High Interspecimen Variability in Nucleic Acid Extraction Efficiency Necessitates the Use of Spike-In Control for Accurate qPCR-based Measurement of Plasma Cell-Free DNA Levels

Grant C. O'Connell, PhD,[1,2]* Paul D. Chantler, PhD,[3,4] Taura L. Barr, RN, PhD[5]

ABSTRACT

Objective: To assess the interspecimen variability associated with plasma DNA extraction in order to provide insight regarding the necessity to use an exogenous spike-in control when measuring cell-free DNA (cfDNA) levels using quantitative polymerase chain reaction (qPCR).

Methods: Plasma specimens were obtained from 8 healthy individuals, 20 patients with cardiovascular disease risk factors, and 54 patients diagnosed with acute stroke. Specimens were spiked with an exogenous oligonucleotide fragment, and total DNA was extracted via automated solid phase anion exchange. We determined recovery of the exogenous fragment via qPCR and used this information to calculate DNA extraction efficiency.

Results: Plasma DNA extraction efficiencies varied dramatically between specimens, ranging from 22.9% to 88.1%, with a coefficient of variance of 28.9%. No significant differences in DNA extraction efficiencies were observed between patient populations.

Conclusions: We strongly recommend the use of an exogenous spike-in control to account for variance in plasma DNA extraction efficiency when assessing cell free DNA (cfDNA) levels by qPCR in future biomarker investigations.

Keywords: normalization, reference, CNAPS, standard, polymerase chain reaction, housekeeping, serum, liquid biopsy, methods, blood

Circulating extracellular DNA, or cell free DNA (cfDNA), is increasingly being cited in the literature as a noninvasive marker of tissue damage in a wide range of disease states and acute traumas.[1] For example, circulating cfDNA levels are elevated in response to traumatic brain injury and stroke, and are predictive of disability and mortality.[2-7] Although biomarker research surrounding cfDNA has exploded during the past decade, there has been little interinvestigation consistency regarding the quantitative polymerase chain reaction (qPCR) methods used for cfDNA quantification.[8-10] In particular, there is a great deal of inconsistency regarding the use of an exogenous oligonucleotide spike-in control to account for interspecimen variability in DNA extraction efficiencies.

Most commonly, when cfDNA levels are quantified via qPCR, total DNA is isolated from a set volume of plasma or serum and eluted in a fixed volume of buffer. qPCR targeting 1 or more genetic loci is then performed using a fixed volume of eluent as input, and the detection of these loci is used as an indication of cfDNA levels.[11] This workflow assumes that DNA extraction efficiencies are similar between specimens. To control for potential confounds introduced by interspecimen variability in DNA extraction efficiency, some investigators have spiked specimens with an exogenous oligonucleotide fragment before DNA isolation and used

Abbreviations

cfDNA, cell free DNA; qPCR, quantitative polymerase chain reaction; CVD, cardiovascular disease; IRBs, institutional review boards; EDTA, ethylenediaminetetraacetic acid; GFP605, 605-base pair green fluorescence protein DNA fragment; PCR, polymerase chain reaction; CT, crossing threshold; ANOVA, analysis of variance; GFP, green fluorescent protein; CV, coefficient of variation

[1]Center for Basic and Translational Stroke Research, Robert C. Byrd Health Sciences Center, West Virginia University, Morgantown, West Virginia, [2]Department of Pharmaceutical Sciences, School of Pharmacy, West Virginia University, Morgantown, West Virginia, [3]Center for Cardiovascular and Respiratory Sciences, Robert C. Byrd Health Sciences Center, West Virginia University, Morgantown, West Virginia, [4]Division of Exercise Physiology, School of Medicine, West Virginia University, Morgantown, West Virginia, [5]Valtari Bio Incorporated, Morgantown, West Virginia

*To whom correspondence should be addressed.
goconnell.wvu@gmail.com

downstream detection of this fragment to normalize cfDNA levels.[7] Similar spike-in strategies are routinely used when quantifying circulating miRNAs;[12] however, they have yet to be widely adopted for quantification of circulating cfDNA.

