Journal of Women's Cardiovascular*
634 *Health* **3**, 44–47 (2013).
- 635 6. E. A. Steegers, P. von Dadelszen, J. J. Duvekot, R. Pijnenborg, Pre-eclampsia.
636 *The Lancet* **376**, 631–644 (2010).
- 637 7. E. Jung, R. Romero, L. Yeo, N. Gomez-Lopez, P. Chaemsaitong, A. Jaovisidha,
638 F. Gotsch, O. Erez, The etiology of preeclampsia. *Am J Obstet Gynecol* **226**,
639 S844–S866 (2022).
- 640 8. O. Erez, R. Romero, E. Jung, P. Chaemsaitong, M. Bosco, M. Suksai, D. M.
641 Gallo, F. Gotsch, Preeclampsia and eclampsia: the conceptual evolution of a
642 syndrome. *Am J Obstet Gynecol* **226**, S786–S803 (2022).
- 643 9. T. M. MacDonald, S. P. Walker, N. J. Hannan, S. Tong, T. J. Kaitu'u-Lino,
644 Clinical tools and biomarkers to predict preeclampsia. *EBioMedicine* **75**, 103780

- 645 (2022).
- 646 10. L. A. Magee, K. H. Nicolaides, P. von Dadelszen, Preeclampsia. *New England*
647 *Journal of Medicine* **386**, 1817–1832 (2022).
- 648 11. T. R. Kolarova, H. S. Gammill, J. L. Nelson, C. M. Lockwood, R. Shree, At
649 Preeclampsia Diagnosis, Total Cell-Free DNA Concentration is Elevated and
650 Correlates With Disease Severity. *J Am Heart Assoc* **10** (2021).
- 651 12. A. C. Palei, Cell-free DNA as a potential biomarker for preeclampsia. *Expert*
652 *Rev Mol Diagn* **21**, 1253–1256 (2021).
- 653 13. Z. Guo, F. Yang, J. Zhang, Z. Zhang, K. Li, Q. Tian, H. Hou, C. Xu, Q. Lu, Z.
654 Ren, X. Yang, Z. Lv, K. Wang, X. Yang, Y. Wu, X. Yang, Whole-Genome
655 Promoter Profiling of Plasma DNA Exhibits Diagnostic Value for
656 Placenta-Origin Pregnancy Complications. *Advanced Science* **7**, 1901819
657 (2020).
- 658 14. R. W. K. Chiu, E. Heitzer, Y. M. D. Lo, F. Mouliere, D. W. Y. Tsui, Cell-Free
659 DNA Fragmentomics: The New “Omics” on the Block. *Clin Chem* **66**,
660 1480–1484 (2020).
- 661 15. K. C. A. Chan, J. Zhang, A. B. Y. Hui, N. Wong, T. K. Lau, T. N. Leung, K.-W.
662 Lo, D. W. S. Huang, Y. M. D. Lo, Size Distributions of Maternal and Fetal DNA
663 in Maternal Plasma. *Clin Chem* **50**, 88–92 (2004).
- 664 16. K. Sun, P. Jiang, S. H. Cheng, T. H. T. Cheng, J. Wong, V. W. S. Wong, S. S. M.
665 Ng, B. B. Y. Ma, T. Y. Leung, S. L. Chan, T. S. K. Mok, P. B. S. Lai, H. L. Y.

- 666 Chan, H. Sun, K. C. A. Chan, R. W. K. Chiu, Y. M. D. Lo, Orientation-aware
667 plasma cell-free DNA fragmentation analysis in open chromatin regions informs
668 tissue of origin. *Genome Res* **29**, 418–427 (2019).
- 669 17. M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, J. Shendure, Cell-free DNA
670 Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin.
671 *Cell* **164**, 57–68 (2016).
- 672 18. P. Jiang, K. Sun, W. Peng, S. H. Cheng, M. Ni, P. C. Yeung, M. M. S. Heung, T.
673 Xie, H. Shang, Z. Zhou, R. W. Y. Chan, J. Wong, V. W. S. Wong, L. C. Poon, T. Y.
674 Leung, W. K. J. Lam, J. Y. K. Chan, H. L. Y. Chan, K. C. A. Chan, R. W. K. Chiu,
675 Y. M. D. Lo, Plasma DNA End-Motif Profiling as a Fragmentomic Marker in
676 Cancer, Pregnancy, and Transplantation. *Cancer Discov* **10**, 664–673 (2020).
- 677 19. L. Serpas, R. W. Y. Chan, P. Jiang, M. Ni, K. Sun, A. Rashidfarrokhi, C. Soni, V.
678 Sisirak, W.-S. Lee, S. H. Cheng, Dnase1l3 deletion causes aberrations in length
679 and end-motif frequencies in plasma DNA. *Proceedings of the National*
680 *Academy of Sciences* **116**, 641–649 (2019).
- 681 20. P. Jiang, K. Sun, Y. K. Tong, S. H. Cheng, T. H. T. Cheng, M. M. S. Heung, J.
682 Wong, V. W. S. Wong, H. L. Y. Chan, K. C. A. Chan, Y. M. D. Lo, R. W. K. Chiu,
683 Preferred end coordinates and somatic variants as signatures of circulating tumor
684 DNA associated with hepatocellular carcinoma. *Proceedings of the National*
685 *Academy of Sciences* **115** (2018).
- 686 21. K. C. A. Chan, P. Jiang, K. Sun, Y. K. Y. Cheng, Y. K. Tong, S. H. Cheng, A. I. C.

