Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Herbal, Dietary, and Non-Prescription Medication Use and Spontaneous Abortion: A Case–Control Study with Exploratory Analysis of Pineapple-Associated Molecular Targets

Reza Saeedinia
IRAN University of Medical Sciences
Haniyeh Rafiepoor
Tehran University of Medical Sciences
Leyla Sahebi
Tehran University of Medical Sciences
Fereshteh Amini
Imam Khomeini Hospital Complex, Tehran University of Medical Sciences
Kyana Jafarabady
Alborz University of Medical Sciences
Ava Hemmat
Mashhad University of Medical Sciences
Zohreh Heidary
z-heidary@sina.tums.ac.ir

Imam Khomeini Hospital Complex

Article

Keywords: Abortion, Spontaneous, Herbal Medicine, Pregnancy, Nonprescription Drugs; Ananas comosus; MAPK; single-cell RNA sequencing

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11282477/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Herbal remedies, selected foods, and non-prescription medications are widely used during pregnancy despite limited evidence regarding their safety. This study examined associations between first-trimester consumption of these products and spontaneous abortion among Iranian women. An exploratory transcriptomic extension investigated predicted pineapple-associated targets and MAPK-related patterns in endometrial datasets from women with recurrent pregnancy loss.

Methods

In this multicenter case-control study (May 2023–March 2024), 377 women aged 18–45 years were recruited from three prenatal clinics in Tehran, including 167 women with physician-confirmed spontaneous abortion and 210 controls with at least one live birth and no history of abortion. Data on demographic, reproductive, and first-trimester exposure to 45 herbal products, foods, and non-prescription medications were collected using a structured questionnaire. Multivariable logistic regression was used to estimate adjusted odds ratios (aORs) and 95% confidence intervals (CIs), with correction for multiple testing using the Benjamini–Hochberg false discovery rate (FDR). To explore underlying mechanisms, predicted pineapple-associated targets were intersected with lists of differentially expressed genes from GEO datasets (GSE165004 and GSE26787), followed by pathway enrichment analysis. These findings were further assessed using decidual single-cell data to evaluate the common-scale MAPK footprint, and branch-expression.

Results

In total, 29.97% of participants reported consuming at least one of the evaluated herbal or food items during the first trimester. After adjustment for confounders and multiple comparisons, pineapple consumption was associated with a higher odds of spontaneous abortion (aOR 2.95, 95% CI 1.55–5.64; FDR adjusted P = 0.0276). Twenty candidate genes were shared between the lists of abortion-associated and pineapple-related targets, including six MAPK-associated genes. Additionally, single-cell analysis suggested that the pooled MAPK footprint was lower in abortions.

Conclusion

Pineapple consumption was associated with higher odds of spontaneous abortion. Exploratory molecular analyses suggested MAPK pathways as a candidate pathway for future investigation, although experimental validation is required to confirm these findings.

Introduction

Traditional medicine is deeply ingrained in the healthcare practices of many populations, particularly in developing nations, and has been employed for centuries to address various ailments [1, 2]. According to the World Health Organization, 80% of the world's population uses traditional medicines for primary healthcare [3]. While these remedies offer potential benefits, their use during pregnancy, a period characterized by substantial physiological changes [4], warrants careful consideration. The delicate balance between maternal health and fetal well-being often leads women to seek alternatives they perceive as safer than prescribed medications, such as herbal remedies [5–7].

A significant challenge is the limited scientific evidence regarding the safety of widely used herbal products during pregnancy, largely due to ethical restrictions preventing controlled clinical trials in pregnant populations [8]. Growing concerns about maternal complications, neonatal morbidity, and the long-term effects of herbal remedies further highlight the importance of addressing this issue. A recent systematic review identified several herbs associated with adverse pregnancy outcomes, including almond oil with preterm birth and heavy licorice intake with a threefold increase in early preterm birth [9]. However, clinical outcomes arising from herb-drug interactions remain under-evaluated. Furthermore, certain dietary items (e.g., pineapple, celery) are frequently consumed for perceived medicinal properties and commonly reported in clinical histories of pregnancy loss, blurring the line between food and herbal medicine.

In Iran, as in many developing countries, the use of non-prescription medications and traditional herbal remedies is highly prevalent [1, 2]. Because each country possesses a distinct repertoire of medicinal plants shaped by cultural practices and regional availability, the safety of locally used herbs cannot be assumed to mirror those reported elsewhere. Assessing patterns of self-medication among Iranian women who have experienced spontaneous abortion provides an opportunity to identify specific regional behaviors that may contribute to adverse outcomes.

Alongside epidemiological investigations, molecular profiling is critical for formulating hypotheses about *how* these exposures might biologically trigger pregnancy loss. By predicting candidate protein targets from the chemical structures of natural products, researchers can identify overlaps with pregnancy-loss-related gene expression. Furthermore, transcriptomics provides a complementary means of examining this cellular context at the maternal-fetal interface [10, 11]. Exploring specific stress and inflammatory pathways, such as MAPK-associated transcriptional outputs, can help determine if the physiological effects of these regional remedies are biologically plausible contributors to abortion [12].

This study examined the prevalence and patterns of first-trimester herbal, dietary, and non-prescription medication use among Iranian women with and without spontaneous abortion. The primary hypothesis was that these exposures were associated with the odds of spontaneous abortion. Following the clinical analysis, pineapple was selected for a post-hoc exploration molecular investigation using computational target prediction, bulk gene-expression datasets, and independent decidual single-cell transcriptomics. This investigation evaluated overlapping candidate genes and MAPK-associated transcriptional patterns.

Because the clinical and molecular datasets were not linked at the participant level, the molecular analyses were intended to generate hypotheses rather than establish a causal mechanism.

Materials and Methods

Study Design and Setting

This multi-center, case-control study was conducted from May 2023 to March 2024. The study population included a sequential sample of all women who visited the prenatal clinics of Imam Khomeini, Shariati, and Mohab Yase hospitals in Tehran, Iran, during the study period; these clinics are recognized as established referral centers.

Participants

This case–control study enrolled women aged 18–45 years with a gynecologist-confirmed history of fewer than three spontaneous abortions. Cases were women who experienced a physician-confirmed spontaneous abortion (pregnancy loss before 20 weeks of gestation) in the index pregnancy. Women with 0, 1, or 2 prior spontaneous abortions were eligible; the analysis focused on the most recent pregnancy loss. Controls were women who had at least one live birth and no history of spontaneous abortion in any pregnancy. Exclusion criteria for both groups were chronic medical conditions (diabetes, hypertension, thyroid disorders, or other systemic diseases that could influence pregnancy outcomes), alcohol or drug abuse, use of prescription medications (except for minor conditions such as the common cold or allergies) within three months before conception or during early pregnancy, multiple gestations, and incomplete medical records. All participants provided informed consent before enrollment.

Data Collection

Data was collected using a structured questionnaire administered in person by trained research staff. The questionnaire captured socioeconomic and demographic characteristics, gestational age, reproductive history, and consumption of 45 specific items during the first trimester (weeks 1–12) of the index pregnancy. For cases, the index pregnancy was the pregnancy that ended in miscarriage; for controls, it was the most recent pregnancy ending in a live birth. This temporal anchoring was emphasized to reduce recall bias. The 45 items were classified a priori into three categories: herbal products (e.g., barberry, saffron, cinnamon, fennel, black seeds, chamomile), foods with traditional medicinal use (e.g., pineapple, celery, parsley, sour grape juice), and non-prescription drugs (e.g., aspirin, vitamin C, laxatives, NSAIDs). Respondents could select multiple options.

Sample size

The sample size was determined based on an anticipated effect size of 0.3, with an alpha error probability of 0.05 and a statistical power of 0.8. The allocation ratio between the groups was set at 1.2. Using these parameters, the total number of participants needed for the study was determined to be 334, with 148 allocated to case and 186 to control groups. This calculation was performed using G*Power software [13].

Statistical Analysis

Descriptive statistics (mean, standard deviation, proportions) were used to summarize participants' demographic, clinical, and medication-related characteristics. Categorical variables were reported as counts and percentages, and continuous variables as means and standard deviations or medians and interquartile ranges, as appropriate. Group differences in categorical variables were assessed with the Chi-square test; for continuous variables, the Student's t-test or Mann–Whitney U test was applied depending on the data distribution. A two-tailed $p < 0.05$ was considered nominally significant. For the primary analysis, each of the 45 items with a prevalence exceeding 5% in either the case or control group was examined in a separate logistic regression model to avoid instability. These models estimated odds ratios (ORs) and 95% confidence intervals (CIs) for the association with abortion, adjusting for age and occupation, which were significant in univariate comparisons. To account for multiple testing, the Benjamini–Hochberg false discovery rate (FDR) correction was applied to all p-values from these logistic regressions; FDR-adjusted $p < 0.05$ was deemed statistically significant. Association rule mining was additionally performed to describe frequently co-consumed products. All analyses were conducted in R version 4.6.1, using the tidyverse package for data management and visualization, the stats package for hypothesis testing and logistic regression, and the arules package for association rule mining.

Ethical Considerations

The study protocol was reviewed and approved by the Institutional Review Board of Tehran University of Medical Sciences. Written informed consent was obtained from all participants prior to data collection. (IR.TUMS.IKHC.REC.1402.030). The study was carried out following the principles established in the Declaration of Helsinki.

Overall analytical workflow

The analysis combined natural-product target prediction, gene-expression analysis, pathway enrichment, and single-cell transcriptomic profiling. Compounds related to pineapple (*Ananas Comosus*) were retrieved from LOTUS natural products datasets [14], and their molecular targets were predicted using Swiss Target Prediction. These targets were compared with differentially expressed gene sets from two abortion-related GEO datasets (https://www.ncbi.nlm.nih.gov/geo/). Pathway enrichment of the overlapping candidates informed the selection of MAPK-associated transcriptional patterns for subsequent descriptive analysis in an independent single-cell dataset. The single-cell analysis provided

cellular context for the prioritized pathway and did not directly test compound exposure or establish a causal effect of pineapple on pregnancy loss.