One reason for this lack of consistency regarding the use of an exogenous spike-in control may be attributed to the fact that, to our knowledge, the interspecimen variability associated with plasma DNA extraction has never been explicitly investigated. In this study, we assessed the variability associated with plasma DNA extraction in order to provide insight regarding whether it is necessary to use a spike-in control when quantifying cfDNA levels using qPCR.

Materials and Methods

Experimental Design

Plasma specimens were obtained from a clinical population consisting of 8 healthy individuals, 20 neurologically normal patients with cardiovascular disease (CVD) risk factors, and 54 patients recently diagnosed with acute stroke (**Figure 1A**). We spiked plasma specimens with a fixed copy number of an exogenous DNA fragment and extracted total plasma DNA via fully automated solid phase anion exchange. The recovery of the exogenous DNA fragment was then assessed via qPCR and used to calculate DNA extraction efficiency. We then assessed the variably in extraction efficiency across all specimens.

Patients

All subjects were recruited at Ruby Memorial Hospital, Morgantown, West Virginia. In the case of patients diagnosed with stroke, cerebrovascular events were of mixed etiology, and diagnosis was confirmed by neuroradiological imaging according to the established criteria for diagnosis of acute ischemic cerebrovascular syndrome;[13] blood specimens were drawn within 24 hours of symptom onset and before the administration of thrombolytics. *Patients with CVD* consisted of individuals who were clinically determined to have positive results for at least 1 major CVD risk factor, such as diabetes, hypertension, or dyslipidemia. *Healthy individuals* were free from injury and disease, as determined by a trained health care professional. All procedures were approved by the institutional review boards (IRBs) of West

Virginia University and Ruby Memorial Hospital (IRB protocol # 1410450461R001). Written informed consent was obtained from all subjects or their authorized representatives before study procedures were performed.

Plasma Isolation

Peripheral venous blood was collected into a K_2 ethylenediaminetetraacetic acid (EDTA) vacutainer (Becton, Dickenson and Company). EDTA-treated blood was immediately spun at 2000 *g* for 10 minutes to separate hemocytes. The resultant plasma was collected and spun at 10,000 *g* for 10 minutes to remove any residual blood cells or debris. Specimens were stored at −80°C until analysis.

Exogenous Spike-In Oligonucleotide Fragment

A nonhuman DNA fragment originating from the green fluorescence protein (GFP)-encoding portion of the *pontellina plumata* genome was generated for use as an exogenous oligonucleotide spike-in, as described previously.[7] This 605-base pair GFP DNA fragment (GFP_{605}) was amplified from purified pGFP-V-RS plasmid template (OriGene Technologies Inc) via polymerase chain reaction (PCR) using sequence-specific primers (forward: 5′ GTTGCTGTGATCCTCCTCCA, reverse: 5′ CCGCCATGGAGATCGAGTG; **Figure 1B**). We electrophoresed the resultant GFP_{605} PCR product via agarose gel and purified it using the QIAquick gel extraction kit (QIAGEN). The mass-unit concentration of purified GFP_{605} was determined via spectrophotometry (NanoDrop; Thermo Fisher Scientific, Inc) and subsequently converted to molar concentration based on the estimated molecular weight of the fragment, which was derived from its sequence using the equation described by Stothard.[14] Plasma specimens were spiked with purified GFP_{605} at a final concentration of 10^5 copies per mL before DNA extraction.