- 687 Wong, I. Hudecova, T. Y. Leung, R. W. K. Chiu, Y. M. D. Lo, Second generation
688 noninvasive fetal genome analysis reveals de novo mutations, single-base
689 parental inheritance, and preferred DNA ends. *Proceedings of the National*
690 *Academy of Sciences* **113** (2016).
- 691 22. P. Jiang, T. Xie, S. C. Ding, Z. Zhou, S. H. Cheng, R. W. Y. Chan, W.-S. Lee, W.
692 Peng, J. Wong, V. W. S. Wong, H. L. Y. Chan, S. L. Chan, L. C. Y. Poon, T. Y.
693 Leung, K. C. A. Chan, R. W. K. Chiu, Y. M. D. Lo, Detection and
694 characterization of jagged ends of double-stranded DNA in plasma. *Genome Res*
695 **30**, 1144–1153 (2020).
- 696 23. Q. Zhou, G. Kang, P. Jiang, R. Qiao, W. K. J. Lam, S. C. Y. Yu, M.-J. L. Ma, L. Ji,
697 S. H. Cheng, W. Gai, W. Peng, H. Shang, R. W. Y. Chan, S. L. Chan, G. L. H.
698 Wong, L. T. Hiraki, S. Volpi, V. W. S. Wong, J. Wong, R. W. K. Chiu, K. C. A.
699 Chan, Y. M. D. Lo, Epigenetic analysis of cell-free DNA by fragmentomic
700 profiling. *Proceedings of the National Academy of Sciences* **119** (2022).
- 701 24. P. Ulz, G. G. Thallinger, M. Auer, R. Graf, K. Kashofer, S. W. Jahn, L. Abete, G.
702 Pristauz, E. Petru, J. B. Geigl, E. Heitzer, M. R. Speicher, Inferring expressed
703 genes by whole-genome sequencing of plasma DNA. *Nat Genet* **48**, 1273–1278
704 (2016).
- 705 25. M. S. Esfahani, E. G. Hamilton, M. Mehrmohamadi, B. Y. Nabet, S. K. Alig, D.
706 A. King, C. B. Steen, C. W. Macaulay, A. Schultz, M. C. Nesselbush, J. Soo, J. G.
707 Schroers-Martin, B. Chen, M. S. Binkley, H. Stehr, J. J. Chabon, B. J. Sworder,

708 A. B.-Y. Hui, M. J. Frank, E. J. Moding, C. L. Liu, A. M. Newman, J. M. Isbell,
709 C. M. Rudin, B. T. Li, D. M. Kurtz, M. Diehn, A. A. Alizadeh, Inferring gene
710 expression from cell-free DNA fragmentation profiles. *Nat Biotechnol* **40**,
711 585–597 (2022).

712 26. Y. M. D. Lo, D. S. C. Han, P. Jiang, R. W. K. Chiu, Epigenetics, fragmentomics,
713 and topology of cell-free DNA in liquid biopsies. *Science (1979)* **372** (2021).

714 27. K. Sun, P. Jiang, A. I. C. Wong, Y. K. Y. Cheng, S. H. Cheng, H. Zhang, K. C. A.
715 Chan, T. Y. Leung, R. W. K. Chiu, Y. M. D. Lo, Size-tagged preferred ends in
716 maternal plasma DNA shed light on the production mechanism and show utility
717 in noninvasive prenatal testing. *Proceedings of the National Academy of*
718 *Sciences* **115** (2018).

719 28. W. Guo, X. Chen, R. Liu, N. Liang, Q. Ma, H. Bao, X. Xu, X. Wu, S. Yang, Y.
720 Shao, F. Tan, Q. Xue, S. Gao, J. He, Sensitive detection of stage I lung
721 adenocarcinoma using plasma cell-free DNA breakpoint motif profiling.
722 *EBioMedicine* **81**, 104131 (2022).

723 29. L. Chen, G. K. Abou-Alfa, B. Zheng, J.-F. Liu, J. Bai, L.-T. Du, Y.-S. Qian, R.
724 Fan, X.-L. Liu, L. Wu, J.-L. Hou, H.-Y. Wang, Genome-scale profiling of
725 circulating cell-free DNA signatures for early detection of hepatocellular
726 carcinoma in cirrhotic patients. *Cell Res* **31**, 589–592 (2021).

727 30. Y. Wang, X. Fan, H. Bao, F. Xia, J. Wan, L. Shen, Y. Wang, H. Zhang, Y. Wei, X.
728 Wu, Y. Shao, X. Li, Y. Xu, S. Cai, Z. Zhang, Utility of Circulating Free DNA

729 Fragmentomics in the Prediction of Pathological Response after Neoadjuvant
730 Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Chem* **69**, 88–99
731 (2023).

732 31. J. Ju, J. Li, S. Liu, H. Zhang, J. Xu, Y. Lin, Y. Gao, Y. Zhou, X. Jin, Estimation of
733 cell-free fetal DNA fraction from maternal plasma based on linkage
734 disequilibrium information. *NPJ Genom Med* **6**, 85 (2021).

735 32. D. S. C. Han, M. Ni, R. W. Y. Chan, V. W. H. Chan, K. O. Lui, R. W. K. Chiu, Y.
736 M. D. Lo, The biology of cell-free DNA fragmentation and the roles of DNASE1,
737 DNASE1L3, and DFFB. *The American Journal of Human Genetics* **106**,
738 202–214 (2020).

