nature portfolio

Peer Review File

cfGWAS reveal genetic basis of cell-free DNA end motifs

Corresponding Author: Dr Huanhuan Zhu

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

A version of this paper was originally rejected for publication by Nature Communications, however that decision was reconsidered after appeal by the authors.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Zhu and colleagues describe a GWAS of cfDNA end motif frequencies, carried out using NIPT data from ~28k pregnant women. The authors identify a number of novel genetic regions that may be involved in cfDNA biology, and linked these to pregnancy phenotypes using genetic correlation, MR, and co-localisation approaches. Currently the manuscript lacks sufficient detail in terms of description and/or reporting for several of the analyses. I also have concerns about the appropriateness of some of the methods utilised. Despite this, I think the finding of the paper are of interest, and would likely hold if the below points were addressed.

1. Sequencing depth was very low, with average depth 0.1525X, and many samples with depths lower than this. In this study, the authors used data from 30 Chinese individuals from the 1KGP downsamples to 0.1X to assess imputation quality, and achieved an R2 of 79.34%. In the STITCH paper, the authors found that a higher sequencing depth of 0.3X achieved a correlation between array genotypes and imputed dosages (r2) of around 0.7.
• What parameters did the authors use for running STITCH, and how were these determined? Specifically, K, the number of ancestral haplotypes, and nGen, the number of generations since the population was founded. It is recommended (https://doi.org/10.1038/ng.3594) to determine these parameters by using an external replication dataset. Were the 30 1KGP individuals used for this purpose? If so, the R2 values may be an overestimate of the imputation quality.
• Was any filtering done in terms of imputation quality (INFO score threshold), before calculating the accuracy?
• Imputation quality is affected by the amount of fetal DNA in a sample, and the low number of sequenced reads in NIPT data (https://doi.org/10.1016/j.celrep.2024.114799). The 1KGP individuals will not be impacted by these NIPT-specific factors, so again, the imputation quality may be an overestimate. Are the authors able to assess imputation quality using individuvals with both NIPT data, and higher coverage non-NIPT data? If such data are not available, this should be noted as a limitation.

2. What were the distributions of the motif scores? Were any trait transformations performed prior to running the GWAS?

3. In a few instances, the authors have grouped the results by the motif start (A,C,G, or T). Eg the results on page 8, where the number of loci for motifs with the same starting nucleotide are given, or for the MR results in Figure S14. What is the biological rationale for this?

4. Key Information is missing from the summary statistics for the GWAS. In order to be useful, the effect, and non-effect allele must be explicitly indicated. What are the allele frequencies for the reported SNPs? It may also be helpful to include imputation quality, if no, or a fairly relaxed INFO score filter were applied.

5. Results for MPCs and MDS phenotypes are not really provided – only summarised by Manhattan plots. Full results for the top SNPs should be included in a supplementary table.

6. Not enough information is provided in the replication results – for each cohort, there is just a SNP and a p-value. It is not therefore clear exactly what is being “replicated” for each locus. Specfically:

• In the Chinese NIPT and natural Chinese replication cohorts, are the same motifs significantly associated as those in the discovery sample?
• For the motif phenotypes, it is not clear whether direction of effect is the same across discovery/replication cohorts.
• The SNPs reported in the replication cohorts is often different to the sample – presumably these are good proxies (r2 > ??) with the smallest p-value at that locus, but this is not explicitly stated.
• For the DutchNIPT samples, the phenotypes not motifs, but related phenotypes; again, for each locus please specify which phenotype(s) were associated.
I think the authors have attempted to show some of these aspects of replication through figure S6; but these plots are too small to be legibile. Instead, I would recommend the authors provide full association results for all motifs at each locus, to allow better interpretation of what truly replicated.

7. ACAT analysis is not described in the methods. From the results, it appears the authors used ACAT to combine p-values across 256 motifs, for each SNP. Normally, ACAT is used for combining p-values of several SNPs (eg in a gene,region, pathway), for a single trait. I note that ACAT can be a general method for combining p-values, but I'm not sure whether the optimal method in this case? Can the authors provide additional justification/references for using it in this context? Alternatively, a method like C-GWAS may perhaps be more appropriate? https://doi.org/10.1038/s41467-022-35328-9

8. For all motifs, a series of analyses were carried out using LDSC: heritability estimated; partioned heritability; genetic correlations.
• For several motifs in Table S3, heritibility is NA. Why is this?
• For the motifs with non-missing heritaability estimates, the h2 in many cases is very low, and not significantly different to 0 ($p<0.05$). For the partioned heritability analyses to be sufficiently powered, it is recomended "the dataset analyzed must have a very large sample size and/or large SNP-heritability, and the trait analyzed must be polygenic" (https://doi.org/10.1038/ng.3404). I do not think this applies to a large number of the motifs analysed. Table S4 reports 52,070 p-values from the partioned heritability analyses for all 256 motifs. If the p-values were uniformly distributed under the null hypothesis, we would expect around 521 of these to be $p < 1e-2$ (ie the p-value threshold used), just by chance. In fact, only 290 p-values meet this threshold. By running the partioned heritability analyses for all motifs, the authors are greatly increasing their multiple testing burden, while obtaining results for several motifs which are not likely to be reliable, due to the low heritabilites. A better approach would be to run these additional anlysis for motifs only where there is evidence of significant SNP-heritability. Furthermore, I'd recommend applying an FDR adjustment (eg Benjamni-Hochberg) to determine significant results, rather than an arbitrirary nominal threshold.
• The above point also applies to the genetic correlation analyses.

9. Regarding the Mendelian Randomisation:
• How were all of the exposures/outcomes selected? Was this based on the genetic correlations, or just the >3 significant SNP requirement?
• If I understood correctly, the pregnancy phenotype GWASs used for the MR used the same individuals as the motifs GWASs, thereby making this a one-sample MR. MR-Egger is an approach developed for two-sample MR, which in one-sample MR may indicate a significant causal effect more often than it should (https://doi.org/10.1093/ije/dyab084). There is also a OneSampleMR R package which has other estimators, specifically for one sample MR: https://doi.org/10.32614/CRAN.package.OneSampleMR
• Given the pregnancy phenotypes and motifs were measured in the same indviduals, it would be useful to compare effect estimates from MR with the observed associations between these variables.

10. The methods overall are generally lacking in some detail:
• As mentioned in comment 1, more details are needed regarding the imputation. What parameters were used for STITCH? (ie k and nGen ??). Was any filtering done in terms of imputation quality (INFO score threshold?)
• For the GWAS, were related individuals removed? Association analyses were undertaken uing Plink, which does not account for realtedness, but it is not stated that the cohort was restricted to an unrelated set.
• Similarly although ancestry PCs were calculated, with 10PCs used as covariates in the GWAS, it is not mentioned whether any exclusions were made based on ancestry?

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This article presents the GWAS results for end motifs in a large dataset of pregnant women. The study identifies 23 study-wide signicant loci, several of which are novel. The study is replicated in three independent cohorts. In addition, one of the genes is further validated by knock out experiments in mice and cell lines. This is a very valuable study, providing novel insights in the factors affecting cfDNA fragmentation. Those insights are of utmost importance for the booming field of cfDNA analyses and cfDNA fragmentomics. The article is well written, the figures are clear and systematic.

Major comment:
The title suggest the genetic basis of cell-free DNA features is studied. However, the study only examines one of the features of the cell free DNA, namely the end motif frequencies and not other aspects e.g. fragment length, jaggedness, ….. Hence, correct the title accordingly to 'cfGWAS reveal genetic bases of cell-free DNA end motifs' would be more suitable.

Minor comments
The study is performed using cfDNA from pregnant women, but the data is analysed as if the cfDNA is derived only from the mother. Since about 10% of the cfDNA is derived from the fetus, which has a different genotype, one could consider there to be an effect on the results/analyses. Would this influence the results/power? How would that affect the heritability scores? Can placenta specific signals be misassigned to the mother? It might be valuable to comment/deliberate this aspect?

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
In this paper, the authors perform GWAS on data from ~28000 NIPT samples. Their main aim is to unravel genome wide associations with the 256 four-nucleotide cfDNA end-motifs. By doing so, they hope to gain more insight into the biology of cfDNA. As they performed single-end sequencing, they could not take into account fragment length association studies. They used one dataset for the primary analysis, and two subsequent non-overlapping datasets to validate their data. These three datasets were from pregnant women. They finally performed an independent study on a small dataset with data from nonpregnant women. As the PANX1 association was considered to be a novel finding, they subsequently performed experimental verification analysis of the biological role of PANX1. Also, the effects of hematological cells on cfDNA end-motifs was studied and a strong causal relationship between neutrophils and end-motifs was noticed.

The study is fully based on statistical analysis, but I did not dive into the statistical methods and trust other reviewers with more relevant expertise will do so.

1. In general, pages 3-20 are main text and the supplemental data, in which the methods are described, consists of an additional 7 pages (text only). With ~400 words per page (estimation), this means the main text of the paper itself is ~7200 words and the supplemental text ~ 2800 words. Moreover, there are 86 references in the main paper and an additional 25 in the supplemental text. I realize the authors present a very extensive study, with a lot of data, but the paper would certainly benefit from less text. The findings of the involvement in cfDNA biology of genes like DNASE1L3 and DFFB is not new, but because of the many different studies discussions on such not-novel findings also take quite some text. Perhaps the authors could focus more on the added value of finding comparable associations with their study and on the novel findings.
2. The authors mention the identification of the involvement of the PANX1 gene is a novel finding, but Linthorst et al. (their reference 49) previously also identified an association with cfDNA biology of this gene in their GWAS study. Furthermore, the authors obtained the complete summary statistics of the study of Linthorst et al. but as far as I can see the Dutch group is not acknowledged. I suppose the authors have an agreement with the group on this, but if not, it feels unjust.
3. In the (recent) past many publications appeared dealing with studies on the biology of cfDNA. This is a very interesting topic, but nevertheless, much is still unknown. I am following these studies with much interest, meanwhile thinking about the clinical relevance / consequences. How can these findings be translated to / be of use in daily practice? Apparently, the authors have some ideas on this, as in the last two sentences of the abstract the authors mention “Novel knowledge uncovered by this study promises to revolutionize liquid biopsy technology” and “the potential of this paradigm is enormous”. Also in the last paragraph of the discussion, the authors mention their research “could contribute to the discovery of therapeutic targets for diseases influenced by abnormal DNA concentrations”, and “the significant discoveries will not only revolutionize our understanding of cfDNA biology but also open new avenues for future clinical applications”. I would like to ask the authors to elaborate more on this.
4. As the authors mention themselves, a limitation of the study is first of all the relatively small initial cohort size, but more importantly the cohorts consisting of data from pregnant women only. Only the data used for the independent study were from non-pregnant individuals, but this was a small cohort size. When replying on comment 3, please take this also into account. When thinking about clinical applications, will these only be applicable to pregnant women? The two other main fields of interest in cfDNA are cancer and transplantation. Will potential “new avenues” also be applicable to these fields, and also to men or only to women? Are differences to be expected? The authors mention that by now worldwide more than 40 million women have undergone NIPT, but what about data from the other fields and data from men?

