Optimizing spin-column elution parameters for efficient HPV DNA recovery from preservative liquid-based cytology cervical scrape samples

NILAMSARI KURNIASIH INDRAWAN[1], WURI HANDAYANI[1], HENNY MEITRI ANDRIE R. PUTRI[2,3], KARTIKA HAPSARI[4,5], PARAMITA SUDIRMAN[1], NUR AL FAIZAH[1], INNAS WIDIASTI[6], MUHAMMAD AL-HADDAD P. IKO[6,7], AJENG KUSUMANINGTYAS PRAMONO[8,9], PERTHO RINALDI MARPAUNG[10], NABILA TSOERAYYA GUSTIA PUDJAS[1], VINCENTIUS SIMEON WEO BUDHYANTO[1] and PUTRI WIKIE NOVIANTI[6]

[1]Laboratory Optimization Team, Yayasan Satriabudi Dharma Setia, Tangerang 15345, Indonesia; [2]Department of Obstetrics and Gynecology, Indonesia Army Hospital, Jakarta 10410, Indonesia; [3]Department of Obstetrics and Gynecology, Universitas Pembangunan Nasional Veteran, Jakarta 12450, Indonesia; [4]Department of Obstetrics and Gynecology, Rumah Sakit Umum Pusat Nasional Dr. Cipto Mangunkusumo, Jakarta 10430, Indonesia; [5]Bintaro Woman and Children Clinic, Tangerang 15229, Indonesia; [6]Siena Clinical, Siena Sains Medika, Jakarta 10340, Indonesia; [7]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia; [8]GENEPICA, London, HA4 7AE, UK; [9]Center for Biomedical Research, Badan Riset dan Inovasi Nasional, Bogor 16912, Indonesia; [10]Faculty of Medicine, Universitas Indonesia, Jakarta 10430, Indonesia

Received April 21, 2026; Accepted August 10, 2026

DOI: 10.3892/wasj.2026.504

Abstract. Optimizing DNA extraction protocols is critical for improving downstream molecular analyses, particularly for low-cellularity clinical samples, such as HPV ThinPrep liquid-based cytology specimens. Variations in pre-analytical extraction parameters substantially influence DNA yield, purity and integrity; yet, these factors remain underexplored beyond default manufacturer guidelines. The present study systematically evaluated the effects of lysis incubation duration (30, 40, 50, 60 and 80 min), elution volume (ranging from 30 to 75 μl) and the number of elution cycles on DNA recovery using a silica spin-column extraction method. The extraction performance was assessed using fluorometry and spectrophotometry, alongside DNA integrity number (DIN) evaluation using a TapeStation. To account for between-sample biological variations and storage temperature confounders, a linear mixed-effects model (LMM) was employed for DNA yield, while a generalized linear mixed model (GLMM) evaluated purity. Response surface methodology was subsequently applied to determine non-linear parameter effects and global optima. The LMM identified the number of elution cycles and its interaction with incubation duration as critical determinants of DNA yield. Response surface methodology optimization identified an individual optimum for maximizing yield, and for optimal purity. A compromise operational point was established to simultaneously satisfy both responses. Furthermore, the first elution cycle consistently recovered the vast majority of DNA mass, rendering a second cycle largely redundant. Quality control analysis demonstrated that an elevated DNA yield does not inherently guarantee a high DNA integrity, emphasizing the compounding impact of storage temperature. On the whole, these findings demonstrate that tailoring extraction parameters and utilizing a single-elution workflow can significantly enhance the quality and quantity of DNA recovered from low-biomass clinical specimens for advanced molecular diagnostics.

Introduction

Molecular biomarkers are rapidly used in clinical practice for disease detection, particularly in oncology, where they often outperform traditional diagnostic methods (for instance: Biopsy examination and visual inspection with acetic acid in cervical cancer screening) in sensitivity and specificity (1). Generating reliable molecular biomarker data requires high-quality clinical specimens, which can be challenging to obtain, particularly in low- and middle-income countries (2,3). Persistent barriers include limited numbers of trained personnel capable of performing proper sample collection, as well as logistical constraints related to transportation, cold-chain maintenance

Correspondence to: Dr Putri Wikie Novianti, Siena Clinical, Siena Sains Medika, Menara Cakrawala 12th Floor Unit 5A, Jl. M.H. Thamrin No. Kav. 9, Kebon Sirih, Menteng DKI, Jakarta 10340, Indonesia
E-mail: pwnovianti@sienaclinical.com

Key words: DNA extraction optimization, lysis incubation time, elution volume, DNA yield and purity, HPV ThinPrep samples

and storage (4). These challenges may compromise specimen integrity, ultimately affecting the accuracy and reliability of downstream molecular analyses.

Liquid-based cytology clinical samples, such as cervical scrapes collected with a brush and preserved in ThinPrep PreservCyt® solution, are widely used in human papillomavirus (HPV) detection for cervical cancer triage screening, as they efficiently preserve cellular material for both cytological and molecular analyses. However, a frequent bottleneck in downstream molecular assays, such as in viral load quantification, genotyping and sequencing, is inefficient DNA extraction, leading to a low yield or suboptimal purity of DNA (5,6).

To address this critical initial step, silica spin-column based extraction kits are widely used in both clinical and research laboratories due to their convenience, scalability and relative consistency. However, these standard protocols may not be optimal for all sample types or downstream applications, a limitation that is particularly acute for low-viral-load HPV samples. Therefore, the selection of elution parameters (particularly volume, incubation duration and the number of elutions) becomes a key optimization step, as they can significantly affect both the final DNA yield and the concentration of the eluate (7).

The optimization of protocol variables, such as incubation time, elution volume and number of elutions can significantly influence DNA yield and quality (8,9). Given that ThinPrep-preserved samples contain preservative agents, variable cell concentrations and potential inhibitors, protocol optimization is particularly relevant. Optimization research reveals trade-offs: Reduced elution volume increases the concentration, but may leave residual DNA on the column, while extended incubation improves recovery (10).

To the best of our knowledge, published data on the optimization of elution variables, specifically for the Monarch gDNA Spin Column kit (New England Biolabs, Inc.) using cervical swab clinical samples is relatively limited. Therefore, the Monarch kit was consciously selected due to its technical advantages in fundamental research, providing superior DNA purity, highly concentrated output and specialized protocols, such as HMW DNA protocols for downstream applications (11,12). The present study aimed to evaluate the effects of the following parameters on DNA concentration and purity from HPV ThinPrep cervicovaginal-scrape samples using the Monarch gDNA Spin Column kit: Incubation duration, elution volume and the number of elutions.