739 33. T. H. T. Cheng, K. O. Lui, X. L. Peng, S. H. Cheng, P. Jiang, K. C. A. Chan, R. W.
740 K. Chiu, Y. M. D. Lo, DNase1 does not appear to play a major role in the
741 fragmentation of plasma DNA in a knockout mouse model. *Clin Chem* **64**,
742 406–408 (2018).

743 34. M. Chen, R. W. Y. Chan, P. P. H. Cheung, M. Ni, D. K. L. Wong, Z. Zhou, M.-J.
744 L. Ma, L. Huang, X. Xu, W.-S. Lee, Fragmentomics of urinary cell-free DNA in
745 nuclease knockout mouse models. *PLoS Genet* **18**, e1010262 (2022).

746 35. S. C. Y. Yu, P. Jiang, W. Peng, S. H. Cheng, Y. T. T. Cheung, O. Y. O. Tse, H.
747 Shang, L. C. Poon, T. Y. Leung, K. C. A. Chan, Single-molecule sequencing
748 reveals a large population of long cell-free DNA molecules in maternal plasma.
749 *Proceedings of the National Academy of Sciences* **118**, e2114937118 (2021).

- 750 36. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ
751 preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
- 752 37. G. Tischler, S. Leonard, biobambam: tools for read pair collation based
753 algorithms on BAM files. *Source Code Biol Med* **9**, 13 (2014).
- 754
- 755

756 Tables

757 **Table 1.** Demographic and clinical characteristics of pregnant women with and
758 without EO-PE of the discovery cohort.

Variables	Second trimester			Third trimester		
	Controls n=56	PE n=71	P value	Controls n=8	PE n=16	P value
Maternal age, years	34(25-43)	34(24-43)	0.779	34(26-38)	33.5(27-41)	0.916
BMI, kg/m ²	23.6±3.1	23.9±3.5	0.596	26.0±2.6	28.4±4.7	0.230
Times of pregnancies	2.8±1.6	2.8±1.5	0.997	2.4±1.8	3.1±1.7	0.267
Times of delivery	0.3±0.4	0.3±0.5	0.843	0.4±0.5	0.9±0.6	0.086
GA at sampling, weeks	19.0±4.9	17.6±1.9	0.861	39.7±1.1	29.9±3.4	<0.0001
GA at PE was diagnosed, weeks	-	30.0±3.0	-	-	29.7±3.3	-
GA at delivery, weeks	-	-	-	39.7±1.2	31.0±3.35	<0.0001
Natural delivery, %(n)	-	-	-	62.5(5)	6.25(1)	0.003
Stillbirth, %(n)	-	-	-	0	6.25(1)	0.470
Male fetus, %(n)	-	-	-	50 (4)	28.6(4)	0.315
Birthweight, g	-	-	-	3187.5±322.7	1342.7±435.0	<0.0001

759 PE, pre-eclampsia; BMI, body mass index; GA, gestational age.

760 Values were given as mean±SD and median(range).

761 P value: The Chi-square test was used to test variables with a 0-1 distribution, such as

762 Natural delivery, Stillbirth, and Male fetus. The remaining continuous or discrete

763 variables, such as Maternal age, Times of pregnancies, and Birthweight, were

764 examined applying the Mann Whitney U test, with the exception of BMI, which was

765 evaluated with the Unpaired T test.

766 **Table S1. List of differential motifs from different comparisons in the discovery**
767 **cohort.**

768

769 **Table S2. Demographic and clinical characteristics of pregnant women with and**

