A combinatorial approach to nucleic acid quality control for efficient workflow standardization and reliable data generation

Katharina Pfeifer-Sancar1, Mirjana Kozulic2, Pierre-Henri Ferdinand1

1 QIAGEN GmbH, Hilden, Germany
2 Hombrechtikon Systems Engineering AG, Switzerland

Abstract: Nucleic acid quality is a critical factor determining the outcome of all molecular workflows. Rigorous quality control (QC) of key steps in the workflow can minimize the risk of failure and maximize the reproducibility and reliability of results. Here, we illustrate how two complementary technologies, UV/Vis spectrophotometry and capillary electrophoresis, can provide both qualitative and quantitative measurement of nucleic acids from diverse sample types used in a multitude of downstream applications. Our data suggest that such a combinatorial approach when implemented along workflows, can conveniently cover most critical sample QC requirements, a prerequisite for delivering success and confidence in data interpretation.

Introduction

The high incidence of irreproducible scientific results is a growing concern, with approximately 28 billion USD spent each year in the United States alone on biomedical research that cannot be reproduced (1). A recent survey of 1,576 researchers conducted by Nature had revealed that more than 70% of the researchers had attempted but failed to reproduce another scientist’s experiments, and more than half even admitted having failed to reproduce their own experiments. Lack of reproducibility in science exists due to diverse reasons ranging from complexity of experiments and statistics, lack of technical expertise, incomplete documentation, poor study design and variability of biological material. Most importantly, a lack of standardization of sample quality at various steps of the molecular workflow generates variability within samples and increases irreproducibility issues (2, 3).

From starting material to final results, every analysis workflow is a journey to unlock the innate information contained within biological samples. Experimental results can be transformed into valuable insights, given that the sample quality has been maintained and the biological message has been unaltered throughout the complete workflow. Nucleic acid samples vary greatly in many aspects and are sensitive to external influence

Contents

Introduction 1
Key sample quality indicators 2
UV/Vis spectrophotometry 2
Capillary electrophoresis 3
Materials and methods 3
gDNA purification and quantification 3
RNA purification, quantification and integrity determination 3
DNA purification from FFPE samples and integrity determination 4
NGS library preparation and analysis by capillary electrophoresis 4
Sample preparation and analysis by capillary electrophoresis prior to Pyrosequencing 4
Results and discussion 4
Quality control of nucleic acids from whole blood, cells and FFPE tissue samples 4
Quality control of DNA products prior to NGS and Sanger sequencing 7
Quality control of nucleic acid samples following an enzymatic digestion 8
Conclusions 9
References 10
from chemicals, nucleases, temperature and radiations, including light. Performing precise qualitative and quantitative measurement early on can prevent precious reagents and samples from being wasted in a failed run.

The current situation warrants the implementation of quality control checkpoints in the research workflow in order to carefully monitor any change in sample parameter or tools employed for producing the sample and the result. Together with superior chemistries, lab automation can eliminate operator-to-operator variations and increase robustness as well as reproducibility of the experiments.

Key sample quality indicators

The key nucleic acid sample quality parameters that can have an impact on your experimental outcome are concentration, purity and integrity. Depending on the nucleic acid type, source and detection technology, sample degradation, concentration or presence of contaminants can significantly impact quality of results and their interpretation. There is no one-for-all solution for assessing all nucleic acid quality parameters in a single measurement run; however, a combination of two technologies, UV/Vis spectrophotometry and capillary electrophoresis, can provide sufficient data for easy and comprehensive as well as time- and cost-efficient quality control (Table 1).

UV/Vis spectrophotometry

While purity of nucleic acid samples can be assessed using a range of analytical methods, spectrophotometry is the most commonly used technology. UV/Vis spectrophotometry measurements enable calculation of nucleic acid concentrations based on the absorbance of the sample at 260 nm. The absorbance at 280 nm and 230 nm can be used to assess the level of contaminating proteins and chemicals, respectively. The absorbance ratio of nucleic acids to contaminants provides a rough estimation of the sample purity as it has been reported that pH and ionic strength can significantly influence these ratios (4). These ratios should nevertheless be considered cautiously when determining the inclusion or exclusion of samples in downstream applications.