Automated DNA Extraction

Total DNA was extracted from 200 µL of spiked plasma using the QIAamp DNA micro kit (QIAGEN) and automated using the QIAcube system (QIAGEN; **Figure 1C**). Purified DNA was eluted in a 35-µL volume of ultrapure water. We performed all extractions in duplicate. Blank DNA extractions were performed in parallel with those of clinical specimens to monitor potentially confounding sources of environmental DNA contamination. The robotics system passed all self-diagnostic checks before the procedures,
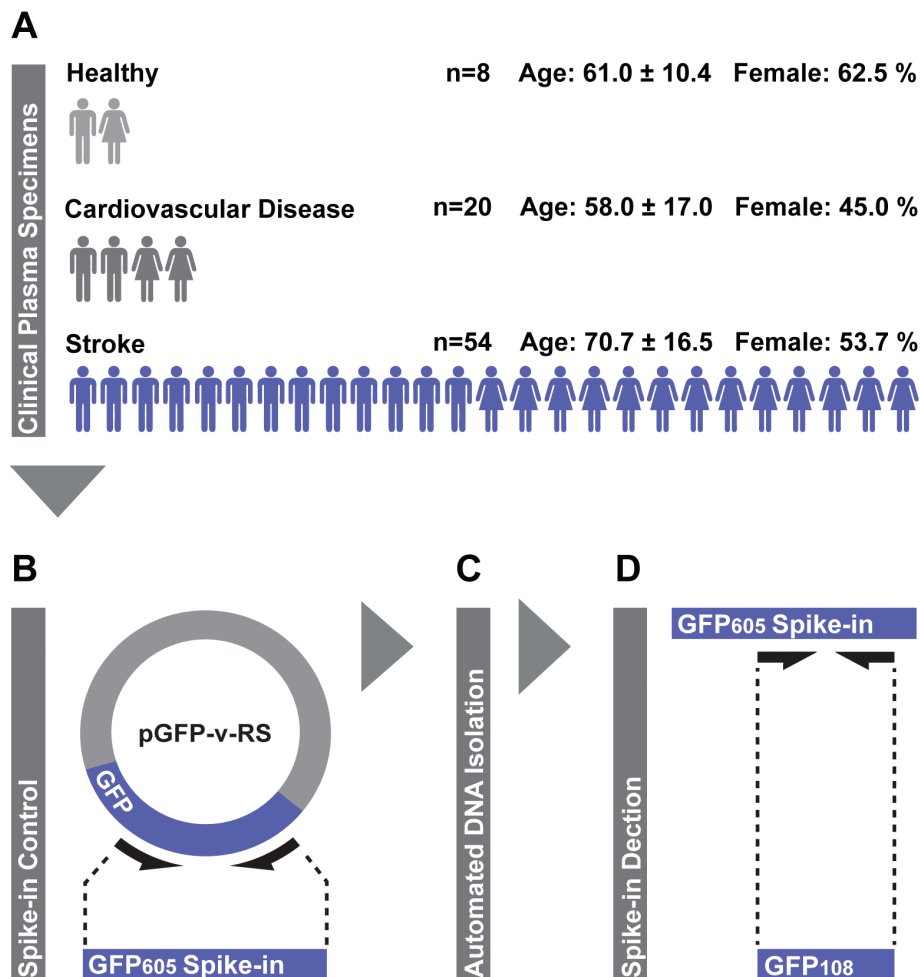
Figure 1

Experimental workflow. **A**, Plasma specimens were collected from a clinical population including healthy individuals, patients with risk factors for cardiovascular disease, and patients diagnosed with acute stroke. Sample sizes and basic demographic information are indicated; each male or female figure represents 4 individuals of that sex (numbers rounded nearest value divisible by 4) and error is presented as standard deviation. **B**, Plasma specimens were spiked with equimolar amounts of a nonhuman 605-base pair green fluorescence protein DNA fragment (GFP_{605}) generated from the green fluorescence protein (GFP)-encoding loci of the pGFP-v-RS plasmid. **C**, Total DNA was extracted from spiked plasma specimens via fully automated solid phase anion exchange. **D**, Spike-in recovery was assessed via quantitative polymerase chain reaction targeting a 108-base pair internal fragment (GFP_{108}) and used to calculate DNA extraction efficiencies.

and all calibration was up to date at the time of experiments, according to the manufacturer-recommended service interval.

