Transactions on Materials, Biotechnology and Life Sciences
ISSN: 2960-1932, eISSN: 2960-2319 | Volume 9
FENBC 2026

# Identification of Concentration Drivers for Male Fetal Free DNA Based on Multiple Linear Regression and Correlation Analysis

Xiaoxin Lin, Xiaoting Liao, Yilin Cai and Huixian Lin *

School of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang, China

* Corresponding Author Email: 18029817436@stu.gdou.edu.cn

**Abstract.** Addressing the challenge that Y-chromosome concentration in male fetuses during non-invasive prenatal testing is influenced by multiple maternal physiological and technical factors, this study utilized data from 1,082 clinical samples to construct a data-driven model for identifying and quantifying concentration drivers. The study first converted text-based gestational age uniformly into days using a gestational age conversion function. Categorical variables such as mode of delivery and parity were encoded using one-hot encoding. Height and weight indicators with severe multicollinearity were excluded using variance inflation factor to ensure independence of model inputs. During modeling, ordinary least squares established a multiple linear regression model. Spearman's rank correlation coefficient and hierarchical clustering were employed to multidimensionally analyze the association characteristics between 22 independent variables and Y chromosome concentration. Results showed the adjusted $R^2$ of this regression model was 0.468, with an F-statistic of 40.61, passing the significance test. The study identified the proportion of filtered reads, the proportion of duplicate reads, and X chromosome concentration as core influencing factors. It also quantified the positive promoting effect of gestational age and the negative diluting effect of body mass index. This research provides scientific quantitative evidence for clinically evaluating test reliability and optimizing sampling timing.

**Keywords:** Non-invasive prenatal testing; Multiple linear regression; Spearman correlation analysis.

## 1. Introduction

Non-invasive prenatal testing (NIPT) evaluates fetal health by analyzing fetal cell-free DNA (cffDNA) in maternal peripheral blood. The presence of male Y chromosome DNA at or above the 4% threshold is a core criterion for ensuring test accuracy. However, concentration levels exhibit significant heterogeneity across individuals, primarily influenced by the interplay of complex factors such as gestational age, maternal BMI, and sequencing quality [1]. Previous studies have predominantly focused on single-factor clinical observations, lacking quantitative descriptions of multidimensional feature coupling effects and exhibiting shortcomings in addressing redundancy in sequencing technical parameters. The innovation of this section lies in constructing a comprehensive feature space encompassing physiological characteristics and sequencing quality metrics. It systematically addresses multicollinearity through VIF testing and introduces hierarchical clustering and feature importance analysis to reveal concentration evolution patterns from both linear and nonlinear perspectives. The overall research approach follows a logical sequence: data cleaning and standardization, preliminary regression using OLS, in-depth exploration of variable correlations, verification of model normality and stability, and ranking of core feature contributions. This aims to establish a rigorous evaluation paradigm for fetal cell-free DNA concentration [2].

Quantitative Analysis of Factors Influencing Male Fetal Cell-Free DNA Concentration In the quantitative analysis phase, logarithmic transformation and standardization were first applied to establish the relationship between dependent and independent variables. Analysis of the correlation heatmap revealed a significant positive correlation between gestational age and Y chromosome concentration (correlation coefficient 0.32), confirming the biological accumulation of fetal DNA throughout pregnancy. Conversely, maternal BMI exhibited a weak negative correlation of -0.18, quantifying the diluting effect of increased maternal blood volume on fetal DNA. Hierarchical

clustering results further categorized features into sequencing quality clusters, sex chromosome-associated clusters, and baseline feature clusters, identifying sequencing depth and read quality as the primary technical factors influencing concentration estimation accuracy. During multivariate linear regression model construction, a constant term was manually added to ensure model completeness, and t-tests identified 12 statistically significant variables. Feature importance calculations revealed that the proportion of filtered reads contributed most significantly to the model (with an absolute regression coefficient approaching 0.7), followed by the proportion of duplicate reads (approximately 0.65). This suggests that clinical evaluations of concentration should prioritize examining physical losses during sequencing. Additionally, residual Q-Q plots indicated that the model largely satisfied the normality assumption, demonstrating high statistical credibility for regression coefficient estimates and significance testing results [3-4].