Materials and methods

Clinical specimen samples. A total of 17 cervical swab specimens were collected from women who underwent independent cervical cancer screening at the researchers' affiliated hospitals in Jakarta, Indonesia, including RSAB Harapan Kita (the author KH was originally affiliated with this hospital and changed affiliations during the study period as part of her professional appointment) and Gatot Soebroto Army Hospital (Indonesia Army Hospital), between August, 2024 and November, 2025. All specimens were processed and analyzed under the approvals granted by the Institutional Review Boards of both participating institutions. Samples were collected at the participating hospitals and subsequently transferred to the designated research laboratory for molecular analyses in accordance with the approved ethical protocols. Written informed consent was obtained from all participants. The cervical scrape samples were preserved in ThinPrep® PreservCyt® Solution (Hologic Inc.) and transported to the centralized main storage facility. Upon arrival, the samples were stored under controlled conditions at 4˚C, -20˚C or -80˚C until nucleic acid extraction. A total of 11 samples were initially maintained at 4˚C for up to 10 months, after which they were transferred to -20˚C for an additional 4 months of storage. In addition, 6 samples were stored at -80˚C for periods ranging from 2 to 10 months. Although the extended storage of PreservCyt aliquots at 4˚C exceeds the manufacturer's (Hologic Inc.) initial 21-day recommendations, previous large-scale evaluations have demonstrated that human genomic DNA remains structurally intact and viable for molecular amplification in PreservCyt at 4˚C for up to 2.5 years, provided that robust, crosslink-reversing extraction methodologies are employed (13). All samples were processed within the predefined acceptable storage timeframe prior to extraction. These samples were then aliquoted to ensure uniformity. The aliquots were used in parallel for the different extraction condition arms.

DNA extractions, quantification and purity assessment. Total genomic DNA was extracted using the Monarch Spin gDNA Extraction kit (cat. no. T3010; New England BioLabs, Inc.) as per the manufacturer's standard protocol through the binding and washing steps. This specific extraction system was selected for its optimized high-temperature lysis chemistry, which is required to effectively reverse the protein-nucleic acid crosslinking inherent to alcohol-based cytology media like PreservCyt (13). The extraction system was selected to facilitate efficient DNA recovery from preserved cervical specimens (14). Previous research has demonstrated that the optimization of digestion and incubation conditions can improve DNA recovery from fixed clinical samples (14). In the present study, extraction performance was evaluated based on DNA yield and purity, while DNA integrity was assessed using TapeStation analysis using a 4200 TapeStation System (Agilent Technologies, Inc.), as a quality control (15).

Samples were initially incubated at room temperature for 2 h. For the lysis step, following the addition of lysis buffer, Proteinase K and RNAse A which is included with the extraction kit Monarch gDNA Spin Column kit (New England Biolabs, Inc.) were added to the samples which are in 1.5 ml-microcentrifuge tube. For optimized lysis conditions, samples were incubated under at 56-60˚C for the specified duration before proceeding to subsequent binding steps; for the two-elution condition, eluates were collected either as separate fraction (to quantify incremental recovery) or pooled when total DNA mass was the endpoint. To determine the optimal conditions for reversing PreservCyt crosslinking and maximizing target recovery, three parameters were systematically evaluated: lysis incubation duration at 56-60˚C (30, 40, 50, 60 and 80 min), elution volume (30, 50, 60 and 75 μl), and the number of elution cycles (single vs. sequential double elution using the provided elution buffer).

Due to limited specimen volume, each treatment condition (incubation time x elution volume x elution count) was performed using replicate aliquots derived from a single

homogenized specimen per patient. For each patient, 4 ml of sample was homogenized and aliquots were randomly assigned to extraction conditions to reduce potential batch and processing biases during DNA extraction. Eluted DNA concentration was quantified using a fluorescence-based assay on the Quantus™ Fluorometer CAT E6150 (Promega Corporation), while purity was assessed by spectrophotometry (NanoDrop) using A260/A280 and A260/A230 ratios; samples were classified as 'pure' when A260/A280 was 1.8-2.0 and A260/A230 was ≥2.0 (16), and samples outside these thresholds were categorized as potentially affected by contaminants.

DNA integrity analysis. DNA quality and quantity were evaluated as part of the quality control assessment for the extraction optimization workflow using the 4200 TapeStation system (Agilent Technologies, Inc.). Samples were prepared following the manufacturer's protocols, utilizing the Genomic DNA ScreenTape assay (cat. no. 5067-5365). Data were analyzed using Agilent TapeStation software version 5.1 (Agilent Technologies, Inc.).

DNA integrity was evaluated by determining the distribution of fragment sizes, expressed as the percentage of DNA within specific size regions relative to the total analyzed DNA. The upper and lower size thresholds for quantification were defined based on the fragment size profile within the instrument's measurement range, with a lower threshold set at 100 bp. The DNA integrity number (DIN) was calculated based on the proportion of DNA fragments above the lower limit within the functional DIN range (17).

Experimental design. The primary response variables evaluated in the present study were DNA yield (ng/μl) and purity status, categorized as either 'pure' or 'contaminated'. The experimental parameters were defined as the independent variables, namely: Incubation duration (t), elution volume (v) and elution cycle (e). The objective of this analysis was to identify the optimal combination of these parameters to maximize both yield and purity levels.

To evaluate the influence of varying experimental conditions within the constraints of limited specimen volumes, a fractional factorial design was employed. This approach utilized specific combinations of the previously defined incubation durations and elution volumes, resulting in a total of nine distinct treatment configurations for 'Yield' and 'Purity' (Table SI). This specific layout was necessitated by the high value and limited availability of the 17 clinical specimens. In clinical optimization studies, the \Sparsity of Effects\ principle suggests that primary system behavior is driven predominantly by main effects and low-order interactions. By selecting cells within the 16-cell grid (focusing on the 40-min and 80-min duration tiers), the characterization of the most promising operational ranges was prioritized, while maintaining ethical and practical stewardship of patient samples. This approach allowed for the exploration of the experimental scope more broadly than a narrow, balanced design of fewer parameters would have permitted.