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

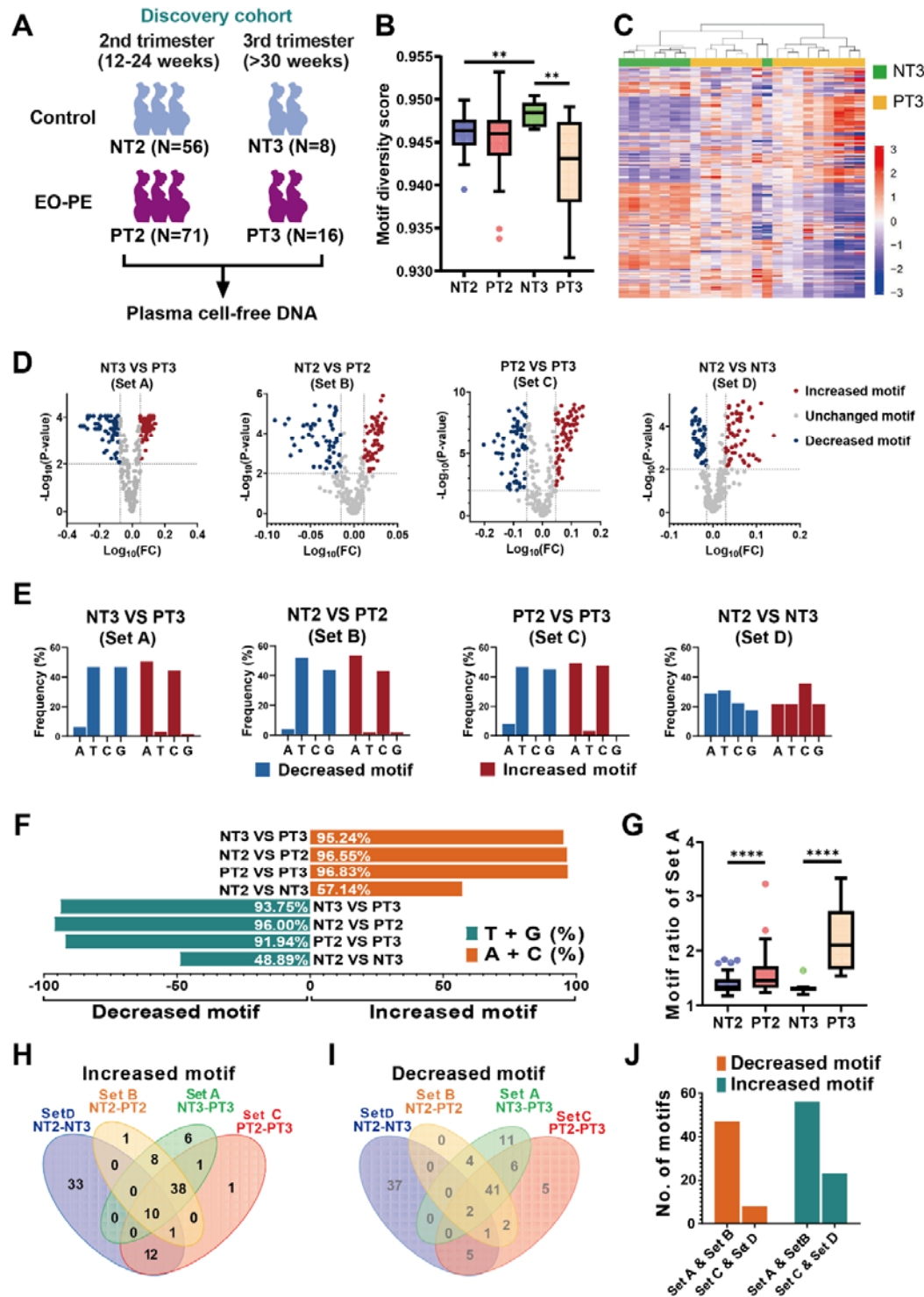
800

801

802

803 **Figure legends**

804 **Figure 1. Identification of PE-specific motif pattern.**



806 (A) Sample collection in the discovery cohort. (B) Boxplot analysis of motif diversity
807 score in plasma DNA from NT2, PT2, NT3 and PT3 groups. (C) Heatmap analysis of
808 frequencies of 256 4mer motifs between NT3 and PT3 groups. (D) Volcano plot
809 analysis of motifs with differential frequencies between different comparisons. The
810 red and blue dots indicate the motifs with a P-value below 0.01 and ranked in the top
811 25% and bottom 25% of the fold change, respectively, which were deemed as
812 specifically increased and decreased motifs in the latter group of the comparisons (e.g.
813 PT3 in the comparison of “NT3 VS PT3”). Gray dots represent non-differential motifs
814 that did not reach the thresholds. (E) Frequency of the first nucleotide at the 5’ start of
815 all differential motifs in each comparison set. (F) The total percentage of T- and
816 G-end motifs in the decreased motifs and the total percentage of A- and C-end motifs
817 in the increased motif group for each group comparison. (G) Comparison of
818 PT3-specific (identified from NT3 VS PT3, Set A) motif ratios between NT2 and PT2
819 groups, and between NT3 and PT3 groups. Venn diagram analysis of increased motifs
820 (H) and decreased motifs (I) among different group comparisons. (J) Number of
821 motifs that overlapped between different group comparisons.

