Liquid biopsy quality control – the importance of plasma quality, sample preparation, and library input for next generation sequencing analysis

Jennifer Whiting, Charmaine San Jose Hinahon, Hannah Saunders, Natalie Supunpong Hernandez, Angie Cheng

Thermo Fisher Scientific, 2130 Woodward St, Austin, TX 78744

ABSTRACT

Liquid biopsy is emerging as a non-invasive companion to traditional solid tumor biopsies. As next generation sequencing (NGS) of circulating cell-free nucleic acids (cfNA = cfDNA and cfRNA) becomes common, it's important to understand the impact of sample preparation on quality, specificity, and sensitivity of liquid biopsy tests. Plasma samples are often limited, and may have undesirable characteristics such as lipemia or hemolysis that contribute unwanted genomic DNA (gDNA) to the sample. Low cfDNA concentration can also limit the amount available for NGS library prep. In this study, we explore the effects of suboptimal plasma and low library input on liquid biopsy NGS, and discuss various techniques for in-process quality control of cfNA samples isolated from plasma.

MATERIALS AND METHODS

In the first half of the study, bulk K2EDTA plasma was collected from four healthy donors, including one normal control and two hemolyzed samples (100 mg/dL). Hemolysis was generated by freeze-thaw. All samples were spun to cell-free and replicate cfNA isolations were performed on 4 mL plasma from each donor using the MagMAX™ Cell-free Total Nucleic Acid Isolation Kit. Yield and quality of isolated cfNA were assessed via Qubit™ High Sensitivity DNA Assay, Agilent 2100 Bioanalyzer™ with a High Sensitivity DNA Kit, and Agilent 4200 Tapestation™ with the High Sensitivity D5000 ScreenTape Assay. Matched NGS libraries were prepared from these samples using 5, 10 and 20 ng nucleic acid input, with the Ion AmpliSeq™ Cancer Hotspot Panel v2. Libraries were sequenced with the Ion Chef™ & Ion GeneStudio S5™ System to compare NGS metrics including read length, molecular coverage, and limit of detection (LOD). Analysis was performed with Ion Reporter™ v5.6 software using the Oncomine™ TagSeq Lung v2 Tumor workflow.

A follow-up study was then conducted to more closely replicate hemolysis generated in true blood collection conditions, using mechanical agitation rather than freeze-thaw. From three healthy donors, whole blood was drawn into a set of BD Vacutainer® K2EDTA tubes. For the normal condition, tubes were spun down to plasma immediately. For the moderate hemolysis condition, tubes were vortexed until hemolysis reached 20-50 mg/dL, and for the high hemolysis condition tubes were vortexted until hemolysis reached 100-200 mg/dL. Plasma hemolysis levels were characterized by the sample vendor, but were also visualized in-house by absorbance using the Nanodrop™ according to a previously published method.[1]

The cell-free nucleic acids were then isolated from the matched samples as described in the above method, using 4 mL plasma per isolation and automated on the Kingfisher™ Flex Purification System. Because the normal plasma samples yielded <20 ng, sequencing libraries were prepared using maximum volume input for all samples (10.4 µl), and subsequently mass normalized when pooling for NGS.