Statistical analyses. The statistical analyses in the present study were designed to achieve two primary objectives. First, mixed-effect models were employed to evaluate the association between the response variables and the independent parameters. This initial step served to validate the hypothesis that DNA 'Yield' and 'Purity' are influenced by the three selected aforementioned parameters. Notably, the focus of this specific analysis was to establish the existence of these associations (i.e., providing evidence of statistical significance) rather than quantifying the magnitude of the effect sizes. To account for variability in storage conditions, this factor was included as a random effect in all mixed models. Second, response surface methodology (RSM) was employed to identify the optimal parameter settings for each individual response variable and, subsequently, to determine the global optimum for both responses simultaneously.

Linear mixed-effect model for 'Yield'. Linear mixed-effects models were employed to estimate the fixed effects of incubation duration (t), elution volume (v), and the number of elutions (e) on the continuous outcome variables: DNA yield (ng/μl). The mixed-effect modeling approach was selected to account for the hierarchical data structure in which multiple technical replicates (k) are nested within each treatment combination (j) applied to the same biological sample (i), thereby requiring mixed-effects to accommodate within-sample correlation (18).

The experimental parameters (t, v, e) were included as fixed effects with a full factorial structure encompassing all two- and three-way interactions. Between-sample biological variation was modeled through a by-sample random intercept b_i, assumed to be normally distributed with zero mean and variance σ_b^2. The model specification for response y_{ijk} (the k-the replicate from the i-th sample under the j-th treatment combination) is:

$$y_{ijk} = \beta_0 + \beta_1 t_j + \beta_2 v_j + \beta_3 e_j + \beta_4 (t \cdot v)_j + \beta_5 (t \cdot e)_j + \beta_6 (v \cdot e)_j + \beta_7 (t \cdot v \cdot e)_j + b_i + \varepsilon_{ijk}.$$

where the global intercept (β_0) represents the baseline response. Coefficients β_1 to β_7 quantify the fixed effects of predictors and their interactions. The random intercept b_i captures sample-specific deviations from the global mean, assumed normally distributed with variance σ_b^2. Residual error ε_{ijk} accounts for unexplained within sample variability, following a normal distribution with variance σ^2.

Observations with missing values were excluded listwise. Model parameters were estimated via restricted maximum likelihood (REML) using the lme4 package (version 1.1-35.1) (19). The significance of fixed-effect coefficients was assessed using t-tests with Satterthwaite's degrees-of-freedom approximation, implemented in the lmerTest package (20).

Generalized linear mixed-effect model for DNA 'Purity'. To analyze the factors influencing extraction purity, a binary composite response variable was constructed based on the previously defined spectrophotometric ratios, categorized as 'Pure' and 'Contaminated'. This binary response was analyzed using a Generalized linear mixed model (GLMM) with a logit link function. This framework is particularly suited for clustered experimental outcomes and non-Gaussian responses, as it explicitly accounts for the nested nature of laboratory measurements (21). Wang *et al* (22), introduced a mixed-effects logistic regression framework designed for large genomic case-control studies, illustrating the capacity of GLMMs to accommodate correlated observations and heterogeneity across experimental units. Similarly, Noma and

Gosho (23), discussed the utilization of logistic mixed-effects models for clustered binary clinical outcomes to emphasize the importance of random effects in accounting for within-subject and within-group dependence.

Let Y_{ijk}^{purity} denotes the purity outcome for the *k*-th observation from the *i*-th biological sample under the *j*-th treatment combination, with $P(Y_{ijk}^{purity} = 1) = \pi_{ijk}^{purity}$. The model is specified as:

$$Y_{ijk}^{purity} \sim Bernoulli(\pi_{ijk}^{purity});$$

$$logit(\pi_{ijk}^{purity}) = ln(\frac{\pi_{ijk}^{purity}}{1-\pi_{ijk}^{purity}}) = \eta_{ijk}^{purity}.$$

The linear predictor η_{ijk}^{purity} included the same fixed-effect factors and interactions as the continuous model, plus a random intercept for biological sample:

$$\eta_{ijk}^{purity} = \gamma_0 + \gamma_1 t_j + \gamma_2 v_j + \gamma_3 e_j + \gamma_4 (t \cdot v)_j + \gamma_5 (t \cdot e)_j + \gamma_6 (v \cdot e)_j + \gamma_7 (t \cdot v \cdot e)_j + u_i,$$

where γ_0 is the fixed global intercept on the log-odds scale, $\gamma_1, \ldots, \gamma_7$ are the fixed-effect coefficients for the main effects and interactions, and $u_i \sim N(0, \sigma_u^2)$ is the random intercept for sample.

The model was fitted via maximum likelihood estimation using the glmer function from the lme4 package in R (19), with a binomial family specified. Wald z-tests were used to evaluate the significance of the fixed effects. The model was fitted via maximum likelihood estimation using the glmer function from the lme4 package in R (19), with a binomial family specified.

RSM optimization. Following the inferential analyses, RSM was employed to visualize the mixed models on each DNA 'Yield' and 'Purity' responses. The response surfaces were graphically represented using two-dimensional contour plots and three-dimensional surface plots. Stationary points within the fitted models were determined through canonical analysis, adhering to established response surface methodology diagnostic protocols (24). Multi-response optimization was carried out by jointly considering the predicted outcomes for yield and purity over the entire design space.

Results

Mixed model analyses revealed that the evaluated experimental factors accounted for a small proportion of the variance in each response, as indicated by the R^2 values shown in Table I. This suggests that other unobserved factors substantially contribute to the overall variability. A notable difference in the modeling results was observed between the LMM for DNA yield and the GLMM for DNA purity. Based on the standardized coefficients and model selection using AIC and likelihood ratio tests (LRT), the removal of individual predictors from the full model did not significantly worsen model fit, as all LRT P-values exceeded the conventional threshold of 0.05. However, among the yield predictors, the number of elutions (*e*) exhibited the largest change in AIC when removed, suggesting its relatively stronger contribution, while the interaction term between incubation duration (*t*) and elution volume (*v*) contributed minimally (Table I).