However, with classic UV/Vis measurements, RNA and DNA cannot be precisely distinguished as both molecules absorb at 260 nm, potentially leading to overestimation of the concentration and underutilization of the nucleic acid sample in an assay. Dye-based fluorometry can overcome this challenge by using a dye selectively binding to either RNA or DNA, but fails to detect and quantify contaminants, be it proteins, chemicals or undyed nucleic acids molecules due to the difference in their structural conformation.

Development of algorithms for deconvolution of absorbance spectra have resulted in the introduction of analysis methods that are able to separate the spectral contribution of DNA and RNA molecules in solution, without using a dye. This feature known as Spectral Content Profiling (SCP) allows dye-free and easy differentiation between DNA, RNA and other contaminants. In a single measurement, it reports total nucleic acid and molecule of interest concentration separately (RNA or DNA) and normalizes the data by subtracting the identifiable contaminants (5).

<table>
<thead>
<tr>
<th>Table 1. Nucleic acid key quality parameters and the two technologies addressing them all.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purity</strong></td>
</tr>
<tr>
<td>Protein contaminants</td>
</tr>
<tr>
<td>UV/Vis spectrophotometry</td>
</tr>
<tr>
<td>Capillary electrophoresis</td>
</tr>
</tbody>
</table>
Capillary electrophoresis

Upstream processing or contaminating nucleases can lead to degradation or fragmentation of nucleic acid molecules. The electrophoretic signatures of samples such as gDNA, RNA or NGS libraries directly correlate with their integrity and quality. Size distribution can be assessed by electrophoresis. While conventional slab-gel electrophoresis methods have suboptimal performance and comprises error-prone manual steps, alternative technologies such as capillary electrophoresis ensure better performance in terms of resolution, sensitivity, time to result and cost per sample.

RNA samples are particularly labile. Quality indicators such as the RIS (RNA Integrity Score, QIAGEN®) and the RIN (RNA Integrity Number, Agilent® Technologies) allow effective and objective assessment of eukaryotic RNA sample integrity and standardization of sample quality. These samples have unique electrophoretic signatures which encompass the abundant RNA subpopulations corresponding to the 5S, 18S and 28S ribosomal RNA (rRNA) subunits as well as the areas between these major peaks corresponding to the distribution of all other RNA types (e.g., mRNA, regulatory RNAs, etc.). RIS and RIN algorithms analyze different electropherogram features, such as peak intensities, signal areas and ratios, including 28S/18S ratio, and compute a value ranging from 1 to 10 (6). For most assessments, values of 8 or above are considered to correspond to high-quality RNA samples suitable for use in downstream applications. These RNA quality indicators can be used to standardize the quality of RNA samples and gauge the integrity of mRNA, lncRNA, siRNA and miRNA fractions within such samples. However, depending on the degradation mechanism whether it is heat, RNase or UV, the impact on RNA integrity varies significantly and the RIS algorithm has shown moderately better performance in predicting results of qRT-PCR experiments when compared to the RIN algorithm (12). Furthermore, in-depth QC of specific RNA molecules can be achieved using validated qPCR assays designed to precisely assess abundance and integrity of their RNA targets (6, 13).

In this study, we have thoroughly evaluated the performance, benefit and complementarity of the two technologies (UV/Vis spectrophotometry with spectral content profiling and capillary electrophoresis) developed to provide both qualitative and quantitative data on key indicators of sample QC. Frequently used starting materials were assessed for their quality: i) gDNA and RNA immediately after sample extraction and ii) NGS libraries and PCR products prior to either NGS or Sanger sequencing.

Materials and methods

gDNA purification and quantification

Genomic DNA was isolated from human blood samples using the QIAamp® DNA Blood Mini Kit automated on the QIAcube® system. DNA samples were analyzed using the QIAxcel® Advanced capillary electrophoresis system and the QIAxcel DNA Screening Kit in combination with the AM900 method and the 15 bp Alignment Marker as described in Schaller et al. (7). Sample quantification and purity assessment based on spectral content profiling were performed using the QIAxpert® UV/Vis spectrophotometer system and the DNA QIAamp App.