822

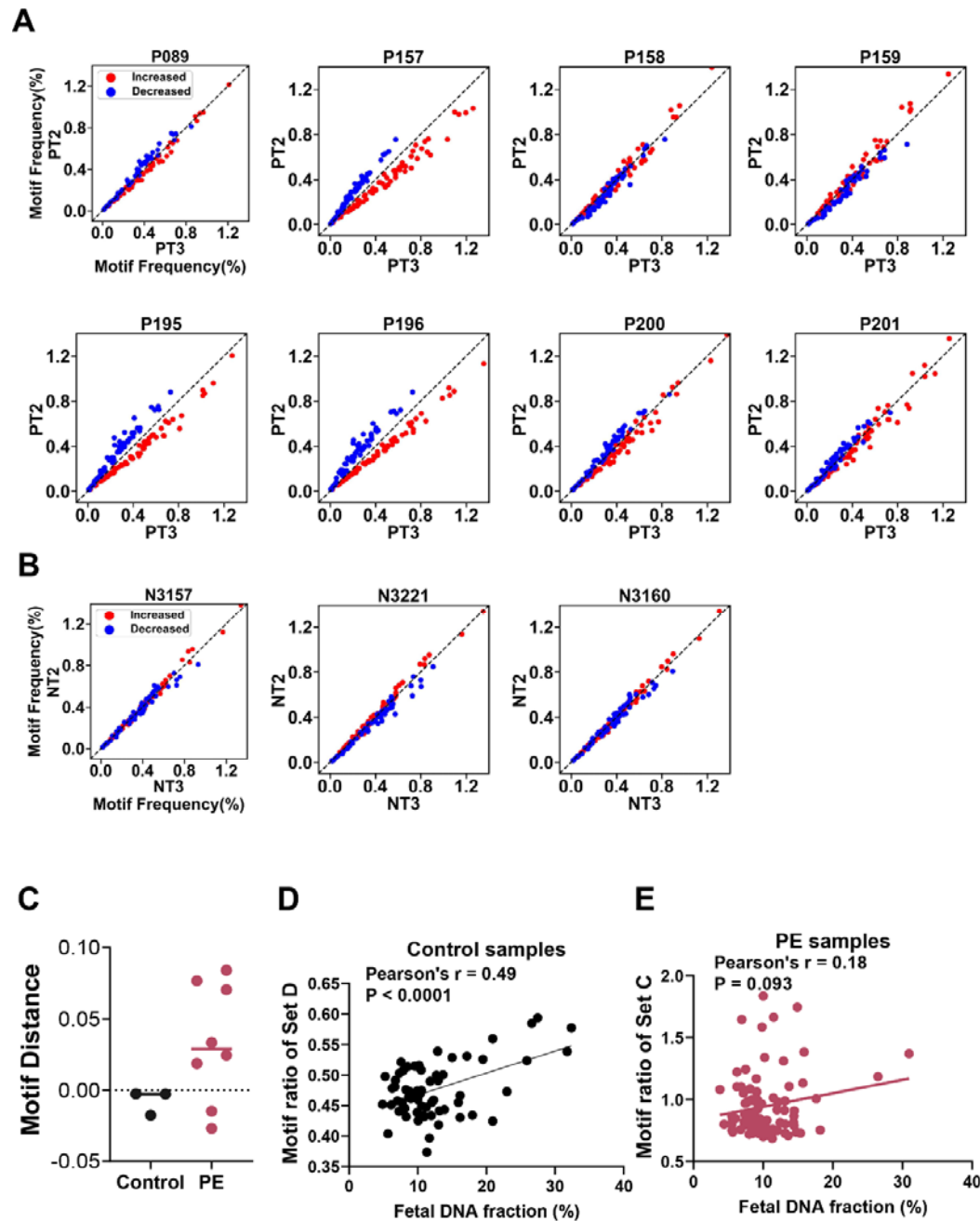
823

824

825

826

827 **Figure 2. Motif distributions in pregnant women with different gestational ages.**



828

829 (A) The frequencies of PE-specific motifs in plasma DNA samples collected from the

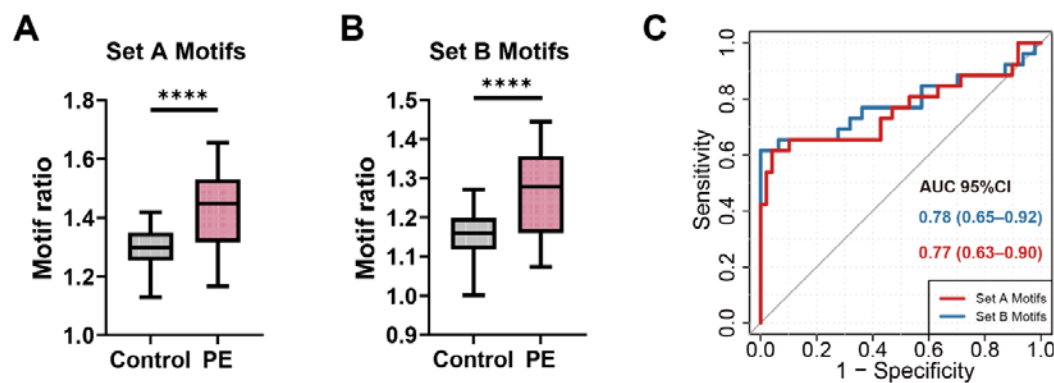
830 same PE pregnant woman during her second and third trimesters. (B) The frequencies

831 of PE-specific motifs in plasma DNA samples collected from the same healthy

832 pregnant woman during her second and third trimesters. (C) Boxplot analysis of motif

distance between plasma DNA of second and third trimesters in control and PE samples. (D) Correlation between NT3-specific (Set D) motif ratio and fetal DNA fraction in maternal plasma of healthy pregnant women. (E) Correlation between PT3-specific (Set C) motif ratio and fetal DNA fraction in maternal plasma of pregnant women with PE.