Spike-In Detection via qPCR

We performed qPCR using the RotorGeneQ (QIAGEN) thermocycling system, and reaction set-up was automated using the Qiagility (QIAGEN) liquid-handling platform. GFP_{605} recovered in plasma eluent was detected via amplification of a 108 base pair internal fragment using sequence-specific primers (forward: 5′ CTCGTACTTCTCGATGCGGG, reverse: 5′ GGCTACGGCTTCTACCACTT; **Figure 1D**). PCR was performed using 5 µL of eluent as input, and products were detected by SYBR green (PowerSYBR, Thermo Fisher
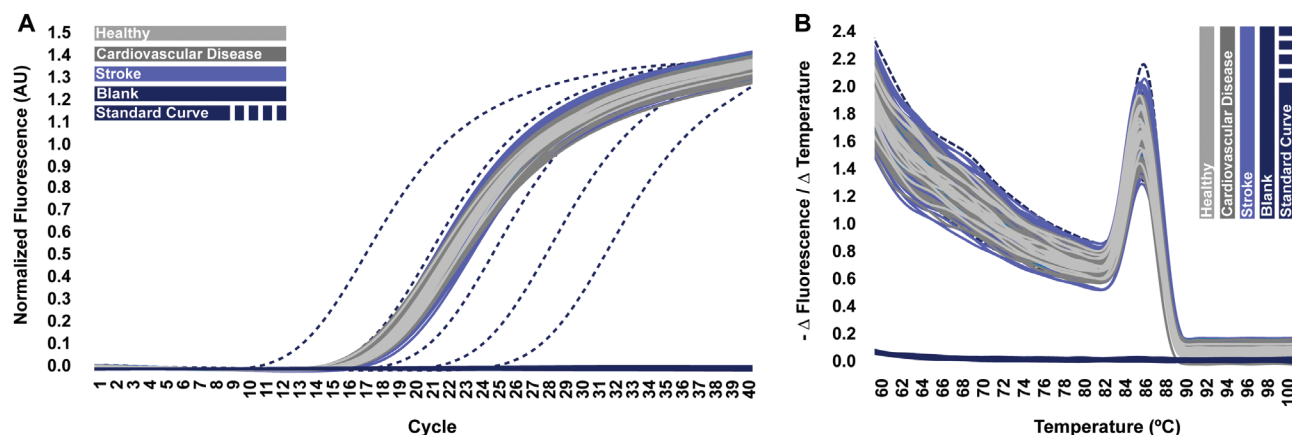
Figure 2

Quantitative polymerase chain reaction (qPCR) data. **A**, Raw amplification plots generated during qPCR targeting the 605-base pair green fluorescence protein DNA fragment (GFP$_{605}$) spike-in in eluents yielded from plasma DNA extraction, overlaid with those of the standard curve. AU indicates arbitrary units. **B**, Melting plots generated during subsequent melting point analysis. Individual lines indicate averaged polymerase chain reaction (PCR) technical replicates associated with each DNA extraction.

Scientific Inc). Raw amplification plots were background corrected, and crossing threshold (CT) values were generated via the Rotor-Gene Q software package. Absolute copy numbers of GFP$_{605}$ in eluents were calculated using a standard curve generated via 10-fold serial dilution of purified GFP$_{605}$ template ultrapure in water. All reactions were performed in triplicate, and the presence of a single PCR product was confirmed with melting-curve analysis.

Calculation of DNA Extraction Efficiency

We calculated DNA extraction efficiency based on GFP$_{605}$ recovery by comparing the absolute copy number of GFP$_{605}$ detected in qPCR to that which was theoretically expected assuming 100% DNA extraction efficiency:

$$DNA\ extraction\ efficiency\ (\%) = 100 * \frac{GFP_{605}\ copies\ detected}{GFP_{605}\ copies\ expected}$$

Statistics

Statistical analysis was performed using SPSS software, version 21.0 (SPSS Inc). We used 1-way analysis of variance (ANOVA) for intergroup comparison of means. The Bartlett test was used for intergroup comparison of variance. Pearson's *r* was used to assess the strength of correlational relationships. We established the level of significance at *P* <.05 for all statistical testing.