RNA purification, quantification and integrity determination

For RNA quantity and contamination determination, RNA was isolated from 5 x 10⁶ Jurkat cells using the RNeasy® Mini Kit manual protocol, according to manufacturer’s recommendations. One part of the sample was treated with DNase I to remove genomic DNA and the other part was left untreated. To obtain phenol-contaminated RNA samples, purified, DNase-treated RNA (~200 ng/µl) was spiked with 5% (v/v) phenol (1:200 dilution with RNase-free water). Subsequently, RNA quantification was performed on the QIAxpert system using the RNeasy spectral content profiling protocol.

For comparison of RIN and RIS values, total RNA was extracted manually from Jurkat cells using the RNeasy Mini and RNeasy Plus Mini Kits, as well as from rat liver and kidney using the RNeasy Plus Mini and RNeasy Plus Universal Kits, following instructions from the protocol. After RNA isolation, samples from each starting material and extraction method were pooled and concentration determined on the NanoDrop ™. To obtain different states of degraded RNA, 15 µl RNA from each sample pool, independent of concentration, were incubated in a Rotilabo®-Block-Heater H250 at 75°C for 10, 20, 30, 40, 50 or 60 minutes. Thereafter, aliquots were stored on ice until further processing. One RNA sample from each pool, which
was not incubated at 75°C but directly placed on ice, served as control. In a second run, RNA samples were incubated for 15 and 25 minutes at 75°C. After heat degradation, RNA samples were analyzed on the QIAxcel Advanced and the Agilent 2100 BioAnalyzer. On the QIAxcel Advanced, the QX RNA QC Kit v2.0 and the CLRNA method was used, following manufacturer’s recommendations. The 15 bp RNA Alignment Marker and the 200–6000 bp RNA Size Marker were run along with the samples. The RIS of each sample was determined using the ScreenGel Software. On the 2100 BioAnalyzer, the RNA Nano Chip Kit was used in combination with the Eukaryote Total RNA Nano Series II Assay, and RIN values were determined using the Agilent 2100 Expert Software. Finally, RIS values of all samples were plotted against RIN values to determine correlation.

DNA purification from FFPE samples and integrity determination

DNA was isolated from FFPE human kidney and liver samples. Approximately 10 µm thick sections were processed and DNA was purified using the GeneRead™ DNA FFPE Kit automated on the QIAcube, according to manufacturer’s recommendations. The protocol included the optional enzymatic step for the removal of artificial C>T mutations caused by formalin fixation. FFPE DNA was analyzed using the QIAxcel Advanced electrophoresis system, the QIAxcel DNA High Resolution Kit and the OM1200 method. The 15 bp Alignment Marker and the 1–20 kb DNA Size Marker were run simultaneously with the samples. Data were analyzed using the default Smear DNA Analysis protocol of the QIAxcel ScreenGel® Software v1.5. FFPE DNA quantification was performed on the QIAxpert system using the DNA QIAamp spectral content profiling protocol.

NGS library preparation and analysis by capillary electrophoresis

MiSeq® NGS library preparation was achieved as described in Gerrard et al. (8). GeneRead NGS libraries were prepared from FFPE DNA using the QIAGEN GeneRead DNA Library Q Kit. The QIAxcel Advanced system and the QX DNA High Resolution Kit in combination with the “GeneRead TE” and “GeneRead LP” standard analysis methods available from the QIAxcel ScreenGel Software v1.5 were used after target enrichment (TE) and final library preparation procedure (LP), respectively.

Sample preparation and analysis by capillary electrophoresis prior to Pyrosequencing®

Genomic DNA was isolated from FFPE samples using the QIAamp DNA FFPE Tissue Kit. Following PCR amplification as described by Ghaderi, M. (9), the PCR products were analyzed by capillary electrophoresis using the QIAxcel Advanced system, the QX DNA High Resolution Kit and the OM800 method. The QX Alignment Marker 15 bp/600 bp and the QX DNA Size Marker 25–500 bp were used to check the size and concentration of the amplicons as well as to detect the presence of potential non-specific nucleic acid fragments. The detailed experiment setup has been described in previous reports (10, 11, 12).