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors provided summary statistics with their revised manuscript. The information provided in these files raises serious concerns regarding the results of the GWAS. While the authors state "all retained SNPs had INFO scores greater than 0.5", in these summary results files, there are many SNPs where the INFO_SCORE column has a value < 0.5. A quick look-up in the uploaded file cfGWAS.m256.T___.csv of a couple of the reported SNPs in Table S1 found at least one with very low

imputation quality (e.g., rs529723778 has INFO_SCORE of 0.027). I haven't done a comprehensive look-up, but would assume this may be the case for others. That isn't to say that all reported variants are likely to be wrong, but there may be a number of false positive hits reported, and the genome-wide results used in their genetic correlations etc. are likely to contain a lot of noise.

Unless the INFO_SCORE column in the uploaded summary stats is incorrect, or I have misinterpreted (hopefully this is the case!), much of the analyses should be repeated after filtering out variants with very low imputation quality.

I address the reviewers' responses to my original comments below:

1. Overall, the additional detail regarding imputation has adequately addressed my comments. In the response about INFO score thresholds, the authors provide a plot (S1C of their previous paper). It is unclear what the plotted INFO_SCORE is showing here; I'd assume median/mean(?) INFO score, per MAF bin. So while the average INFO score across all SNPs was >0.5, it isn't clear that each SNP had INFO > 0.5. Indeed, the summary statistics uploaded by the authors seemed to confirm that many variants had INFO < 0.5.
2. Thank you for the additional clarification regarding trait transformation.
3. The additional explanation is appreciated.
4. The extra fields requested are available now in the uploaded summary stats but should also be included in the paper (e.g., table S1).
5. ..
6. The additional information is welcome. Betas should be provided in table S3 in addition to p-values so consistency of direction of effect can be seen. It appears from Figures S8 and S9 that some variants have opposite directions of effect in the replication cohorts, compared to the discovery. If directions of effects are inconsistent, these shouldn't really be classified as a "replicated" association.
7. ..
8. ..
9. The comparison of MR vs observed observations is interesting. Thank you for providing this; perhaps worth considering for a supplementary figure!
10. Perhaps worth trying to repeat relatedness estimation after filtering for poorly imputed variants.

(Remarks on code availability)
Scripts for running several analyses have been uploaded. No major comments.

Reviewer #2

(Remarks to the Author)
I am happy with the answers to my comments.
Just a minor comment:
In you answer you state 'QUITL2, which allows simultaneous imputation of maternal and fetal genomes'.
I believe this method is QUILT2.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Even though the paper was rejected in the first place, the authors submitted a revised version. The authors did a tremendous piece of work, and thoroughly revised the paper. I would like to thank the authors for their efforts!

As far as my previous comments are concerned: the authors responded sufficiently to these. My first comment, however, is, despite the response of the authors, not completely resolved, but I leave it to the editor how to proceed with this. The paper is well-written, but it is still very lengthy. Moreover, the results section is full of background information and partly even discussion. I understand why the authors have chosen this format, as otherwise the rationale behind the different analyses would not be understood and the reader would be completely lost, but the fact is that this makes the paper very lengthy and introduction, results and discussion are not fully separated.

Even though in general, the paper is well-written, throughout the paper there are a some typos and grammar errors, e.g. (but not limited to):
Line 87: ….only the three genes DFFB, …
Line 98: locus = loci
Line 441-442: Consequently, the origin ….. to the plasma. Incorrect sentence.
Lines 448-449: ….in which the cfDNA motifs…. Change into: …with the cfDNA motifs….
Lines 593-595: Recently, a newly….fetal genomes. Incorrect sentence.
Please go through the paper thoroughly to correct these.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you for addressing the concerns raised in my previous review. I'm pleased to see that the analyses have been re-done with the INFO filter, along with the updated MAF and HWE p-value filters. These changes, as anticipated, have indeed resulted in a more robust study with the increased replication rate. It's great to see that the majority of findings still hold. This revised version provides a much stronger proof of concept for the utility of cfDNA in genetic studies.

I do have a couple of final minor points the authors should check:

On page 9, lines 240-41: The phrase "among others" following the list of replicated genes (PANX1, DFFB, and PADI4) seems superfluous and should be removed.

On page 10, lines 255-258 (and in the response to previous comment 5): The statement about consistent effect directions across studies isn't fully supported by Figure S8 and Table S3. There appear to be a small number of significant associations showing opposite effect directions in one of the replication cohorts (I think all are the natural population cohort) compared to the discovery study. This nuance should be accurately reflected in the text.

(Remarks on code availability)

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

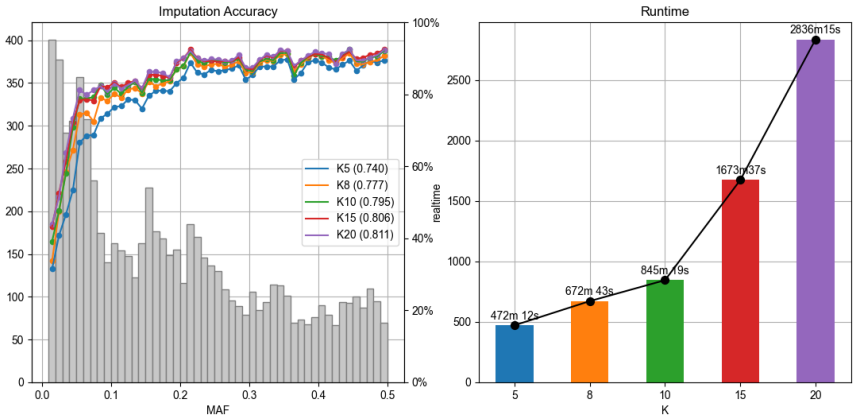
Zhu and colleagues describe a GWAS of cfDNA end motif frequencies, carried out using NIPT data from ~28k pregnant women. The authors identify a number of novel genetic regions that may be involved in cfDNA biology, and linked these to pregnancy phenotypes using genetic correlation, MR, and co-localisation approaches. Currently the manuscript lacks sufficient detail in terms of description and/or reporting for several of the analyses. I also have concerns about the appropriateness of some of the methods utilised. Despite this, I think the finding of the paper are of interest, and would likely hold if the below points were addressed.

Response: We sincerely appreciate the reviewer's thorough evaluation of our manuscript and the invaluable feedback provided. Below are our responses and corresponding revisions:

1. Sequencing depth was very low, with average depth 0.1525X, and many samples with depths lower than this. In this study, the authors used data from 30 Chinese individuals from the 1KGP downsamples to 0.1X to assess imputation quality, and achieved an R2 of 79.34%. In the STITCH paper, the authors found that a higher sequencing depth of 0.3X achieved a correlation between array genotypes and imputed dosages (r2) of around 0.7.

• What parameters did the authors use for running STITCH, and how were these determined? Specifically, K, the number of ancestral haplotypes, and nGen, the number of generations since the population was founded. It is recommended (https://doi.org/10.1038/ng.3594) to determine these parameters by using an external replication dataset. Were the 30 1KGP individuals used for this purpose? If so, the R2 values may be an overestimate of the imputation quality.

Response: We thank the reviewer for this comment. We used K = 10 and nGen = 16,000, where nGen was determined based on the formula nGen = 4 × sample size / K, with a sample size of approximately 40,000. To select an appropriate value for K, we conducted an evaluation using around 20,000 individuals (including 30 high-depth WGS samples as true set), testing K values of 5, 8, 10, 15, and 20. The results showed that while imputation accuracy plateaued at $K \geq 10$, computational time increased substantially with higher K values. Therefore, we selected K = 10 as a balance between accuracy and efficiency.



We have added and highlighted the following content in the Method section:

To ensure the data was suitable for genetic analysis, we applied genotype imputation using STITCH (with parameters K=10, nGen=16,000), which provided allele dosages for the imputed genotypes.

• Was any filtering done in terms of imputation quality (INFO score threshold), before calculating the accuracy?

Response: We thank the reviewer for this comment. For the calculation of imputation accuracy and subsequent GWAS analysis, we included only SNPs with a minor allele frequency (MAF > 0.05) and Hardy-Weinberg equilibrium p-value > 1e-6. After applying these filters, all retained SNPs had an INFO score greater than 0.5. The following figure, which illustrates this, was included in our published work (Supplementary Figure S1C, https://www.cell.com/cell-genomics/fulltext/S2666-979X(24)00237-4):

[editorial note: third party material redacted]

• Imputation quality is affected by the amount of fetal DNA in a sample, and the low number of sequenced reads in NIPT data (https://doi.org/10.1016/j.celrep.2024.114799). The 1KGP individuals will not be impacted by these NIPT-specific factors, so again, the imputation quality may be an overestimate. Are the authors able to assess imputation quality using individuvals with both NIPT data, and higher coverage non-NIPT data? If such data are not available, this should be noted as a limitation.