Results

All specimens exhibited similar amplification efficiencies in qPCR, as evidenced by homogeneous slopes in the linear phase of amplification plots (**Figure 2A**); this indicated that any interspecimen differences in GFP$_{605}$ detection occurred as a result of differences in DNA extraction efficiency and were not driven by differences in PCR amplification efficiency. Further, CT values for all patient specimens fell within those of the standards (**Figure 2A**), which indicated that the resultant standard curve (**Figure 3**) was suitable for absolute quantification of GFP$_{605}$. No PCR product was detected in reactions containing eluents from blank DNA extractions (**Figure 2A**), and all clinical specimens produced PCR products with a single melting point (**Figure 2B**), indicating that our findings were not influenced by exogenous DNA contamination or nonspecific amplification events.

While levels of recovered GFP$_{605}$ were relatively consistent between DNA extraction technical replicates, they were extremely variable between specimens. CT values from qPCR targeting GFP$_{605}$ ranged across 2 full thermocycles, suggesting a high degree of interspecimen variability in DNA extraction efficiency (**Figure 4A**). Specially, extraction efficiencies calculated based on GFP$_{605}$ recovery ranged from 22.9% to 88.1% across all specimens, with a coefficient of
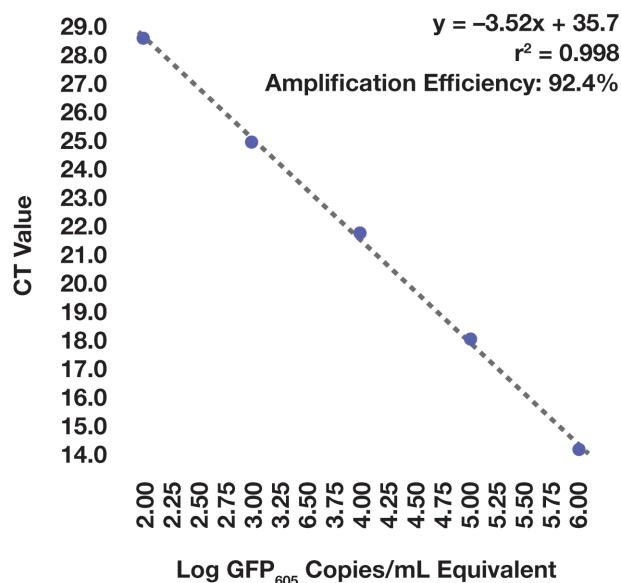
Figure 3

Standard curve generation. Standard curve and resultant linear equation used for absolute quantification of the 605-base pair green fluorescence protein DNA fragment (GFP$_{605}$) spike-in. The correlation coefficient was calculated using Pearson's *r*.

variance of 28.9% (**Figure 4B**). No statistically significant differences in DNA extraction efficiency were observed between groups. Specimens obtained from patients with CVD and stroke appeared to exhibit greater levels of variability in extraction efficiencies than those obtained from healthy individuals; however, this difference was not statistically significant.

Discussion

The purpose of this study was to assess the variability associated with plasma DNA extraction, in order to provide insight regarding the necessity of using an exogenous spike-in control when quantifying cfDNA levels using qPCR. We observed substantial interspecimen variability in DNA extraction efficiency; therefore, we recommend the use of an exogenous spike-in control to account for such variance when assessing cfDNA levels by qPCR in future biomarker investigations.

The levels of interspecimen variability in DNA extraction efficiency that we observed in this investigation would have the potential to introduce significant artificial variance in terms of downstream cfDNA quantification. Depending on the study population, the differences in extraction efficiencies that we observed across specimens would have the potential to manifest in several-fold increases in variance with regards to detected cfDNA levels. An artificial increase in variance of this magnitude would be adequate to substantially reduce the power of intra- and intergroup statistical testing, requiring increases in sample size to observe statistically significant effects. This finding is most concerning with regard to investigations examining conditions such as ischemic stroke in which only modest increases in cfDNA levels are observed in pathologic testing.[2,7] Thus, we strongly recommend the use of an exogenous spike-in control to account for variability introduced during DNA extraction. We believe that doing so has the potential to dramatically improve data fidelity, especially in applications in which highly accurate cfDNA quantification is required.