RESULTS

Figure 1 – Total yield from normal and hemolyzed plasma

Qubit Yield
Replicate Average
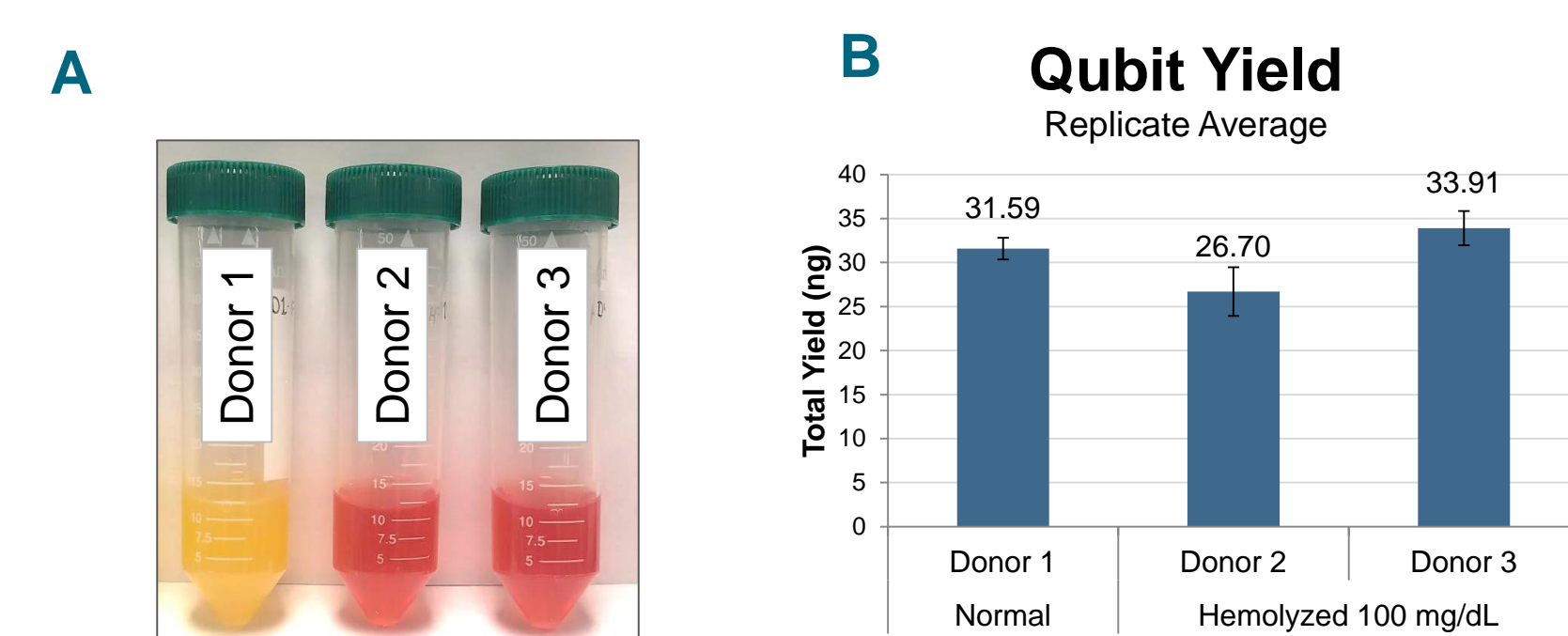
A) Bulk plasma samples were collected from 3 separate donors. Donor 1 had no hemolysis, and is considered "Normal." Donors 2 and 3 had 100 mg/dL hemolysis generated by freeze-thaw. Photo of representative aliquots from each donor. B) Total cfNA yield from replicate 4mL isolations measured by Qubit High-Sensitivity DNA Assay. For each Donor, six replicate isolations were performed to demonstrate assay reproducibility. Error bars = SD.