Response: We thank the reviewer for this comment. Indeed, the unique characteristics of cfDNA, fetal DNA, and short sequencing reads can influence imputation performance, and genomic DNA data from the 1000 Genomes Project (1KGP) may not

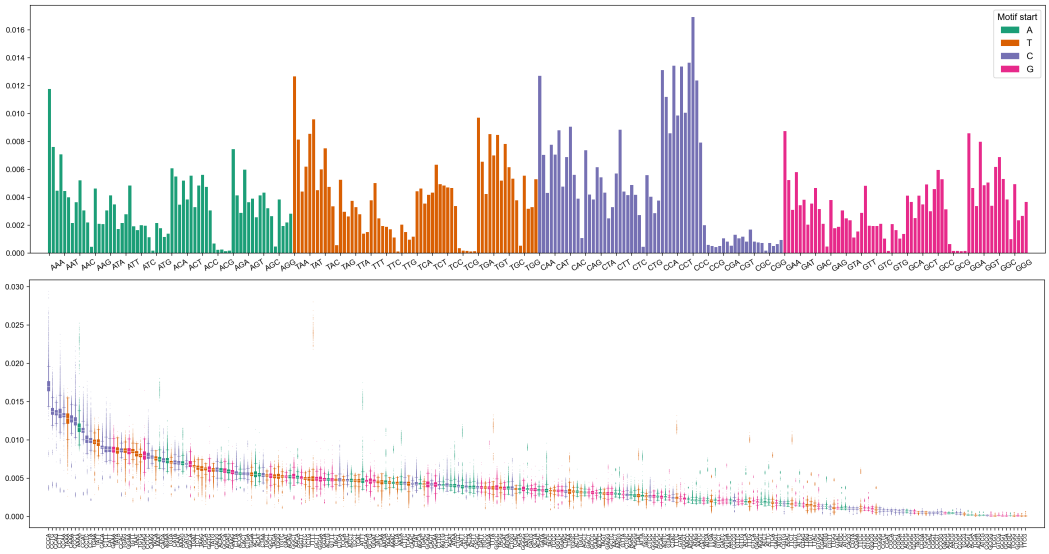
fully capture these properties. To address this limitation, in our previous work, we collected data from 14 pregnant women with both high-depth cfDNA sequencing (45.39×) and corresponding real NIPT data. The high-depth cfDNA data were used as the ground truth for genotypes. Under this setting, the imputation accuracy reached approximately 85%. We acknowledge that the high proportion of common variants (and higher imputation accuracy) is mainly due to the limited size (14 samples) of the high-depth sequence data, leading to the exclusion of low-frequency variants from the imputation accuracy calculation. The following figure, which illustrates this, was included in our published work (Supplementary Figure S1D, https://www.cell.com/cell-genomics/fulltext/S2666-979X(24)00237-4):

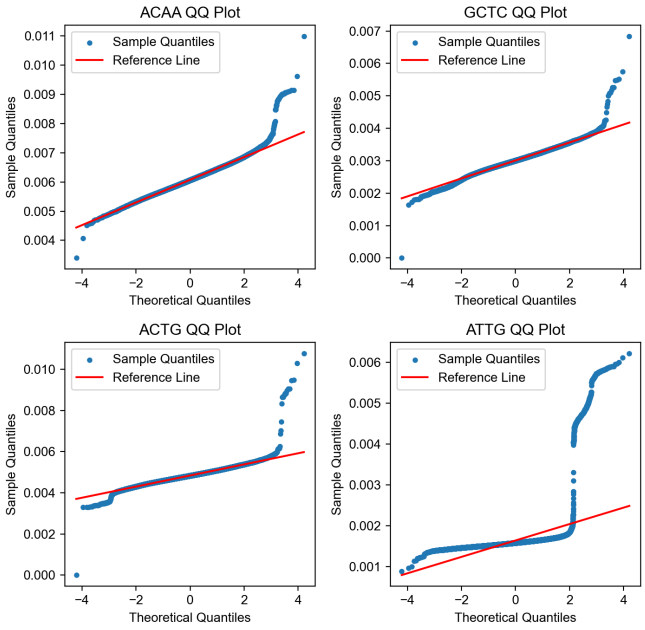
[editorial note: third party material redacted]

2. What were the distributions of the motif scores? Were any trait transformations performed prior to running the GWAS?

Response: We thank the reviewer for this comment. We provided the distributions of the 256 motif frequencies using bar plots and box plots, as shown in Supplementary Figure S2. To assess whether these distributions follow a normal distribution, we conducted Kolmogorov–Smirnov (K-S) tests. However, due to the high sensitivity of the K-S test in large samples, none of the motif frequencies passed the normality test. As an alternative, we generated Q–Q (quantile–quantile) plots for several representative motifs, including those with the minimum, two randomly-selected medium, and maximum K-S p-values. Q–Q plots of selected 4-mer motif frequencies suggest that most motif frequencies are approximately normally distributed, although mild deviations—particularly in the tails—are observed. Motifs such as ACAA and GCTC show relatively good alignment with the theoretical normal distribution, whereas ACTG and ATTG exhibit heavier tails and deviations, especially in higher quantiles. These patterns reflect the variability in motif frequency distributions across individuals in a large cohort. We have added and highlighted the following content in the Results section (Phenotype data):

We presented the distributions of the 256 motif frequencies using bar plots and box plots (Figure S2). Overall, the distributions of motif frequencies are consistent with previously reported patterns. For example, the motif CCCA exhibits the highest frequency, C-end motifs tend to have higher average frequencies than others, and motifs containing CG dinucleotides generally show lower frequencies (https://pmc.ncbi.nlm.nih.gov/articles/PMC10151549/).





Additionally, each motif frequency was normalized to a standard normal distribution for GWAS using the --pheno-quantile-normalize option in PLINK2. The corresponding PLINK2 commands used in the analysis are detailed in the Methods section (Genome-wide association study):

We filter in SNPs with minor allele frequency (MAF) greater than 0.05 (--maf 0.05), p-value for testing Hardy-Weinberg equilibrium (HWE) greater than 1e-6 (--hwe 1e-6), and genotype missingness less than 10% (--geno 0.1). Additional arguments included --pheno-quantile-normalize and --covar-variance-standardize, which were used to

transform motif frequencies to a standard normal distribution and to standardize quantitative covariates to have a mean of zero and variance of one.

3. In a few instances, the authors have grouped the results by the motif start (A,C,G, or T). Eg the results on page 8, where the number of loci for motifs with the same starting nucleotide are given, or for the MR results in Figure S14. What is the biological rationale for this?

Response: We thank the reviewer for this insightful comment. We acknowledge that our manuscript did not sufficiently elaborate on the biological significance of cfDNA end motifs, beyond the brief mention in the Introduction: "End motifs refer to the nucleotide composition at the 5' end of cfDNA, generated during cfDNA digestion. Analyzing these motifs helps trace cfDNA origin and provides insights into physiological and pathological states." As end motifs are the primary phenotypes in our study, they indeed merit a more thorough introduction. cfDNA end motifs reflect the activity and cleavage preferences of nucleases. For instance, the nucleases DFFB, DNASE1L3, and DNASE1 preferentially generate cfDNA fragments ending in A, C, and T, respectively. Therefore, the abundance of specific end motifs can indicate the activity of corresponding nucleases. In recent years, cfDNA end motifs have emerged as promising biomarkers in liquid biopsy applications, especially in oncology, as their frequencies have been shown to differ between healthy individuals and cancer patients. Moreover, cfDNA end motifs are tissue-specific and can inform on tissue-of-origin, which has important applications in areas such as transplantation monitoring. We have added and highlighted the following content in the Results section (Phenotype data):

Different cfDNA end motifs reflect nuclease activity, pathophysiological processes, and tissue-specific DNA fragmentation patterns. Specifically, DFFB, DNASE1L3, and DNASE1 preferentially generate A-end, C-end, and T-end motifs, respectively. Recent studies have highlighted cfDNA end motifs as emerging biomarkers for oncology and transplantation monitoring due to their frequency shifts under pathological conditions and their tissue-specific profiles.

As for our Mendelian randomization analysis, while we observed that T-end motifs had the greatest number of causal associations with pregnancy-related phenotypes, we currently lack direct biological evidence to fully explain this observation. One possibility is that the unique physiological context of pregnancy may influence placenta-specific nuclease activity, hormonal regulation, or maternal immune adaptation, thereby favoring T-end motif production. We believe this finding has potential clinical significance, suggesting that cfDNA end motif profiling could serve

as a novel biomarker for pregnancy-related physiological states, immune responses, or complications. We have added and highlighted the following content in the Discussion section:

CfDNA end motifs have been increasingly explored as biomarkers in cancer detection and transplantation monitoring. Our team have recently uncovered their role in predicting preeclampsia during early pregnancy. Pregnancy also offers a valuable model to study tissue-of-origin dynamics (maternal vs. fetal). To our knowledge, this is the first study to explore the causal relationships between cfDNA end motifs and physiological and biochemical phenotypes in a pregnant population. We found that T-end motifs exhibited the most causal associations and colocalizations with pregnancy phenotypes. This may be related to placenta-specific nuclease activity, hormonal changes, or maternal immune adaptation during pregnancy. While our current analysis cannot fully explain this observation, it highlights a promising area for further investigation into the biological and clinical significance of cfDNA end motif dynamics in pregnancy.

4. Key Information is missing from the summary statistics for the GWAS. In order to be useful, the effect, and non-effect allele must be explicitly indicated. What are the allele frequencies for the reported SNPs? It may also be helpful to include imputation quality, if no, or a fairly relaxed INFO score filter were applied.

Response: We thank the reviewer for this comment. We will release the full summary statistics for all 256 cfDNA end motif frequencies, including key information such as effect and non-effect alleles, allele frequencies, and imputation quality metrics (e.g., INFO scores). We have submitted a data registration application for these GWAS summary statistics to the Ministry of Science and Technology of China. Upon approval, we will upload the full dataset to the GWAS Catalog, where it will be freely available for visualization and download by the research community. For review purposes, we have uploaded the summary statistics for all 256 GWAS results (organized into four files, each containing 64 motifs) to Google Drive (https://drive.google.com/drive/folders/1Um6bPoBgw4Xokl677T0CDjD_8wxUjTvF?usp=sharing).

5. Results for MPCs and MDS phenotypes are not really provided – only summarised by Manhattan plots. Full results for the top SNPs should be included in a supplementary table.

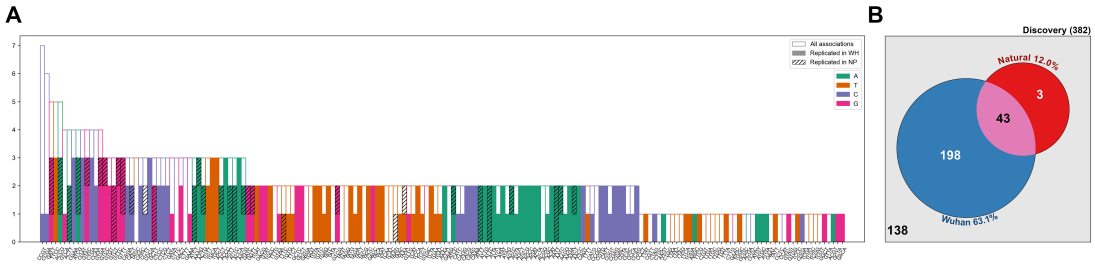
Response: We thank the reviewer for this comment. After careful consideration, we have decided to focus this manuscript on the 256 distinct cfDNA end motifs, which yielded the most statistically significant GWAS associations in our analysis. In contrast, the MPC and MDS phenotypes showed fewer genome-wide significant associations. To maintain clarity and ensure a focused narrative, we have removed these results from both the main text and the supplementary materials. However, we recognize the potential value of these results and will consider presenting them in a future study or data release for researchers interested in further exploration.