The fact that we observed dramatically higher levels of variability in extraction efficiencies between specimens than between technical replicates suggests that variability in DNA extraction efficiency is driven primarily by differences in specimen composition, rather than by technical inconsistencies. DNA binding in solid-phase extraction is sensitive to numerous considerations, including specimen ion concentrations, pH, protein levels, and lipid content.[15-18] These factors are inevitably variable in human blood, and such interindividual differences may underlie the high levels of interspecimen variability that we observed in plasma DNA extraction efficiencies. Future work that aims to identify the plasma properties that most substantially influence DNA extraction efficiencies could yield information that would allow for the development of improved, less variable extraction methods.

One limitation of this particular study was the fact that we only explored the variability associated with 1 particular plasma DNA extraction method. Solid phase anion exchange is the most commonly implemented method of plasma DNA extraction, and the kit we used has been implemented in numerous previous investigations;[19-23] nonetheless, a wide array of other methods and commercial products are available to researchers. Thus, we believe it is important to recognize that the extreme variability we observed in this investigation may not be generalizable in terms of other methods or other kits. However, our findings
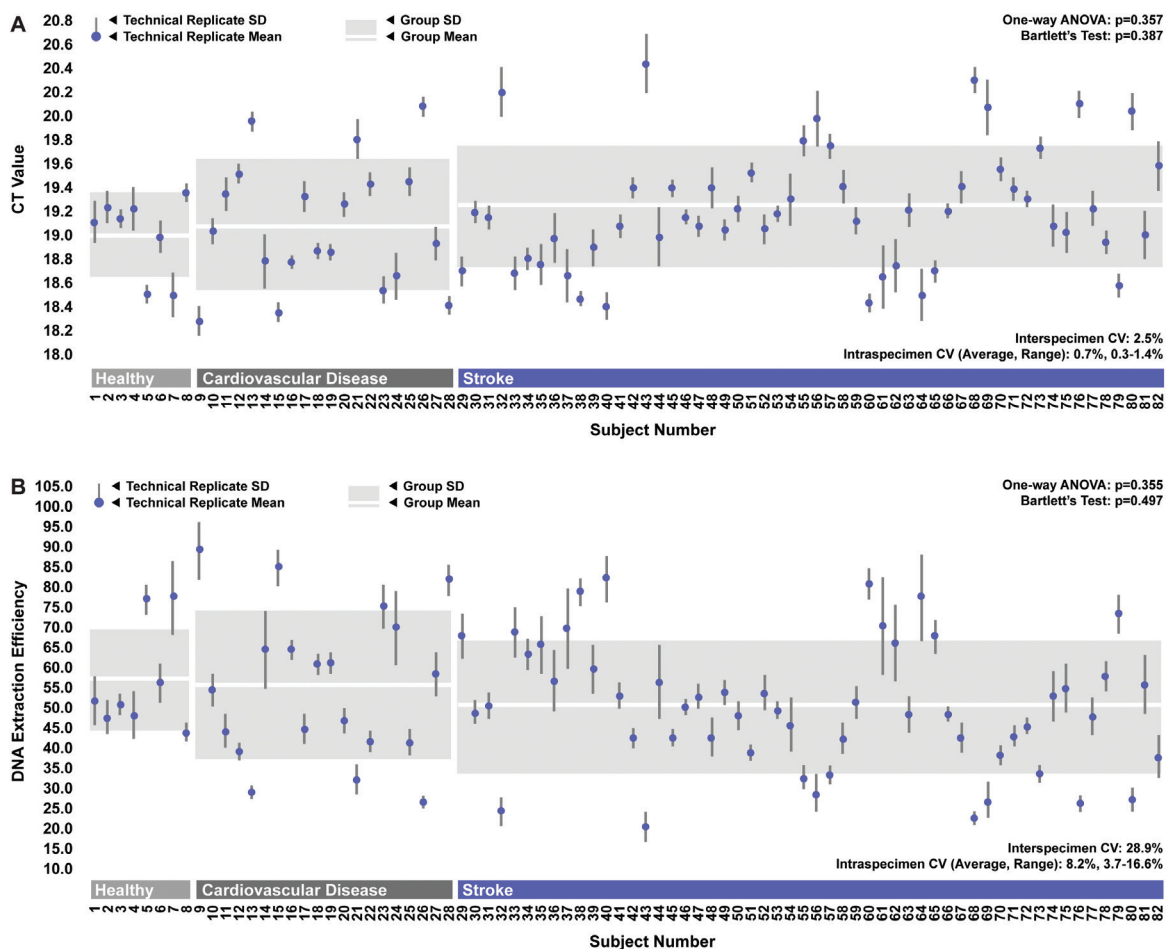
Figure 4