For DNA purity, the number of elutions exhibited the largest standardized coefficient, indicating a potential role in explaining purity, although its contribution was not statistically significant. The R^2 values further indicated that fixed effects explained only a small to moderate proportion of the variance, with marginal R^2 values ranging from 0.098 to 0.0267, while the inclusion of random effects substantially increased the explained variance, with conditional R^2 values ranging from 0.418 to 0.485 (Table I). This suggests that unmeasured factors beyond the experimental parameters evaluated in the present study may play a substantial role in determining both DNA yield and purity.

Jackknife resampling analyses were performed to evaluate the stability of both mixed-effects models, with detailed resampling results presented in Tables SII and SIII. These analyses demonstrated high agreement with the models derived from the full dataset; specifically, 17 out of 17 resampled models for DNA yield, and 10 out of 17 models for DNA purity, were consistent with the primary full models. These analyses demonstrate that despite the limited sample size and the use of second-order modeling, the results remain robust and stable.

Subsequently, the previously described mixed-effects models were mapped using RSM. It is important to emphasize that, herein, RSM was utilized strictly as an analytical tool to visualize and extract optimal points from the existing second-order mixed models, rather than to fit new independent models. Both response variables were first optimized individually based on the two continuous parameters using their corresponding mixed-effects models. For DNA yield, the RSM analysis identified a stationary optimum at 77 min of incubation duration and an elution volume of 42 μl, producing a predicted yield of 4,420 ng/μl. Although the mixed model for DNA purity lacked statistically significant predictors, its optimum parameters were still assessed via RSM visualization. The optimum point for DNA purity was achieved at 80 min of incubation and an elution volume of 75 μl, resulting in a predicted purity probability of 0.99, indicating a very high likelihood of achieving acceptable DNA purity under these conditions (Fig. 1).

Identifying experimental parameters that satisfy both responses simultaneously, the mean values of the individual optima were calculated and presented in Table II, resulting in a 'compromised' condition of a 59-μl elution volume and 79 min of incubation. This averaged parameter set represents a shift from the yield optimum, specifically an increase of 17 μl in elution volume and an increase of 2 min in incubation time, as well as a shift from the purity optimum, namely a decrease of 17 μl in elution volume and a decrease of 1 min in incubation time, thereby providing a balanced trade-off between DNA yield and purity (Fig. 1). Providing a quantitative basis for the optimization, the final fitted second-order polynomial equations for DNA yield and purity obtained from the mixed-effects RSM method are given below:

$$Yield = 3353.55 + 904.37t - 373.30v - 130.32tv - 242.41t^2 - 443.31v^2 - 1731.79elogit(Purity)$$
$$= 0.2133 - 0.1398t - 0.0989v - 0.0087tv - 0.0026t^2 - 0.0017v^2 + 0.02133e$$

Table I. Summary of the results of mixed models on 'Yield' and 'Purity' responses.

Response variable	Fixed-effect variable	Std. coefficient	P-value	AIC	P-value (LRT)	R^2
Yield	Incubation duration (t)	0.096	0.283	557.627	0.216	R^2m=0.242, R^2c=0.479
	Elution volume (v)	-0.061	0.587	556.335	0.579	R^2m=0.266, R^2c=0.467
	No. of elutions (e)	-0.262	0.174	559.177	0.137	R^2m=0.239, R^2c=0.434
	Incubation duration^2 (t^2)	-0.04	0.586	555.468	0.528	R^2m=0.260, R^2c=0.482
	Elution volume^2 (v^2)	-0.072	0.167	556.462	0.141	R^2m=0.244, R^2c=0.418
	Incubation duration x elution volume ($t.v$)	-0.021	0.912	557.121	0.904	R^2m=0.267, R^2c=0.482
Purity	Incubation duration (t)	-0.615	0.538	47.748	0.776	R^2m=0.181, R^2c=0.483
	Elution volume (v)	-0.539	0.59	47.734	0.796	R^2m=0.175, R^2c=0.485
	No. of elutions (e)	1.426	0.154	49.857	0.139	R^2m=0.098, R^2c=0.418
	Incubation duration^2 (t^2)	-1.133	0.257	48.652	0.321	R^2m=0.129, R^2c=0.453
	Elution volume^2 (v^2)	-0.877	0.381	49.244	0.209	R^2m=0.144, R^2c=0.455
	Incubation duration x elution volume ($t.v$)	1.613	0.107	48.257	0.443	R^2m=0.157, R^2c=0.441

Standardized coefficients are shown for Yield. P-values for Yield were obtained using Wald t-tests with Satterthwaite's degrees-of-freedom approximation, while P-values for Purity were obtained using Wald z-tests. AIC and R^2 values are from reduced models with the corresponding term removed. R^2m=marginal R^2 (fixed effects only); R^2c=conditional R^2 (fixed + random effects). Storage temperature was included as a random effect.

Table II. Optimization results based on the mixed models for the response variables.

Optimization approach	Response	Volume (v)	Duration (t)	Prediction response and its 95% confidence interval (CI)
Single-response	Yield	42	77	5,247.567 (859.110; 9,503.83)
	Purity	75	80	0.999 (0.177; 1.000)
Multi-responses	Yield	59	79	4433.583 (-142; 9099.336)
	Purity			0.966 (0.154, 0.999)

Optimization approach: Single-response optimization refers to the determination of optimal conditions for each response variable independently based on its respective RSM model. Multi-response optimization was applied to satisfy both responses simultaneously by calculating the mean values of the individual optimal settings, resulting in a compromised dual-response optimum. Confidence intervals: Values in parentheses represent the 95% confidence intervals (CI) for the predicted response values at the corresponding optimum points. Elution condition: All predictions for DNA yield and purity were generated using Elution = 1, which was identified as the optimal elution condition based on model estimates.

DNA integrity quality control. The quality assessment of DNA extracted from HPV samples (001, 002, 003, 004, 005A, 005B, 005C, 006, 007, 009, 010, 011, 014, 015, 016, 017, 018 and 019) using the TapeStation platform demonstrated considerable variability in DNA integrity and fragment size distribution among samples. Based on the DIN values presented in Table SIV, samples HPV 002, HPV 003, HPV 004, HPV 005A, HPV 014, HPV 017 and HPV 018 exhibited relatively high DNA integrity (DIN ≥7.0). Among these, HPV 003 showed the highest DIN value (8.4), followed by HPV 014 and HPV 017 (DIN 7.9), indicating excellent preservation of genomic DNA. HPV 002 and HPV 018 also demonstrated high integrity with DIN values of 7.6 and 7.0, respectively. By contrast, HPV 001 exhibited a comparatively low DIN value (4.5), suggesting substantial DNA degradation. Samples HPV 007, HPV 011, HPV 015, HPV 016 and HPV 019 displayed moderate DNA integrity, with DIN values ranging from 5.1 to 6.4, indicating partial fragmentation of genomic DNA.