Results and discussion

Quality control of nucleic acids from whole blood, cells and FFPE tissue samples

Downstream applications can be severely affected by impurities present in the nucleic acid samples, e.g., proteins, salts or solvents, etc., thus jeopardizing the sensitivity and efficiency of the experiment. Furthermore, the presence of contaminants possibly absorbing at 260 nm can lead to an overestimation of the nucleic acid concentration when considering the $A_{260}$ value. The precise composition of the DNA extracted either from human whole blood, Jurkat cells or human FFPE tissue samples were analyzed on the QIAxpert UV/Vis spectrophotometry system. This system allows spectral content profiling of samples, enabling unbiased quantification of nucleic acid of interest by subtracting the spectral contribution of absorbing contaminants, such as proteins, chemicals or other residues. Figure 1 shows the total and normalized nucleic acid quantification and presence of contaminants.

Complementary to yield and purity, an integrity analysis of nucleic acid samples can be carried out by capillary electrophoresis system. It allows monitoring of sample fragmentation and delivers insights into sample’s suitability prior to downstream detection step. Excessive degradation of samples can lead to underestimation of target quantity, thus jeopardizing applications such as long-range PCR and generating false-negative results. Degradation of a target below the limit of detection can
subsequently reduce the fidelity of experimental results. Figure 2 depicts the electrophoretic profiles of gDNA samples mechanically degraded by sonication. Degradation mechanisms lead to the creation of a tail of low-molecular weight fragments observable only by electrophoresis. Samples of high integrity show a high representation of intact nucleic acid molecules, forming a long tail of high-molecular weight fragments. Degradation of gDNA molecule can be easily observed in electropherograms, with increasing degradation upon prolonged sonication reflected by a shift in size distribution of the sample to smaller fragments.

<table>
<thead>
<tr>
<th>gDNA</th>
<th>Total RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultra pure</td>
<td>From FFPE</td>
</tr>
<tr>
<td>Absorbance spectra</td>
<td></td>
</tr>
<tr>
<td>Total absorbance (A_{260})</td>
<td>2.02</td>
</tr>
<tr>
<td>Total nucleic acids (ng/µl)</td>
<td>100.7</td>
</tr>
<tr>
<td>Nucleic acids of interest (ng/µl)</td>
<td>100.7</td>
</tr>
<tr>
<td>Impurities (A_{260})</td>
<td>0.00</td>
</tr>
<tr>
<td>Background (A_{260})</td>
<td>0.00</td>
</tr>
<tr>
<td>Residues (%)</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

Figure 1. Representative QIAxpert data with spectral content profiling from gDNA and RNA samples. gDNA were extracted either from human blood (QIAamp DNA Mini Kit without RNase digest, automated on the QIAcube) or human FFPE kidney sample (GeneRead DNA FFPE Kit automated on the QIAcube). RNA was extracted from Jurkat cells (RNeasy Mini Kit, manual purification with and without DNase digest). Nucleic acid of interest (gDNA or RNA) is indicated by the blue absorbance line while contaminating nucleic acids and all detectable impurities are depicted by the orange line. Sample background is depicted by a grey line while residues (uncharacterized impurities that cannot be attributed to reference profiles used in the algorithm) are depicted by a yellow line. The black curve represents the total absorbance spectra of the sample (nucleic acid, impurities and residues).

Figure 2. Integrity check of gDNA. A Agarose gel showing gDNA in various states of degradation. Lane 1: non-degraded gDNA sample. Lanes 2 and 3: gDNA samples partially degraded by sonication with 7 and 14 pulses, respectively. Lanes 4 and 5: gDNA samples fully degraded with 70 and 105 pulses, respectively.

B Corresponding QIAxcel gel image for samples in lanes 1, 3 and 5.

C From left to right, corresponding QIAxcel electropherograms of samples in lanes 1, 3 and 5, respectively.