Figure 2 – Nucleic acid size profiles from normal and hemolyzed plasma
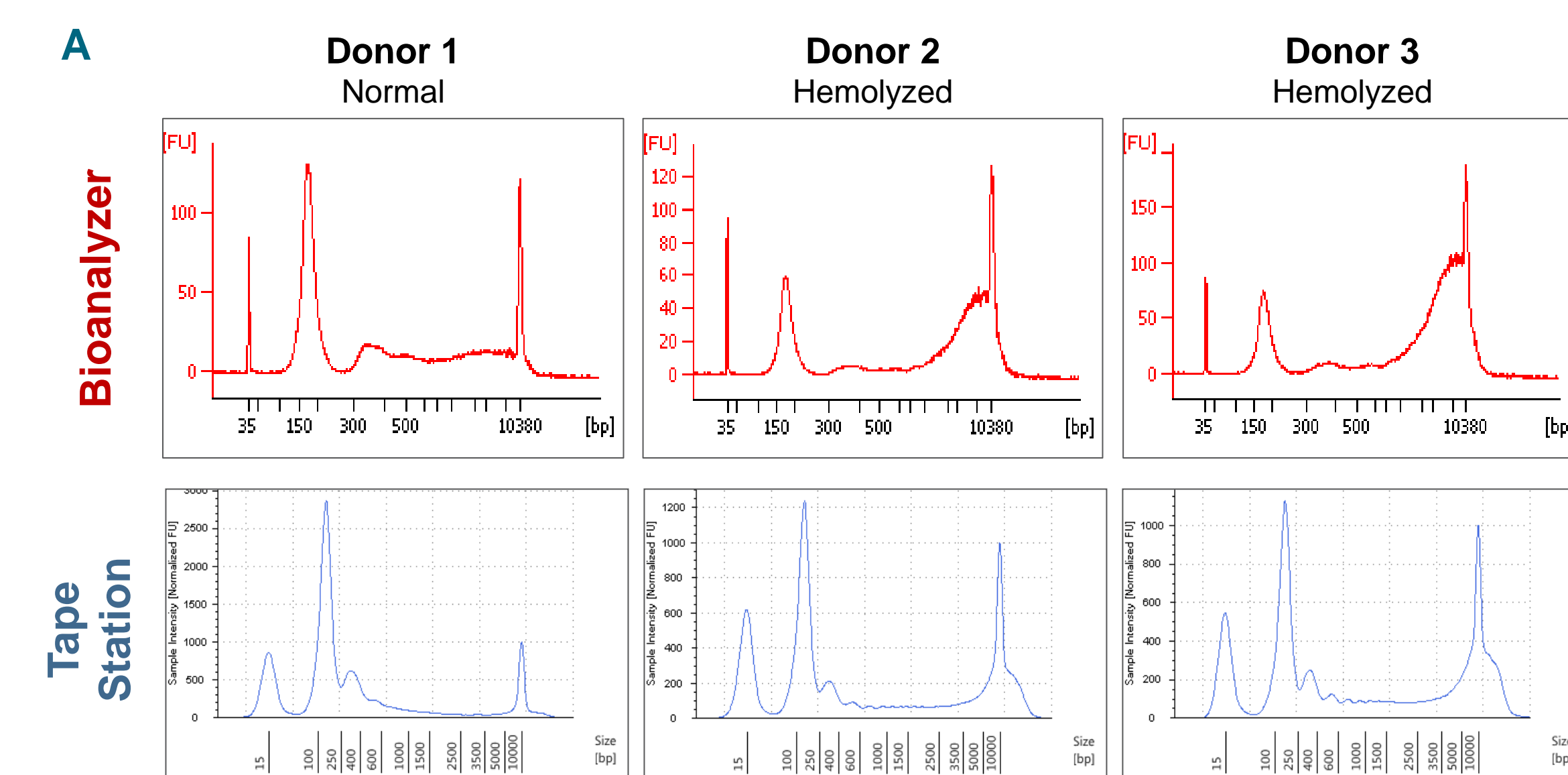
A) Nucleic acid size profiles from each isolation were analyzed on both the Agilent 2100 Bioanalyzer using the High Sensitivity DNA Kit (top row, red lines), and the Agilent 4200 Tapestation using the High Sensitivity D5000 ScreenTape Assay (bottom row, blue lines). Comparison of results revealed a difference in mean cfDNA peak