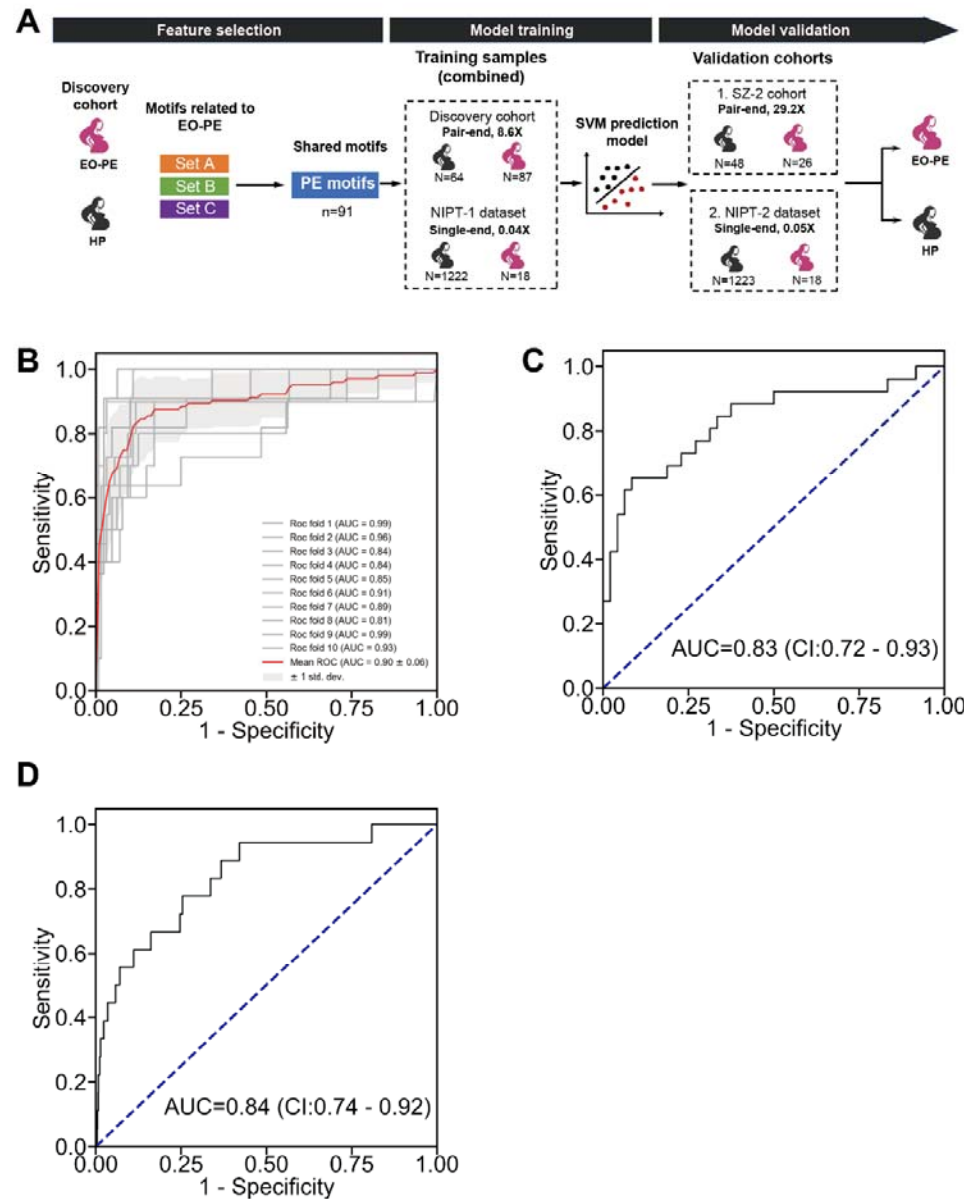
Figure 3. PE-specific motif analysis in plasma DNA of SZ-2 validation cohort.



Boxplot of the motifs ratio in PE and control samples based on the PE-specific motifs of Set A (A) and Set B (B). (C) ROC curve analysis between PE and control groups with the use of Set A and Set B motifs ratios.

850

851 **Figure 4. PE prediction in SZ-2 and NIPT cohort based on SVM model**



852

853 (A) Illustration of PE prediction using PE-specific motifs by SVM model. (B) ROC

854 curve analysis between PE and control groups in pooled training samples based on the
855 prediction model using 91 PE-specific motifs and the 10-fold cross-validation ROC
856 curve of EO-PE detection in SZ-2 cohort (C) and NIPT-2 sample set (D) based on PE
857 prediction model.