Notably, the DNA concentration was not consistently associated with DNA integrity. Several samples with a high DNA yield, including HPV 019 (8,150 ng), HPV 016 (7,075 ng) and HPV 018 (8,400 and 7,300 ng), did not necessarily exhibit the highest DIN values or the most intact fragment profiles. By contrast, samples HPV 003, HPV 004 and HPV 014, which exhibited comparatively moderate DNA concentrations, demonstrated superior DNA integrity characterized by higher DIN values and dominant HMW DNA fragments. These findings indicate that elevated DNA concentration alone is insufficient to predict DNA quality, as high DNA yield may include fragmented or partially degraded DNA

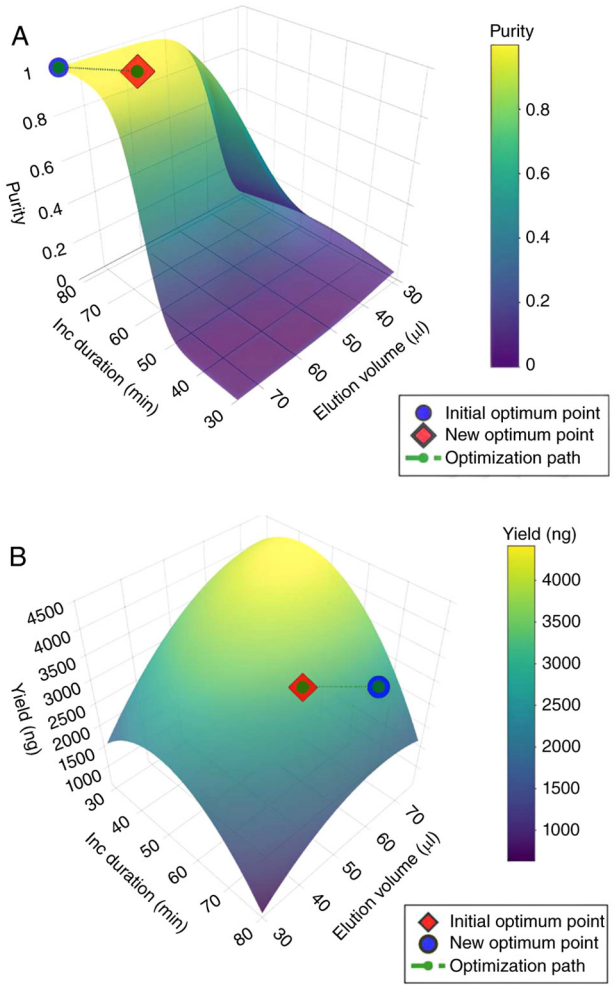

Figure 1. Visualization of response surface methodology results. Columns represent the visualization type: 2D contour plots (left panels) and 3D response surface plots (right panels). Rows represent the optimized parameters: (A) 'Yield' and (B) 'Purity'. The green dotted lines indicate the predicted optimum values for each response variable.

molecules. Overall, these results indicate that DNA integrity is more strongly associated with the preservation of long DNA fragments and elevated DIN values than with total DNA concentration alone. Samples demonstrating a high DNA integrity, particularly HPV 003, HPV 004, HPV 014, HPV 017 and HPV 018, may therefore be more suitable for downstream long-read sequencing applications. By contrast, samples with moderate DNA integrity may remain appropriate for amplification-based molecular assays, such as PCR or qPCR targeting short amplifications. Samples with a lower DNA integrity, including HPV 001, may require further optimization of extraction procedures and storage conditions to improve DNA preservation for advanced molecular analysis.

Discussion

The present study aimed to identify the optimal DNA extraction parameters (incubation duration (*t*), elution volume (*v*) and the number of elutions (*e*) and DNA integrity analysis from cervicovaginal specimens in ThinPrep buffer, a matrix often characterized by low cellularity and high dilution. The experimental data, albeit derived from a limited sample size, suggest that incubation time, elution volume and the number of elutions each contributed to extraction performance, with the number of elutions exhibiting the most consistent statistical association with DNA yield. These preliminary results may have implications for supporting more consistent DNA recovery in downstream molecular diagnostics. In clinical screening workflows, particularly for specimens with low cellularity, even minor variations in extraction efficiency can influence the success of downstream molecular detection. This is particularly relevant in resource-limited settings where specimen quality is often compromised, though further validation with larger sample cohorts is warranted.

The RSM fitted using the mixed-effects framework, evaluated the effects of incubation duration (*t*), elution volume (*v*) and the number of elution cycles (*e*) on DNA yield and purity. As the predicted optimum conditions for DNA yield and DNA purity occurred at different combinations of incubation time and elution volume, an operational compromise condition was determined by averaging the coordinates of the two individual optima. This approach resulted in a compromise condition of approximately 79 min of incubation and 59 μl of elution volume (Table II), providing a balanced operating condition that maintained satisfactory DNA yield while preserving DNA purity. The higher efficiency of the first elution observed in the present study is consistent with the desorption kinetics of silica membranes, where DNA release is greatest during the initial contact with the elution buffer and decreases in subsequent elution cycles. Although additional elution cycles may increase total DNA recovery the incremental gain becomes progressively smaller indicating diminishing returns (8).

Furthermore, the present study observed a trade-off between total yield and concentration governed by elution volume. Previous research has reported that low-volume elution can produce more concentrated nucleic acid eluates and may improve recovery of low-abundance targets (9). The data presented herein suggest that within the 30-80 μl range, these effects act independently, allowing for flexibility based on the specific requirements of downstream applications. Lower elution volumes which yield higher DNA concentrations are more advantageous for next-generation sequencing (NGS) workflows, whereas higher elution volumes producing more dilute eluates are better suited for PCR-based applications (25,26). This optimization aligns with a broad body of literature [e.g., Akahane *et al* (27); Shibata *et al* (6) and Naegele *et al* (5)] showing that the delicate balance between DNA quantity, purity and concentration is the primary determinant of success in sensitive molecular assays.