size, which was 174 bp with the Bioanalyzer, and 148 bp with the Tapestation. The Bioanalyzer peak size more closely matches the expected size of mono-nucleosomal cfDNA length. Additional differences between the platforms can be seen for high molecular weight DNA, which is more apparent on Bioanalyzer traces. Detection of genomic DNA (gDNA) in the sample is an important quality control metric, as gDNA can interfere with library amplification. B) Smear analysis quantification of cfDNA concentration (50-275 bp) demonstrates equivalent results between platforms. Although total Qubit yields were similar for all three donors, the hemolyzed donors yield significantly less desirable cfDNA and more contaminating gDNA.
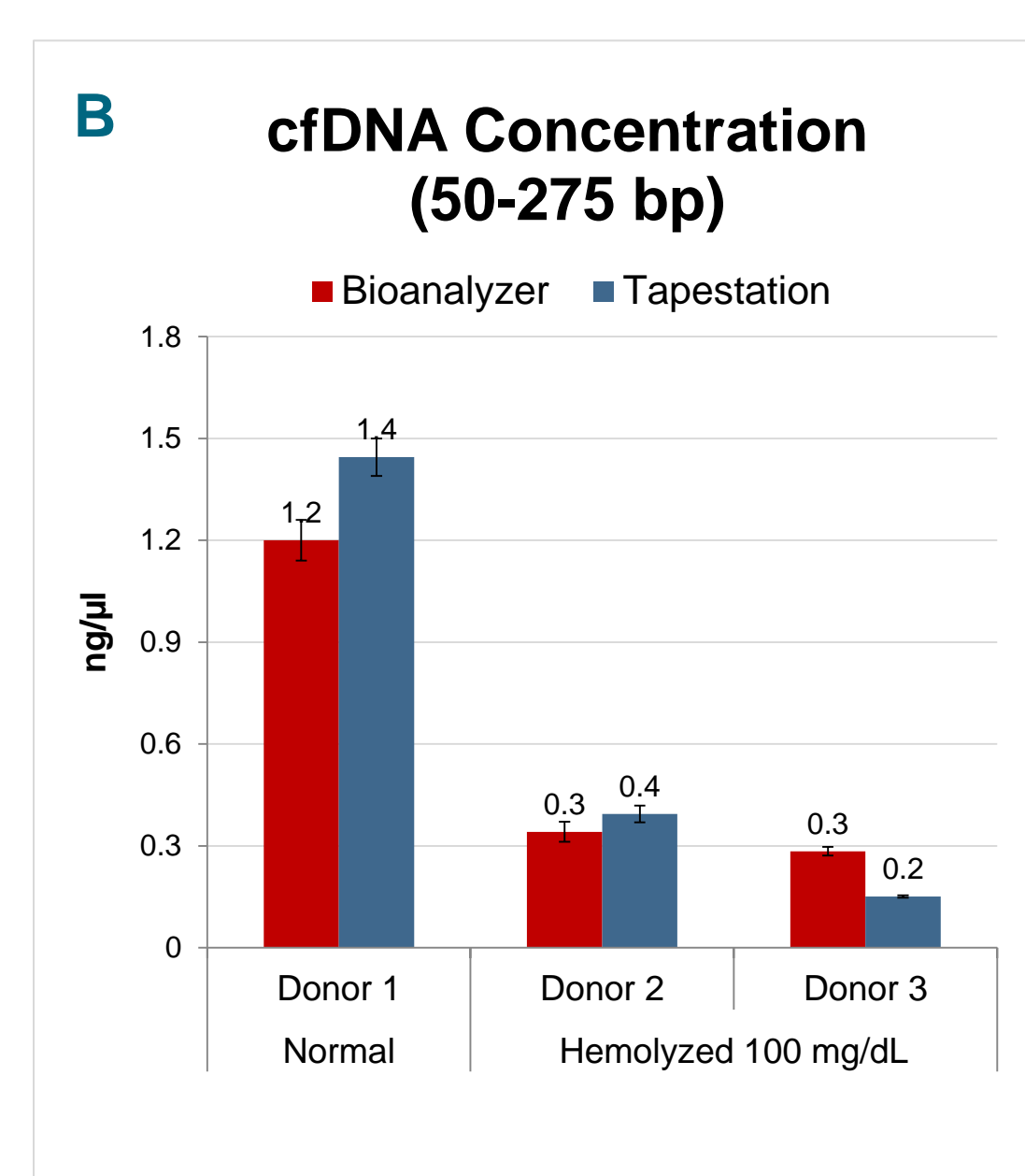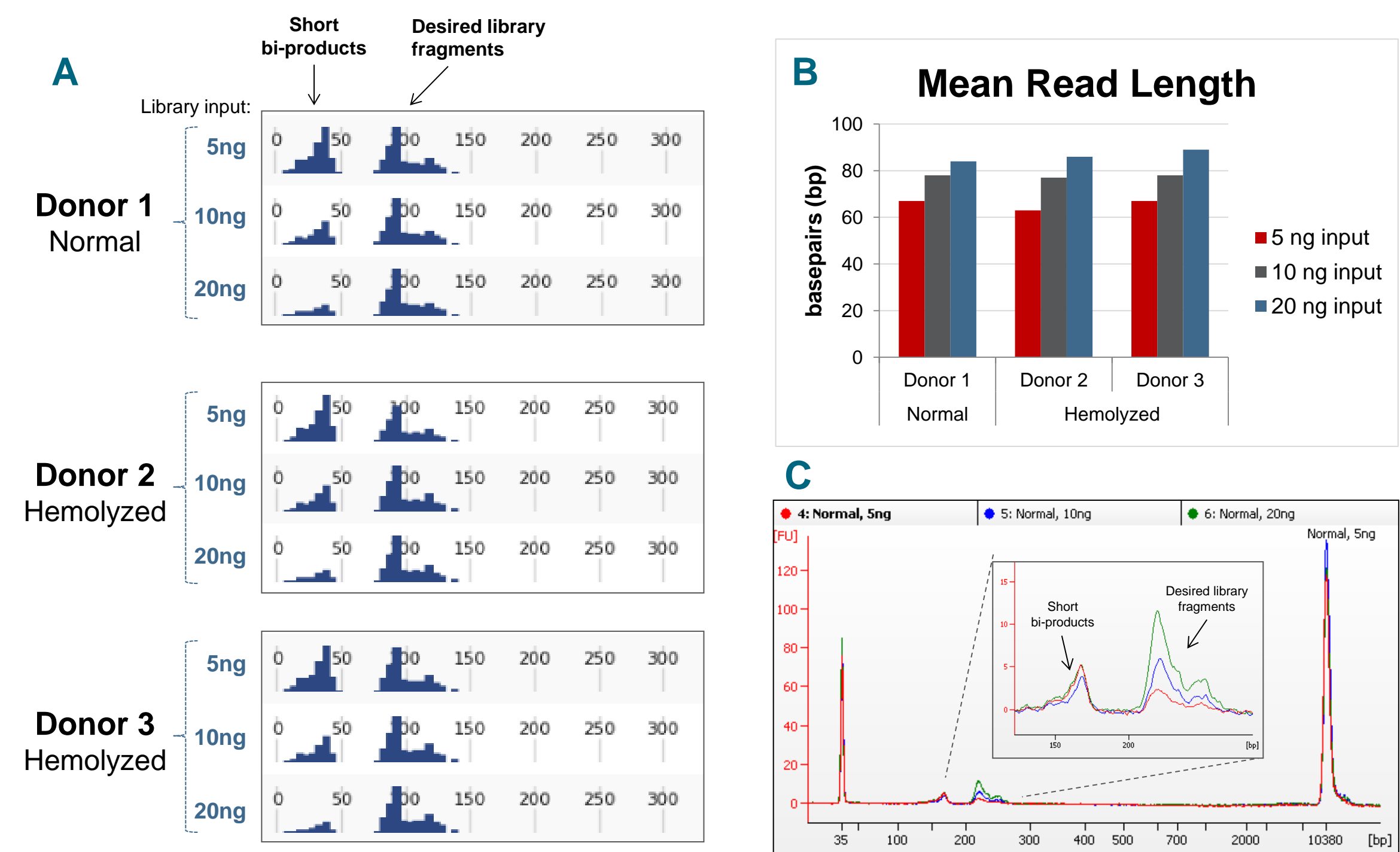
Figure 3 – Effect of decreasing sample input on library size and read length