Pineapple compound and target prediction

Compounds reported for Ananas comosus were retrieved from the LOTUS natural-products database (https://lotus.naturalproducts.net/) [14], yielding 166 distinct compound names. Putative targets were obtained using SwissTargetPrediction (https://www.swisstargetprediction.ch/) [15]. Predictions with reported probabilities > 0.50 were retained, resulting in a set of 173 candidate target genes. No absorption, distribution, metabolism, and excretion (ADME) or oral-bioavailability filters were applied, and constituent concentrations in consumed pineapple products were unavailable. Target prioritization therefore did not account for dietary dose or achievable tissue exposure. Bromelain, a proteolytic enzyme preparation, was outside the scope of this small-molecule target-prediction analysis. Predictions were considered computational hypotheses rather than experimentally confirmed interactions, and inclusion in LOTUS indicated reported occurrence in A. comosus, not specificity to pineapple.

Identification of abortion associated differentially expressed genes

Gene-expression datasets GSE165004 [16] and GSE26787 [17] were retrieved from the Gene Expression Omnibus and analyzed separately using limma [18]. Both datasets comprised endometrial biopsies obtained during non-conception cycles rather than tissue collected during abortion. GSE165004 included mid-secretory endometrial samples collected on menstrual-cycle days 19–21 and profiled using the Agilent SurePrint G3 Human GE v2 8 × 60K microarray. The comparison comprised 24 women with recurrent pregnancy loss and 24 fertile controls. GSE26787 comprised mid-luteal endometrial biopsies profiled using the Affymetrix Human Genome U133 Plus 2.0 array; five women with recurrent miscarriage were compared with five fertile controls. The unexplained-infertility and implantation-failure groups, respectively, were excluded from these comparisons. Differentially expressed genes were selected using an FDR-adjusted P value < 0.05 and an absolute $\log_2$ fold change ≥ 1. Selected genes were subsequently intersected with predicted targets of pineapple-associated compounds.

Gene overlap and pathway prioritization

Overlap between the predicted pineapple-associated target genes and the DEG sets was examined using Venny diagram (https://www.interactivenn.net/). The overlapping candidate genes were submitted to EnrichR for pathway enrichment analysis against KEGG 2026 [19–21] and WikiPathways 2024 [22] gene-set libraries [23]. Enrichment results guided pathway selected for subsequent single-cell transcriptomic analysis.

Single cell data source and study design

An exploratory secondary analysis of publicly available human single-cell RNA sequencing (scRNA-seq) data was performed to describe decidual cell composition, MAPK-associated transcriptional patterns. Data were obtained from Gene Expression Omnibus accession GSE214607, generated by Wei et al. [11]. The analysis included first-trimester samples from five women undergoing elective termination of normal pregnancies and three women experienced abortion decidual tissue. According to the deposited sample metadata, the source data were generated using the 10x Genomics Chromium Single Cell 3' v2 protocol and Illumina NovaSeq 6000 sequencing, with upstream processing performed using Cell Ranger v3.1.1 and the hg38 reference assembly. The present analysis began with deposited gene-by-cell count matrices and accompanying barcode and feature annotations.

Count matrix processing and quality control

Count matrices were imported with Read10X in Seurat, using Ensembl identifiers as feature names. Gene-expression features were retained when multimodal matrices were present. A Seurat object was created for each library, retaining genes detected in at least three cells and initially retaining cells with at least 200 detected genes. Objects were merged within the normal-pregnancy and abortion groups separately, with sample-specific barcode prefixes preserving cell identity.

Ensembl-to-symbol mappings were obtained from the corresponding feature annotation files. Ensembl version suffixes were removed, duplicate identical mappings were collapsed, and identifiers associated with conflicting symbols were excluded from mapping-dependent analyses. Mitochondrial genes were identified by the MT- prefix in their mapped gene symbols. Subsequent cell-level filtering retained cells with more than 200 and fewer than 2,500 detected genes and less than 5% mitochondrial counts. Sample-level cell numbers were tabulated before and after this filtering step. No additional doublet detection, ambient RNA correction, or cell-cycle regression was implemented in this study.

Normalization dimensionality reduction and clustering

Within each clinical group, retained cells were normalized using Seurat LogNormalize with a scale factor of 10,000 (Seurat documentation). The 2,000 most variable genes were selected using the variance-stabilizing transformation method in FindVariableFeatures, centered and scaled with ScaleData, and used for principal component analysis. Up to 30 principal components were computed, with the number restricted to the smaller of 30, the number of retained cells minus one, and the number of selected variable genes minus one. Harmony was applied to the principal component representation using sample identity (orig.ident) as the integration variable [24].

The first 20 Harmony dimensions, or all available dimensions when fewer than 20 were computed, were used for neighborhood graph construction, graph-based clustering at resolution 0.5, and uniform manifold approximation and projection (UMAP). Random seeds were set to 123 for the specified stochastic steps, including UMAP and clustering. Separate embeddings were generated for the two groups; positions in these embeddings were not interpreted as directly aligned coordinates across groups. Harmony-adjusted embeddings were used for visualization and clustering, whereas RNA counts and normalized RNA expression were retained for annotation and scoring.

Reference based cell type annotation

Cells were annotated using SingleR with the Vento-Tormo maternal–fetal interface reference atlas [10, 25], provided as vento18_10x.processed.h5ad. The sparse reference expression matrix and associated cell-type and tissue metadata were imported using rhdf5 and stored in a SingleCellExperiment object.

Query and reference expression matrices were matched using unique shared gene symbols. Features with missing or ambiguous gene-symbol mappings were excluded, and at least 1,000 shared genes were required for annotation. To reduce computational demands, up to 300 reference cells per cell type were randomly sampled without replacement using a fixed seed of 123. Pruned SingleR labels were used for downstream grouping, and cells without a retained label were classified as “Uncertain.” Cell-type counts and percentages within each cluster were tabulated, and cell-type annotations were displayed on group-specific UMAP embeddings.

Common expression scales for between group descriptions

For descriptive comparisons between groups, RNA count layers were joined where necessary, and retained counts were log-normalized using the same procedure for both groups. Normalized expression for gene g in cell c was defined as:

$$x_{gc} = ln(1 + \frac{10{,}000 \times y_{gc}}{L_c})$$

In this equation, y denotes the retained RNA count for the gene, and L denotes the total RNA counts across all features retained in the cell’s RNA assay, calculated before selecting MAPK-related genes. Common scores were calculated using only genes with unambiguous mappings in both groups. Expression and score tables from both groups were combined for scaling, while the Seurat objects and UMAP embeddings were kept separate. Scaling parameters were estimated across all retained cells from both groups to preserve differences that would otherwise be removed by centering within each group or cell type.

PROGENy derived MAPK transcriptional footprint score

MAPK-associated transcriptional output was summarized using the human PROGENy perturbation-response model, model_human_full [26]. Records for the MAPK pathway with finite weights and model p-values were retained, ordered by increasing model p-value and then gene symbol, and deduplicated by gene symbol. The top 500 genes were selected before intersection with the expression data. These model p-values were used solely to rank genes in the published footprint and were not calculated from the pregnancy groups.

The selected footprint was restricted to genes with unambiguous mappings in both cohorts. At least 250 of the original 500 genes were required to be shared, and at least 250 shared footprint genes were

required to have nonzero expression in one or more cells within each cohort. These were technical coverage requirements, not biological decision thresholds. Shared mapped genes with zero expression in an individual cell remained in the calculation. The raw score was calculated as the sum of log-normalized expression multiplied by the corresponding published signed PROGENy weights:

$$A_c = \sum_{g \in G} w_g x_{gc}; \quad M_c = \frac{A_c - \mu_A}{s_A}$$

G denotes the shared retained footprint and w denotes the published gene weight. The common MAPK score M was obtained by subtracting the mean raw score across all retained cells from both groups and dividing by its sample standard deviation. Footprint genes were not individually z-standardized before this weighted sum. Higher scores indicate higher relative MAPK-associated transcriptional output under the PROGENy model. The implementation used a direct weighted matrix product rather than a fitted model or an enrichment test. It does not directly measure MAPK protein abundance, kinase activity, or phosphorylation.

MAPK branch gene expression summaries

Three compact gene panels were summarized separately: ERK1/2 (MAP2K1, MAP2K2, MAPK1, and MAPK3), JNK (MAP2K4, MAP2K7, MAPK8, and MAPK9), and p38 (MAP2K3, MAP2K6, and MAPK14). These panels were treated as core gene-expression summaries rather than independently validated branch-activity footprints or comprehensive KEGG 2026 pathway scores.

For each available panel gene, log-normalized expression was z-standardized using its mean and sample standard deviation across all retained cells in both groups. The branch score was the arithmetic mean of these gene-level z-scores:

$$B_{bc} = \frac{1}{|G_b|} \sum_{g \in G_b} \frac{x_{gc} - \mu_g}{s_g}$$

Only shared genes with finite standard deviations greater than 10^{-8} were eligible. A branch required at least two eligible genes and at least 75% of its specified panel, corresponding to at least three genes for each of the three panels. Scores for branches failing this requirement were recorded as missing rather than replaced with zero. The genes used and branch availability were exported in a coverage audit.