Consequently, for laboratories processing ThinPrep specimens with silica-spin kits, prioritizing an extended incubation of ~77 min over increasing elution volumes or adding redundant elution steps to maximize DNA yield may be recommended. For maximizing DNA purity, an extended incubation of ~80 min is recommended instead. While a second elution remains a viable strategy when maximum DNA mass is required for replicate assays, it should be noted that this approach provides diminishing returns and does not substantially alter purity metrics. Ultimately, elution volumes in the mid-range should be selected based primarily on the concentration requirements of the specific downstream workflow, such as NGS or qPCR, as this flexibility allows for

standardized, cost-effective protocols that are compatible with existing laboratory infrastructure in resource-limited settings. By maintaining measurable DNA quality and integrity profiles suitable for molecular applications, these optimized extraction conditions may support standardized extraction workflows compatible with downstream molecular analyses.

The optimization results demonstrated that achieving high DNA purity required a relatively larger elution volume compared to conditions that maximized DNA yield or those recommended by the Monarch extraction protocol. This observation is likely influenced by both intrinsic sample quality and the physicochemical behavior of silica-based extraction columns. Low-biomass samples require higher elution volumes to improve DNA release from silica membranes, while insufficient volumes result in incomplete recovery and reduced yield and purity (28,29).

Previous research has reported that increasing elution volume enhances DNA desorption efficiency under low-salt conditions, particularly in silica column-based extraction systems, albeit at expense of DNA concentration (8). Higher elution volumes enhance DNA purity by minimizing co-elution of residual contaminants (28,30). During centrifugation, a larger elution volume can promote more effective spatial separation eluted DNA and contaminants retained within the silica membrane or column matrix, thereby minimizing carryover into the final eluate (29).

In the present study, the purity cut-off values applied for DNA quality assessment were not strictly based on the minimum input specifications recommended by the NGS library preparation kit, but were instead determined through a combination of empirical considerations and evidence from recent literature. Although manufacturers recommend A260/280 (1.8-2.0) and A260/230 (2.0-2.2) ranges, recent studies show long-read sequencing performance depends more on DNA integrity and fragment length than strict spectrophotometric thresholds (12,31). This approach aligns with recent reports suggesting that functional validation through sequencing performance provides a more robust assessment of DNA suitability than reliance on purity ratios alone (5,27).

Furthermore, the findings of the present study highlight that utilizing DNA integrity as a fundamental quality control step reveals a lack of consistent correlation with total DNA concentration. As a quality control indicator, high DIN values are highly preferred; previous research indicates that DIN strongly predicts the success of downstream analyses, especially NGS (32). On the other hand, quality control assessments demonstrating a predominance of short fragments reflect DNA degradation driven by biological and environmental processes, leading to progressive fragmentation (33).

The extraction process parameters, particularly lysis duration and temperature, play a decisive role in preserving long DNA fragments. These factors are particularly critical in PreservCyt samples, where protein-DNA crosslinking requires controlled lysis conditions to prevent additional fragmentation (34,35). In addition, the findings of the present study revealed cases of high DNA concentration that do not exhibit broad fragment size distribution, suggesting that DNA quantity does not necessarily reflect its functional quality. This condition is likely due to the accumulation of degraded DNA, emphasizing the importance of evaluating fragment size distribution alongside quantification (33,36).

Moreover, the observation that increasing elution volume and the number of elution cycles does not consistently improve DNA integrity highlights the limitations of conventional approaches that focus primarily on maximizing DNA recovery. Although yield optimization is often used as a key indicator of extraction success, particularly in limited clinical samples (37), the findings of the present study suggest that such strategies may overlook critical aspects of DNA quality. As previously reported by Simbolo *et al* (36) short DNA fragments may dominate even when the total yield is high. Therefore, the presnet study expands the current perspective by emphasizing that extraction optimization should balance both DNA quantity and integrity.

The observed variability in DNA quality also points to the influence of pre-analytical factors, such as sample storage and handling, which can accelerate DNA degradation and affect final analytical outcomes (38). In practice, this has a direct impact on the efficiency of downstream application, including sequencing performance. High-integrity DNA is essential for achieving optimal analytical results (35). The heterogeneous DNA integrity observed in the present study may be attributed to variations in storage conditions (4, -20 and -80˚C) and storage durations. These findings are consistent with and further extend those of previous reports indicating that DNA degradation is cumulative and strongly influenced by storage conditions and sample handling (33,38); in the other words, the stability reported in the literature does not entirely eliminate the risk of fragmentation if not accompanied by optimal extraction conditions.

Overall, DNA integrity assessment provides complementary quality-control information beyond yield and purity measurements and may assist in evaluating sample suitability for downstream molecular applications. The optimization analyses suggested that elution conditions were more strongly associated with DNA recovery than incubation duration or elution volume within the tested range. Collectively, the findings support the use of a single-elution workflow and provide a framework for optimizing DNA extraction from ThinPrep specimens, while acknowledging that further validation using downstream molecular assays is required (8). Collectively, these findings confirm that elution protocol primarily determines DNA yield while storage temperature serves as a confounding factor affecting both yield and purity. Based on these results, the recommended protocol for optimal HPV DNA extraction is using the elution 1 method, with an incubation time of 77 min and an elution volume of 41 μl to maximize yield, while storage at -80˚C is advised to maintain DNA purity. While a second elution remains a viable strategy when maximum DNA mass is required for replicate assays, it should be noted that this approach provides diminishing returns and does not substantially alter purity metrics. Ultimately, elution volumes in the mid-range should be selected based primarily on the concentration requirements of the specific downstream workflow, such as NGS or qPCR, as this flexibility allows for standardized, cost-effective protocols that are compatible with existing laboratory infrastructure in resource-limited settings (39,40). By maintaining measurable DNA quality and integrity

profiles suitable for molecular applications, these optimized extraction conditions may support standardized extraction workflows compatible with downstream molecular analyses.