6. Not enough information is provided in the replication results – for each cohort, there is just a SNP and a p-value. It is not therefore clear exactly what is being "replicated" for each locus. Specfically:

• In the Chinese NIPT and natural Chinese replication cohorts, are the same motifs significantly associated as those in the discovery sample?

Response: We thank the reviewer for this comment. To provide a clearer picture of the replication performance, we have added a new supplementary table (Table S3) containing detailed information on the replication results. This includes all 382 study-wide significant motif–locus association pairs identified in the discovery cohort, along with the top associated SNPs in the replication cohorts, their matched motifs, genetic effect sizes, p-values, and linkage disequilibrium with the originally identified SNPs. In addition, we included a bar plot and a Venn diagram (Supplementary Figure S7) to visually summarize the replication results. Among the 382 study-wide significant motif–locus associations identified in the discovery cohort, 241 (63.1%) were replicated in the Chinese NIPT cohort and 46 (12.0%) in the natural Chinese population cohort, with a total of 244 unique associations replicated in at least one cohort. We have added and highlighted the following content in the Results section (GWAS replication studies):

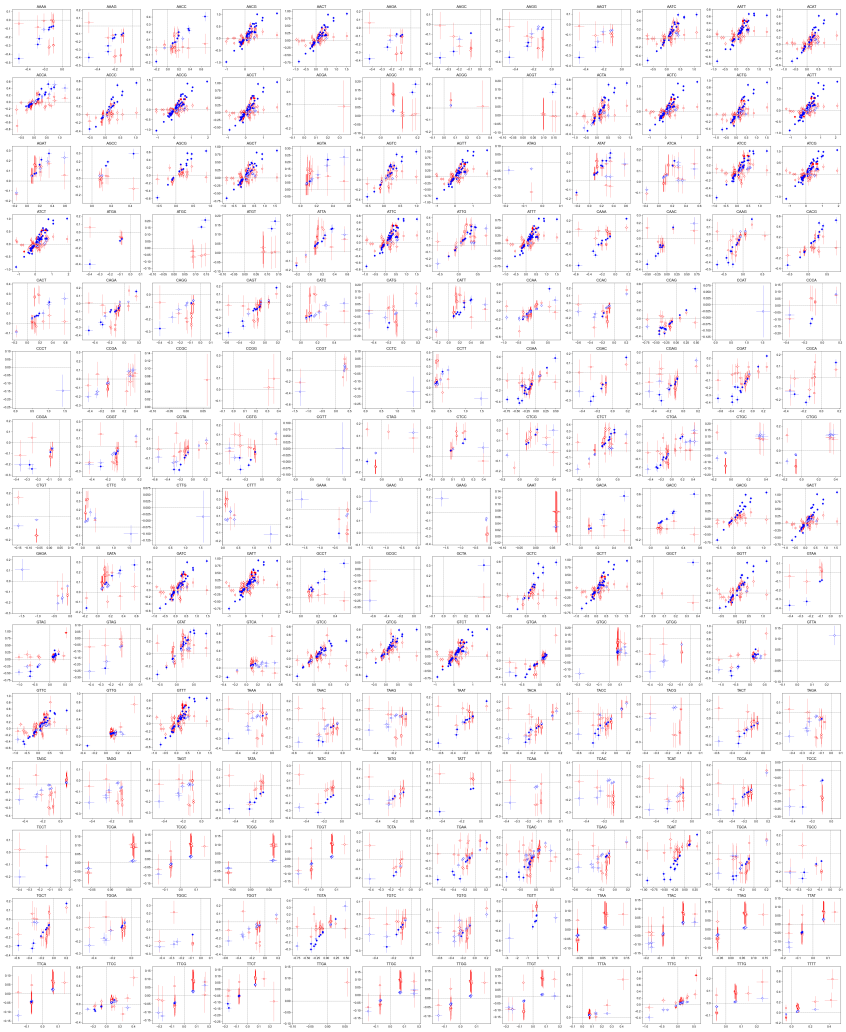
In Table S3, we present detailed replication results for the 382 study-wide significant motif–locus association pairs. Using the full GWAS summary statistics of the 256 motif frequencies from both the NIPT and natural population cohorts, we evaluated the replication of each association. For every motif–locus pair, we report the lead SNP identified in the replication cohort, along with its GWAS p-value and linkage disequilibrium information relative to the corresponding discovery SNPs. To visually summarize the findings, we included a bar plot and Venn diagram (Figure S7). Overall, 241 associations (63.1%) were replicated in the NIPT cohort and 46 (12.0%) in the natural population cohort, with a total of 244 unique associations replicated in at least one cohort.

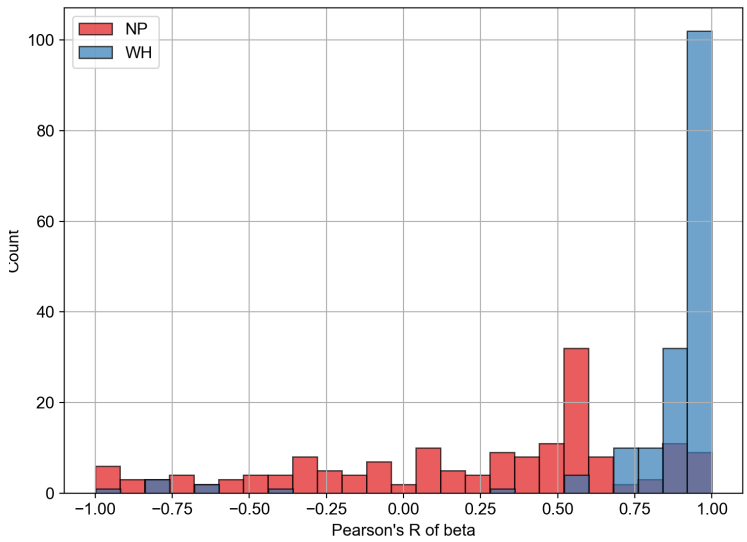


• For the motif phenotypes, it is not clear whether direction of effect is the same across discovery/replication cohorts.

Response: We thank the reviewer for this comment. To assess the consistency of effect directions between the discovery and replication cohorts (NIPT and natural population), we generated scatter plots comparing the genetic effects of significant SNPs associated with the 256 end motif frequencies in the discovery dataset and those of the same SNPs in the replication cohorts. Additionally, we calculated Pearson's correlation coefficients (R) to quantify the concordance of genetic effects between the studies. Among the 256 motifs, 158 showed R > 0.5 and 146 showed R > 0.75 when comparing the discovery and NIPT cohorts. For the natural population cohort, 69 motifs had R > 0.5 and 23 had R > 0.75. These results indicate a substantial degree of consistency in the direction and magnitude of genetic effects across datasets. We have added and highlighted the following content in the Results section (GWAS replication studies):

To further assess the consistency of effect directions between the discovery and these two replication cohorts, we generated scatter plots comparing the genetic effects of significant SNPs associated with each of the 256 end motifs in the discovery GWAS and the corresponding effects in the replication cohorts (Figure S8). For each motif, we also calculated Pearson's correlation coefficients (R) between genetic effects in the discovery study and replication cohorts (Figure S9). Among the 256 motifs, 158 showed R > 0.5 and 146 showed R > 0.75 in the NIPT cohort, while 69 and 23 met these thresholds in the natural population cohort, respectively. The NIPT cohort exhibited a higher degree of consistency in both the direction and magnitude of genetic effects, whereas the natural population cohort showed lower concordance, likely due to its smaller sample size and the non-pregnant samples.





• The SNPs reported in the replication cohorts is often different to the sample –

presumably these are good proxies (r2 > ??) with the smallest p-value at that locus, but this is not explicitly stated.

Response: We thank the reviewer for this comment. To define a successfully replicated signal, we extended each significantly associated locus identified in the discovery cohort by 500 kb on both sides and examined whether any SNPs within this extended region in the replication cohorts met the following criteria: (1) a p-value < 1e-5 and (2) linkage disequilibrium ($r^2 > 0.1$) with a significant SNP within the corresponding discovery locus. The r^2 values between the associated SNPs in the discovery cohort and the lead SNPs in the replication cohorts are provided in Supplementary Table S3. We have added and highlighted these replication processes in the Method section (Independent cohort for replication study):

To define a successfully replicated signal, we extended each significantly associated locus identified in the discovery cohort by 500 kb on both sides and examined whether any SNPs within this extended region in the replication cohorts met the following criteria: (1) a p-value < 1e-5 and (2) linkage disequilibrium ($r^2 > 0.1$) with a significant SNP within the corresponding discovery locus.

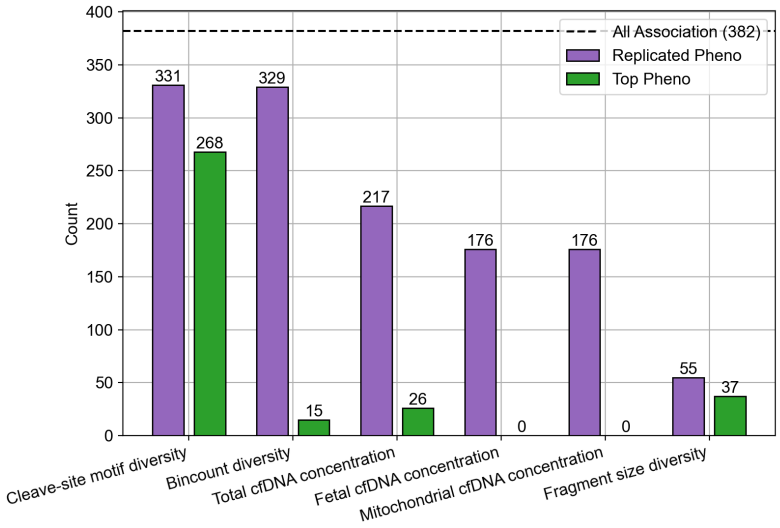
• For the Dutch NIPT samples, the phenotypes not motifs, but related phenotypes; again, for each locus please specify which phenotype(s) were associated.
I think the authors have attempted to show some of these aspects of replication through figure S6; but these plots are too small to be legibile. Instead, I would recommend the authors provide full association results for all motifs at each locus, to allow better interpretation of what truly replicated.