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

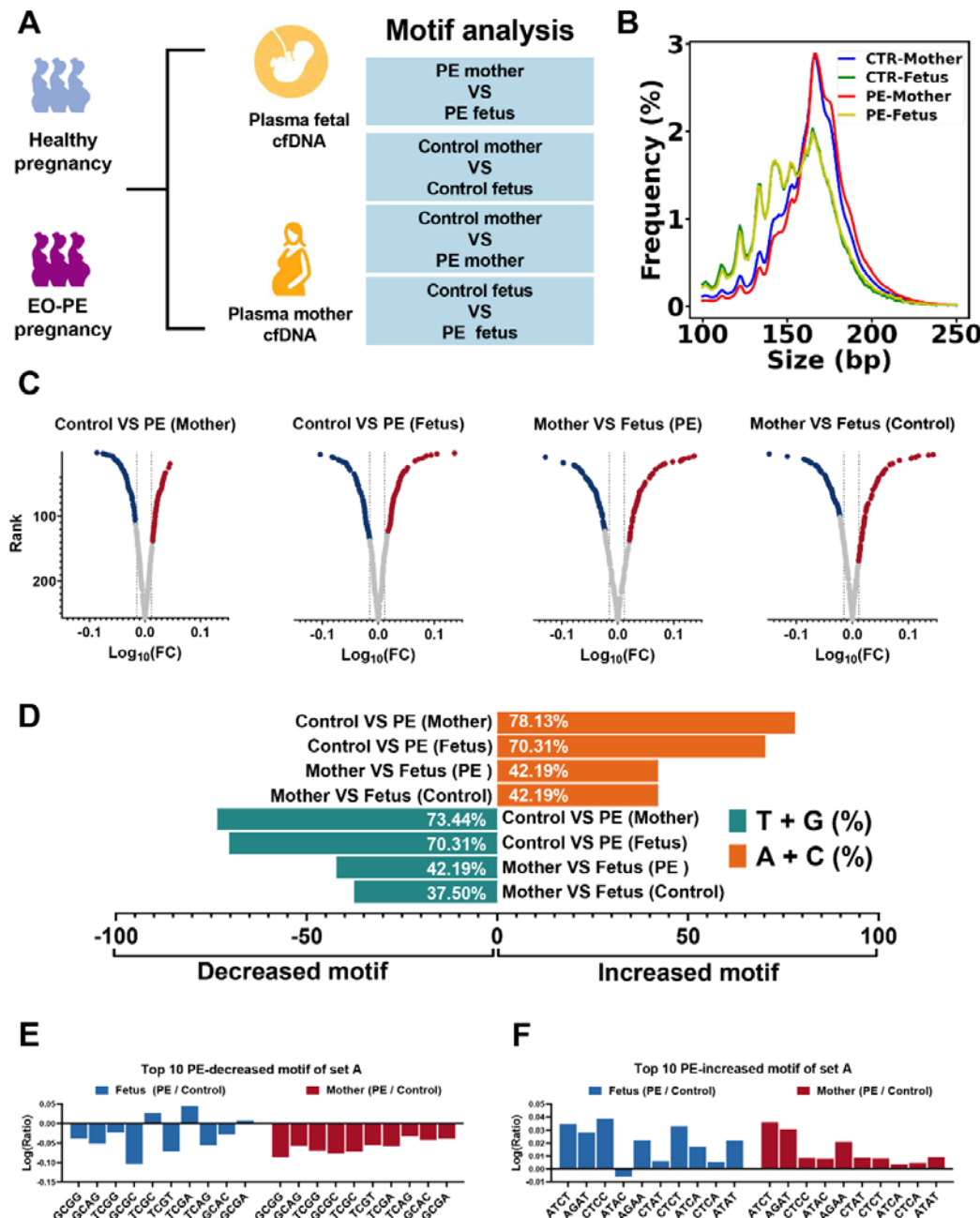
873

874

875

876 **Figure 5. Ends motif of fetal- and maternal-derived cfDNA in a healthy and a PE**

877 **pregnant woman.**



878

879 (A) Illustration of motif analysis in fetal and maternal cfDNA in pregnant women

880 with and without PE. Four sets of motif comparisons were performed between fetal-

881 and maternal-derived cfDNA and between PE and healthy pregnant women. (B) Size
882 profiles of fetal- and maternal-derived cfDNA in healthy and PE pregnant women. (C)
883 Volcano plot analysis of motifs with differential frequencies between different
884 comparisons. Y-axis indicates the rank of the motifs according to their frequency in
885 descending order. Red and blue dots indicate the motifs ranked in the top 25% and
886 bottom 25% of the fold change, respectively, which were deemed as specifically
887 increased and decreased motifs in the latter group of the comparisons. Gray dots
888 indicate those motifs below the thresholds, which were deemed as non-differential
889 motifs. (D) The total percentage of T- and G-end motifs in the decreased motifs and
890 the total percentage of A- and C-end motifs in the increased motif group for each
891 group comparison. Logarithmic scale representation of the ratio between frequencies
892 of top 10 PE-decreased motifs (E) and top 10 PE-increased motifs (F) of set A (based
893 on the discovery cohort) in plasma fetal- and maternal-derived cfDNA of PE and
894 healthy pregnant women.