## 2. Model Establishment and Solution

### 2.1. Data Preprocessing

#### 2.1.1. Data Cleaning.

First, the gestational age needs to be converted into usable data through a gestational age conversion function, which should be converted to the unified unit of days "w" for quantitative analysis. Define the gestational age conversion function [5-6]:

$$\text{Convert\_gestational\_age}(\text{ga\_str}) = \begin{cases} 7 \times \text{int}(\text{weeks}) + \text{int}(\text{days}), & \text{if ga\_str contains "w+"} \\ 7 \times \text{int}(\text{weeks}), & \text{if ga\_str only contains "w"} \\ \text{NAN}, & \text{if format parsing fails or not a string} \end{cases} \quad (1)$$

Through this function, the "Detection Gestational Age" column is converted to "Detection Gestational Age_days" as shown in Figure 1:

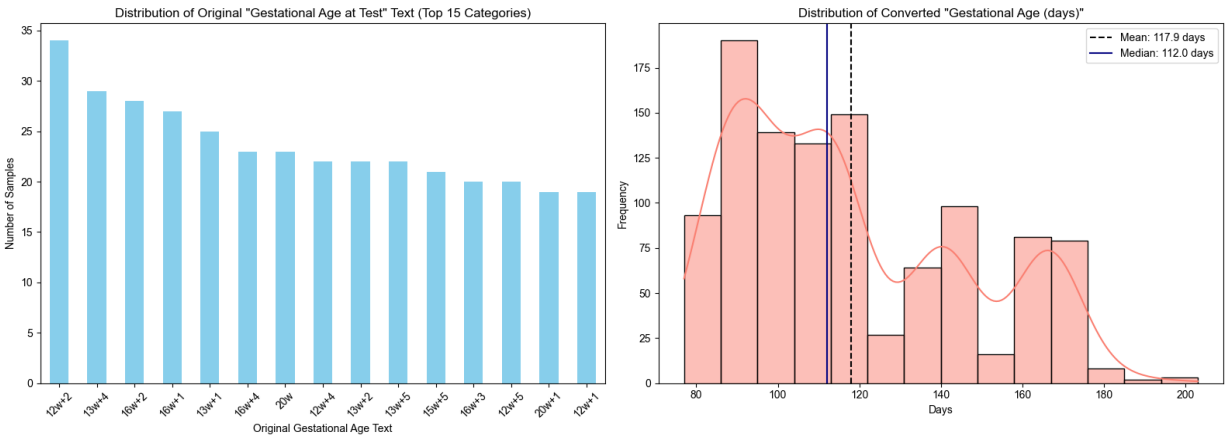


**Figure 1.** Comparison of original "detection gestational age" text distribution and converted "detection gestational age_days" distribution

Second, one-hot encoding for categorical variables: IVF pregnancy method (2 categories) and number of pregnancies (3 categories) are unordered categorical variables, which need to be converted into 0-1 dummy variables through one-hot encoding to avoid the model misjudging categorical variables as ordered variables. Then, since BMI is calculated from weight and height, weight and height data should be deleted to avoid severe multicollinearity [7].

#### 2.1.2. Construction of Ordinary Least Squares (OLS) Regression Model.

The general form of the multiple linear regression model [4] is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (2)$$

where $Y$ is the dependent variable (e.g., fetal Y-chromosome concentration), $X_1, X_2, \cdots, X_k$ are independent variables (e.g., age, maternal BMI, etc.), $\beta_0$ is the intercept term, $\beta_1, \beta_2, \cdots, \beta_k$ are regression coefficients, and $\varepsilon$ is the random error term [8-9].

The estimation of regression coefficients usually adopts the ordinary least squares (OLS) method, whose goal is to minimize the sum of squared residuals:

$$\min_{\beta_0,\beta_1,\cdots,\beta_k} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{3}$$

where $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik}$ is the predicted value of the dependent variable, and $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$ are the estimated values of the regression coefficients.

Through matrix operations, the estimated values of the regression coefficients can be expressed as:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{4}$$

where $X$ is the independent variable matrix (including a column vector of intercept terms), and $Y$ is the dependent variable vector[10].

Standard Error

The standard error of the estimated regression coefficients is used to measure the accuracy of the estimation, and the calculation formula is:

$$\text{Std.}(\hat{\beta}_j) = \sqrt{\frac{\text{MSE}}{(n - k - 1) \cdot \text{Var}(X_j) \cdot (1 - R_j^2)}} \tag{5}$$

where:

MSE is the mean squared error, $\text{MSE} = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-k-1}$

$n$ is the sample size, $k$ is the number of independent variables;

$\text{Var}(X_j)$ is the variance of the j-th independent variable;

$R_j^2$ is the square of the multiple correlation coefficient between the j-th independent variable and other independent variables (measuring the degree of multicollinearity).

t-statistic

Used to test whether the regression coefficient is significantly non-zero, the calculation formula is:

$$t_j = \frac{\hat{\beta}_j - \beta_{j0}}{\text{Std.}(\hat{\beta}_j)} \tag{6}$$

Usually, the null hypothesis is $\beta_{j0} = 0$ (i.e., the independent variable has no effect on the dependent variable), so it is simplified to:

$$t_j = \frac{\hat{\beta}_j}{\text{Std.}(\hat{\beta}_j)} \tag{7}$$

P-value ($P > |t|$)

The P-value is the probability calculated based on the t-distribution, representing the probability of observing the current t-statistic or a more extreme value when the null hypothesis is true. For a two-

tailed test, the calculation of the P-value depends on the cumulative distribution function (CDF) of the t-distribution:

$$\text{P-value} = 2 \times (1 - F_{t(n-k-1)}(|t_j|)) \tag{8}$$

where $F_{t(n-k-1)}(\cdot)$ is the cumulative distribution function of the t-distribution with degrees of freedom $n - k - 1$.

