Quality Control of Genomic DNA for The French Kidney Disease Study by the Biobanque de Picardie

Abstract

Biobanks play a crucial role in medical research and particularly in modern genetic studies, providing end users with well-characterized and professionally annotated collections of biospecimens. Scientific communities rely heavily on the biobank quality standards, as high-quality input material contributes significantly to overall success of downstream applications.

Having evolved into specialized infrastructures, biobanks not only tend to facilitate collaborations between medical, biological, and research partners, but also ensure streamlined biomaterial processing from sample collection to distribution. They aim to control external and internal sample processing operations and identify factors that can potentially impact sample quality. Care must be taken at biospecimen acquisition, short- and long-term storage, handling of material during analysis, and when transporting samples to other collaborating or requesting laboratories. It is important for biobanks to adopt a reliable and reproducible quality metric to individually qualify each sample.

The French Chronic Kidney Disease - Renal Epidemiology and Information Network (CKD-REIN) project is a prospective cohort study that was established to identify the determinants, biomarkers, and practice patterns associated with chronic kidney disease outcomes. The CKD-REIN study included collection and processing of biomaterials, especially to enable their future use in genome-wide association studies (GWAS). Whole blood samples of the cohort study were processed into buffy coats followed by genomic DNA (gDNA) extraction and quality control (QC) by the Biobanque de Picardie.

Authors

Yves-Edouard Herpe¹, Céline Lange², Anne-Sophie Bruyer¹, David Berthelot¹, Anne Boland³, Bertrand Fin³, Stéphanie Guénin⁴, Gaelle Mongeard⁴, Dorothée Cannet⁵, Bénédicte Stengel⁶, Eva Graf⁷, and Vera Rykalina⁷

¹ Biobanque de Picardie
² Agence de la Biomédecine
³ Centre National de Recherche en Génomique Humaine
⁴ Centre de Ressources Régional en Biologie Moléculaire
⁵ Etablissement Français du Sang Bourgogne Franche-Comté
⁶ INSERM U1018
⁷ Agilent Technologies
Introduction

The CKD-REIN project is an ongoing study, aiming for better understanding and management of CKD progression and outcomes. The CKD-REIN is a clinical-based prospective cohort that has enrolled 3,033 adult patients with CKD Stage 3–4 receiving nephrologist-led care in France since 2013. The data and research samples have been available for scientific collaborations since 2016. The initial results of the prospective observational cohort study have already been summarized by numerous French institutions and discussed in detail in several comprehensive reviews

The Biobanque de Picardie is a certified biological resource center (ISO 9001 and NFS 96900) accountable for centralization of all CKD-REIN samples. The main tasks of the Biobanque de Picardie for the project are long-term storage of biological specimens, such as plasma, buffy coats, serum, and urine, extraction of nucleic acids from buffy coats, and provision of high-quality gDNA samples to a requesting laboratory for genome-wide association studies (GWAS).

Assessment of gDNA integrity based on the electrophoretic separation in agarose slab gels has typically been the primary technique utilized by most research biobanks. However, this type of DNA qualification depends largely on the evaluator’s experience and can be especially tedious for large-scale projects such as cohort studies in terms of time and labor. The 4200 TapeStation system together with its quality metric for gDNA, the DIN, offer a fully automated solution for an objective assessment of sample integrity. The TapeStation instrument supports sample loading from a 96-well plate and minimizes user intervention by fully automated sample analysis. The DIN quality metric can be applied to develop thresholds that meet quality standards required for specific applications. The objective QC results using DIN can be easily communicated between biobanks and their end partners that proceed with downstream sample analysis.

In relation to the research goals and large scope of the CKD-REIN study, the Biobanque de Picardie developed its own strategy for conservation of biomaterials and assessment of their quality.

The blood specimens were processed into buffy coats and stored long term for up to 5 years. Quality of gDNA extracted from buffy coats was assessed on the 4200 TapeStation system with the Genomic DNA ScreenTape assay over the entire storage duration. Various conditions that can impact gDNA quality, such as a delay in blood processing and consistency in buffy coat isolation by subordinated biological resources centers, were also investigated using the objective DIN quality metric. In addition, data on reproducibility of DIN measurement between the biobank and their end user enriched the study.