D Superimposed electropherogram view of two FFPE gDNA samples (red and green) with a customized high-molecular weight size marker ranging from 1–20 kb (blue).
The quality of the template material used for a PCR or RT reaction has a profound effect on the success and reliability of a resulting qRT-PCR and microarray data. It is thus crucial to maintain a high RNA integrity until its conversion into more stable cDNA. RNA extracted from rat kidney and subjected to heat degradation was analyzed on the QIAxcel Advanced system. A gel image of the samples is shown in Figure 3. The individual and superimposed electropherogram of samples in lanes 1, 7 and 11 indicate the shift in size distribution of RNA molecules toward lower molecular weight. The 28S rRNA shows highest sensitivity to degradation and the corresponding peak in electropherogram disappears with increasing level of degradation.

Nucleic acid integrity was analyzed on both the Agilent BioAnalyzer 2100 and the QIAGEN QIAxcel Advanced system. BioAnalyzer RIN (RNA Integrity Number) was plotted against QIAxcel RIS (RNA Integrity Score) for comparison (Figure 4).

The correlation of RIN and RIS values is fairly high with a $R^2$ of 92.29%. Both values are very good quality indicators of RNA samples. However, when RNA samples are degraded by UV or other cross-linking mechanisms that alter the RNA integrity, such as formaldehyde used for fixation of FFPE samples, the QIAGEN RIS has been considered more reliable in predicting the outcome of qRT-PCR assays (12).

FFPE samples, that can be stored for years before use, are stabilized using a formaldehyde solution which causes chemical modification by reaction between formaldehyde and nucleic acids, including crosslinking of molecules (other nucleic acids, proteins, lipids, etc.) as well as deamination of cytosine, leading to artificial C>T mutations that result in errors in replication and sequencing. After DNA or RNA purification from FFPE tissues, QC is a highly recommended step in order to identify critical quality parameters that can jeopardize the success of the

![Figure 3. Integrity check of RNA.](image)

![Figure 4. Correlation of RIN and RIS.](image)
downstream application. Chemical contamination of FFPE samples can be observed and quantified using SCP to evaluate sample quality (Figure 1). Due to degradation, the electropherogram of DNA FFPE samples shows a broad size distribution. The position in the electropherogram translates its integrity, while the use of reference samples (i.e., size markers of known concentration) allows accurate quantification and size determination of nucleic acids (Figure 2).

While purity and integrity values are strong indicators of FFPE sample quality, they do not provide information about chemical modifications and crosslinks between molecules. Enzymatic treatments included in commercially available kits, such as the GeneRead DNA FFPE Kit used in this study, eliminate artificial C>T mutations from cytosine deamination caused by formalin fixation and aging, that may otherwise result in sequencing errors. However, even after this enzymatic reaction, traces of nucleic acids molecules that are still cross-linked can impair downstream enzymatic assays, such as RT-PCR (13). Therefore, comprehensive assessment of the quality of nucleic acid from FFPE samples should also include a set of qPCR assays for DNA and RT-qPCR assays for RNA in order to determine the upper limit with regard to amplicon size. Alternatively, in the case of oligo-dT-primed cDNA, real-time RT-PCR assays can be carried out using different primer pairs to generate similar-sized amplicons located at an increasing distance from the 3’ end of a RNA transcript. The degree of success of the amplification reactions would indicate the extent of RNA degradation along the entire transcript.

Quality control of DNA products prior to NGS and Sanger sequencing

Next-generation sequencing encompasses multi-step, complex and resource-intensive procedures, including the library preparation steps. The quality of the NGS library is the most critical factor influencing the success of the sequencing run, affecting both the sequence validity and the number of reads. The QIAxcel Advanced capillary electrophoresis system was used to monitor the performance and efficiency at each step of the library preparation process: after shearing of gDNA, amplification of the adapter ligated libraries and amplification of the captured libraries (Figure 5).

Figure 5. MiSeq NGS library preparation steps analyzed on the QIAxcel Advanced system. A Gel image of library samples along the library preparation steps. B Sample information table. C Single electropherogram view of samples in lanes A2, A9, A10 and A12. Black arrows indicate presence of adapters and primer dimers in the samples after ligation and amplification.
This library comprises 80 ribosomal proteins (RP) encoding genes for screening of rare mutations associated with Diamond-Blackfan Anemia (DBA) using the Illumina® MiSeq platform [8]. At an intermediate stage of library preparation, unligated adapters and primer dimers must be cleared from the final library by size selection. The presence of adapters, primer dimers and/or unwanted large fragments are indicative of an unsuccessful library preparation, rendering complete failure of a sequencing run.