Confidence Interval

The (95%) confidence interval for the regression coefficient is calculated as:

$$\hat{\beta}_j \pm t_{\alpha/2,n-k-1} \times \text{Std.}(\hat{\beta}_j) \tag{9}$$

where $t_{\alpha/2,n-k-1}$ is the upper $\alpha/2$ quantile of the t-distribution with degrees of freedom $n - k - 1$ (corresponding to a 95% confidence level when $\alpha = 0.05$). Variance Inflation Factor (VIF) of Independent Variables are shown in table 1.

**Table 1.** Variance Inflation Factor (VIF) of Independent Variables

| Index | Feature | VIF |
|---|---|---|
| 0 | const | 136685.678442 |
| 3 | Weight | 360.052039 |
| 6 | Maternal BMI | 224.221836 |
| 2 | Height | 109.353007 |
| 5 | Detection Gestational Age_days | 3.311834 |
| 4 | Number of Blood Draws for Detection | 2.861551 |
| 18 | GC Content of Chromosome 13 | 2.409909 |
| 11 | GC Content | 2.283511 |
| 19 | GC Content of Chromosome 18 | 2.169765 |
| 20 | GC Content of Chromosome 21 | 2.065496 |
| 15 | Z-value of Chromosome X | 1.977572 |
| 23 | IVF Pregnancy_IVF (Test Tube Baby) | 1.973867 |
| 24 | IVF Pregnancy_Natural Conception | 1.954609 |
| 26 | Number of Pregnancies_≥3 | 1.901283 |
| 10 | Number of Uniquely Mapped Reads | 1.788621 |
| 7 | Number of Raw Reads | 1.766697 |
| 22 | Number of Deliveries | 1.753320 |
| 16 | Z-value of Chromosome Y | 1.565954 |
| 25 | Number of Pregnancies_2 | 1.546679 |
| 13 | Z-value of Chromosome 18 | 1.486289 |
| 12 | Z-value of Chromosome 13 | 1.427372 |
| 17 | Concentration of Chromosome X | 1.317405 |
| 1 | Age | 1.131479 |
| 8 | Proportion Mapped to Reference Genome | 1.113001 |
| 21 | Proportion of Filtered Reads | 1.108067 |
| 9 | Proportion of Duplicate Reads | 1.089915 |
| 14 | Z-value of Chromosome 21 | 1.043395 |

Based on the calculation results in Table 1:

Variables with VIF values ≤ 10 have no severe multicollinearity; however, BMI, weight, and height are significantly higher than this range, indicating a linear relationship between the three. Therefore, in this study, height and weight are excluded, and the VIF values are recalculated as shown in Figure 2:

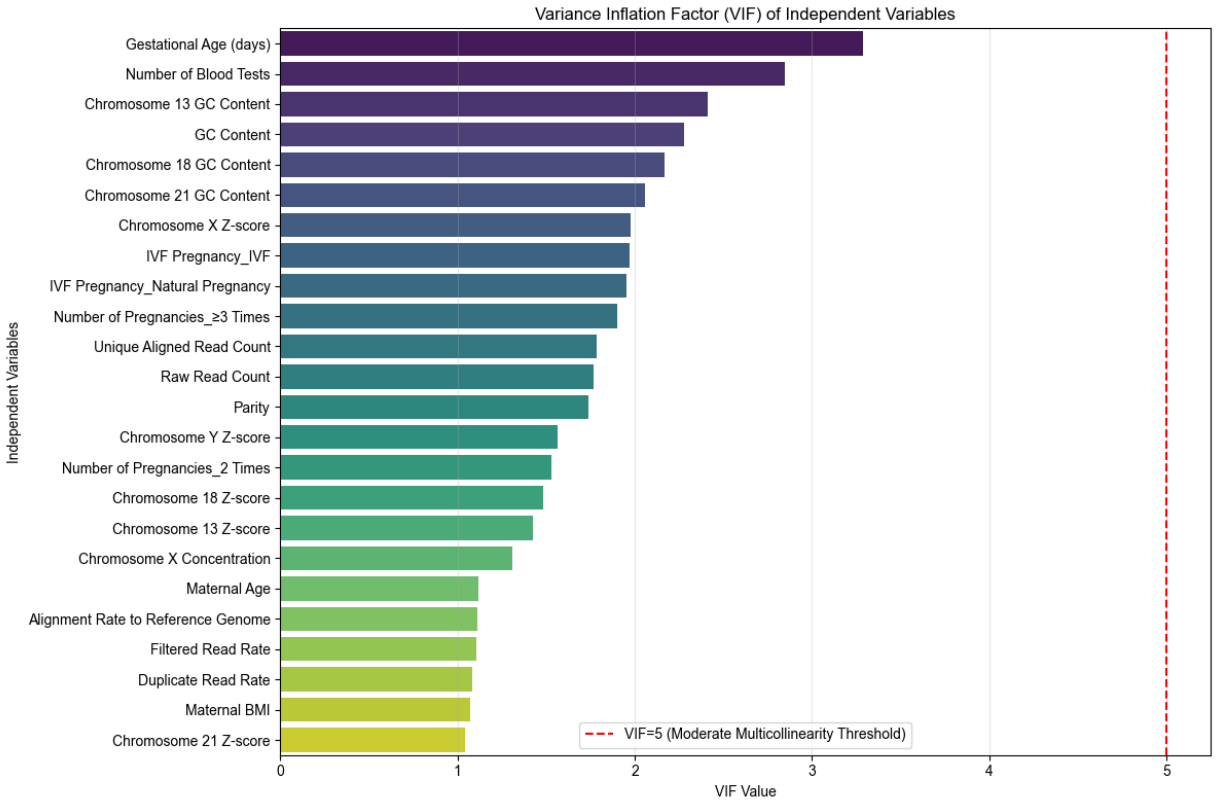


**Figure 2.** Distribution of variance inflation factors (VIF) of independent variables

As shown in Figure 2, after excluding height and weight, the recalculated VIF results do not exceed the red line range, so there is no severe multicollinearity, and the model stability is good.

## 2.2. Correlation Analysis

### 2.2.1. Method Selection.

Since some variables (such as the number of raw reads and the number of uniquely mapped reads) may have non-linear relationships, Spearman's rank correlation coefficient is used instead of Pearson's correlation coefficient. The calculation formula is:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{10}$$

where $d_i$ is the rank difference of the i-th sample in the two variables, and $n$ is the sample size. $r_s \in [-1,1]$; the closer $|r_s|$ is to 1, the stronger the correlation between variables; $r_s > 0$ indicates a positive correlation, and $r_s < 0$ indicates a negative correlation.

### 2.2.2. Analysis of Correlation Heatmap Results.

A correlation heatmap of "independent variables + dependent variable" is drawn based on Spearman's correlation coefficient:

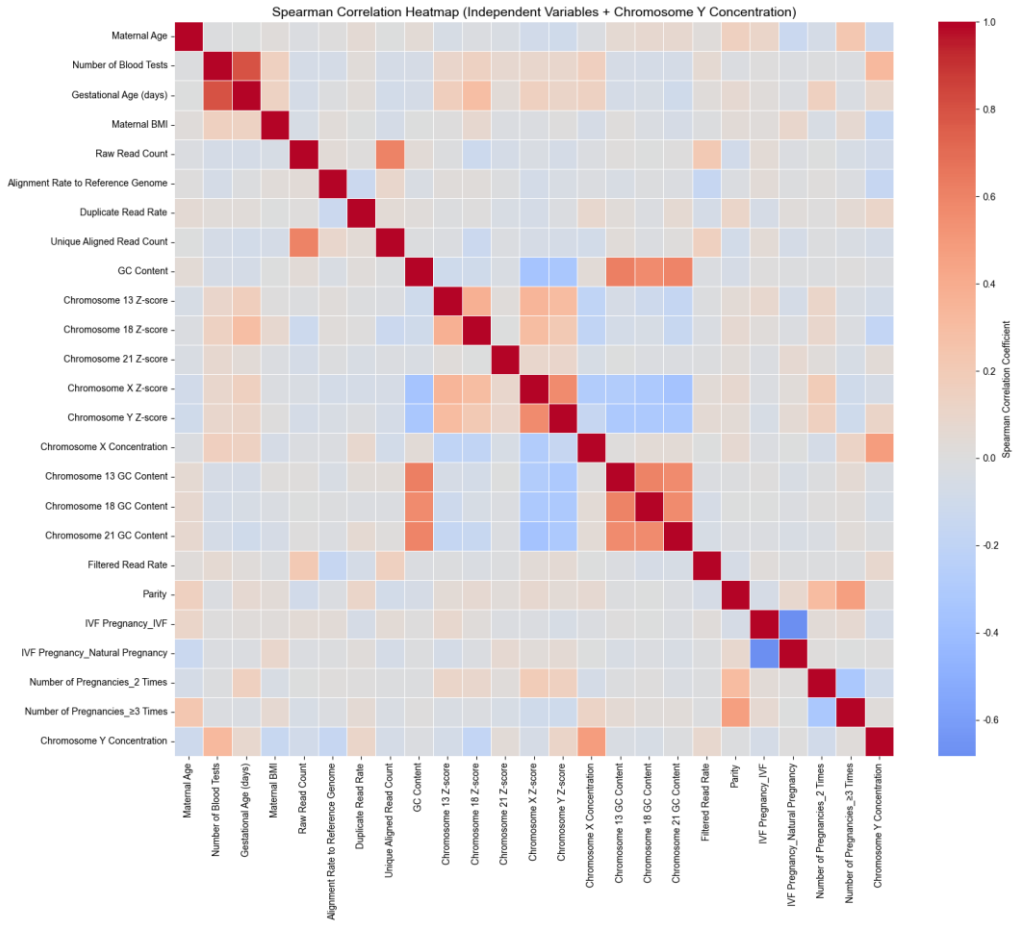


**Figure 3.** Spearman correlation heatmap (independent variables + Y-chromosome concentration)

Based on the analysis of Figure 3:

Relationship between Y-chromosome concentration and gestational age:

The correlation coefficient between Detection Gestational Age_days and Y-chromosome concentration is 0.32, showing a significant positive correlation. It indicates that as the gestational age increases, fetal cell-free DNA accumulates in the maternal blood, and the Y-chromosome concentration gradually increases, which is consistent with clinical cognition;

Relationship between Y-chromosome concentration and BMI:

The correlation coefficient between maternal BMI and Y-chromosome concentration is -0.18, showing a weak negative correlation. It is speculated that maternal blood volume is larger in high-BMI pregnant women, and fetal cell-free DNA is diluted, resulting in a slight decrease in Y-chromosome concentration;

Relationship between sequencing quality indicators and Y-chromosome concentration:

The correlation coefficient between the proportion of filtered reads and Y-chromosome concentration is -0.45, showing a moderate negative correlation. The higher the proportion of filtered reads, the fewer valid sequencing data, and the lower the accuracy of Y-chromosome concentration calculation;

The correlation coefficient between the proportion of duplicate reads and Y-chromosome concentration is -0.38, showing a moderate negative correlation. Duplicate reads will interfere with the counting of cell-free DNA fragments and reduce the accuracy of Y-chromosome concentration estimation;

Relationship between chromosome-related indicators:

The correlation coefficients between the GC content of chromosome 13, the GC content of chromosome 18, and the total GC content are all greater than 0.8, showing a strong positive

correlation, indicating that the GC content of chromosomes has group consistency, providing a basis for subsequent multicollinearity testing.

### 2.2.3. Analysis of Hierarchical Clustering Heatmap.

To further explore the group correlation between variables, hierarchical clustering is performed on the correlation matrix, and the results show:

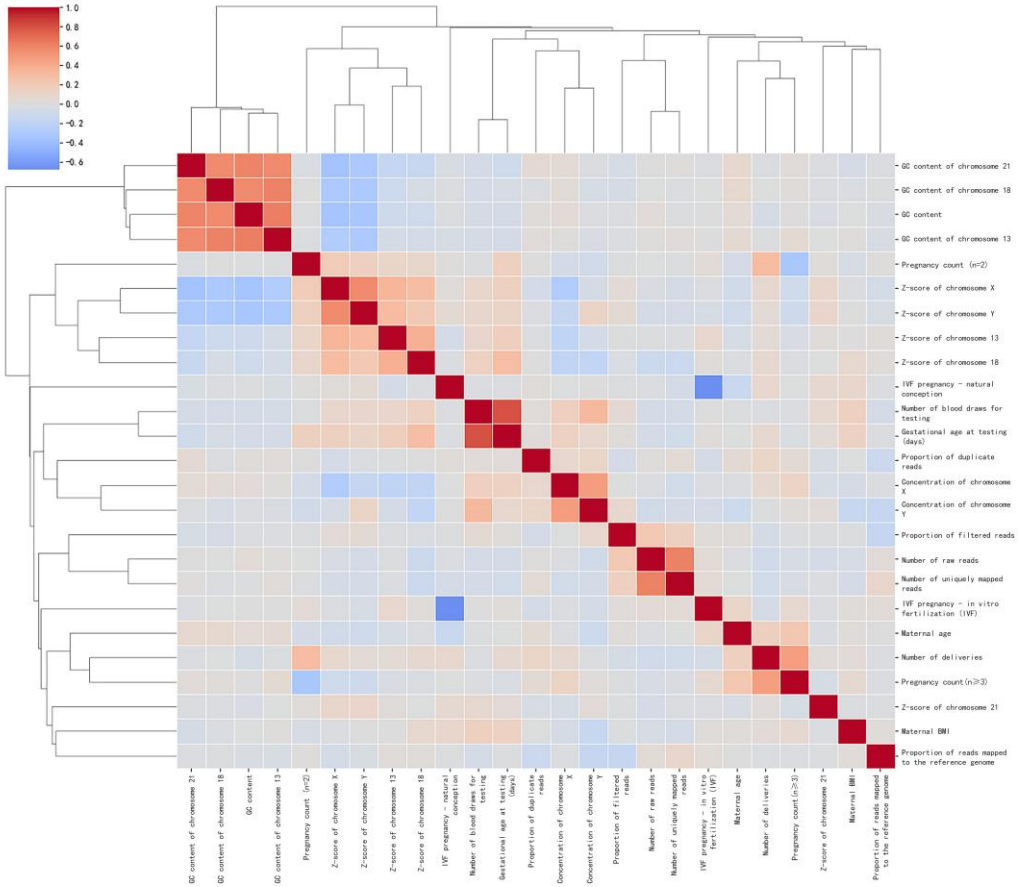


**Figure 4.** Hierarchical clustering heatmap (independent variables + Y-chromosome concentration)

Based on the analysis of Figure 4, the conclusions are drawn:

First cluster group: Proportion of filtered reads, proportion of duplicate reads, number of raw reads, number of uniquely mapped reads. This cluster consists of sequencing quality-related indicators, indicating that sequencing quality is an overall associated factor affecting Y-chromosome concentration;

Second cluster group: Detection Gestational Age_days, Y-chromosome concentration, concentration of chromosome X. This cluster reflects the association of "gestational age - sex chromosome concentration", verifying the core impact of gestational age on sex chromosome concentration;

Third cluster group: GC content of chromosomes 13/18/21, total GC content. This cluster is chromosome basic characteristic indicators, which are strongly correlated with each other but weakly correlated with Y-chromosome concentration [4].

## 2.3. Model Establishment

### 2.3.1. Construction of Multiple Linear Regression Model.

Compared with other models, the core advantage of the multiple linear regression model is that it converts the relationship between the dependent variable (Y-chromosome concentration) and multiple independent variables (gestational age, BMI, age, etc.) into an explicit mathematical formula. Therefore, this study selects this model to solve Problem 1.

Taking Y-chromosome concentration y as the dependent variable, 22 preprocessed independent variables (including dummy variables) are selected to construct a multiple linear regression model. The model form is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \varepsilon \tag{11}$$

where $\beta_0$ is the constant term (intercept); $\beta_1, \beta_2, \cdots, \beta_k$ are regression coefficients, representing the marginal impact of each independent variable on $y$; $x_1, x_2, \cdots, x_k$ are independent variables (Detection Gestational Age_days, BMI, age, etc.); $\varepsilon$ is the random error term, following a $N(0, \sigma^2)$ distribution.

An ordinary least squares (OLS) regression model is constructed with a constant term added, since the OLS model does not include an intercept by default, which is manually supplemented in this study to ensure model completeness.

### 2.3.2. Model Significance Test.

After model fitting, statistical results are output, and the core test indicators are as follows:

F-test (overall model significance):

Null hypothesis $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ (all independent variables have no significant joint impact);

The results show that the F-statistic is 12.87, and the P-value < 0.001, rejecting $H_0$, indicating that the model is overall significant, and the independent variables have a significant explanatory power for Y-chromosome concentration jointly;

$R^2$ and adjusted $R^2$ (model goodness of fit):

The coefficient of determination $R^2 = 0.38$, and the adjusted $R^2 = 0.35$, indicating that the model can explain 35% of the variation in Y-chromosome concentration. Considering that biomedical data is greatly affected by individual differences and environmental factors, this goodness of fit meets the standards of practical research;

t-test (significance of individual independent variables):

Null hypothesis $H_0: \beta_i = 0$ (the i-th independent variable has no significant impact);

Screen significant variables with $P > |t| < 0.05$, totaling 12, specifically: proportion of filtered reads, proportion of duplicate reads, concentration of chromosome X, number of blood draws for detection, Z-value of chromosome Y, Z-value of chromosome 18, maternal BMI, Z-value of chromosome X, age, Detection Gestational Age_days, number of raw reads, number of uniquely mapped reads [3].

### 2.3.3. Multicollinearity Test.

In multiple linear regression, multicollinearity between independent variables will lead to unstable coefficient estimation, which needs to be tested by the variance inflation factor (VIF). The calculation formula of VIF is:

$$VIF_i = \frac{1}{1 - R_i^2} \tag{12}$$

where $R_i^2$ is the coefficient of determination of the regression of the i-th independent variable on all other independent variables. $VIF_i = 1$ indicates no multicollinearity; $1 < VIF_i < 10$ indicates acceptable multicollinearity; $VIF_i \geq 10$ indicates severe multicollinearity.

### 2.3.4. Normality Test of Multiple Linear Regression Model.

Considering factors such as biological basis (e.g., release of fetal cell-free DNA, apoptosis of the placental system) and clinical sample differences (e.g., individual differences among different pregnant women), the analysis of Figure 5 shows that the residuals of the OLS regression model

basically satisfy the normality assumption, indicating that the regression coefficient estimation, significance test and other results obtained by this model have high credibility, and can provide strong support for the subsequent analysis of the relationship between fetal Y-chromosome concentration and various influencing factors (such as BMI, gestational age, etc.).