Response: We thank the reviewer for this valuable comment. In response, we have added a comprehensive supplementary table (Table S3) that provides the full replication results across all cohorts. For the NIPT and natural population replication studies, we included complete association statistics for each of the 382 study-wide significant motif-locus pairs, along with visual summaries such as bar plots, Venn diagrams, and scatter plots.

Regarding the Dutch cohort, where replication was based on related cfDNA phenotypes rather than end motifs, we specified in the supplementary table which phenotype(s) were associated at each replicated locus (Table S3). For each significant motif-locus pair, we listed all Dutch phenotypes that supported replication, as well as the top associated phenotype (i.e., the one with the smallest p-value). Furthermore, we summarized the involvement of each phenotype using bar plots: one indicating the total

number of motif-locus associations each phenotype was involved in, and another indicating how often each phenotype served as the top phenotype (Figure S10). Notably, the “cleave-site motif diversity” phenotype appeared most frequently in both metrics, suggesting it has the strongest associations with cfDNA end motif frequencies among the six Dutch features analyzed.

The above paragraph has been incorporated and highlighted in the revised Results section (GWAS replication studies).



7. ACAT analysis is not described in the methods. From the results, it appears the authors used ACAT to combine p-values across 256 motifs, for each SNP. Normally, ACAT is used for combining p-values of several SNPs (eg in a gene,region, pathway), for a single trait. I note that ACAT can be a general method for combining p-values, but I’m not sure whether the optimal method in this case? Can the authors provide additional justification/references for using it in this context? Alternatively, a method like C-GWAS may perhaps be more appropriate? https://doi.org/10.1038/s41467-022-35328-9

Response: We thank the reviewer for this comment. In the Discussion section of the ACTA paper (https://pmc.ncbi.nlm.nih.gov/articles/PMC6407498/), the authors noted that “As an omnibus testing procedure, ACAT in principle can be applied to combine complementary methods in nearly all kinds of genetic studies, including single-variant analysis, multiple-traits analysis, and set-based analysis such as that considered in this paper.” Based on this, we understood that ACAT is applicable to multiple-trait analyses, which aligns with our study involving 256 end motif traits. However, after careful

consideration, we have chosen to focus our manuscript on the 256 individual end motifs, which we believe represent the most statistically robust and interpretable GWAS associations in our analysis. Given that the ACAT analysis is not the primary focus of our study, we have removed these results from the main materials to maintain clarity.

8. For all motifs, a series of analyses were carried out using LDSC: heritability estimated; partioned heritability; genetic correlations.
• For several motifs in Table S3, heritibility is NA. Why is this?

Response: We thank the reviewer for this comment. In supplementary table, the heritability values are marked as "NA" for certain motifs because their estimated SNP heritability was negative when using LDSC. While negative heritability is not biologically meaningful, it can arise due to statistical estimation error. Common causes include small sample size, noisy summary statistics, or inadequate control of population stratification. In our study, the sample size and covariate adjustment were consistent across all 256 motifs. Specifically, the same number of samples were used for all motifs, and population structure was controlled for using the top 10 principal components. Since the majority of motifs showed positive heritability estimates, we believe that small sample size and population stratification are unlikely to explain the observed negative estimates. As mentioned in our response to Question 2, we conducted Kolmogorov–Smirnov (K–S) tests to assess the normality of motif frequency distributions. To explore whether deviations from normality might be related to negative heritability estimates, we generated a scatter plot of K–S test statistics (used instead of p-values due to their near-zero values) against the raw heritability values. However, we did not observe a significant correlation between the two ($p = 0.57$, linear regression). At this stage, we acknowledge that we have not identified a definitive reason why a small subset of motifs yielded negative heritability estimates. We have added a statement discussing this observation in the Limitations section of the manuscript.

Moreover, we observed that a small subset of motifs yielded negative SNP heritability estimates using LD Score Regression. While this is likely due to statistical noise rather than biological factors, the exact cause remains unclear and warrants further investigation in larger cohorts or with alternative methods.

• For the motifs with non-missing heritaability estimates, the h2 in many cases is very low, and not significantly different to 0 ($p<0.05$). For the partioned heritability analyses to be sufficiently powered, it is recomended “the dataset analyzed must have a very large sample size and/or large SNP-heritability, and the trait analyzed must be

polygenic" (https://doi.org/10.1038/ng.3404). I do not think this applies to a large number of the motifs analysed. Table S4 reports 52,070 p-values from the partioned heritability analyses for all 256 motifs. If the p-values were uniformly distributed under the null hypothesis, we would expect around 521 of these to be p < 1e-2 (ie the p-value threshold used), just by chance. In fact, only 290 p-values meet this threshold. By running the partioned heritability analyses for all motifs, the authors are greatly increasing their multiple testing burden, while obtaining results for several motifs which are not likely to be reliable, due to the low heritabilites. A better approach would be to run these additional anlysis for motifs only where there is evidence of significant SNP-heritability. Furthermore, I'd recommend applying an FDR adjustment (eg Benjamni-Hochberg) to determine significant results, rather than an arbitrirary nominal threshold.

Response: We thank the reviewer for this insightful comment. We fully agree that the relatively low SNP heritability estimates for many motifs limit the statistical power of the partitioned heritability analyses, as highlighted in the referenced study (https://doi.org/10.1038/ng.3404). As the reviewer pointed out, if the p-values were uniformly distributed under the null hypothesis, we would expect approximately 521 p-values < 0.01, whereas only 290 such p-values were observed, suggesting limited signal. We had indeed attempted several multiple testing correction approaches, including Bonferroni correction and Benjamini–Hochberg (BH) FDR correction. However, no associations remained significant after correction. Despite this, we believed that reporting nominally significant results (p < 0.05) could still provide useful hints about potential motif–cell type associations, while clearly acknowledging their exploratory nature. In light of the reviewer's suggestion, we have explicitly stated this limitation in the Results section ("Enrichment in hematological cell types"):

It should be noted that these results were based on a nominal significance threshold of 0.05 without strict multiple testing correction. Therefore, the findings should be interpreted with caution and regarded as exploratory, providing preliminary indications of potentially associated tissues and cell types.

• The above point also applies to the genetic correlation analyses.

Response: We thank the reviewer for this insightful comment. As we stated in the manuscript, "At the strictest Bonferroni-corrected significance threshold of 1.88e-6 (=0.05/ (256*104)), no motif-phenotype pairs reached significance. However, when we relaxed the threshold to a more lenient suggestive level of 1e-3, a total of 61 motif-phenotype pairs showed significant correlations (Table S9). To further ensure the

reliability of the results, we retained only pregnancy phenotypes with more than five associated motifs." As the reviewer pointed out, the low SNP-heritability estimates for many motifs also limit the statistical power of the genetic correlation analyses, resulting in no motif-phenotype pairs passing the stringent multiple testing correction threshold. Additionally, we would like to clarify that LDSC produces p-values with a minimum resolution of 1e-4; therefore, for four pairs with reported p-values of 1e-4, the true p-values might be even smaller, but the precision of LDSC is limited at this level. We have now added and highlighted the following limitation at the end of this section:

These genetic correlation results were based on a suggestive significance threshold of 1e-3 without strict multiple testing correction. Thus, the findings should be interpreted with caution and considered exploratory.

9. Regarding the Mendelian Randomisation:

• How were all of the exposures/outcomes selected? Was this based on the genetic correlations, or just the >3 significant SNP requirement?

Response: We thank the reviewer for this comment. As noted in the previous response, although some phenotype–motif pairs showed high genetic correlation coefficients (rg), none passed the study-wide significance threshold. Therefore, the MR analyses were not based on the genetic correlation results. Instead, we systematically tested the causality of all phenotype–motif pairs. For the selection of exposure phenotypes, we included only those with more than three independent significant SNPs after LD clumping, ensuring adequate instrument strength. This criterion was applied consistently in both directions of the MR analysis (i.e., phenotype-to-motif and motif-to-phenotype). For the outcome variables, no additional selection criteria were applied—as long as the selected instrumental variables were available in the outcome GWAS dataset and not significantly associated with the outcome, the outcome phenotype was retained for analysis.

• If I understood correctly, the pregnancy phenotype GWASs used for the MR used the same individuals as the motifs GWASs, thereby making this a one-sample MR. MR-Egger is an approach developed for two-sample MR, which in one-sample MR may indicate a significant causal effect more often than it should (https://doi.org/10.1093/ije/dyab084). There is also a OneSampleMR R package which has other estimators, specifically for one sample MR: https://doi.org/10.32614/CRAN.package.OneSampleMR

Response: We thank the reviewer for this insightful comment. We fully agree that the MR-Egger method, while useful for detecting and adjusting for horizontal pleiotropy, can yield biased estimates in the context of one-sample MR due to sample overlap, as highlighted in the cited study (https://doi.org/10.1093/ije/dyab084). To mitigate this, we complemented MR-Egger with the IVW method and only considered results as robustly causal when both methods reached study-wide significance ($p < 5e\text{-}6$). While MR-Egger is sensitive to bias in one-sample settings, its results remain informative if the variability in instrument strength is high. Meanwhile, the IVW method is generally robust in overlapping-sample scenarios and provides a more conservative estimate.

We also appreciate the reviewer's suggestion regarding the OneSampleMR package. Although the exposure (pregnancy phenotypes) and outcome (cfDNA motif frequencies) datasets partially overlap rather than representing a true one-sample setting, we agree that additional validation is necessary. Therefore, we applied the MRlap method (https://github.com/n-mounier/MRlap), which is specifically designed for two-sample MR with potentially overlapping samples and accounts for multiple sources of bias, including weak instruments, winner's curse, and sample overlap. We re-analyzed the 156 associations that were significant in both MR-Egger and IVW using MRlap. The MRlap results remained consistent, yielding significant associations at the same Bonferroni-corrected threshold ($p < 5e\text{-}6$), with p-value magnitudes comparable to those obtained from the IVW method. These results have been added to Supplementary Table S12. We have included and highlighted the following statements in the Methods section (Mendelian randomization):

Due to sample overlap between the cfDNA motif GWAS and pregnancy phenotype GWAS, and to minimize potential bias in MR-Egger results, we additionally performed the MRlap method. MRlap is a two-sample MR method that accounts for sample overlap, weak instrument bias, and winner's curse. We applied MRlap to the 156 associations identified by both MR-Egger and IVW, and confirmed their significance.