This application note demonstrates how an objective and widely recognized DIN quality metric can be implemented at different levels into many biobank quality control operations. It also highlights the importance of quality assurance for biobanks so that they can gain desired confidence and peace of mind regarding the quality of their samples. The DIN quality metric can ensure biobanks that their operations reach or even exceed standards of the research partners with which they collaborate.
Experimental

Project logistics

The Chronic Kidney Disease - Renal Epidemiology and Information Network (CKD-REIN) large-scale sample processing protocol was made possible through cooperation between several key partners (Figure 1). In brief, whole blood was collected, refrigerated, and shipped to the project partners by nephrology hospital departments or clinics across France. Treatment of blood samples, buffy coat generation, and short-term storage were carried out by the Centres de Ressources Biologiques (CRB) (13 in total), affiliated with university hospitals or the Etablissement français du Sang (EFS). The long-term storage of buffy coats, extraction of gDNA, and quality control analysis were performed by the Biobanque de Picardie. Prior to QC analysis, gDNA samples were diluted and normalized by the Centre de Ressources Régionales en Biologie Moléculaire.

The Centre National de Recherche en Génomique Humaine (CNRGH, CEA) was in charge of generating the genotyping data required for genome-wide association studies (GWAS). All pre-analytical data of the samples were recorded using a secure, centralized, and dedicated professional biobank database, Databiotec, with a GFI software editor.

Figure 1. Schematic representation of project logistics for the CKD-REIN study.
Sample preparation

Whole blood from each CKD-REIN study participant was collected in four Vacutainer K$_2$EDTA tubes (BD Biosciences, p/n 367873) resulting in a total volume of 24 mL. The tubes were centrifuged at 2,000 x g for 10 minutes at room temperature (18 - 25 °C) to separate buffy coats from peripheral blood. The buffy coat content of two Vacutainer tubes, including plasma and red blood cell fractions, was collected (500 µL in total) and transferred to a Matrix 2D Barcoded ScrewTop Storage Tube (Thermo Scientific, p/n 3743) for long-term storage. For safety purposes regarding the risks associated with possible extraction failure, a second backup storage tube was prepared. To optimize the sample processing and traceability, the Barcoded Latch Racks (Thermo Scientific, p/n 4900), meeting the Society for Biomolecular Screening (SBS) microplate dimensional standard, were utilized. The buffy coat tubes were stored long term in –80 °C mechanical freezers, equipped with temperature sensors and a redundant alarm system supported by a 24/7 on-call technician. All frozen cryovials were transferred to the Biobanque de Picardie for long-term storage followed by subsequent centralized extraction. The gDNA was extracted from buffy coats using the QIAamp DNA Blood Midi kit (Qiagen, p/n 51185) with the QIAvac Connecting System (Qiagen, p/n 19419) and eluted in 200 µL of AE buffer. 96 samples were extracted each day for 1.5 months (2,689 samples in total).

DNA analysis

Initial quantification of gDNA samples was performed on the Infinite M1000 multimode flagship microplate reader (Tecan, p/n 1009007090) with a Quant-iT PicoGreen dsDNA reagent (Thermo Fisher Scientific, p/n P7581) (520 nm). The samples with a concentration above 300 ng/µL were diluted to meet the DIN functional range using the automated liquid handling and robotics workstation Freedom EVO 150 base unit (Tecan, p/n 1007006958) complemented with the EVOware software. The 4200 TapeStation system (p/n G2991AA) and the Genomic DNA ScreenTape (p/n 5067-5365) with Genomic DNA reagents (p/n 5067-5366) from Agilent Technologies were used for analysis of gDNA integrity. gDNA sample preparation and electrophoretic analysis were performed according to the manufacturer’s instructions.4

The number of samples used for DNA analysis (n = 2,636) differs from the number of the study participants (n = 3,033) for the following reasons: (i) some study participants were not sampled; (ii) some samples did not yield sufficient biomaterial for extraction; or (iii) some samples did not yield sufficient extracted DNA for subsequent qualitative evaluation.