The present study provides a framework for optimization, several limitations remain that offer clear avenues for future research. First, while the present study focused on yield and purity, it did not directly measure downstream performance metrics, such as read length distribution, mapping rates, or HPV detection sensitivity. Establishing this direct link via qPCR Ct shifts and NGS coverage remains a critical next step to validate the clinical utility of these optimized parameters. Additionally, although spectrophotometric ratios (namely: A260/A280 and A260/A230) provide essential purity data, they do not fully account for the presence of PCR inhibitors. Future studies incorporating qPCR inhibition assays, spike-in controls, or the testing of supplemental techniques, such as elution buffer warming and the addition of carrier nucleic acids would provide more in-depth biochemical insight into enhancing recovery from samples with critically low cellularity. Finally, as laboratories move toward high-throughput diagnostic settings, a formal cost-benefit analysis will be essential for economic optimization. As noted by Neal *et al* (41), such low-cost optimizations are increasingly vital for preserving the high-quality DNA required for advanced molecular testing in resource-constrained environments, ensuring that procedural efficiency does not come at the expense of diagnostic accuracy.

In conclusion, the present study demonstrated that the optimization of pre-analytical extraction parameters using linear mixed models identified incubation time as a contributor to DNA yield variation from cervicovaginal samples in ThinPrep buffer without compromising purity. Incubation time and elution volume independently influenced DNA recovery, suggesting opportunities to optimize extraction conditions according to downstream concentration requirements. By prioritizing procedural optimization over increased buffer volume, this approach may improve recovery efficiency, particularly in low-cellularity samples. However, as downstream functional performance was not evaluated, these findings should be considered preliminary recommendations pending validation in diagnostic workflows. Overall, the results presented herein provide a potentially cost-effective strategy to improve DNA extraction efficiency and inform future studies on clinical and molecular performance.

Acknowledgements

The authors wish to express their gratitude to Professor Renske Steenbergen and Ms. Annina van Splunter (VU University Medical Center, Amsterdam, The Netherlands) for their insightful initial discussions regarding the optimization of DNA extraction. The authors would also like to thank Dr Ahmad R. Utomo and Mrs. Kinasih Prayuni (Yarsi University, Jakarta, Indonesia) for their valuable feedback on the optimization results and their suggestions for evaluating DNA quality. In addition, the authors would like to thank Mr. Abdul Azizul Hakim and Mrs. Christine Sembiring (Yayasan Satriabudi Dharma Setia, Tangerang, Indonesia) for their exceptional management of research logistics and for facilitating cross-institutional discussions.

Funding

The wet-lab experiments, subsequent laboratory analyses and data interpretations were supported by research funding from PaninBank through the Yayasan Satriabudi Dharma Setia initiative to advance genomic and molecular diagnostic research in Indonesia. Downstream data analyses and part of manuscript preparation were supported by Siena Clinical under Internal grant no. MCI-XI26-002.

Availability of data and materials

The data generated in the present study may be requested from the corresponding author.

Authors' contributions

NKI, NAF, HMARP, KH and PWN conceptualized the study. AKP conceptualized the laboratory experiments and experimental design. NKI and PWN were involved in the study methodology. NKI, WH, PS and NTGP were involved in the investigative aspects of the study. PRM contributed to sample collection. NKI, WH, PS, PRM and NTGP were involved in data curation. NKI, MAHPI, NAF, IW and PWN were involved in the formal analysis. MAHPI was involved in visualization. PRM and IW were involved in project administration. MAHPI and PWN were involved in the writing of the original draft of the manuscript. IW and PWN were involved in the writing, review and editing of the manuscript. VSWB, HMARP, KH and PWN supervised the study. VSWB and AKP was involved in data validation. VSWB and NKI confirm the authenticity of all the raw data presented in this study. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

The study was approved by the Medical Ethics Review Committee of the RSAB Harapan Kita (Jakarta Indonesia; approval no. IRB/04/02/ETIK/2023) and the Indonesia Army Hospital (Jakarta, Indonesia; approval no. 127/XI/KEPK/2024). Writen informed consent was obtained from all participants.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Yusuf M: Perspectives on cervical cancer: Insights into screening methodology and challenges. Cancer Screen Prev 3: 47-55, 2024.
2. Jayathissa P and Rupasinghe DA: Review on biomarkers for disease diagnosis and disease prevention: Use case in low and middle-income countries (LMICs). Int J Res Publication and Reviews Journal homepage: www.ijrpr.com [Internet]. 2024. www.ijrpr.com.