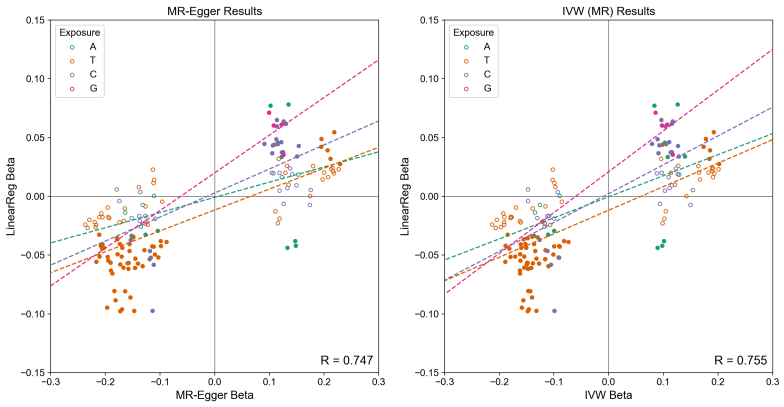
And in the Results section (Post-GWAS analyses with one-sample pregnancy phenotypes):

At the Bonferroni-corrected significance threshold of 5e-6 based on both MR-Egger method and inverse variance weighted method, our P->M analysis identified 156 potentially causal associations involving 5 pregnancy phenotypes, all belonging to the hematological category; these associations were further confirmed using MRlap, which accounts for sample overlap and other biases.

• Given the pregnancy phenotypes and motifs were measured in the same indviduals, it

would be useful to compare effect estimates from MR with the observed associations between these variables.

Response: We thank the reviewer for this insightful comment. In response, we performed linear regression analyses regressing motif frequencies on pregnancy phenotypes for the 156 motif-phenotype pairs identified as significant in the MR analyses. We found that 141 pairs (90.38%) exhibited concordant directions of effect estimates between the MR results and the observed phenotypic associations. Furthermore, the Pearson's correlation coefficients (R) between the MR-predicted and observed effect estimates were 0.747 for the IVW method and 0.755 for the MR-Egger method, indicating a high degree of consistency.



10. The methods overall are generally lacking in some detail:

• As mentioned in comment 1, more details are needed regarding the imputation. What parameters were used for STITCH? (ie k and nGen ??). Was any filtering done in terms of imputation quality (INFO score threshold?)

Response: We thank the reviewer for this comment. As detailed in our response to Question 1, we have provided a comprehensive explanation for the selection of STITCH parameters, including the values used for K and nGen. These values have also been added to the revised Methods section for clarity. Regarding imputation quality control, we did not apply an explicit INFO score threshold because, after filtering SNPs with MAF > 0.05 and Hardy–Weinberg equilibrium p-value > 1e-6, all retained SNPs had INFO scores greater than 0.5—commonly considered an acceptable benchmark for downstream analyses.

• For the GWAS, were related individuals removed? Association analyses were

undertaken uing Plink, which does not account for realtedness, but it is not stated that the cohort was restricted to an unrelated set.

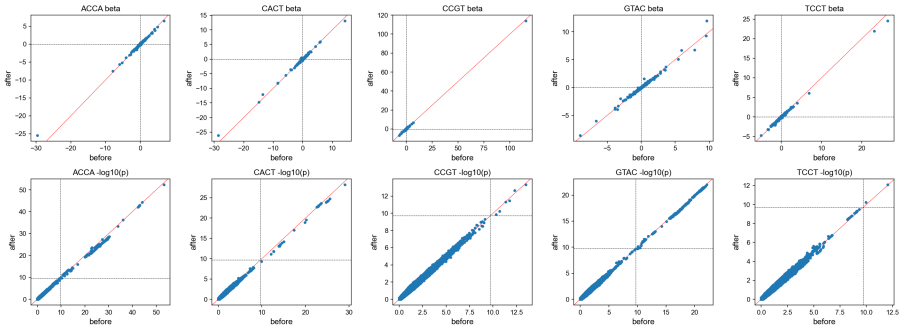
Response: We thank the reviewer for this valuable comment. In our study, we did not remove related individuals from the cohort. We generated two types of variant files at the population level: one was based on variants called by BaseVar, and the other on variants imputed by STITCH. The BaseVar-called variant files exhibited extensive missingness and were predominantly haplotypic in nature; using these files, the KING software outputted NA values, and PLINK produced -INF kinship estimates, making relatedness inference infeasible. We also attempted kinship estimation based on the STITCH-imputed variants. However, both KING and PLINK reported an unexpectedly large number of related individuals, which we considered likely to be artifacts, given that our cohort primarily consisted of unrelated pregnant women aged between 20 and 40 years.

```
Summary of Kinship Analysis
---------------------------
Total number of samples: 28016
Total number of pairs analyzed: 392434120

Kinship Coefficient > 0.354 (Identical twins or clones): 220 pairs, 440 samples
Kinship Coefficient 0.177 - 0.354 (First-degree relatives): 259 pairs
Kinship Coefficient 0.0884 - 0.177 (Second-degree relatives): 49764924 pairs
Kinship Coefficient 0.0442 - 0.0884 (Third-degree relatives): 237790277 pairs
Kinship Coefficient < 0.0442 (Unrelated or distant relatives): 104878440 pairs

Distribution of Kinship Coefficients:
- Mean: 0.06004523781616244
- Median: 0.0580052
- Standard Deviation: 0.023710003586716426
```

To assess the potential impact of relatedness on our GWAS results, we randomly selected 10 motif frequency phenotypes (covering a range of signal strengths) and reran the GWAS after removing identical twins and first-degree relatives identified above. The p-values were found to be highly consistent between analyses with and without relatedness filtering. This suggests that, given the relatively small proportion of potentially related individuals compared to the overall sample size (>20,000), the presence of relatedness had minimal influence on the results.



We have added and highlighted the following content in the Limitation section:

Currently, no robust approaches are available for inferring kinship relationships based on NIPT-like ultra-low-depth sequencing data. Neither BaseVar-called variants nor STITCH-imputed genotypes yielded reliable kinship estimates. Given that the cohort consisted predominantly of middle-aged pregnant women, we expect the proportion of close relatives to be very small, and their influence on the GWAS results to be negligible. Nevertheless, the development of appropriate methods for kinship estimation under ultra-low sequencing depth remains an important area for future research.

• Similarly although ancestry PCs were calculated, with 10PCs used as covariates in the GWAS, it is not mentioned whether any exclusions were made based on ancestry?

Response: We thank the reviewer for this comment. Based on the PCA results, we assessed ancestry outliers by examining whether each sample fell within six standard deviations from the mean along both PC1 and PC2. We found no outliers under this criterion and therefore did not exclude any samples based on ancestry. In our previous work (Figure S3d, https://pmc.ncbi.nlm.nih.gov/articles/PMC11602577/), we presented the PC1–PC2 plot, which showed that all NIPT samples clustered centrally, consistent with the recruitment site in Wuhan, China.

[editorial note: third party material redacted]

Reviewer #2 (Remarks to the Author):

This article presents the GWAS results for end motifs in a large dataset of pregnant women. The study identifies 23 study-wide signicant loci, several of which are novel. The study is replicated in three independent cohorts. In addition, one of the genes is further validated by knock out experiments in mice and cell lines. This is a very valuable study, providing novel insights in the factors affecting cfDNA fragmentation. Those insights are of utmost importance for the booming field of cfDNA analyses and cfDNA fragmentomics. The article is well written, the figures are clear and systematic.

Response: We sincerely thank the reviewer for the positive and encouraging comments on our work. We greatly appreciate the time and effort the reviewer dedicated to carefully evaluating our study and providing thoughtful feedback. Below are our responses and corresponding revisions:

Major comment:
The title suggest the genetic basis of cell-free DNA features is studied. However, the study only examines one of the features of the cell free DNA, namely the end motif frequencies and not other aspects e.g. fragment length, jaggedness, ….. Hence, correct the title accordingly to 'cfGWAS reveal genetic bases of cell-free DNA end motifs' would be more suitable.

Response: We thank the reviewer for this comment. Following the reviewer's suggestion, we have revised the title to "cfGWAS reveal genetic basis of cell-free DNA end motifs" to more accurately reflect the scope of our study.

Minor comments
The study is performed using cfDNA from pregnant women, but the data is analysed as if the cfDNA is derived only from the mother. Since about 10% of the cfDNA is derived from the fetus, which has a different genotype, one could consider there to be an effect on the results/analyses. Would this influence the results/power? How would that affect the heritability scores? Can placenta specific signals be misassigned to the mother? It might be valuable to comment/deliberate this aspect?

Response: We thank the reviewer for this important comment. Indeed, during the NIPT testing window (gestational weeks 12–22), the fetal fraction typically ranges from 5% to 10%. Considering that approximately half of the fetal DNA is maternally inherited, the effective proportion of "non-maternal" DNA contamination is estimated to be around 2.5%–5%. In our previous work, we demonstrated that jointly modeling

maternal and fetal genotypes during imputation is theoretically feasible and can improve the accuracy of maternal genotype imputation while simultaneously estimating fetal genotype dosages (https://pmc.ncbi.nlm.nih.gov/articles/PMC11602596/). However, we acknowledge that treating all cfDNA as maternal in downstream statistical analyses, without distinguishing the fetal contribution, could introduce some degree of bias. We have now added and highlighted this point as a limitation of the study, as follows:

Notably, at the time of NIPT testing, the fetal DNA fraction in maternal plasma is approximately 10%, with half of it being identical to the maternal DNA. In our genotype imputation and GWAS analyses, we did not distinguish between maternal and fetal DNA, but rather treated all cfDNA as maternal. While this approach simplifies the analysis, it may introduce minor biases, particularly for loci with strong placental or fetal-specific signals. Recently, a newly developed genotype imputation method, QUITL2, which allows simultaneous imputation of maternal and fetal genomes. In ongoing work, we are applying this method to our NIPT dataset to separate maternal and fetal genomes for future GWAS and post-GWAS analyses, which will help further refine the findings.

Reviewer #3 (Remarks to the Author):

In this paper, the authors perform GWAS on data from ~28000 NIPT samples. Their main aim is to unravel genome wide associations with the 256 four-nucleotide cfDNA end-motifs. By doing so, they hope to gain more insight into the biology of cfDNA. As they performed single-end sequencing, they could not take into account fragment length association studies. They used one dataset for the primary analysis, and two subsequent non-overlapping datasets to validate their data. These three datasets were from pregnant women. They finally performed an independent study on a small dataset with data from nonpregnant women. As the PANX1 association was considered to be a novel finding, they subsequently performed experimental verification analysis of the biological role of PANX1. Also, the effects of hematological cells on cfDNA end-motifs was studied and a strong causal relationship between neutrophils and end-motifs was noticed.