NGS libraries were prepared using either 5, 10, or 20 ng nucleic acid input for each donor, to observe effects of input mass on sequencing results. A) Histograms representing read lengths for each library. The presence of small bi-products (from 0-50 bp) increased with decreasing library input. B) No difference was observed in mean read length between normal and hemolyzed plasma samples. However, mean read length was approximately 25% less when libraries prepared with 5 ng nucleic acid input compared to 20 ng input, due to the increase in small bi-products. C) Prepared libraries run on the Agilent Bioanalyzer 2100 as an in-process quality control check. Small bi-products and main amplicon peaks are visible. Representative overlay traces from Donor 1 (normal plasma) for libraries generated with 5 ng (red), 10 ng (blue), and 20 ng input (green).

### Figure 4 – Effect of library input on NGS metrics
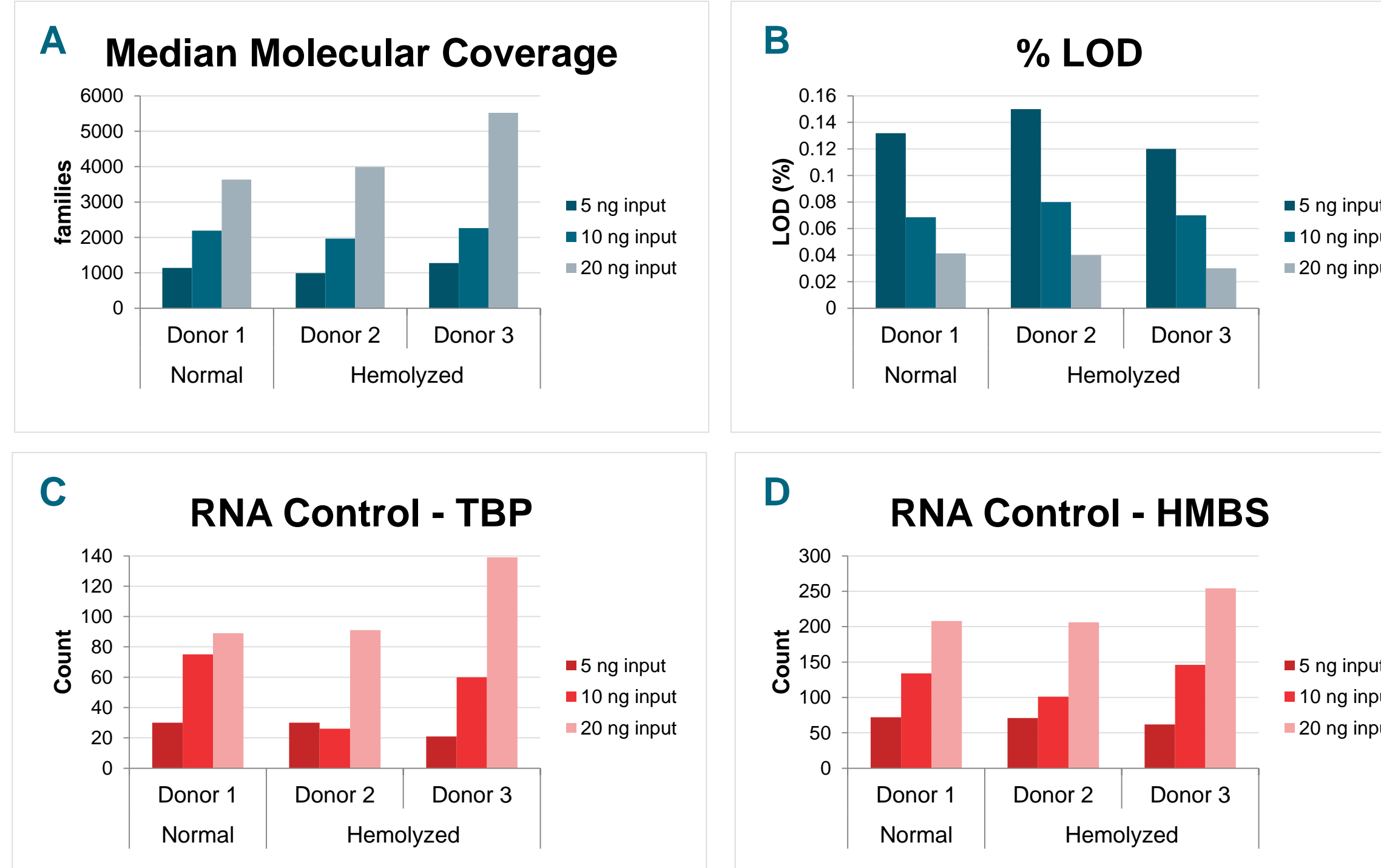
A) Median molecular coverage also decreased with decreasing library input mass. B) Low molecular coverage results in higher limits of detection (% LOD), illustrating the effect of low nucleic acid input on assay sensitivity. No differences were detected between the normal and hemolyzed plasma samples in the % LOD reported by the Ion Reporter software. Read counts for cfRNA controls TBS and HMBS (C and D) increased with increasing sample input, but again no significant difference was observed between normal and hemolyzed plasma donors.