Gene expression and cell composition

To examine MAPK-related signaling at the gene level, we assessed a targeted panel of 20 genes covering the ERK (EGFR, MAP2K1, MAP2K2, MAPK1, MAPK3), JNK (MAP2K4, MAP2K7, MAPK8, MAPK9), and p38 (MAP2K3, MAP2K6, MAPK14) cascades, negative regulators (DUSP4, DUSP6, SPRY2, SPRY4, PPM1A, PPM1B, PPP5C), and the NF-κB subunit RELA. Expression was recalculated from the RNA assay counts after joining count layers where necessary. Counts were divided by each cell's total RNA counts across all retained features, multiplied by 10,000, and transformed using ln(1 + x). Gene symbols were matched to dataset features using Ensembl-to-symbol mappings after removing Ensembl

version suffixes; direct gene-symbol matches were also accepted. Features with conflicting mappings or multiple features corresponding to the same symbol were excluded. Gene availability was checked separately in each group.

For each annotated cell type within the abortion and normal-pregnancy groups, mean log-normalized expression was calculated across all cells, including those with zero expression. The percentage of expressing cells was defined as the proportion with expression greater than zero. Dot plots represented expressing percentages by dot area on a shared 0–100% scale and mean log-normalized expression by color. A common color scale spanned the observed range of cell-type mean expression values across both groups.

Cell composition was calculated as the number of retained cells assigned to each cell type divided by the total number of retained cells within each sample or clinical group, multiplied by 100. Sample-specific and pooled group percentages were displayed as stacked bar charts. Pooled percentages were calculated from cell counts across all samples within each group, so samples with more retained cells contributed more to group summaries. Cell types were defined using pruned SingleR annotations. Cells labeled “Uncertain” or having missing or empty annotations were grouped as “Uncertain” and included in all denominators. These percentages describe captured, quality-filtered cells rather than absolute cell abundance in the original tissue.

Descriptive analysis and visualization

Between-group comparisons were descriptive and included three abortion samples and five normal-pregnancy samples. Cells were summarized individually but were not treated as independent biological replicates. No hypothesis tests or confidence intervals were calculated. For each of the four scores, MAPK transcriptional footprint, ERK1/2 core gene expression, JNK core gene expression, and p38 core gene expression, we calculated the number of cells, number of contributing samples, mean, median, standard deviation, and first and third quartiles within each cell type and clinical group and across all retained cells per group. All four scores were finite for all remained cells; and no cells were excluded because of non-finite scores. Each cell contributed equally to these summaries.

Score distributions were displayed as violin plots with equal maximum widths and density curves trimmed to the observed score ranges. Box plots showed medians and interquartile ranges, and overlaid points indicated pooled-cell means. Cell counts were labeled, and a common y-axis scale was used across cell-type panels for each score. Findings were interpreted as descriptive patterns in the analyzed dataset rather than estimates of population-level effects.

Analyses were performed in R (version 4.6.1) using Seurat (version 5.5.1), SeuratObject (version 5.4.0), harmony (version 2.0.5), progeny (version 1.34.0), and ggplot2 (version 4.0.3). Cell-type annotation was performed using SingleR (version 2.14.2).

Results

Participant’s characteristics

The mean age of the participants was 32.24 years. An overall pregnancy-ending gestational age of 30.42 weeks was reported in the original analysis; its denominator and interpretation require clarification because the case definition specifies loss before 20 weeks. Each woman in the study had been pregnant approximately two times (1.99 ± 1.19), A notable proportion of the women (29.97%) reported using natural herbal remedies, with about two-thirds of them utilizing these remedies at least once a week. Self-medication was less common, with only 8.75% of the women engaging in this practice. The demographic characteristics of the participants are elaborated in Table 1. Barberry, saffron, and cinnamon were the most frequently consumed herbs among the participants, with usage rates of 74.27%, 71.62%, and 54.91%, respectively. In terms of self-medication, vitamin C and aspirin were the most popular choices, with prevalence rates of 30.77% and 26.26%, as indicated in Table 2. The workflow of the study presented in Fig. 1.

Table 1

Patient Demographic Characteristics: Comparison Between Case and Control Groups. Participants were considered cases if they had a confirmed diagnosis of spontaneous abortion. Controls were women who had at least one live birth and no history of spontaneous abortion. P-value less than 0.05 is considered significant.

Demographic factors	**Women with abortion (N = 167) (%)**	**Women without abortion history (N = 210) (%)**	**Total (N = 377)**	***p* value**
Age M ± SD	33.13 ± 6.78	31.11 ± 6.47	32.24 ± 6.72	**< 0.001**
Weight M ± SD	74.08 ± 14.74	74.05 ± 14.45	74.01 ± 14.56	0.492
Height M ± SD	161.11 ± 5.88	161.35 ± 5.82	161.24 ± 5.84	0.656
BMI M ± SD	28.57 ± 5.68	28.47 ± 5.55	28.51 ± 5.59	0.429
Number of living children M ± SD	0.96 ± 0.94	0.98 ± 1.09	0.973 ± 1.03	0.584
Eduction	62 (46.62%)	71 (53.38%)	133	0.797
High school	67 (43.23%)	88 (56.77%)	155	
Diploma	38 (42.70%)	51 (57.30%)	89	
College education				
Location	101 (41.22%)	144 (58.78%)	245	0.102
Town	66 (50.00%)	66 (50.00%)	132	
City				
Occupation	149 (42.69%)	200 (57.31%)	349	**0.030**
Homemaker	18 (64.29%)	10 (35.71%)	28	
Employed				
History of abortion in family	167 (93.30%)	12 (6.70%)	179	**< 0.001**
Self-medication	27 (81.82%)	6 (18.18%)	33	**< 0.001**
Natural herbs used	69 (61.06%)	44 (38.94%)	113	**< 0.001**
Frequency of herbs taking in a week	83 (38.79%)	131 (61.21%)	214	0.521
0 times	58 (49.57%)	59 (50.43%)	117	

Demographic factors	Women with abortion (N = 167) (%)	Women without abortion history (N = 210) (%)	Total (N = 377)	*p* value
1–2 times	15 (48.39%)	16 (51.61%)	31	
3–4 times	11 (73.34%)	4 (26.66%)	15	
5 or more				
Bleeding in the first trimester	43 (84.31%)	8 (15.69%)	51	**< 0.001**
Bleeding in the second trimester	1 (25.00%)	3 (75.00%)	4	0.633

Table 2
Comparison of Self-Prescribed and Herbs Use Between Women with and Without History of Abortion

self-prescribed and herbs	Women with abortion history (N = 167) Number (%)	Women without abortion history (N = 210) Number (%)	Total (N = 377)	Adjusted p value
Cohosh	0 (0.00%)	2 (100.00%)	2	0.6593
Celery	85 (51.20%)	81 (48.80%)	166	0.0269
Goji berry	1 (100.00%)	0 (0.00%)	1	0.6309
Cinnamon	114 (55.07%)	93 (44.93%)	207	0.0036
Fennel	33 (76.74%)	10 (23.26%)	43	0.0036
Hyacinth of error	0 (0.00%)	2 (100.00%)	2	0.6593
Curd	1 (50.00%)	1 (50.00%)	2	1
Saffron	133 (49.26%)	137 (50.74%)	270	0.0055
The Black Seeds	44 (67.69%)	21 (32.31%)	65	0.0036
Herbal aloe vera	4 (44.44%)	5 (55.56%)	9	1
Barberry	141 (50.36%)	139 (49.64%)	280	0.0036
Sourgrape juice	75 (54.35%)	63 (45.65%)	138	0.0074
Thyme	106 (55.50%)	85 (44.50%)	191	0.0036
Sesame seeds	84 (61.76%)	52 (38.24%)	136	0.0036
Angelica root	1 (100.00%)	0 (0.00%)	1	0.6309
Marshmallow	16 (80.00%)	4 (20.00%)	20	0.0055
Licorice	6 (50.00%)	6 (50.00%)	12	0.885
Bean egg	11 (68.75%)	5 (31.25%)	16	0.1247
Angelica	56 (61.54%)	35 (38.46%)	91	0.0036
Violets	5 (62.50%)	3 (37.50%)	8	0.6566
Various spices	85 (52.80%)	76 (47.20%)	161	0.0094
Asparagus	3 (33.33%)	6 (66.67%)	9	0.885
Various seeds and grains	6 (50.00%)	6 (50.00%)	12	0.885
Parsley	108 (57.14%)	81 (42.86%)	189	0.0036

self-prescribed and herbs	Women with abortion history (N = 167) Number (%)	Women without abortion history (N = 210) Number (%)	Total (N = 377)	Adjusted p value
Oregano	97 (61.39%)	61 (38.61%)	158	0.0036
Pineapple	88 (65.67%)	46 (34.33%)	134	0.0036
Papaya	1 (50.00%)	1 (50.00%)	2	1
Evening primrose oil	1 (25.00%)	3 (75.00%)	4	0.8041
Clove tea	15 (68.18%)	7 (31.82%)	22	0.0509
Mountain tea	34 (62.96%)	20 (37.04%)	54	0.0074
Vitamin C	65 (56.03%)	51 (43.97%)	116	0.0055
Laxatives	8 (66.67%)	4 (33.33%)	12	0.24
Aspirin tablets	52 (52.53%)	47 (47.47%)	99	0.1034
Khakshir flower	89 (54.60%)	74 (45.40%)	163	0.0036
King like flower	1 (100.00%)	0 (0.00%)	1	0.6309
Fenugreek	20 (74.07%)	7 (25.93%)	27	0.0055
Night flower	5 (100.00%)	0 (0.00%)	5	0.0363
Primrose oil	1 (50.00%)	1 (50.00%)	2	1
Chamomile tea	19 (65.52%)	10 (34.48%)	29	0.0388
Cumin tea	21 (77.78%)	6 (22.22%)	27	0.0036
Peppermint tea	11 (55.00%)	9 (45.00%)	20	0.564
Anise tea	1 (33.33%)	2 (66.67%)	3	1
Marjoram	2 (100.00%)	0 (0.00%)	2	0.3258
NSAIDs	3 (42.86%)	4 (57.14%)	7	1
Henna (topical)	21 (77.78%)	6 (22.22%)	27	0.0036

Participant characteristics and product use associated with spontaneous abortion

Among the 377 participants, 167 had experienced spontaneous abortion and 210 were controls. Women in the spontaneous-abortion group were older than controls (33.13 ± 6.78 versus 31.11 ± 6.47 years; P <

0.01), and occupation differed between groups (P = 0.030). Natural-herbal-product use and self-medication also differed between groups (both P < 0.001; Table 1). Differences in product consumption were observed across multiple items after correction for multiple comparisons; complete results are presented in Table 2.