DNA reanalysis

For GWAS, the gDNA samples were transferred to the Centre National de Recherche en Génomique Humaine (CNRGH, CEA), where they were again quantified and qualified. Quantification was performed for all samples in duplicate using the Molecular Device Spectramax plate reader with the Quant-iT Broad Range dsDNA Assay Kit (Thermo Fisher Scientific, p/n Q331320). In contrast, qualification of the samples was carried out at random for approximately 15% of the entire collection. Integrity of gDNA for the selected samples was also assessed on the 4200 TapeStation system with Genomic DNA ScreenTape assay. To exclude any error during sample handling, approximately 20% of the samples were subjected to additional in-house tests by PCR to verify the sex of the study participant.
Results and discussion

DNA quality metric

Determining the quality of gDNA input material is an essential step for both biobanks and research communities, as downstream analysis can be time consuming, labor intensive, and expensive. Identifying poor-quality samples prior to committing them to subsequent analysis not only conserves precious time and resources, but also helps to maintain the collective data generated by the study. Accurate assessment of gDNA stability, throughout individual stages of a biospecimen processing workflow, with a standardized quality metric is a powerful tool to reveal potential factors affecting sample quality. This strategy can also be applied to preselect samples that cannot meet study thresholds. A reliable quality metric can ensure that high-quality samples proceed to downstream applications and the data generated by those samples will not jeopardize the cumulative experimental efforts.

The 4200 TapeStation system together with the Genomic DNA ScreenTape assay provide fast and high-throughput electrophoretic separation of gDNA. The quality metric of the Genomic DNA ScreenTape assay, the DIN, allows for objective and reliable assessment of gDNA quality. The DIN score is calculated on a scale of 1 to 10 and automatically generated by the Agilent TapeStation analysis software.

The DIN quality metric was applied to the CKD-REIN samples as an optimal QC tool to evaluate the level of gDNA integrity. The CKD-REIN samples reported high DIN values, confirming high quality of gDNA (Table 1). Sample qualification using DIN enabled development of a quality threshold, which helped to communicate and reproduce the results between the project partners. A random set of six samples analyzed with the gDNA ScreenTape assay is shown in Figure 2. The gDNA presented as a single band on the gel image view (Figure 2A) and a clear smear on the electropherogram overlay (Figure 2B). A report function of the TapeStation analysis software was used to document the results. Easily developed reports allowed fast and convenient data transfer between the CKD-REIN partners.

Table 1. Statistical summary of sample quality evaluation using DIN quality metric.

<table>
<thead>
<tr>
<th>Number of samples</th>
<th>DIN</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,636</td>
<td>8.3 ± 0.4</td>
<td>8.4</td>
<td>6.4</td>
<td>9.7</td>
</tr>
</tbody>
</table>

Figure 2. An example of six gDNA samples extracted from buffy coats and analyzed on the Agilent 4200 TapeStation system with the Genomic DNA ScreenTape assay. The samples can be visualized as (A) gel image and (B) electropherogram. The automatically determined DIN is displayed under the individual gel lane. A high DIN indicates highly intact gDNA, and a low DIN a strongly degraded gDNA.
QC of gDNA extracted from long-term stored buffy coats

Extraction of gDNA directly from blood specimens is a common practice in molecular biology laboratories, as large quantities can be obtained. Another popular approach is to purify DNA from buffy coats, a packed fraction of white blood cells. In contrast with whole blood samples, buffy coats samples have a relatively small volume but can still provide high yields of DNA sufficient for modern methods of genetic analysis. The latter approach can be more cost and space efficient and is highly recommended for many genomic and epidemiological studies, if DNA cannot be extracted within a few days of collection.

The CKD-REIN blood samples were processed to yield buffy coats, which were stored frozen at –80 °C for up to 5 years. This type of biospecimen storage was a sensible alternative for the large CKD-REIN project, as gDNA could not be extracted within recommended time upon sample collection due to logistics and economics reasons. All buffy coat samples were centralized by the Biobanque de Picardie, resulting in seamless gDNA extraction in a specialized laboratory by professional technicians. It was important for the Biobanque de Picardie to investigate whether gDNA extracted from buffy coats independent of storage duration is of high quality and therefore suitable for GWAS.