3. Sayed S, Cherniak W, Lawler M, Tan SY, El Sadr W, Wolf N, Silkensen S, Brand N, Looi LM, Pai SA, *et al*: Improving pathology and laboratory medicine in low-income and middle-income countries: Roadmap to solutions. Lancet 391: 1939-1952, 2018.
4. Albano PM, Notarte KI, Macaranas I and Maralit B: Cross-contamination in molecular diagnostic laboratories in low- and middle-income countries. Philippine J Pathol 5: 7-11, 2020.
5. Naegele K, Weissbach FH, Leuzinger K, Gosert R, Bubendorf L and Hirsch HH: Impact of nucleic acid extraction procedures on human papillomavirus (HPV) detection and genotyping. J Med Virol 95: e28583, 2023.
6. Shibata T, Nakagawa M, Coleman H, Owens S, Greenfield W, Sasagawa T and Robeson MS II: Evaluation of DNA extraction protocols from liquid-based cytology specimens for studying cervical microbiota. PLoS One 16: e0237556, 2021.
7. Donà MG, Benevolo M, Pimpinelli F, Battista M, Rollo F, Stivali F, Moscarelli A, Giuliani M, Di Carlo A and Vocaturo A: Comparative evaluation of different DNA extraction methods for HPV genotyping by linear array and INNO-LiPA. J Med Virol 83: 1042-1047, 2011.
8. Dilley K, Pagan F and Chapman B: Methods for ensuring the highest DNA concentration and yield in future and retrospective trace DNA extracts. Sci Justice 61: 193-197, 2021.
9. Vutukuru MR, Sharma DK, Chakraborty I, Mukhopadhyay D and Mitra N: A rapid and high-yield method for nucleic acid extraction. Sci Rep 15: 12479, 2025.
10. Van Biesen N, Cools P and Meyers E: Comparison and optimization of DNA extraction methods for human DNA from dried blood spot samples. Pediatr Rep 17: 30, 2025.
11. Zamuner FT, Ramos-López A, García-Negrón A, Purcell-Wiltz A, Cortés-Ortiz A, Cuevas AR, Gosala K, Winkler E, Sidransky D and Guerrero-Preston R: Evaluation of silica spin-column and magnetic bead formats for rapid DNA methylation analysis in clinical and point-of-care settings. Biomed Rep 21: 112, 2024.
12. Trigodet F, Lolans K, Fogarty E, Shaiber A, Morrison HG, Barreiro L, Jabri B and Eren AM: High molecular weight DNA extraction strategies for long-read sequencing of complex metagenomes. Mol Ecol Resour 22: 1786-1802, 2022.
13. Agreda PM, Beitman GH, Gutierrez EC, Harris JM, Koch KR, LaViers WD, Leitch SV, Maus CE, McMillian RA, Nussbaumer WA, *et al*: Long-term stability of human genomic and human papillomavirus DNA stored in BD surePath and hologic preservCyt liquid-based cytology media. J Clin Microbiol 51: 2702-2706, 2013.
14. Steinau M, Patel SS and Unger ER: Efficient DNA extraction for HPV genotyping in formalin-fixed, paraffin-embedded tissues. J Mol Diagn 13: 377-381, 2011.
15. Padmanaban A, Inche A, Gassmann M and Salowsky R: High-throughput DNA sample QC using the agilent 2200 tapestation system. J Biomol Tech 24: S41, 2013.
16. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA and Aguilar-Quesada R: DNA source selection for downstream applications based on DNA quality indicators analysis. Biopreserv Biobank 14: 264-270, 2016.
17. Lu X, Wei Y, Sun J, Xiao B, Zhang X, Li W, Chen Y, Lin F, Zhang L, Wang Y, *et al*: A comparative study of three nucleic acid integrity assay systems. Biopreserv Biobank 21: 624-630, 2023.
18. Gelman A and Hill J: Data analysis using regression and multilevel/hierarchical models. New York: Cambridge University Press; 2007.
19. Bates D, Mächler M, Bolker B and Walker S: Fitting linear mixed-effects models using lme4. J Stat Softw 67: 1-48, 2015.
20. Kuznetsova A, Brockhoff PB and Christensen RHB: lmerTest package: Tests in linear mixed effects models. J Stat Softw 82: 1-26, 2017.
21. Agresti A: Foundations of linear and generalized linear models. WILEY; 2015.
22. Wang JT, Chang XY, Zhao Q and Zhang YM: FastBiCmrMLM: A fast and powerful compressed variance component mixed logistic model for big genomic case-control genome-wide association study. Brief Bioinform 25: bbae290, 2024.
23. Noma H and Gosho M: Logistic mixed-effects model analysis with pseudo-observations for estimating risk ratios in clustered binary data analysis. Stat Med 44: e70280, 2025.
24. Chen YF, Lin PW, Chen WH, Yen FY, Yang HS and Chou CT: Biogas upgrading by pressure swing adsorption with design of experiments. Processes 9: 1325, 2021.
25. McNulty SN, Mann PR, Robinson JA, Duncavage EJ and Pfeifer JD: Impact of reducing DNA input on next-generation sequencing library complexity and variant detection. J Mol Diagn 22: 720-727, 2020.
26. McKee AM, Spear SF and Pierson TW: The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. Biol Conserv 183: 70-76, 2015.
27. Akahane T, Yamaguchi T, Kato Y, Yokoyama S, Hamada T, Nishida Y, Higashi M, Nishihara H, Suzuki S, Ueno S and Tanimoto A: Comprehensive validation of liquid-based cytology specimens for next-generation sequencing in cancer genome analysis. PLoS One 14: e0217724, 2019.
28. Desjardins P and Conklin D: NanoDrop microvolume quantitation of nucleic acids. J Vis Exp 22: 2565, 2010.
29. Fiedorová K, Radvanský M, Němcová E, Grombiříková H, Bosák J, Černochová M, Lexa M, Šmajs D and Freiberger T: The impact of DNA extraction methods on stool bacterial and fungal microbiota community recovery. Front Microbiol 10: 821, 2019.
30. Téblick L, Van Keer S, De Smet A, Van Damme P, Laeremans M, Cortes AR, Beyers K, Vankerckhoven V, Matheeussen V, Mandersloot R, *et al*: Impact of collection volume and DNA extraction method on the detection of biomarkers and HPV DNA in first-void urine. Molecules 26: 1989, 2021.
31. Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P and Snutch TP: MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. Genome Res 28: 266-274, 2018.
32. Hiramatsu K, Matsuda C, Masago K, Toriyama K, Sasaki E, Fujita Y, Haneda M, Ebi H, Shibata N and Hosoda W: Diagnostic utility of DNA integrity number as an indicator of sufficient DNA quality in next-generation sequencing-based genomic profiling. Am J Clin Pathol 160: 261-267, 2023.
33. Pajnič IZ: Analysis of human degraded DNA in forensic genetics. Genes(Basel) 16: 1375, 2025.
34. Santaus TM, Zhang F, Li S, Colin Stine O and Geddes CD: Effects of Lyse-It on endonuclease fragmentation, function and activity. PLoS One 14: e0223008, 2019.
35. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR and Ordoukhanian P: Library construction for next-generation sequencing: Overviews and challenges. Biotechniques 56: 61-77, 2014.
36. Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A, Malpeli G, Malpeli G, Lawlor RT and Scarpa A: DNA qualification workflow for next generation sequencing of histopathological samples. PLoS One 8: e62692, 2013.
37. Choudhary A, Mambo E, Sanford T, Boedigheimer M, Twomey B, Califano J, Hadd A, Oliner KS, Beaudenon S, Latham GJ and Adai AT: Evaluation of an integrated clinical workflow for targeted next-generation sequencing of low-quality tumor DNA using a 51-gene enrichment panel. BMC Med Genomics 7: 62, 2014.
38. Schumacher S, Lauesgaard JM, Carlsson T, Linder A and Sundfeldt K: Optimization of pre-analytical handling to maintain DNA integrity in diagnostic papanicolaou tests. J Mol Diagn 27: 199-208, 2025.
39. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, *et al*: Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36: 338-345, 2018.
40. van Dijk EL, Jaszczyszyn Y, Naquin D and Thermes C: The third revolution in sequencing technology. Trends Genet 34: 666-681, 2018.
41. Neal CJ, Zbinden ZD, Douglas ME and Douglas MR: Reducing DNA extraction costs through factorial design for the DNAdvance Kit. BMC Res Notes 17: 397, 2024.