In logistic regression models adjusted for age and occupation, pineapple consumption was associated with higher odds of spontaneous abortion (adjusted OR 2.95, 95% CI 1.55–5.64) and was the only exposure that remained statistically significant after false discovery rate correction (FDR-adjusted P = 0.0276; Fig. 2). Barberry showed higher nominal odds (adjusted OR 2.16, 95% CI 1.07–4.37), whereas celery showed lower nominal odds (adjusted OR 0.57, 95% CI 0.32–0.99); neither association remained significant after correction (FDR-adjusted P = 0.2213 and 0.2542, respectively). Thus, the initial group comparisons identified several exposure differences, but only pineapple met the FDR threshold in the covariate-adjusted regression analyses.

Association analysis of herbs

The association analysis method was utilized to uncover correlations between items, focusing on identifying frequently occurring combinations in transactions. Notably, a high association was observed between the consumption of Mountain tea and barberry among women, with a co-occurrence rate of 98%. Furthermore, when women consumed celery, there was a strong likelihood of also consuming barberry, with a rate of 93%. Detailed co-occurrence patterns of individual herb consumption with confidence > 0.75 can be found in Table 3, while Table 4 presents the association analysis outcomes with a confidence level of 100%, indicating consistent co-consumption of items across all instances analyzed.

Table 3
The association analysis of individual herbs

Condition	Consequent	Confidence	Lift
Mountain tea	barberry	98%	1.314
Celery	barberry	93%	1.242
Sesame seeds	barberry	93%	1.241
Fennel	cinnamon	91%	1.643
Angelica	barberry	90%	1.207
Sesame seeds	saffron	90%	1.246
The Black Seeds	cinnamon	88%	1.597
The Black Seeds	saffron	88%	1.224
Celery	saffron	87%	1.213
Mountain tea	saffron	87%	1.209
Mountain tea	thyme	85%	1.672
Pineapple	parsley	85%	1.688
The Black Seeds	thyme	85%	1.664
Angelica	thyme	85%	1.661
Angelica	parsley	84%	1.657
Mountain tea	oregano	83%	1.978
Mountain tea	parsley	81%	1.617
Mountain tea	cinnamon	81%	1.476
Oregano	thyme	81%	1.591
The Black Seeds	parsley	80%	1.581
Sesame seeds	parsley	79%	1.576
The Black Seeds	celery	78%	1.761
Oregano	parsley	78%	1.545
pineapple	thyme	78%	1.524
Sour grape juice	parsley	78%	1.538
Pineapple	cinnamon	76%	1.379
Thyme	cinnamon	76%	1.375

Condition	Consequent	Confidence	Lift
Sesame seeds	thyme	76%	1.487
Sour grape juice	thyme	75%	1.48
Sesame seeds	cinnamon	75%	1.359

Table 4
association analysis of items with highest confidence

Condition	Consequent	Confidence	Lift
pineapple, Sesame seeds, vitamin C	parsley	100%	1.984
angelica, celery, oregano	thyme	100%	1.963
angelica, cinnamon, oregano	thyme	100%	1.963
angelica, oregano, pineapple	thyme	100%	1.963
angelica, oregano, Sesame seeds	thyme	100%	1.963
angelica, oregano, Sour grape juice	thyme	100%	1.963
angelica, Sesame seeds, vitamin C	thyme	100%	1.963
celery, Sesame seeds, vitamin C	thyme	100%	1.963
oregano, pineapple, vitamin C	thyme	100%	1.963
celery, Mountain tea	barberry	100%	1.339
cinnamon, Mountain tea	barberry	100%	1.339
Mountain tea, oregano	barberry	100%	1.339
Mountain tea, parsley	barberry	100%	1.339
Mountain tea, saffron	barberry	100%	1.339
cinnamon, Mountain tea, oregano	barberry	100%	1.339
cinnamon, Mountain tea, parsley	barberry	100%	1.339
cinnamon, Mountain tea, saffron	barberry	100%	1.339
cinnamon, Mountain tea, thyme	barberry	100%	1.339
Mountain tea, oregano, parsley	barberry	100%	1.339
Mountain tea, oregano, saffron	barberry	100%	1.339
Mountain tea, oregano, thyme	barberry	100%	1.339
Mountain tea, parsley, saffron	barberry	100%	1.339
Mountain tea, parsley, thyme	barberry	100%	1.339
Mountain tea, saffron, thyme	barberry	100%	1.339

Pineapple target overlap and prioritization of MAPK signaling

The pineapple compound list comprised 167 entries corresponding to 166 distinct compound names after removal of a duplicate ferulic acid entry. Target prediction yielded 173 target-name annotations. The selected gene lists from GSE165004 and GSE26787 contained 3,817 and 12,451 distinct DEGs, respectively, of which 1,669 were shared between datasets (Fig. 3A). Predicted pineapple targets overlapped with 22 candidates from GSE165004 and 102 from GSE26787, with 20 genes common to all three sets, including ATP2A2, CYP4A11, DYRK1B, EGFR, FYN, GSK3B, HSP90AA1, KDM1A, KDM5C, LAP3, MAPK1, NOS2, NR3C1, PPARA, PPM1A, PPM1B, PPP1CC, PPP5C, PTPN2, and RELA. (Fig. 3B, C)

Pathway enrichment analysis of these 20 genes identified MAPK signaling as the term with the lowest nominal P value in both the KEGG 2026 and WikiPathways 2024 results which presented in Fig. 3D. Six genes, PPM1A, PPP5C, PPM1B, MAPK1, EGFR, and RELA, mapped to the KEGG MAPK pathway, which contained 296 annotated genes ($P = 3.25 \times 10^{-7}$; adjusted $P = 2.26 \times 10^{-5}$; enrichment odds ratio = 29.10). The same six genes mapped to the WikiPathways MAPK signaling pathway (WP382), comprising 246 annotated genes ($P = 1.09 \times 10^{-7}$; adjusted $P = 8.60 \times 10^{-6}$; enrichment odds ratio = 35.25).

Other enriched KEGG pathways included chemical carcinogenesis–receptor activation and focal adhesion (both adjusted $P = 3.91 \times 10^{-5}$). The aryl hydrocarbon receptor pathway was also enriched in WikiPathways (adjusted $P = 8.60 \times 10^{-6}$). Based on the MAPK enrichment observed in both libraries, MAPK-related transcriptional patterns and the six overlapping genes were examined in the subsequent single-cell analysis.

Single-cell dataset and cell composition

The single-cell analysis included 26,735 cells after quality filtering: 14,744 from five normal-pregnancy samples and 11,991 from three abortion samples. Individual samples contributed 1,257–4,658 cells in the normal-pregnancy group and 1,591–5,210 cells in the abortion group. Reference-based annotation identified 36 cell categories in each group. An additional 379 normal-pregnancy cells (2.57%) and 179 abortion cells (1.49%) were classified as “Uncertain.” (Fig. 4A, B)

Decidual natural killer subtype 2 (dNK2) was the most abundant annotated population in both groups, comprising 24.34% of normal-pregnancy cells and 16.70% of abortion cells. The abortion group also had lower pooled proportions of dNK1 (1.48% versus 7.87%) and decidual stromal subtypes dS1 (1.33% versus 7.81%) and dS3 (6.14% versus 10.40%). Conversely, cells annotated as villous cytotrophoblasts (VCT) accounted for 14.49% of abortion cells compared with 2.60% of normal-pregnancy cells, while decidual macrophage subtype 3 (dM3) accounted for 11.20% versus 0.42%. The proportions of dM1, syncytiotrophoblasts (SCT), and monocytes (MO) were also higher in the abortion group. Extravillous trophoblast (EVT) proportions were more similar between groups (6.08% versus 5.22%). (Fig. 4C)

Cell composition varied substantially between samples (Fig. 4D). One abortion sample, GSM6613043, contributed 1,736 of the 1,737 VCT-annotated cells, 495 of the 498 SCT-annotated cells, and 1,301 of the 1,343 dM3 cells in that group. These populations represented 33.32%, 9.50%, and 24.97% of retained cells in this sample, respectively. In the other two abortion samples, the corresponding proportions were

0–0.06% for VCT, 0.04–0.06% for SCT, and approximately 0.62–0.63% for dM3. Similarly, 373 of the 384 VCT-annotated cells in the normal-pregnancy group originated from GSM6613032. The pooled differences therefore reflected substantial contributions from individual samples and were not consistent across all samples.

MAPK score coverage and pooled transcriptional patterns

Of the 500 selected PROGENy MAPK footprint genes, 486 were unambiguously mapped in both groups, corresponding to 97.2% shared coverage. Group-level gene-detection assessment identified 486 footprint genes with nonzero expression in normal pregnancy and 485 in the abortion group. All prespecified genes were available for the ERK1/2, JNK, and p38 expression panels. All 20 genes in the targeted expression panel were mapped in both groups, and all four comparative scores had valid numerical values, with no missing or infinite values, for every retained cell.