A shift in the size of target-enriched samples and libraries after adapter ligation can be observed during a QC check of libraries prepared using the GeneRead DNA Library Q Kit on the GeneReader NGS system and analyzed on the QIAxcel Advanced system (Figure 6).

![Figure 6. Gel image of GeneRead NGS libraries before and after adapter ligation. Target-enriched libraries (lanes A5–A8) and final libraries after adapter ligation and size selection (lanes B1–B4). Lane 5: DNA size marker 50–800 bp at 5 ng/µl concentration.](image)

Just as for NGS, a quality assessment of products prior to either Sanger or Pyrosequencing is of utmost importance in order to verify the correctness of their size and determine their purity. While SCP of samples can be performed as shown in Figure 1, capillary electrophoresis allows critical assessment by detecting undesirable nucleic acid contaminants that can interfere with the sequencing run, such as primer dimers or non-specific DNA fragments. An example of successful PCR screening of samples after the amplification of c-kit gene at exon 11 is demonstrated in Figure 7 (10). c-kit genes are found in gastrointestinal stromal tumors (GIST) and non-small cell lung cancers (NSCLC).

After the amplification, PCR products were analyzed on the QIAxcel Advanced system to identify samples showing the presence of duplications and/or deletions compared to the amplification length of a healthy genotype (wild-type) prior to Pyrosequencing for detailed sequence analysis (9).

![Figure 7. PCR screening of deletion and duplication mutations in c-kit exon 11 prior to Pyrosequencing. Lanes A2, A7 and A10: Wild-type exon 11 allele, Lane A11: Exon 11 homozygous deletion, Lane A12: Exon 9 duplication (p.A502 Y503dup).](image)

Quality control of nucleic acid samples following an enzymatic digestion

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) system constitutes an RNA-guided gene-editing technology that makes use of a bacterially-derived protein (Cas9) and a synthetic guide RNA (gRNA) to introduce a double-stranded break at a specific location within the genome. Editing is achieved by transfecting a cell with the Cas9 protein along with a specially designed gRNA that directs the cut through hybridization with its matching genomic sequence. Quantitative measurement of CRISPR/Cas9 gene editing can be achieved using a simple and reproducible T7EI assay as described by Fu, Y. et al and illustrated in Figure 2.3 “Representative capillary electrophoresis traces from a T7EI experiment” of the article (14). Specific restriction fragments can be observed in the edited genome, whereas no fragments are detected in a negative control, thus indicating a successful assay.
Conclusions

There are multiple factors influencing the successful outcome of a molecular workflow. One major factor determining the success of an assay and its reproducibility is the quality of the sample itself. Performing quality control early on enables you to make well-informed decision on samples prior to processing them through extensive workflows, thus saving precious samples and resources. There are a number of techniques that may be used to assess nucleic acid quantity and quality, but none of these alone can provide all of the information required to fully describe a sample. In this study, we have demonstrated a combinatorial approach toward nucleic acid quality control for efficient workflow standardization and reliable data generation.

The results presented in this study establish that the QIAxcel Advanced capillary electrophoresis system in combination with the QIAxpert UV/Vis spectrophotometer can cover the comprehensive assessment of key quality parameters of nucleic acid samples. The two techniques complement each other with regard to the insights they provide into sample quality and the minimum resources they require for the analysis, such as faster time to result, low cost per sample and reduced training effort, thus making them ideal for nucleic acid QC. They can be seamlessly integrated into every molecular workflow and can maximize the robustness and reproducibility of the experiments performed. Moreover, modern lab automation involving the QIAxpert and the QIAxcel Advanced systems support additional features, such as SCP and RIS, respectively that leverage better quality assurance and traceability owing to unified digital data analysis, management, reporting and exchange format.
References

5. QIAxpert Spectral Content Profiling – Technical Information [Link].
7. Schaller, A. et al. (2016) Quality control of genomic DNA using the QIAxcel system.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at www.qiagen.com or can be requested from QIAGEN Technical Services or your local distributor.