895

896

897

898

899

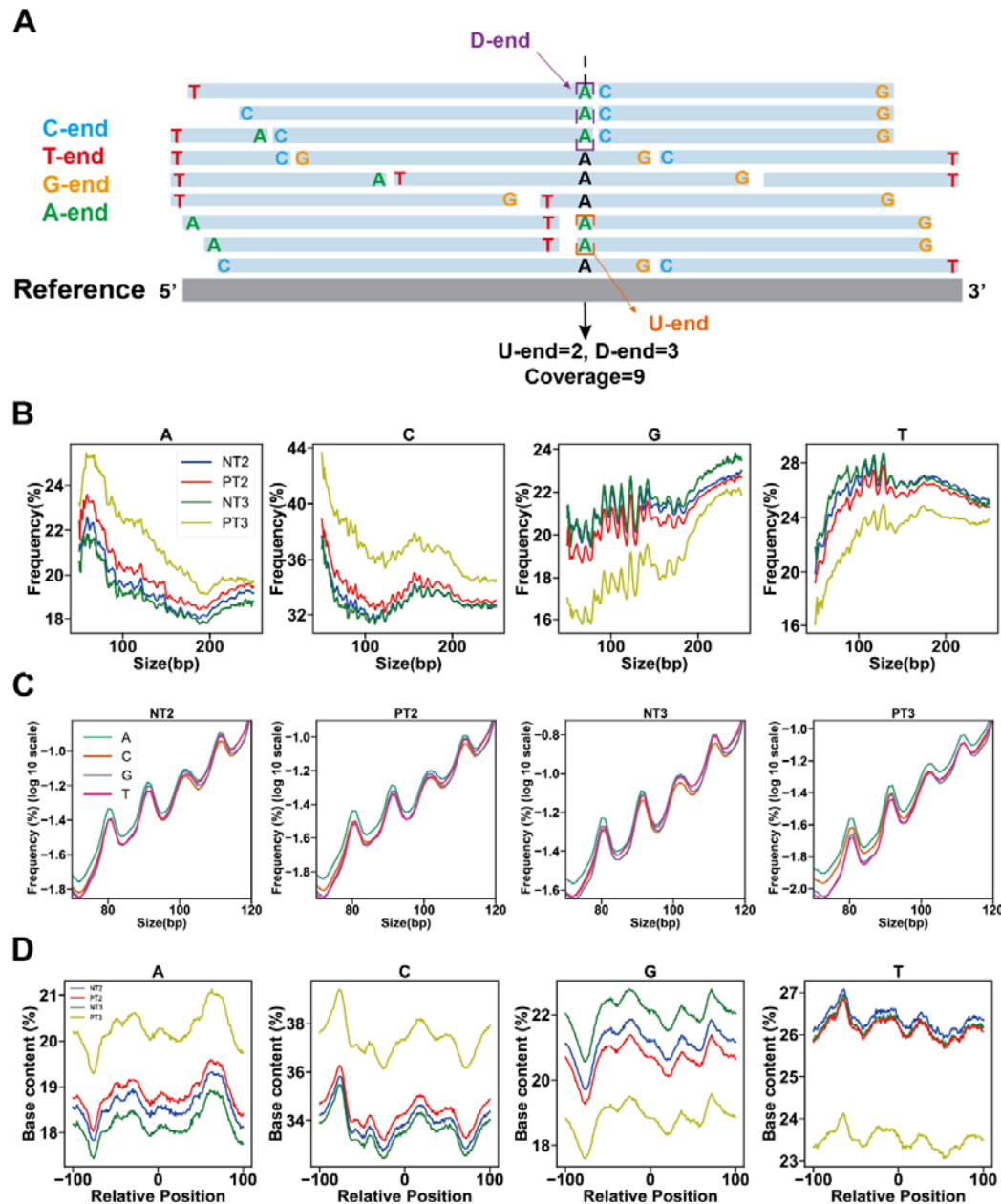
900

901

902

903 **Figure 6. Size and nucleosome footprint analysis for cfDNA carrying different**

904 **U-ends.**



905

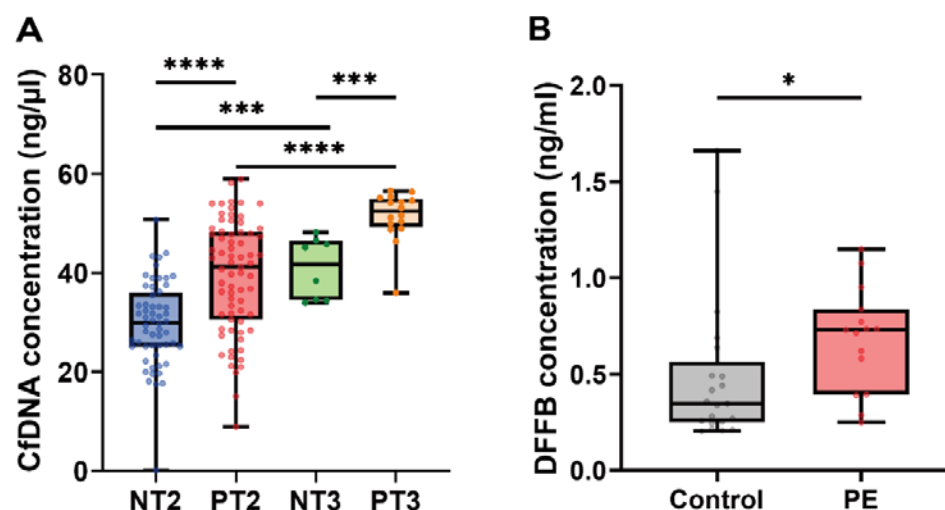
906 (A) Illustration of ends occurrence and coverage of cfDNA with different ends across

907 the genome. (B) Composition of plasma cfDNA with different U-ends in NT2, PT2,

908 NT3 and PT3 groups at each of the cfDNA lengths between 50 and 250bp. The

frequency of cfDNA at each individual size was determined by calculating the number of cfDNA molecules with a specific ending nucleotide and normalizing it to the total number of cfDNA molecules with the same size. (i.e. in each size, the summed frequency of cfDNA with A, T, C and G ends was 100%). (C) Size profiles of cfDNA with different U-ending nucleotides across 50 and 120bp in different groups (i.e the frequency of cfDNA with particular ending nucleotides at each individual cfDNA length was normalized by the amount of cfDNA with the same ending nucleotides between 50 and 120bp). (D) Composition of cfDNA with different U-ending nucleotides at each genomic location relative to nucleosome structure in NT2, PT2, NT3 and PT3 groups. The occurrence of each type of cfDNA ends in each relative position of nucleosome was normalized by the total number of ends in this location.

Figure 7.



(A) cfDNA concentration in the plasma of NT2, PT2, NT3 and PT3 groups. (B) The

925 concentrations of DFFB in the plasma of NT2, PT2, NT3 and PT3 groups.

926

927 **Figure S1.**

928 Top 10 motifs with specific increased frequencies (red) and top 10 motifs with

929 specific decreased frequencies (blue) in different motif sets identified in the discovery

930 cohort.

931

932 **Figure S2.**

933 Frequencies of motifs with different starting nucleotides among the increased motifs

934 (red) and decreased motifs (blue) of (A) maternal-derived cfDNA of the PE

935 pregnancy compared to those of the healthy pregnancy, (B) fetal-derived cfDNA of

936 the PE pregnancy compared to those of the healthy pregnancy, (C) fetal-derived

937 cfDNA compared to maternal-derived cfDNA of the PE pregnancy, (D) fetal-derived

938 cfDNA compared to maternal-derived cfDNA of the healthy pregnancy.

939

940 **Figure S3. Size and nucleosome footprint analysis for cfDNA carrying different**

941 **D-ends.**

942 (A) Composition of plasma cfDNA with different D-ends in NT2, PT2, NT3 and PT3

943 groups at each of the cfDNA lengths between 50 and 250bp. (B) Size profiles of

944 cfDNA with different D-ending nucleotides across 50 and 120bp in different groups.

945 (C) Composition of cfDNA with different D-ending nucleotides at each genomic

946 location relative to nucleosome structure in NT2, PT2, NT3 and PT3 groups.

947 **Figure S4.**

948 The concentrations of DNASE1 in the plasma of NT2, PT2, NT3 and PT3 groups.