On the common scale, the pooled, cell-weighted MAPK footprint score was descriptively lower in the abortion group; however, the donor-level comparison did not provide statistical evidence of a between-group difference (P = 0.339; BH-adjusted P = 1.00; Fig. 5A). The mean was − 0.259 in the abortion group and 0.210 in normal pregnancy, giving an observed mean difference of − 0.469 score units. The corresponding medians were − 0.279 and 0.210, with interquartile ranges of − 0.948 to 0.397 and − 0.402 to 0.823, respectively. The violin distributions showed a lower central position in the abortion group with substantial overlap between groups (Fig. 5A). Sample-specific mean MAPK scores ranged from − 0.669 to 0.108 in the abortion group and from − 0.356 to 0.437 in normal pregnancy, demonstrating within-group variation.

The core branch-expression summaries showed smaller pooled differences: mean differences calculated as abortion minus normal pregnancy were − 0.080 for ERK1/2, − 0.032 for JNK, and − 0.013 for p38. ERK1/2 medians were − 0.159 in the abortion group and − 0.038 in normal pregnancy. JNK and p38 had identical group medians of − 0.240 and − 0.290, respectively, despite differences in their means. These expression-panel summaries therefore did not reproduce the magnitude of the MAPK footprint difference (Table 5).

Table 5

Pooled single cell score summaries on the common scale Normal pregnancy: 14,744 cells from five samples; RM: 11,991 cells from three samples. Differences are RM minus normal pregnancy. Each cell contributes equally; summaries are not estimates based on independent cell replicates.

Score	Normal mean	Abortion mean	Mean difference	Normal median	Abortion median
MAPK footprint	0.210	-0.259	-0.469	0.210	-0.279
ERK1/2 expression	0.036	-0.044	-0.080	-0.038	-0.159
JNK expression	0.014	-0.018	-0.032	-0.240	-0.240
p38 expression	0.006	-0.007	-0.013	-0.290	-0.290

Cell-type-specific MAPK transcriptional patterns

Within-cohort standardized MAPK footprint distributions are shown by annotated cell type in Fig. 5B. In the pooled, cell-weighted analysis, mean MAPK footprint scores were lower in the abortion group in 30 of the 36 annotated cell categories (Fig. 5C). These descriptive differences may reflect unequal donor contributions and sparse cell representation in some categories and do not establish a consistent pattern across donors. Cells annotated as VCT had a mean score of − 0.802 in the abortion group compared with 0.235 in normal pregnancy (mean difference, − 1.037). Corresponding SCT scores were − 0.476 and 0.070 (mean difference, − 0.547). These comparisons largely reflected cells contributed by individual samples, as described above. All mean differences were calculated as abortion minus normal pregnancy.

Among decidual natural killer cell subtypes, mean differences were − 0.262 for dNK1, − 0.232 for dNK2, and − 0.447 for dNK3. Lower mean scores were also observed in decidual macrophage subtypes dM1 (mean difference, − 0.345), dM2 (− 0.237), and dM3 (− 0.272), monocytes (MO; −0.439), and decidual stromal subtypes dS1 (− 0.349) and dS3 (− 0.235). The difference for dS2 was smaller (− 0.059).

Some cell categories showed the opposite direction. EVT had a slightly higher mean MAPK footprint score in the abortion group (− 0.447 versus − 0.527; mean difference, 0.080). Higher scores were also observed in maternal endothelial and lymphatic endothelial cells, with mean differences of 0.212 and 0.467, respectively. Several categories with large negative differences contained few cells, including fetal fibroblast subtype fFB1 (four normal-pregnancy cells and ten abortion cells), fFB2 (one cell per group), and fetal endothelial cells (two cells per group). These small cell counts limited interpretation of the differences.

Branch-expression scores also varied by cell type. VCT-annotated cells had lower mean ERK1/2, JNK, and p38 expression scores in the abortion group, whereas dM3 cells had higher means for all three panels despite their lower MAPK footprint score.

Expression of the six overlapping MAPK-associated genes

Among the six genes identified in both MAPK enrichment results, pooled MAPK1 expression was lower in the abortion group than in normal pregnancy. Mean log-normalized expression was 0.297 versus 0.431, and the percentage of expressing cells was 21.02% versus 29.74%. PPM1A, PPP5C, and PPM1B also had lower pooled mean expression and lower percentages of expressing cells in the abortion group. EGFR expression was higher, with mean log-normalized expression of 0.139 versus 0.037 and expressing percentages of 7.61% versus 2.29%. Mean RELA expression was similar between groups (0.132 in the abortion group versus 0.133 in normal pregnancy; Table 6). Cell-type-specific expression levels and expressing-cell percentages for the six overlapping MAPK-associated genes are presented in Fig. 5D.

Table 6

Pooled expression of the six overlapping MAPK associated genes Expression values are means of log-normalized RNA expression, including zeros. Expressing percentages denote cells with nonzero expression. Pooled estimates were reconstructed by weighting the cell-type summaries by their cell counts.

Gene	Normal mean expression	Abortion mean expression	Normal expressing %	Abortion expressing %
PPM1A	0.166	0.128	12.50	9.57
PPP5C	0.086	0.069	6.80	5.17
PPM1B	0.183	0.137	14.03	10.27
MAPK1	0.431	0.297	29.74	21.02
EGFR	0.037	0.139	2.29	7.61
RELA	0.133	0.132	10.17	10.00

Higher percentages of EGFR-expressing cells were also observed within trophoblast-annotated populations. The percentages in the abortion and normal-pregnancy groups were 23.32% versus 9.38% for VCT, 23.29% versus 10.19% for SCT, and 17.42% versus 2.21% for EVT, respectively. In contrast, the percentage of MAPK1-expressing cells was lower in the abortion group within VCT (3.40% versus 9.64%) and EVT (6.17% versus 13.25%). These results describe individual-gene expression and should be interpreted separately from the weighted MAPK transcriptional footprint (Fig. 5E). All comparisons were descriptive, and no statistical significance was inferred.

Discussion

This multicenter case–control study examined the association between herbal medicine use and spontaneous abortion, identifying pineapple as the only assessed exposure that remained statistically significant after adjustment for age and occupation and correction for multiple testing. This association motivated an exploratory investigation of potential molecular mechanisms through compound-target prediction and analysis of independent bulk and single-cell transcriptomic datasets. The intersection of predicted pineapple-associated targets with pregnancy-loss-associated gene lists yielded 20 shared genes, with MAPK signaling enriched in both KEGG 2026 and WikiPathways 2024. Single-cell analysis further identified lower pooled MAPK transcriptional footprint scores in the abortion group, accompanied by variation across cell populations. These findings identify a clinical association and a candidate pathway for further investigation, while leaving their causal relationship unresolved.

First-trimester herbal-remedy use was reported by 29.97% of participants, broadly comparable to the 33% reported in a Saudi Arabian survey [27]. A meta-analysis reported a pooled prevalence of self-medication during pregnancy of 44.50% [28]. However, this broader exposure category is not directly comparable to herbal-remedy use. In the present study, self-medication was reported separately by 8.75% of participants. Differences in exposure definitions, populations, and observation periods should

therefore be considered when comparing prevalence estimates. Previous research has also documented herbal-medicine use during pregnancy and associations between particular preparations and obstetric outcomes, including raspberry-leaf use and caesarean delivery [29]. Such observations emphasize the importance of assessing specific products rather than treating all herbal preparations as a single exposure.

Medication and herbal-product use may be influenced by cultural practices, socioeconomic circumstances, previous treatment experience, and access to healthcare information [30–32]. Broader discussions of self-medication also describe the importance of knowledge, accessibility, and perceived treatment needs [33]. These considerations provide context for the observed use patterns but were not directly evaluated as explanations in this study. In particular, the lower reported frequency of self-medication does not establish greater awareness of medication risks or more cautious behavior.

Women in the abortion group were older than controls (mean age, 33.13 versus 31.11 years), and a higher proportion were employed (10.78% versus 4.76%). These differences informed the inclusion of age and occupation in the exposure-specific regression models. Employment-related stress and lifestyle circumstances have been discussed as possible influences on pregnancy outcomes [34], but employment status alone cannot identify these mechanisms. Direct measurement of work demands, working hours, stress, and socioeconomic conditions would be needed to assess their contribution. More detailed adjustment for reproductive, clinical, dietary, and occupational factors would also help determine the robustness of the pineapple association.

Pineapple consumption was associated with higher odds of spontaneous abortion (adjusted OR 2.95, 95% CI 1.55–5.64; FDR-adjusted P = 0.0276). Although barberry and celery showed nominal associations, neither remained significant after multiple-testing correction. Association-rule analysis identified frequent co-consumption of pineapple and parsley, demonstrating that participants encountered products in combinations. These patterns describe concurrent use but do not establish interaction effects or separate the contributions of individual products. Because the regression models adjusted for age and occupation without mutually adjusting for all co-consumed products, residual confounding by accompanying exposures remains possible.

The nominal barberry association can be considered alongside experimental work reporting adverse fetal-growth effects following high-dose barberry exposure [35]. A review of Berberis vulgaris and berberine also described dose-dependent toxicity and potential interactions with drug-metabolizing enzymes [36]. These reports provide context for further investigation but do not overcome the statistical uncertainty of the barberry finding in the present study. The results therefore support prioritizing pineapple while treating the other exposure estimates as exploratory.

Previous experimental studies provide biological context for investigating pineapple preparations. Monji et al. (2016) [37] likewise reported that an aqueous pineapple-extract fraction stimulated isolated pregnant rat and human uterine muscle, with inhibitor experiments suggesting serotonergic involvement. Subsequent work showed that responses varied with preparation and experimental setting, including an

absence of uterine stimulation by the unripe fraction in the reported in vivo pregnancy and postpartum experiments [38–40]. These findings support preparation-dependent biological activity, but they do not establish that dietary consumption produces equivalent exposure. They also address uterine contractility rather than the decidual MAPK mechanism explored here.