DNA extraction efficiency. **A,** Crossing threshold (CT) values generated during quantitative polymerase chain reaction (qPCR) targeting the 605-base pair green fluorescence protein DNA fragment (GFP$_{605}$) spike-in in DNA extraction eluents across all specimens for each group. ANOVA indicates analysis of variance; CV, coefficient of variation. **B,** Corresponding DNA extraction efficiencies. Points and error bars depict averages, and SDs associated with DNA extraction technical replicates for each specimen. Boxes depict averages and SDs across biological replicates for each group.

strongly imply that the variability associated with these other methods and products should be vetted via preliminary experiments to avoid potential confounds in situations in which omission of an exogenous spike-in control is being considered.

We believe it is important to note that the use of an exogenous spike-in control is a means of limiting artificial variance induced during DNA extraction specifically and does not account for other potential sources of artificial variance that may arise during other parts of the larger overall workflow.

Most notably, DNA degradation and leukocyte enucleation during blood collection, processing, and storage before DNA isolation can also introduce significant variability that is not accounted for via the use of a spike-in control.[24,25] Thus, it is important to take steps to limit these potential confounds, such the use of blood preservatives designed specifically for cfDNA analysis and the institution of consistent specimen handling protocols for clinical blood collection.[26,27] The use of an exogenous spike in-control is only 1 of several measures that should be used to ensure high-fidelity final data.

Collectively, our findings suggest that interspecimen differences in plasma DNA extraction efficiency represent a significant potential confound in cfDNA quantification. Thus, we strongly recommend the use of an exogenous spike-in control in future biomarker studies that investigate circulating cfDNA levels via qPCR.**LM**

Acknowlegments

Personal and Financial Conflicts of Interest

G.C.O. and T.L.B. have a patent pending regarding markers of stroke and stroke severity. T.L.B. serves as chief scientific officer for Valtari Bio Incorporated. Work by G.C.O. is part of a pending licensing agreement with Valtari Bio Incorporated. The remaining authors report no potential conflicts of interest.