The CKD-REIN gDNA samples were analyzed on the 4200 TapeStation system with the Genomic DNA ScreenTape assay. The DIN provided by the TapeStation analysis software was applied as an objective quality metric to observe a correlation between duration of buffy coat storage and gDNA quality. The DIN values were plotted against the time between buffy coat freezing and gDNA extraction (Figure 3). Most DIN data points were tightly packed between 8 and 9, confirming high quality of gDNA extracted from buffy coats. A DIN score of 7 is routinely used by the CNRGH for high-throughput sequencing projects requiring high-quality DNA. As a point of comparison, this score was referenced by the Biobanque de Picardie. A relatively small number of stored samples were observed with DIN scores below the threshold.

The samples with DIN number below the threshold were preliminary identified as those with a potentially higher risk of low call rates in subsequent genotyping analysis. However, these samples with a DIN between 6 and 7 were not excluded from the CKD-REIN collection and genotyping project.

Distribution of DIN values at shorter storage periods did not differ from the DIN spread after the prolonged storage time. A linear regression model was applied to the data to investigate the relationship between buffy coat storage and gDNA quality. It was concluded that there was no major correlation between gDNA integrity and storage time of the buffy coats. Even after 5 years of storage at –80 °C, extracted gDNA remained intact.

Figure 3. Integrity of gDNA extracted from buffy coats stored at –80 °C for up to 5 years (n = 2,636). The blue horizontal line represents the DIN threshold used by the Biobanque de Picardie for gDNA qualification.
Influence of delay in buffy coat generation on gDNA quality

Whole blood sample collection, storage, and processing conditions can have a significant impact on the immediate sample quality. These conditions may also negatively affect the quality of the nucleic acids extracted from blood samples. In the framework of the CKD-REIN project, the blood specimens collected at the clinical sites were transported to the regional CRBs. This resulted in variable periods of refrigeration time until further sample processing. Although the biobanking standard protocol threshold of 6 hours was met for the majority of the blood specimens (69%), some samples could only be processed into buffy coats after 12 or even after 24 hours. It was essential for the Biobanque de Picardie to qualify the samples that had a delay in treatment and compare them with those fully satisfying the protocol for blood processing. The relationship between DIN values and the time between blood collection and buffy coat freezing is shown in Figure 4. The DIN value of 7 was used by the Biobanque de Picardie as a preliminary QC threshold for sufficient gDNA integrity. Only two samples processed into buffy coats after 6 hours had DIN values less than 7. The mean, median, and standard deviation for DIN depending on time between blood collection and buffy coat freezing remained the same (Table 2). The samples processed into buffy coats after 12 and even 24 hours retained a high level of integrity (DIN > 7). By comparing DIN values, it was verified that a delay in blood processing of up to 32 hours did not influence gDNA quality.

![Figure 4](image)

**Figure 4.** Integrity of gDNA in relation to time after which the buffy coats were isolated and stored at −80 °C (n = 2,636). Blue vertical and horizontal lines represent the thresholds used by the Biobankque de Picardie for blood processing and gDNA qualification, respectively.

**Table 2.** Statistical summary of sample quality evaluation using DIN quality metric depending on the delay in blood processing.

<table>
<thead>
<tr>
<th>Delay in blood processing (hours)</th>
<th>Number of samples (%)</th>
<th>DIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>≤ 6</td>
<td>1820 (69%)</td>
<td>8.3 ± 0.4</td>
</tr>
<tr>
<td>6 to 12</td>
<td>742 (28%)</td>
<td>8.3 ± 0.4</td>
</tr>
<tr>
<td>≥ 12</td>
<td>74 (3%)</td>
<td>8.3 ± 0.4</td>
</tr>
</tbody>
</table>
Comparison of DIN values between CRB partners