Figure 5 – Analysis of plasma hemolyzed in K2EDTA blood tubes
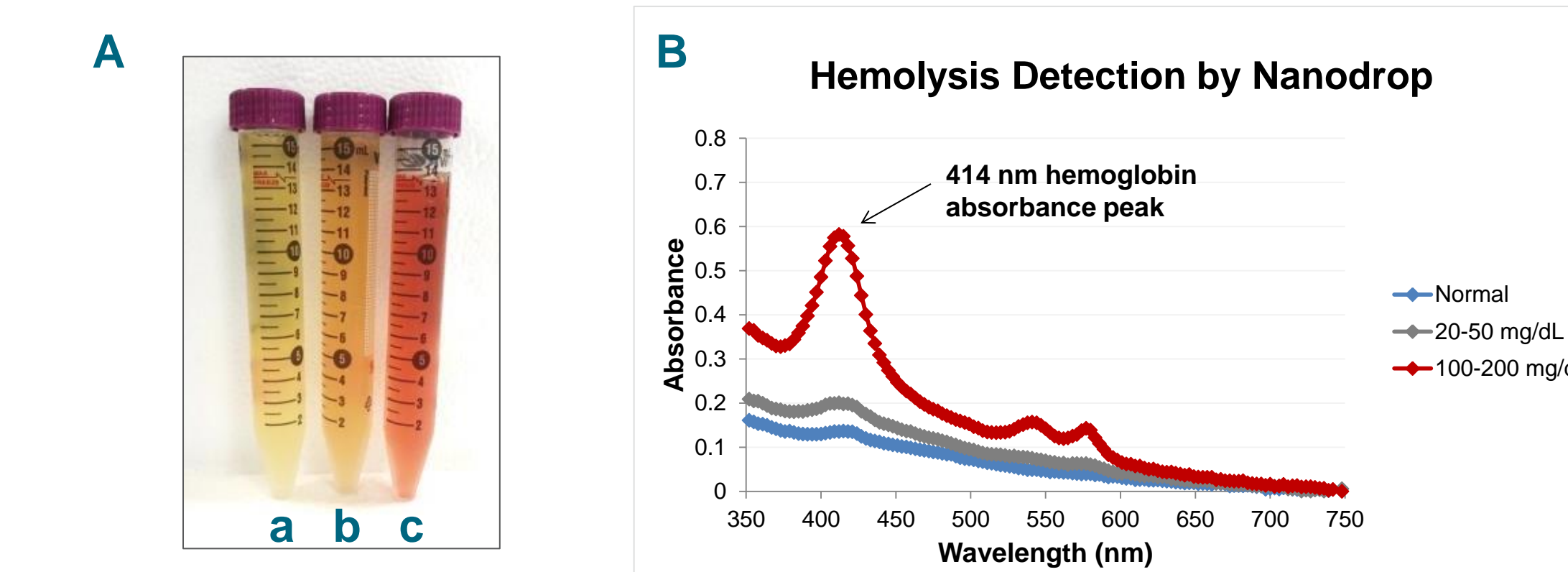