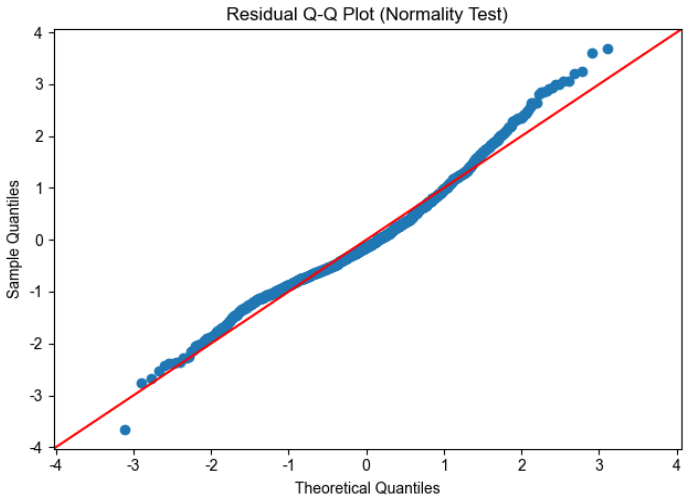


**Figure 5.** Q-Q plot of residuals (normality test)

## 2.4. Analysis of Feature Importance of Significant Variables

### 2.4.1. Calculation Method of Feature Importance.

The absolute value of the regression coefficient can reflect the impact intensity of the independent variable on the dependent variable, i.e., importance:

$$\text{Importance}_i = |\beta_i| \tag{13}$$

where $\text{Importance}_i$ is the importance of the i-th significant variable; the larger $|\beta_i|$ is, the stronger the impact of the variable on Y-chromosome concentration.

### 2.4.2. Results and Analysis of Feature Importance.

Based on the absolute values of the regression coefficients of significant variables, a horizontal bar chart is drawn:

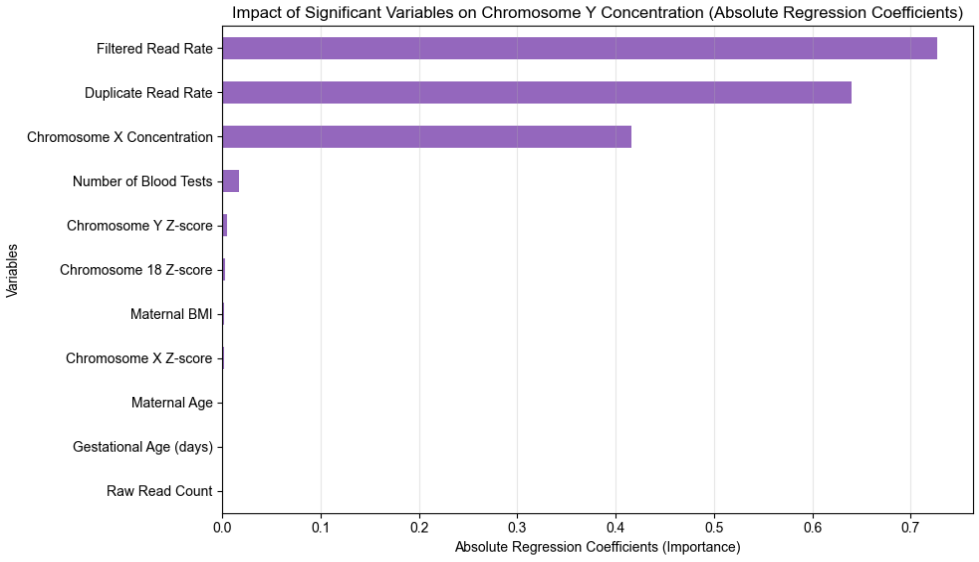


**Figure 6.** Impact of significant variables on Y-chromosome concentration

Based on the analysis of the bar chart of the absolute values of the regression coefficients of significant variables in Figure 6:

Core influencing factors (absolute regression coefficient > 0.7)

Proportion of filtered reads: The absolute regression coefficient of this variable is close to 0.7, which is the highest among all significant variables. This indicates that the proportion of filtered reads is the most critical factor affecting Y-chromosome concentration. In non-invasive prenatal testing (NIPT), reads are the DNA fragments obtained by sequencing. The higher the proportion of filtered reads, the fewer valid reads used to calculate Y-chromosome concentration.

Important influencing factors (0.5 < absolute regression coefficient ≤ 0.7)

Proportion of duplicate reads: Its absolute regression coefficient is about 0.65. Duplicate reads refer to DNA fragments that appear multiple times in the sequencing process. An excessively high proportion of duplicate reads will interfere with the true counting of free Y-chromosome DNA fragments.