Variation in experimental findings for other herbal extracts further emphasizes the importance of preparation, dose, and model system when interpreting potential effects during pregnancy [39, 40]. Potential interactions with cytochrome P450 enzymes add another consideration, particularly when herbal products are consumed alongside medications [41]. pharmacological issues support more detailed exposure assessment, although the present study did not measure extract composition, enzyme activity, or herb–drug interactions.

The compound analysis provided a means of selecting molecular candidates for investigation. Predicted targets of constituents reported for pineapple were compared with expression-associated gene lists from GSE165004 and GSE26787, yielding 20 shared genes. This overlap connects computational predictions with disease-associated expression but does not identify the constituents responsible for the clinical association. Compounds reported in pineapple are not necessarily pineapple-specific, and distinct names do not establish unique chemical structures. Their concentrations in consumed products and availability in maternal or placental tissues were also unknown. Predicted interactions therefore remain candidates for experimental evaluation.

MAPK signaling was enriched in both the KEGG 2026 and WikiPathways 2024 analyses through six shared genes: PPM1A, PPP5C, PPM1B, MAPK1, EGFR, and RELA. Consistent identification of the pathway supported its prioritization for single-cell analysis. These genes represent distinct signaling components and regulators; their enrichment implicates MAPK-related processes without establishing the direction of pathway activity.

The relevance of MAPK signaling to pregnancy maintenance is supported by Bader et al. [42] who reported lower ERK1/2 mRNA in villous cytotrophoblasts, extravillous cytotrophoblasts, and decidual stromal cells from women with recurrent miscarriage, together with lower ERK and phosphorylated ERK protein measurements. These observations are broadly consistent with the lower pooled MAPK footprint in our abortion group. However, our EVT-annotated population showed a slightly higher mean footprint score. Differences in cell identification and in the measurements themselves, transcript abundance, protein phosphorylation, and a weighted transcriptional footprint, limit direct equivalence.

The lower MAPK footprint observed in our study is also consistent with experimental evidence linking reduced Ras–MAPK signaling to impaired trophoblast function. Zhang et al. [43] reported reduced MEG3 expression in villous tissues from women with unexplained recurrent spontaneous abortion. In primary trophoblast experiments, MEG3 promoted EZH2-mediated repression of RASA1, increased Raf, MEK, and ERK phosphorylation, enhanced proliferation and invasion, and reduced apoptosis. Manipulation of RASA1 supported its role in these effects. This provides a mechanistic context for the transcriptional

pattern observed in our analysis, although whether pineapple constituents influence this regulatory mechanism remains unresolved.

Our identification of MAPK-related targets can also be considered alongside Insuan et al. [44]. In LPS-stimulated RAW264.7 mouse macrophages, purified pineapple-rhizome bromelain reduced ERK1/2, JNK, and p38 phosphorylation and suppressed inflammatory mediators, while total MAPK protein levels remained unchanged. These results demonstrate that a pineapple-derived preparation can modulate MAPK signaling under defined conditions. However, suppression accompanied an anti-inflammatory response and therefore does not, by itself, imply cellular injury. Bromelain is a proteolytic enzyme preparation, so its effects cannot establish the activity of the small-molecule constituents evaluated in our target-prediction analysis.

The single-cell results reinforce the importance of cellular context. Mean MAPK footprint scores were lower in the abortion group in 30 of the 36 annotated categories, including several decidual immune and stromal populations. Nevertheless, higher scores in EVT and some endothelial categories argue against uniform suppression across the maternal–fetal interface. The ERK1/2, JNK, and p38 expression panels showed smaller pooled differences than the footprint. PROGENy uses perturbation-responsive genes to infer pathway-associated transcriptional activity, providing a different measure from the expression of core signaling components [12]. Consequently, the footprint and branch panels need not show identical patterns. High coverage of the footprint gene set supports technical completeness but does not establish its biological validity in every decidual population.

Individual-gene findings further illustrate this distinction. MAPK1, PPM1A, PPP5C, and PPM1B showed lower pooled expression in the abortion group, whereas EGFR expression was higher and RELA expression was similar between groups. Lower phosphatase transcript abundance does not demonstrate increased dephosphorylation, and higher receptor expression does not establish stronger downstream signaling. Likewise, unchanged RELA expression cannot exclude changes in NF-κB phosphorylation or localization. These findings identify components for functional investigation rather than directly measuring their activity.

The World Health Organization has identified challenges in herbal-medicine regulation, quality assurance, safety evaluation, and monitoring (World Health Organization, 2005). Earlier discussions of medication exposure during pregnancy similarly emphasize the need to assess potential maternal and fetal effects [45]. In this context, the clinical implication of our study is the importance of recording specific products, their preparation, timing, frequency, and combinations. The observed association does not establish a harmful dietary threshold for pineapple or justify treating all herbal products as having equivalent effects.

Several limitations should be considered. The case–control design and retrospective self-report permit recall bias, exposure misclassification, and residual confounding. Pineapple intake was not quantified in terms of amount, frequency, ripeness, or form of consumption, precluding assessment of dose–response relationships and differences between preparations. Although exposure was assigned to the

first trimester, its timing relative to symptoms or recognition of pregnancy complications may remain uncertain. The estimated association between pineapple consumption and spontaneous abortion was based on a small number of exposed participants, resulting in a wide confidence interval (adjusted OR 2.95; 95% CI 1.55–5.64) and limited precision. Controls were selected from women who had a live birth and no history of abortion; this restriction may introduce selection bias and limit their comparability with cases. Adjustment was limited to age and occupation, with no adjustment for other co-consumed items or clinical confounders, leaving potential confounding by reproductive, clinical, socioeconomic, and dietary factors. The questionnaire's inclusion of foods, herbal preparations, supplements, and medications also necessitates consistent exposure definitions.

The molecular analyses depend on compound identification, gene mapping, selection thresholds, and database coverage. Bulk-expression differences may reflect tissue composition as well as changes within cells. The public datasets lacked pineapple-exposure measurements and were independent of the clinical cohort. Recurrent pregnancy loss in these datasets also differs from the index spontaneous-abortion phenotype assessed clinically. These analyses therefore provide disease-associated molecular context rather than direct evidence that pineapple exposure produced the observed changes.

The single-cell comparison included eight donors, despite containing 26,735 cells. Unequal cell recovery influenced pooled summaries, several populations contained few cells, and individual samples contributed most of the VCT-, SCT-, and dM3-annotated cells in their groups. Trophoblast assignments in decidual material require confirmation using canonical markers and sampling information. The absence of additional doublet detection and ambient-RNA correction further limits interpretation of unusual populations. Because the comparisons were descriptive and cell-weighted, independent validation using donor-level analyses is required. The MAPK transcriptional footprint also does not directly measure kinase activity or phosphorylation.

Future studies should evaluate the clinical association prospectively in larger, diverse populations, with detailed exposure measurement and assessment of co-consumption. Complementary experiments should test structurally defined pineapple constituents at concentrations supported by exposure and bioavailability data in relevant trophoblast and decidual systems. Measurements of total and phosphorylated ERK, JNK, and p38 should be combined with transcriptional responses, cell survival, invasion, and inflammatory outcomes. Target perturbation and rescue experiments would help determine whether any functional effects depend on the proposed signaling components and whether the clinical and molecular observations share a mechanism.

This study underscores the complex relationship between herbal medicine use and spontaneous abortion. The reported estimates included positive and inverse nominal associations, pineapple was the only assessed exposure associated with higher odds of spontaneous abortion after covariate adjustment and correction for multiple testing. Exploratory target-prediction and bulk-expression analyses prioritized MAPK signaling, while independent single-cell data revealed lower pooled MAPK transcriptional footprint scores in the abortion group, with variation across cell populations. These

findings identify a clinical association and a candidate pathway for mechanistic investigation. Future studies should validate the association in larger, diverse prospective cohorts with detailed assessment of exposure timing, dose, preparation, and co-consumption. Complementary experiments should test defined pineapple constituents at biologically relevant concentrations to determine whether they alter MAPK phosphorylation and trophoblast or decidual cell function, thereby assessing a potential mechanistic link to pregnancy loss.

Declarations

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate:

Ethical Approval: The study protocol was reviewed and approved by the Institutional Review Board of Tehran University of Medical Sciences. Written informed consent was obtained from all participants prior to data collection. (IR.TUMS.IKHC.REC.1402.030).

Informed consent:

All the individuals had been informed of the purposes of the study and written informed consent was obtained from all participants.

Funding

Author Contribution

R.S and H.R carried out the study starting from conception, analysis and interpretation of data and drafting of the manuscript. L.S, F.A participated in analysis and interpretation of data for important intellectual content, H.R, K.J, A.H participated in critical review of the manuscript, data analysis and interpretation of the finding. Z.H contributed in supervision. All author contribute in drafting. All authors have read and approved the final version of the manuscript.

Acknowledgements:

Not applicable.

Clinical trial number: Not applicable.

Data Availability

The dataset used and analyzed during the current study are available from the corresponding author on reasonable request. Public transcriptomic datasets are available under GEO accessions GSE165004, GSE26787, and GSE214607.