In the CKD-REIN cohort study, buffy coats were generated by the 13 different partners of the *Biobanque de Picardie*, each representing an individual region in France. The DIN metric was implemented for a comparative analysis to assess possible variation in gDNA integrity between the *Centres de Ressources Biologiques* (CRB). A boxplot method was applied to compare both the distribution of DIN depending on the region and the performance of each center in relation to the buffy coat generation protocol (Figure 5). The DIN medians of all CRBs were similar, ranging between 8 and 9 and confirming high sample quality. A threshold of 7 was established for the comparative study. The DIN minimum values for all CRBs were higher or equal to 7. Analysis of gDNA collection revealed that several CRBs had a few samples (10 in total) with DIN values less than 7 (indicated as dots in Figure 5). Based on the small number of such samples, it was concluded that their lower DIN values were associated with the samples themselves, rather than with buffy coat isolation or other sample processing conditions. The TapeStation system with DIN metric enabled the *Biobanque de Picardie* to accurately assess all samples and preselect those requiring ancillary investigation. The results of the boxplot method demonstrated that there was no significant difference between the distributions of DIN values across the CRBs. High quality of the samples verified that biospecimen handling and transportation were consistent in all 13 centers and did not influence gDNA integrity.

![Figure 5](image)

**Figure 5.** Comparative analysis of DIN values between the 13 CRBs. The boxplot method demonstrated that there was no significant difference between the distributions of DIN values across the CRBs.
Consistency of DIN between laboratories

A random subset of 394 CKD-REIN samples (about 15% of the total collection) was reanalyzed by the CNRGH for gDNA integrity assessment on the 4200 TapeStation system with the Genomic DNA ScreenTape assay. This additional QC step was necessary to verify the fulfillment of the substantial quality requirements in relation to initial sample handling, storage organization, and subsequent transfer. A scatterplot displaying a correlation between the DIN values obtained by the Biobanque de Picardie and those generated by the CNRGH is shown in Figure 6. A total of 32 (8.1%) of the reanalyzed samples delivered DIN values identical to those reported by the Biobanque de Picardie, whereas 94 (24%) samples had a difference in DIN values between 0 and 0.2. For 197 (50%) samples, the DIN difference ranged between 0.2 and 0.6. Similarly, 61 (15.5%) samples gave DIN differences between 0.6 and 1.0. Overall, both laboratories independently obtained very comparable results, with differences in DIN exceeding 1 only for 10 (2.5%) samples.

Previously, Agilent showed that analysis of gDNA integrity with the 4200 TapeStation system and Genomic DNA ScreenTape assay is highly reproducible and user independent. Thus, it can be assumed that variation in DIN between the two laboratories can be explained either by additional sample handling prior to the integrity re-assessment or other factors to which the analysis is sensitive. Instrument-to-instrument, run-to-run, or reagent lot-to-lot deviations can all, to some extent, contribute to overall analysis precision. For instance, an instrument-to-instrument variation in DIN was assessed earlier by Agilent using three different 4200 TapeStation systems. The study conducted on 18 technical replicates resulted in standard deviation of 0.2, indicating that the instrument has minimal contribution to DIN variation.

In the case of two different laboratories, slightly higher variability in DIN measurements is expected, considering the possible contribution of other analysis and sample conditions. An average difference in DIN values calculated for paired data points was equal to 0.2 with a standard deviation of 0.4. The low average of DIN differences indicated that shipping and handling of samples between the laboratories was performed successfully in a way that did not affect gDNA integrity.

Figure 6. Comparison of DIN values between the Biobanque de Picardie and CNRGH.
Conclusion
Assessing the quality of DNA extracted from biospecimens is high priority, not only for biobanks but also for scientific communities that rely on sample supply from biobanks. Whole blood collection, processing, logistics, and storage may all impact the integrity of gDNA, necessitating a reliable quality metric to monitor sample quality. The DNA integrity number (DIN) objective quality metric was used by the Biobanque de Picardie to demonstrate that long-term storage of buffy coats was a reliable approach for the studies, when gDNA extraction could not be performed within a few days after blood collection. It was also noted that a delay in blood processing can be extended from 6 to at least 32 hours without impacting gDNA quality, thus providing more flexibility for future projects. Furthermore, a similar distribution of DIN was shown for the samples collected from the different French regions. None of the local CRBs provided samples that resulted in gDNA degradation. In addition, comparison of the DIN numbers obtained by the Biobanque de Picardie and their research partner, the CNRGH, revealed a very good laboratory-to-laboratory correlation. The DIN reanalysis confirmed high reproducibility of the results and stable shipment conditions.