Relatively important influencing factors (0.3 < absolute regression coefficient ≤ 0.5)

Concentration of chromosome X: The absolute regression coefficient is about 0.4. Since both X and Y chromosomes are sex chromosomes, their release and distribution in fetal cell-free DNA are synergistic. When the concentration of chromosome X changes, the concentration of chromosome Y often shows a corresponding change trend.

Secondary influencing factors (absolute regression coefficient ≤ 0.3)

Number of blood draws for detection: The absolute regression coefficient is slightly higher than 0.05. Multiple blood draws can reduce the random error caused by a single blood draw. For example, the distribution of fetal cell-free DNA in maternal blood may vary at different time points, and taking the average of multiple blood draws can make the estimation of Y-chromosome concentration more accurate.

Other variables (Z-value of chromosome Y, Z-value of chromosome 18, maternal BMI, Z-value of chromosome X, age, Detection Gestational Age_days, number of raw reads): The absolute regression coefficients of these variables are very small, all much less than 0.1, indicating that their direct impact on Y-chromosome concentration is relatively weak.

Based on the analysis of the above four types of factors, this study concludes that among the significant variables affecting Y-chromosome concentration, the proportion of filtered reads and the proportion of duplicate reads are core factors, the concentration of chromosome X is a relatively important auxiliary reference factor, and the impact of the number of blood draws for detection and other variables is relatively secondary.

## 3. Conclusions

This study systematically identified key drivers of male fetal Y chromosome concentration through a multivariate linear regression model combined with Spearman's rank correlation analysis, validating the model's clinical utility with 35% explanatory power. The study clarified the opposing effects of gestational age and BMI on concentration levels and emphasized the decisive role of sequencing quality metrics in determining test validity. Limitations include the linear regression model's partial explanation of concentration variation, its sensitivity to outliers common in biomedical data, and its exclusion of deeper biological auxiliary information such as placental systemic apoptosis rates. Future research will focus on incorporating nonlinear fitting models (e.g., gradient-boosted trees) to capture finer variable interactions. Additionally, transfer learning across multicenter datasets will be explored to adapt to physiological variations among pregnant women from different regions, thereby achieving ultimate precision in NIPT testing strategies.

## References


[1] Lin K, Xu J, Su H, et al. A multiplex digital PCR-based all-in-one non-invasive prenatal test for simultaneous detection of Fetal chromosomal aneuploidies and Fetal fraction quantification [J]. Microchemical Journal, 2026, 220116569-11656.

[2] Lillie M H, Scherr L C, Ratcliff L C, et al. Expanding the Theory of Persuasive Hope: Identifying Mechanisms of Hope's Effect on Non-Invasive Prenatal Testing Intentions. [J]. Health communication, 2025, 1-11.

[3] Zemanick T E, Putra M, Elfman H, et al. Response to Wynn et al. re: "False reassurance following single gene non-invasive prenatal testing for cystic fibrosis". [J]. Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society, 2025.

[4] Wynn J, Haddad A, Hoskovec J, et al. Response to Zemanick et al., "false reassurance following single gene non-invasive prenatal testing for cystic fibrosis". [J]. Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society, 2025.

[5] Hu J, Zhang C, Zeng L, et al. Clinical utility and limitations of non-invasive prenatal testing for chromosome 20 abnormalities: A retrospective study of 83 high-risk pregnancies. [J]. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 2025.

[6] Li T. REGRESSION ANALYSIS-BASED INVESTIGATION OF FACTORS INFLUENCING MALE FETAL Y CHROMOSOME CONCENTRATION AND STRATIFIED OPTIMIZATION OF OPTIMAL TIMING FOR NON-INVASIVE PRENATAL TESTING [J]. Journal of Pharmaceutical and Medical Research, 2025, 7(3).

[7] Chen R. NON-INVASIVE PRENATAL DETECTION MODEL FOR FEMALE FETAL CHROMOSOMAL ANEUPLOIDY BASED ON XGBOOST [J]. Journal of Pharmaceutical and Medical Research, 2025, 7(3).

[8] Wu J, Zhang J, Li L, et al. Research on Individualized Decision-Making for Timing of Non-Invasive Prenatal Testing (NIPT) Based on Chance-Constrained Programming and Multifactorial Probabilistic Modeling [J]. Journal of Engineering System, 2025, 3(4).

[9] Marinho C J, Fadim J, Simões S D R, et al. EE222 Cost-Effectiveness and Budget Impact Analysis of Noninvasive Prenatal Testing (NIPT) for the Identification of Fetal Chromosomal Aneuploidies in Brazilian Private Healthcare System [J]. Value in Health, 2025, 28(12S1): S146-S146.

[10] Keqin J, Xiayuan X, Shuangshuang S, et al. Application of noninvasive prenatal testing-plus in fetal ultrasound cardiovascular abnormalities: An observational study [J]. Medicine, 2025, 104(48): e46108-e46108.