References

1. Sattari, M. et al. Self-reported Use and Attitudes Regarding Herbal Medicine Safety During Pregnancy in Iran. *Jundishapur J. Nat. Pharm. Prod.* **7** (2), 45–49 (2012).
2. Yazdi, N. et al. Use of complementary and alternative medicine in pregnant women: A cross-sectional survey in the south of Iran. *J. Integr. Med.* **17** (6), 392–395 (2019).
3. Hoenders, R. et al. A review of the WHO strategy on traditional, complementary, and integrative medicine from the perspective of academic consortia for integrative medicine and health. *Front. Med. (Lausanne)*. **11**, 1395698 (2024).
4. John, L. J. & Shantakumari, N. Herbal Medicines Use During Pregnancy: A Review from the Middle East. *Oman Med. J.* **30** (4), 229–236 (2015).
5. Cuzzolin, L. et al. Use of herbal products among 392 Italian pregnant women: focus on pregnancy outcome. *Pharmacoepidemiol Drug Saf.* **19** (11), 1151–1158 (2010).
6. Illamola, S. M. et al. Use of Herbal Medicine by Pregnant Women: What Physicians Need to Know. *Front. Pharmacol.* **10**, 1483 (2019).
7. Sarecka-Hujar, B. & Szulc-Musioł, B. *Herbal Medicines-Are They Effective and Safe during Pregnancy?*. *Pharmaceutics*, **14**(1). (2022).
8. Lyerly, A. D., Little, M. O. & Faden, R. The second wave: Toward responsible inclusion of pregnant women in research. *Int. J. Fem Approaches Bioeth.* **1** (2), 5–22 (2008).
9. Muñoz Balbontín, Y. et al. Herbal Medicinal Product Use During Pregnancy and the Postnatal Period: A Systematic Review. *Obstet. Gynecol.* **133** (5), 920–932 (2019).
10. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563** (7731), 347–353 (2018).
11. Wei, P. et al. Identification and validation of a signature based on macrophage cell marker genes to predict recurrent miscarriage by integrated analysis of single-cell and bulk RNA-sequencing. *Front. Immunol.* **13**, 1053819 (2022).

12. Schubert, M. et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9** (1), 20 (2018).
13. Faul, F. et al. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods*. **39** (2), 175–191 (2007).
14. Rutz, A. et al. *The LOTUS initiative for open knowledge management in natural products research*. *Elife*, 11. (2022).
15. Daina, A., Michielin, O. & Zoete, V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res.* **47** (W1), W357–w364 (2019).
16. Keleş, I. D. et al. Gene pathway analysis of the endometrium at the start of the window of implantation in women with unexplained infertility and unexplained recurrent pregnancy loss: is unexplained recurrent pregnancy loss a subset of unexplained infertility? *Hum. Fertil. (Camb)*. **26** (5), 1129–1141 (2023).
17. Lédée, N. et al. Specific and extensive endometrial deregulation is present before conception in IVF/ICSI repeated implantation failures (IF) or recurrent miscarriages. *J. Pathol.* **225** (4), 554–564 (2011).
18. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43** (7), e47 (2015).
19. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28** (11), 1947–1951 (2019).
20. Kanehisa, M. et al. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53** (D1), D672–d677 (2025).
21. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28** (1), 27–30 (2000).
22. Agrawal, A. et al. WikiPathways 2024: next generation pathway database. *Nucleic Acids Res.* **52** (D1), D679–d689 (2024).
23. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44** (W1), W90–W97 (2016).
24. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*. **16** (12), 1289–1296 (2019).
25. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20** (2), 163–172 (2019).
26. Schubert, M. D.A., *Calculate PROGENy pathway scores from gene expression [Internet]. PROGENy documentation; [date unknown] [cited 2026 Oct 4]. Available from*: https://saezlab.github.io/progeny/reference/progeny.html
27. Aljofan, M. & Alkhamaiseh, S. Prevalence and Factors Influencing Use of Herbal Medicines During Pregnancy in Hail, Saudi Arabia: A cross-sectional study. *Sultan Qaboos Univ. Med. J.* **20** (1), e71–

e76 (2020).

28. Bouqoufi, A. et al. Prevalence and associated factors of self-medication in worldwide pregnant women: systematic review and meta-analysis. *BMC Public. Health*. **24** (1), 308 (2024).
29. Nordeng, H. et al. Use of herbal drugs during pregnancy among 600 Norwegian women in relation to concurrent use of conventional drugs and pregnancy outcome. *Complement. Ther. Clin. Pract.* **17** (3), 147–151 (2011).
30. Holst, L. et al. The use and the user of herbal remedies during pregnancy. *J. Altern. Complement. Med.* **15** (7), 787–792 (2009).
31. Rizk, M. A. et al. Knowledge and practices of pregnant women in relation to the intake of drugs during pregnancy. *J. Egypt. Public. Health Assoc.* **68** (5–6), 567–591 (1993).
32. Verstappen, G. M. et al. Prevalence and predictors of over-the-counter medication use among pregnant women: a cross-sectional study in the Netherlands. *BMC Public. Health*. **13**, 185 (2013).
33. Bennadi, D. Self-medication: A current challenge. *J. Basic. Clin. Pharm.* **5** (1), 19–23 (2013).
34. Katz, V. L. Work and work-related stress in pregnancy. *Clin. Obstet. Gynecol.* **55** (3), 765–773 (2012).
35. Mehrabani, D. et al. The Effect of Aqueous Extract of Berberis Vulgaris on Fetal Height and Weight during Pregnancy. *Int. J. Nutr. Sci.* **4** (2), 83–87 (2019).
36. Rad, S. Z. K., Rameshrad, M. & Hosseinzadeh, H. Toxicology effects of Berberis vulgaris (barberry) and its active constituent, berberine: a review. *Iran. J. Basic. Med. Sci.* **20** (5), 516–529 (2017).
37. Monji, F. et al. Investigation of uterotonic properties of Ananas comosus extracts. *J. Ethnopharmacol.* **193**, 21–29 (2016).
38. Monji, F. et al. Role of the serotonergic pathway in uterotonic activity of Ananas comosus (L.) Merr. - An in vitro and in vivo study. *Phytomedicine* **48**, 32–42 (2018).
39. Lewicki, S. et al. Morphofunctional Renal Alterations in Progeny of Mice Fed Rhodiola kirilowii Extracts or Epigallocatechin During Pregnancy and Lactation. *J. Med. Food*. **20** (1), 86–92 (2017).
40. Tian, X. Y. et al. The effects of Boehmeria nivea (L.) Gaud. on embryonic development: in vivo and in vitro studies. *J. Ethnopharmacol.* **134** (2), 393–398 (2011).
41. Langhammer, A. J. & Nilsen, O. G. In vitro inhibition of human CYP1A2, CYP2D6, and CYP3A4 by six herbs commonly used in pregnancy. *Phytother Res.* **28** (4), 603–610 (2014).
42. Bader, G. et al. *Role of MAPK signaling pathway in recurrent pregnancy loss*. (2025).
43. Zhang, J., Liu, X. & Gao, Y. The long noncoding RNA MEG3 regulates Ras-MAPK pathway through RASA1 in trophoblast and is associated with unexplained recurrent spontaneous abortion. *Mol. Med.* **27** (1), 70 (2021).
44. Insuan, O. et al. Anti-Inflammatory Effect of Pineapple Rhizome Bromelain through Downregulation of the NF-κB- and MAPKs-Signaling Pathways in Lipopolysaccharide (LPS)-Stimulated RAW264.7 Cells. *Curr. Issues Mol. Biol.* **43** (1), 93–106 (2021).
45. Hudson, R. E. et al. Drug exposure during pregnancy: Current understanding and approaches to measure maternal-fetal drug exposure. *Front. Pharmacol.* **14**, 1111601 (2023).