The Agilent 4200 TapeStation system together with the Genomic DNA ScreenTape assay provided the Biobanque de Picardie with a fully automated solution for reliable evaluation of gDNA integrity, offering sufficient throughput for the large CKD-REIN cohort study. It was demonstrated that the objective DIN quality metric can be successfully implemented at different levels to various biobank quality control operations. Quality control based on DIN number enabled definition of application-specific thresholds and helped the Biobanque de Picardie achieve peace of mind regarding the quality of each individual sample. All CKD-REIN samples were verified as high quality and further processed for genome-wide association studies.

References
Acknowledgements

The authors would like to thank the teams of all the biological resources centers that participated in the CKD-REIN project:

Biobanque de Picardie, CRB du Centre Hospitalier Universitaire Amiens Picardie, 1 Rond-Point du Pr Christian Cabrol, 80054 Amiens Cedex 1 (BRIF number: BB-0033-00017)

NeuroBioTec, CRB des Hospices Civils de Lyon Groupement Hospitalier Est, Hôpital Neurologique, 59 Boulevard Pinel, 69677 Bron Cedex (BRIF number: BB-0033-00046)

Centre de ressources biologiques du Centre Hospitalier Universitaire de Nantes Hôtel Dieu, Institut de biologie, 9, quai Moncousu, 44093 Nantes Cedex 1 (BRIF number: BB-0033-00040)

Centre de ressources biologiques du Centre Hospitalier Universitaire Grenoble Alpes, Boulevard de la Chantoune, CS 10217, 38700 La Tronche (BRIF number: BB-0033-00069)

Centre de ressources biologiques du Centre Hospitalier Régional Universitaire de Nancy, Hôpitaux de Brabois, Bâtiment Recherche Rue du Morvan, 54500 Vandoeuvre-les-Nancy (BRIF number: BB-0033-00035)

Servicer de Néphrologie, Centre Hospitalier de Perpignan, 20 Avenue du Languedoc, 66046 Perpignan Cedex 9

Plateforme de Ressources Biologiques, Hôpital Henri Mondor, 51 avenue du Maréchal de Lattre de Tassigny, 94000 Créteil (BRIF number: BB-0033-00021)

CIC-1435, Centre d’Investigation Clinique Plurithématique, Centre Hospitalier Universitaire de Limoges, 2 Avenue Martin Luther King, 87042 Limoges Cedex

Plateforme de Ressources Biologiques de l’Hôpital européen Georges-Pompidou, 20-40 rue Leblanc, 75015 Paris (BRIF number: BB-0033-00063)

Etablissement Français du sang Hauts de France – Normandie, Site de Bois-Guillaume 609, chemin de la Bretèche, 76235 Bois-Guillaume

Etablissement Français du sang Nouvelle Aquitaine, site Pellegrin, Place Amélie Raba Léon, CS 21010, 33075 Bordeaux Cedex

Etablissement Français du sang Hauts de France – Normandie, Site de Loos-Eurasanté, Avenue Pierre Mauroy, Parc Eurasante Epi-de Soil, 59120 Loos

Etablissement Français du sang Île de France, Site Avicenne, Hopital Avicenne porte 8, 125 route de Stalingrad, 93009 Bobigny

Etablissement Français du sang Occitanie, Site de Toulouse, 75 rue de Lisieux, 31300 Toulouse

Etablissement Français du sang Grand-Est, Site de Colmar, 6 rue du Hohnack, 68025 Colmar Cedex

Etablissement Français du sang Grand-Est, Site de Metz, 6 rue des Dames de Metz, 57000 Metz

Etablissement Français du sang PACA-Corse, Site de Marseille, 149, boulevard Baille, 13392 Marseille Cedex 05