The study is fully based on statistical analysis, but I did not dive into the statistical methods and trust other reviewers with more relevant expertise will do so.

Response: We sincerely thank the reviewer for taking the time to carefully review our manuscript and provide invaluable feedback. Below are our responses and corresponding revisions:

1. In general, pages 3-20 are main text and the supplemental data, in which the methods are described, consists of an additional 7 pages (text only). With ~400 words per page (estimation), this means the main text of the paper itself is ~7200 words and the supplemental text ~ 2800 words. Moreover, there are 86 references in the main paper and an additional 25 in the supplemental text. I realize the authors present a very extensive study, with a lot of data, but the paper would certainly benefit from less text. The findings of the involvement in cfDNA biology of genes like DNASE1L3 and DFFB is not new, but because of the many different studies discussions on such not-novel findings also take quite some text. Perhaps the authors could focus more on the added value of finding comparable associations with their study and on the novel findings.

Response: We thank the reviewer for this valuable comment. After careful consideration, we have decided to focus our manuscript on the analysis of the 256 distinct end motifs, which we believe represent the most statistically robust GWAS associations. Accordingly, we have removed the sections related to MPCs, MDS, and ACAT analyses from both the main text and supplementary materials to enhance the clarity and focus of the manuscript. In addition, we have condensed and streamlined the descriptions throughout the manuscript to improve conciseness. However, we would like to note that

due to the incorporation of additional analyses and expanded discussions as requested by the reviewers, the overall length of the manuscript remains relatively substantial.

2. The authors mention the identification of the involvement of the PANX1 gene is a novel finding, but Linthorst et al. (their reference 49) previously also identified an association with cfDNA biology of this gene in their GWAS study. Furthermore, the authors obtained the complete summary statistics of the study of Linthorst et al. but as far as I can see the Dutch group is not acknowledged. I suppose the authors have an agreement with the group on this, but if not, it feels unjust.

Response: We thank the reviewer for this valuable comment. In the study by Linthorst et al., six cfDNA features were analyzed, including bincount diversity, fetal concentration, mitochondrial DNA concentration, fragment size diversity, motif diversity score, and total cfDNA concentration; however, cfDNA end motif frequencies were not among the phenotypes they investigated. We have carefully reviewed our statement and clarified that the association between the *PANX1* gene and cfDNA end motif frequencies was newly identified in our analysis. Meanwhile, we also examined the correlation between the six previously studied cfDNA features and the cfDNA end motif frequencies, suggesting that if Linthorst et al. had also analyzed end motif frequencies, *PANX1* might have been implicated in their study as well. We have properly cited the work of Linthorst et al. in our manuscript, and their summary statistics are publicly available. In addition, our study represents the first cfDNA end motif GWAS conducted in an Asian population, providing a valuable complement to the Dutch study in terms of both phenotype focus and population diversity.

3. In the (recent) past many publications appeared dealing with studies on the biology of cfDNA. This is a very interesting topic, but nevertheless, much is still unknown. I am following these studies with much interest, meanwhile thinking about the clinical relevance / consequences. How can these findings be translated to / be of use in daily practice? Apparently, the authors have some ideas on this, as in the last two sentences of the abstract the authors mention “Novel knowledge uncovered by this study promises to revolutionize liquid biopsy technology” and “the potential of this paradigm is enormous”. Also in the last paragraph of the discussion, the authors mention their research “could contribute to the discovery of therapeutic targets for diseases influenced by abnormal DNA concentrations”, and “the significant discoveries will not only revolutionize our understanding of cfDNA biology but also open new avenues for future clinical applications”. I would like to ask the authors to elaborate more on this.

Response: We thank the reviewer for this thoughtful comment. Our study primarily focuses on statistical genetic analyses to identify genetic variants associated with cfDNA end motif frequencies, as well as the phenotypic correlations and causal relationships between cfDNA end motifs and pregnancy-related traits. Although we performed experimental validations in knockout mice and cell lines, confirming that cfDNA motif frequencies differ between wild-type and knockout conditions, we did not fully elucidate the underlying biological pathways or mechanisms leading to these differences. Nevertheless, we believe that our findings provide important genetic insights into cfDNA generation and clearance processes. As we discussed in the manuscript, with large sample sizes, cfGWAS has the potential to identify more credible genes associated with cfDNA molecular features, thereby enhancing our understanding of cfDNA biology. This knowledge could ultimately contribute to future clinical applications of cfDNA, such as improving liquid biopsy technologies. For instance, since circulating tumor DNA (ctDNA) often exists in low concentrations, strategies aimed at protecting cfDNA from degradation or reducing its clearance could help maintain sufficient ctDNA levels, allowing sensitive detection even with minimal blood volumes. In addition, we have expanded our discussion regarding the biological significance of cfDNA end motifs. In the revised manuscript, we included the following content:

In the Result section:

We presented the distributions of the 256 motif frequencies using bar plots and box plots (Figure S2). Overall, the distributions of motif frequencies are consistent with previously reported patterns. For example, the motif CCCA exhibits the highest frequency, C-end motifs tend to have higher average frequencies than others, and motifs containing CG dinucleotides generally show lower frequencies. Different cfDNA end motifs reflect nuclease activity, pathophysiological processes, and tissue-specific DNA fragmentation patterns. Specifically, DFFB, DNASE1L3, and DNASE1 preferentially generate A-end, C-end, and T-end motifs, respectively. Recent studies have highlighted cfDNA end motifs as emerging biomarkers for oncology and transplantation monitoring due to their frequency shifts under pathological conditions and their tissue-specific profiles.

In the Discussion section:

CfDNA end motifs have been increasingly explored as biomarkers in cancer detection and transplantation monitoring. Our team have recently uncovered their role in predicting preeclampsia during early pregnancy. Pregnancy also offers a valuable

model to study tissue-of-origin dynamics (maternal vs. fetal). To our knowledge, this is the first study to explore the causal relationships between cfDNA end motifs and physiological and biochemical phenotypes in a pregnant population. We found that T-end motifs exhibited the most causal associations and colocalizations with pregnancy phenotypes. This may be related to placenta-specific nuclease activity, hormonal changes, or maternal immune adaptation during pregnancy. While our current analysis cannot fully explain this observation, it highlights a promising area for further investigation into the biological and clinical significance of cfDNA end motif dynamics in pregnancy.

4. As the authors mention themselves, a limitation of the study is first of all the relatively small initial cohort size, but more importantly the cohorts consisting of data from pregnant women only. Only the data used for the independent study were from non-pregnant individuals, but this was a small cohort size. When replying on comment 3, please take this also into account. When thinking about clinical applications, will these only be applicable to pregnant women? The two other main fields of interest in cfDNA are cancer and transplantation. Will potential “new avenues” also be applicable to these fields, and also to men or only to women? Are differences to be expected? The authors mention that by now worldwide more than 40 million women have undergone NIPT, but what about data from the other fields and data from men?

Response: We thank the reviewer for raising this important point. In our study, we introduced an independent validation cohort composed of non-pregnant individuals and successfully replicated several significant signals such as *PANX1*, *ABO*, and *MSR1*, suggesting that there may be a shared genetic background influencing cfDNA end motif frequencies across both pregnant and non-pregnant populations. At the same time, we also identified signals that were replicated only in the pregnant cohort but not in the natural population cohort, including *FCHO2* and *PSMD3*. Further analysis showed that *FCHO2* is highly expressed in female reproductive organs such as the uterus, vagina, and cervix, indicating that it may represent a pregnancy-specific cfDNA signature. As described in the manuscript:

"Four loci—*FCHO2*, *PSMD3*, and two long non-coding RNA genes—were replicated by two NIPT studies but not in the natural population, whereas *NAMPTP2* was replicated only in the natural population and not in the two NIPT studies. Based on these results, we can conclude that the majority of cfDNA-associated genes are shared between pregnant and non-pregnant individuals. If we define pregnancy-specific loci as those replicated by NIPT studies but not in the natural population, then there are four such loci. Notably, *FCHO2* is highly expressed in female reproductive organs, including the uterus, vagina, and cervix."

We acknowledge, however, that the sample size of our non-pregnant validation cohort (~400 individuals) was relatively small compared to the tens of thousands of cases in the pregnant discovery cohorts, leading to lower statistical power in the non-pregnant group. We have clearly stated this limitation in the manuscript: "Furthermore, our discovery set relied on a cohort of pregnant women, so it’s unclear how many findings are general for all people or specific to pregnancy. Large-scale independent studies on non-pregnant cohorts are necessary to resolve this uncertainty."

Regarding clinical applications, while some signals may be pregnancy-specific, the majority of genetic associations we uncovered are likely to be relevant across different populations, including men and non-pregnant women. Therefore, we believe that the “new avenues” mentioned in our discussion could potentially extend to other major cfDNA research fields, such as oncology and transplantation medicine, although further studies will be necessary to comprehensively characterize any differences.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors provided summary statistics with their revised manuscript. The information provided in these files raises serious concerns regarding the results of the GWAS. While the authors state "all retained SNPs had INFO scores greater than 0.5", in these summary results files, there are many SNPs where the INFO_SCORE column has a value < 0.5. A quick look-up in the uploaded file cfGWAS.m256.T___.csv of a couple of the reported SNPs in Table S1 found at least one with very low imputation quality (e.g., rs529723778 has INFO_SCORE of 0.027). I haven't done a comprehensive look-up, but would assume this may be the case for others. That isn't to say that all reported variants are likely to be wrong, but there may be a number of false positive hits reported, and the genome-wide results used in their genetic correlations etc. are likely to contain a lot of noise.

Unless the INFO_SCORE column in the uploaded summary stats is incorrect, or I have misinterpreted (hopefully this is the case!), much of the analyses should be repeated after filtering out variants with very low imputation quality.

Responses: We sincerely thank the reviewer for the careful examination of our summary statistics and for pointing out this important issue. The INFO score shown in our previous figure represents the average imputation quality within each MAF interval, rather than the INFO score for individual variants. Although this was described in the Methods, it was not clearly emphasized in our previous response, which led to this misunderstanding.

In the revised version, we have explicitly added the INFO score as a variant-level filtering criterion. Specifically, in addition to the thresholds of MAF > 0.05 and HWE p > 1e-6, we now apply an INFO score > 0.4 cutoff in all GWAS analyses. Variants below any of these thresholds have been completely excluded. The updated GWAS summary statistics have been uploaded and are available at: https://drive.google.com/drive/folders/1Um6bPoBgw4Xokl677T0CDjD_8wxUjTvF?usp=sharing.