Figures

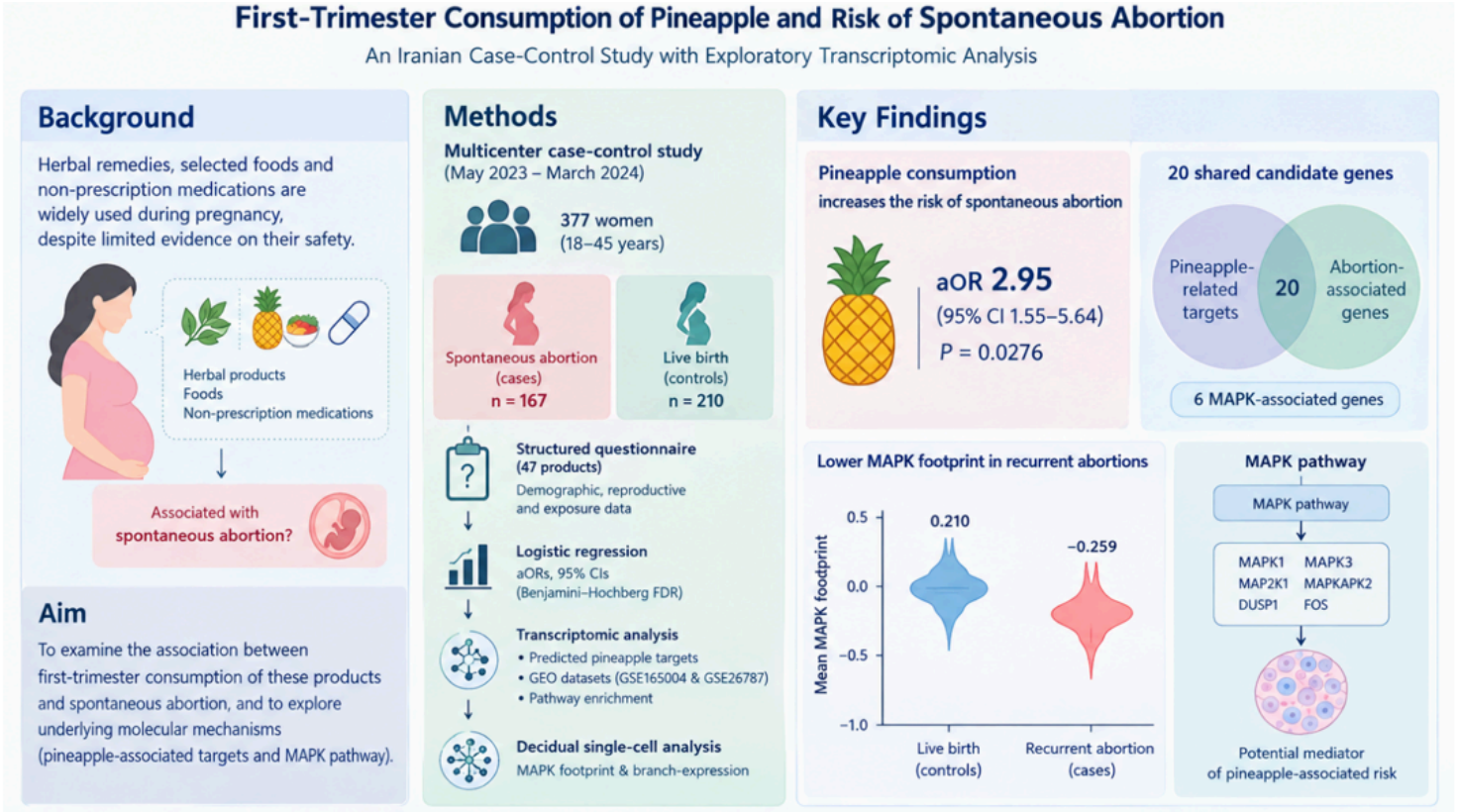
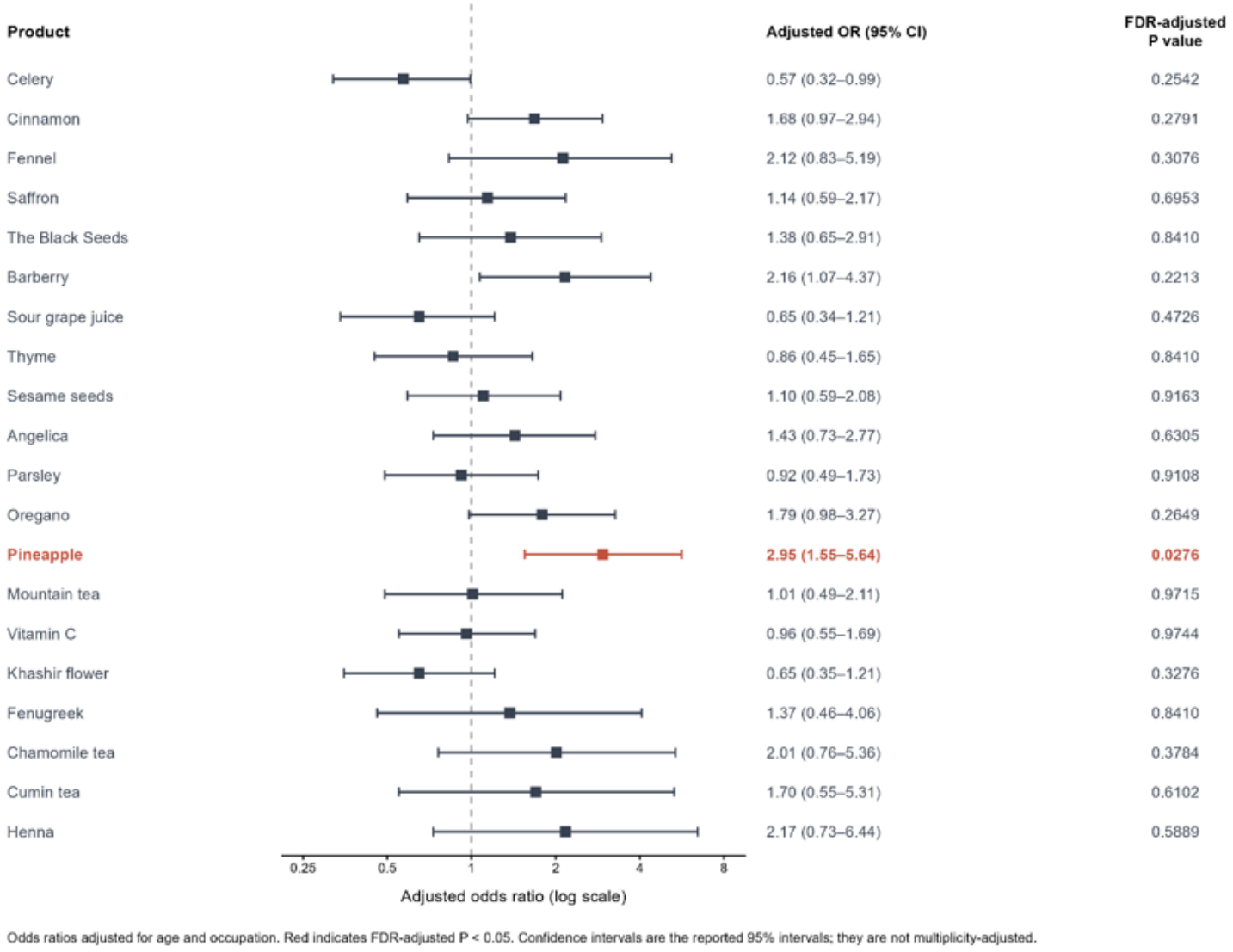


Figure 1

Overview of the clinical study and exploratory transcriptomic analyses.



Odds ratios adjusted for age and occupation. Red indicates FDR-adjusted P < 0.05. Confidence intervals are the reported 95% intervals; they are not multiplicity-adjusted.

Figure 2

Product consumption and odds of spontaneous abortion. Squares and horizontal lines represent odds ratios and 95% confidence intervals adjusted for age and occupation. The dashed line indicates OR = 1. Red denotes Benjamini–Hochberg FDR-adjusted P < 0.05; confidence intervals are not multiplicity-adjusted.

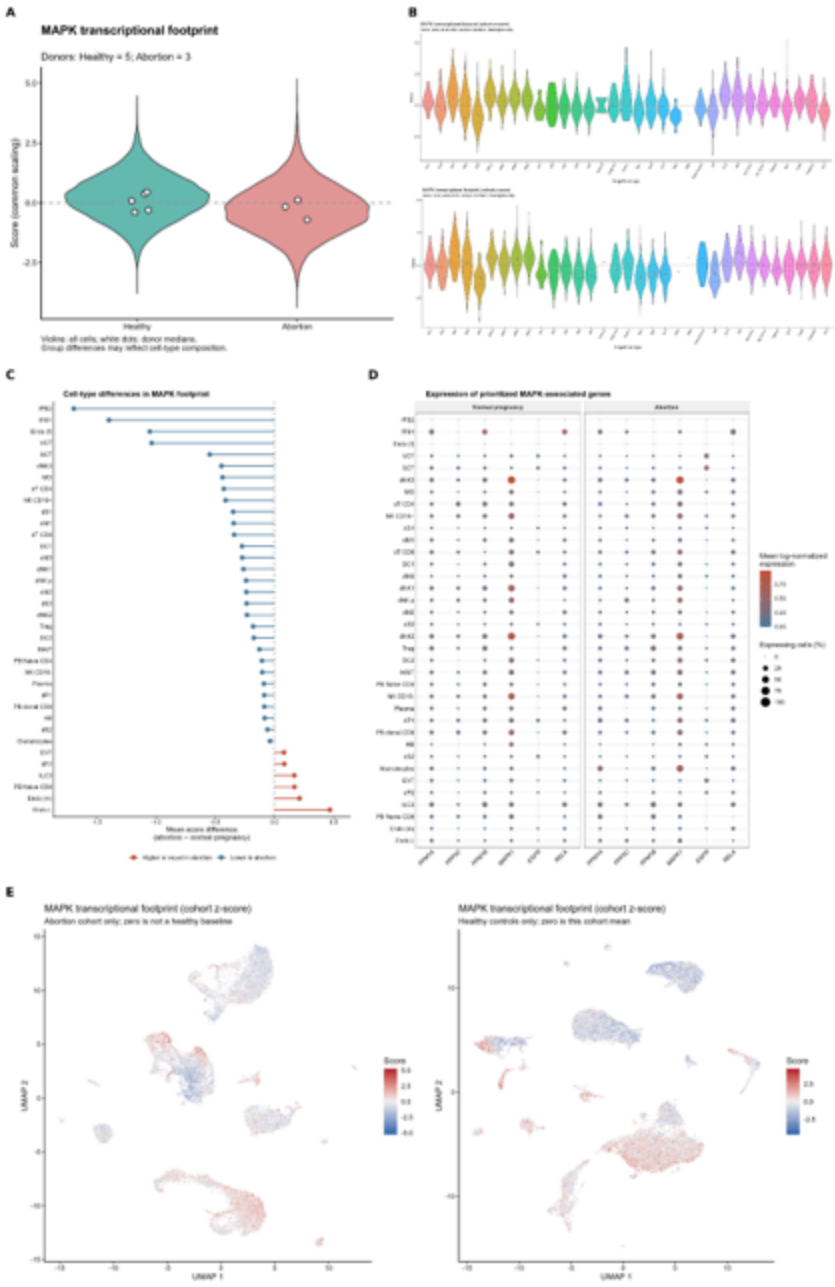


Figure 5

Decidual MAPK footprint scores and prioritized gene expression. **(A)**Common-scale MAPK score distributions; white points indicate donor medians **(B)**Cell-type score distributions standardized within each cohort; white points indicate sample medians. **(C)** Mean common-scale differences (abortion minus normal pregnancy); blue indicates lower and red higher or equal scores. Lines connect estimates to zero, not confidence limits. **(D)** Six-gene expression profiles; dot color indicates mean log-normalized expression and size indicates expressing-cell percentage. **(E)** Within-cohort standardized scores on separate UMAPs: abortion, left; normal pregnancy, right. Panels B and E show within-cohort variation rather than between-group differences.