References

1. Swarup V, Rajeswari MR. Circulating (cell-free) nucleic acids—A promising, non-invasive tool for early detection of several human diseases. *FEBS Lett*. 2007;581(5):795–799.
2. Tsai N-W, Lin T-K, Chen S-D, et al. The value of serial plasma nuclear and mitochondrial DNA levels in patients with acute ischemic stroke. *Clin Chim Acta*. 2011;412(5–6):476–479.
3. Macher H, Egea-Guerrero JJ, Revuelto-Rey J, et al. Role of early cell-free DNA levels decrease as a predictive marker of fatal outcome after severe traumatic brain injury. *Clin Chim Acta*. 2012;414:12–17.
4. Campello Yurgel V, Ikuta N, Brondani da Rocha A, et al. Role of plasma DNA as a predictive marker of fatal outcome following severe head injury in males. *J Neurotrauma*. 2007;24(7):1172–1181.
5. Rainer TH, Wong LKS, Lam W, et al. Prognostic use of circulating plasma nucleic acid concentrations in patients with acute stroke. *Clin Chem*. 2003;49(4):562–569.
6. Lam NYL, Rainer TH, Wong LKS, et al. Plasma DNA as a prognostic marker for stroke patients with negative neuroimaging within the first 24 h of symptom onset. *Resuscitation*. 2006;68(1):71–78.
7. O'Connell GC, Petrone AB, Tennant CS, et al. Circulating extracellular DNA levels are acutely elevated in ischaemic stroke and associated with innate immune system activation. *Brain Inj*. 2017; http://dx.doi.org/10.1080/02699052.2017.1312018.
8. Devonshire AS, Whale AS, Gutteridge A, et al. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Anal Bioanal Chem*. 2014;406(26):6499–6512.
9. Tsang JCH, Lo YMD. Circulating nucleic acids in plasma/serum. *Pathology*. 2007;39(2):197–207.
10. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer—A survey. *Biochim Biophys Acta*. 2007;1775(1):181–232.
11. Gahan PB, ed. *Circulating Nucleic Acids in Early Diagnosis, Prognosis and Treatment Monitoring*. Vol 5. Dordrecht: Springer Netherlands; 2015.
12. Kroh EM, Parkin RK, Mitchell PS, et al. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). *Methods*. 2010;50(4):298–301.
13. Kidwell CS, Warach S. Acute ischemic cerebrovascular syndrome: diagnostic criteria. *Stroke*. 2003;34(12):2995–2998.
14. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*. 2000;28(6):1102–1104.
15. Guo W, Jiang L, Bhasin S, et al. DNA extraction procedures meaningfully influence qPCR-based mtDNA copy number determination. *Mitochondrion*. 2009;9(4):261–265.
16. Boom R, Sol CJ, Salimans MM, et al. Rapid and simple method for purification of nucleic acids. *J Clin Microbiol*. 1990;28(3):495–503.
17. Aplenc R, Orudjev E, Swoyer J, et al. Differential bone marrow aspirate DNA yields from commercial extraction kits. *Leukemia*. 2002;16(9):1865–1866.
18. Thompson JD, Cuddy KK, Haines DS, et al. Extraction of cellular DNA from crude cell lysate with glass. *Nucleic Acids Res*. 1990;18(4):1074.
19. Legendre C, Gooden GC, Johnson K, et al. Whole-genome bisulfite sequencing of cell-free DNA identifies signature associated with metastatic breast cancer. *Clin Epigenetics*. 2015;7(1):100.
20. Kim K, Shin DG, Park MK, et al. Circulating cell-free DNA as a promising biomarker in patients with gastric cancer: diagnostic validity and significant reduction of cfDNA after surgical resection. *Ann Surg Treat Res*. 2014;86(3):136–142.
21. Chicard M, Boyault S, Daage LC, et al. Genomic copy number profiling using circulating free tumor DNA highlights heterogeneity in neuroblastoma. *Clin Cancer Res*. 2016;22(22):5564–5573.
22. Kuang Y, Rogers A, Yeap BY, et al. Noninvasive detection of EGFR T790M in gefitinib or erlotinib resistant non–small cell lung cancer. *Clin Cancer Res*. 2009;15(8):2630–2636.
23. Gorges TM, Schiller J, Schmitz A, et al. Cancer therapy monitoring in xenografts by quantitative analysis of circulating tumor DNA. *Biomarkers*. 2012;17(6):498–506.
24. Sozzi G, Roz L, Conte D, et al. Effects of prolonged storage of whole plasma or isolated plasma DNA on the results of circulating DNA quantification assays. *J Natl Cancer Inst*. 2005;97(24):1848–1850.
25. Jung M, Klotzek S, Lewandowski M, et al. Changes in concentration of DNA in serum and plasma during storage of blood samples. *Clin Chem*. 2003;49(6):1028–1029.
26. Wong D, Moturi S, Angkachatchai V, et al. Optimizing blood collection, transport and storage conditions for cell free DNA increases access to prenatal testing. *Clin Biochem*. 2013;46(12):1099–1104.
27. Toro PV, Erlanger B, Beaver JA, et al. Comparison of cell stabilizing blood collection tubes for circulating plasma tumor DNA. *Clin Biochem*. 2015;48(15):993–998.

HOLOGIC®

First and Only FDA Cleared
Digital Cytology System

Genius™ Cervical AI

Genius™ Review Station

Genius™ Digital Imager

Empower Your Genius With Ours

Make a Greater Impact on Cervical Cancer

with the Advanced Technology of the
Genius™ Digital Diagnostics System

Click or Scan
to discover more

_genius™
DIGITAL DIAGNOSTICS