Furthermore, we have re-performed all downstream analyses, including genetic correlation, Mendelian randomization, replication, and pathway-based analyses, based on the new GWAS results. The updated results have been incorporated into the revised manuscript. Importantly, the main findings and overall conclusions remain highly consistent with the original version. For example, the most significant associations continue to involve *PANX1*, *DNASE1L3*, and *DFFB*; replication rates have slightly

improved; and the key enriched pathways remain those related to DNA fragmentation, caspase activity, and apoptosis. In addition, causal relationships inferred by MR analysis still remained between hematological phenotypes, particularly leukocyte traits, and cfDNA motif frequencies.

To facilitate comparison, we have provided a separate file “05-What was changed in this revision.docx” presenting the updated key results and figures alongside the previous versions, enabling a clearer visualization of the changes made.

I address the reviewers' responses to my original comments below:

1. Overall, the additional detail regarding imputation has adequately addressed my comments. In the response about INFO score thresholds, the authors provide a plot (S1C of their previous paper). It is unclear what the plotted INFO_SCORE is showing here; I'd assume median/mean(?) INFO score, per MAF bin. So while the average INFO score across all SNPs was >0.5, it isn't clear that each SNP had INFO > 0.5. Indeed, the summary statistics uploaded by the authors seemed to confirm that many variants had INFO < 0.5.

Responses: We sincerely thank the reviewer for this comment. We apologize for the lack of clarity in our previous description. In the revised version, we have explicitly added the INFO score as an additional variant-level filtering metric. Specifically, we now require INFO > 0.4 for all retained variants, in addition to MAF > 0.05 and HWE p > 1e-6. All subsequent analyses, including GWAS, replication, Mendelian randomization, and genetic correlation, have been updated accordingly.

2. Thank you for the additional clarification regarding trait transformation.

Responses: Thank you.

3. The additional explanation is appreciated.

Responses: Thank you.

4. The extra fields requested are available now in the uploaded summary stats but should also be included in the paper (e.g., table S1).

Responses: We thank the reviewer for this comment. These extra fields of SNPs have now been included in Supplementary Table S3 of the revised manuscript.

5. The additional information is welcome. Betas should be provided in table S3 in addition to p-values so consistency of direction of effect can be seen. It appears from Figures S8 and S9 that some variants have opposite directions of effect in the replication cohorts, compared to the discovery. If directions of effects are inconsistent, these shouldn't really be classified as a "replicated" association.

Responses: We thank the reviewer for this comment. Accordingly, beyond assessing p-values, we have verified that the effect direction (beta) of the top SNP was consistent across discovery and replication studies for all replicated associations. The updated results have been incorporated and highlighted in the main text, and the beta values for these top SNPs in discovery and replication studies have been provided in Supplementary Table S3.

To further assess the consistency of effect directions between the discovery and these two replication studies, we generated scatter plots comparing the genetic effects of significant SNPs associated with each of the 176 end motifs that contains significant associated loci in the discovery GWAS and the corresponding effects in the replication studies (Figure S8). We note that for the 256 and 44 motif-locus associations replicated in the Chinese NIPT and the natural population study, respectively, the effect directions of their top SNPs were consistently identical between the discovery study and the replication study (Table S3).

6. The comparison of MR vs observed observations is interesting. Thank you for providing this; perhaps worth considering for a supplementary figure!

Responses: We thank the reviewer for this suggestion. We have added this figure as Supplementary Figure S17.

7. Perhaps worth trying to repeat relatedness estimation after filtering for poorly imputed variants.

Responses: We thank the reviewer for this comment. We have recalculated the relatedness after filtering out variants with MAF, HWE, and INFO score. Below are the results:

Relatedness (pairs)	Identical twins	1-degree relatives	2-degree relatives	3-degree relatives	Unrelated relatives
Genotype data without filtering	220	259	49,764,924	237,790,277	104,878,440

Genotype data after filtering	220	207	8,748,364	161,135,145	222,550,184

The new results showed that the numbers of predicted identical twins and 1-degree relatives barely changed, but the numbers of 2-degree and 3-degree relatives were dramatically reduced, and conversely, the number of unrelated relatives were significantly increased. The new results are more reasonable and improved, although an excess of related pairs still remains. We have added and highlighted the following sentence in the Discussion section:

When applying additional quality control, using more reliable variants, such as those filtered by INFO score, results in more reasonable estimates of relatedness.

Reviewer #2 (Remarks to the Author):

I am happy with the answers to my comments.
Just a minor comment:
In you answer you state 'QUITL2, which allows simultaneous imputation of maternal and fetal genomes'.
I believe this method is QUILT2.

Responses: We thank the reviewer for the meticulousness. We have revised it as “QUILT2”

Reviewer #3 (Remarks to the Author):

Even though the paper was rejected in the first place, the authors submitted a revised version. The authors did a tremendous piece of work, and thoroughly revised the paper. I would like to thank the authors for their efforts!

As far as my previous comments are concerned: the authors responded sufficiently to these. My first comment, however, is, despite the response of the authors, not completely resolved, but I leave it to the editor how to proceed with this. The paper is well-written, but it is still very lengthy. Moreover, the results section is full of background information and partly even discussion. I understand why the authors have chosen this format, as otherwise the rationale behind the different analyses would not be understood and the reader would be completely lost, but the fact is that this makes the paper very lengthy and introduction, results and discussion are not fully separated.

Responses: We thank the reviewer for this valuable comment. We have carefully reviewed both the main text and the supplementary materials and have condensed the content accordingly. Specifically, we have removed repetitive descriptions across sections and transferred detailed methodological explanations to the Supplementary Information. Although the main text has been substantially shortened, the overall manuscript remains relatively long due to the comprehensive nature of the analyses and results.

Even though in general, the paper is well-written, throughout the paper there are a some typos and grammar errors, e.g. (but not limited to):
Line 87: ….only the three genes DFFB, …
Line 98: locus = loci
Line 441-442: Consequently, the origin ….. to the plasma. Incorrect sentence.
Lines 448-449: ….in which the cfDNA motifs…. Change into: …with the cfDNA motifs….
Lines 593-595: Recently, a newly….fetal genomes. Incorrect sentence.
Please go through the paper thoroughly to correct these.

Responses: We thank the reviewer for this comment. We have carefully rechecked the entire manuscript to ensure that all typographical and formatting errors have been corrected.

Reviewer #1 (Remarks to the Author)

Thank you for addressing the concerns raised in my previous review. I'm pleased to see that the analyses have been re-done with the INFO filter, along with the updated MAF and HWE p-value filters. These changes, as anticipated, have indeed resulted in a more robust study with the increased replication rate. It's great to see that the majority of findings still hold. This revised version provides a much stronger proof of concept for the utility of cfDNA in genetic studies.

Responses: We sincerely thank the reviewer for all the time and efforts dedicating to our work. With the constructive and insightful comments, our work has been significantly improved. We are deeply grateful for the support and recognition.

I do have a couple of final minor points the authors should check:

1. On page 9, lines 240-41: The phrase "among others" following the list of replicated genes (PANX1, DFFB, and PADI4) seems superfluous and should be removed.

Responses: We thank the reviewer for this comment. We have revised and condensed the sentence as follows:

Overall, three loci (*PANX1*, *DFFB*, and *PADI4*) were consistently replicated across all three studies, indicating shared genetic associations between pregnant and non-pregnant individuals.

We also noticed that the preceding paragraph (Lines 236–244) could be further condensed. The original version is as follows:

In the Chinese NIPT study, 14 out of 15 (93.3%) loci were replicated at a replication significance threshold of 1e-5, including PANX1, DFFB, DNASE1L3, and MSR1 (Figure 2C, Table S2). Among the 15 significant loci, the Dutch study successfully replicated 14 (93.3%), including PANX1, DFFB, DNASE1L3, and DNASE1L1 (Figure 2C, Table S2). Furthermore, 3 loci (20.0%), including PANX1, DFFB, and PADI4, were successfully replicated in the natural population study (Figure 2C, Table S2), demonstrating that the associations between cfDNA motif features and these genes are not limited to the pregnant population but are also present in the general population.

We have revised and highlighted this paragraph as follows:

At a replication significance threshold of 1e-5, 14 (93.3%), 14 (93.3%), and 3 (20.0%) loci were replicated in the Chinese NIPT, Dutch, and natural population cohorts, respectively (Figure 2C, Supplementary Data 2). Specifically, *PANX1*, *DFFB*, *DNASE1L3*, and *MSR1* were replicated in the Chinese NIPT cohort; *PANX1*, *DFFB*, *DNASE1L3*, and *DNASE1L1* in the Dutch cohort; and *PANX1*, *DFFB*, and *PADI4* in the natural population cohort. The replication rates were notably higher in the pregnant cohorts than in the non-pregnant cohort, likely due to larger sample sizes. These findings suggest that the identified genetic associations are not limited to pregnant population but also in the general population.

2. On page 10, lines 255-258 (and in the response to previous comment 5): The statement about consistent effect directions across studies isn't fully supported by Figure S8 and Table S3. There appear to be a small number of significant associations showing opposite effect directions in one of the replication cohorts (I think all are the natural population cohort) compared to the discovery study. This nuance should be accurately reflected in the text.

Responses: We thank the reviewer for this valuable comment. We completely understand the confusion arising from our statement that "the effect directions of their top SNPs were consistently identical between the discovery study and the replication study", as Figure S8 and Table S3 show a few points in the second and fourth quadrants. To clarify, this statement referred specifically to the **top SNPs** within each **replicated locus**. The points appearing in the second and fourth quadrants correspond either to non-replicated SNPs (hollow dots in Figure S8) or to non–top SNPs within replicated loci. For the 256 and 44 top SNPs (representing motif–locus associations) replicated in the Chinese NIPT and natural population cohorts, respectively, the effect directions were indeed consistent between the discovery and replication studies.
To avoid further confusion, we have revised Figure S8 to display only the top SNPs for the significant association in discovery study (solid dots indicate replicated, while hollow dots indicate non-replicated) and have bolded the beta values of the replicated motif-locus associations (top SNPs) in Table S3. We have also clarified the corresponding description in the main text as follows:

For the 256 and 44 replicated motif-locus associations in the Chinese NIPT and natural population cohorts, respectively, we further examined whether their top SNPs showed

consistent effect directions between the discovery and replication studies (Supplementary Figure 8, Supplementary Data 3). The scatter plots demonstrated that all replicated motif–locus associations (solid dots) exhibited consistent effect directions across the corresponding studies.