The effects of hemolysis were further explored on plasma derived from whole blood drawn into blood collection tubes and hemolyzed via mechanical agitation, rather than bulk plasma collection with freeze-thaw. A) Blood from each donor was drawn into K2-EDTA tubes and either processed to plasma immediately (a, normal), or vortexed to induce moderate hemolysis (b, 20-50 mg/dL) or high hemolysis (c, 100-200 mg/dL). B) Detection of hemolysis by hemoglobin absorbance at 414 nm (Nanodrop™, UV-Vis)[1]. Representative traces shown for Donor 1, with a visible peak at 414 nm in the highly hemolyzed sample.

Figure 6 – Yield and cfDNA size profiles from normal, moderate, and high hemolysis plasma samples
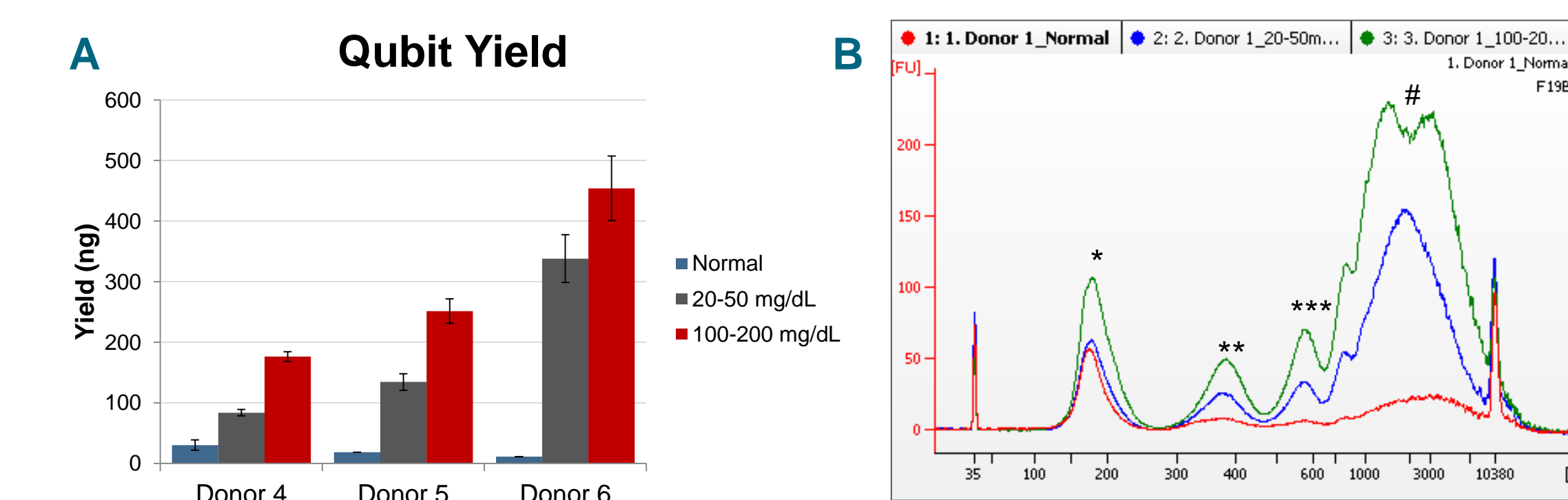
A) From the matched plasma samples, cfDNA was isolated and quantified using Qubit. Hemolyzed samples had significantly higher yields than matched normal samples for all donors. It was observed that bead collection time was slower in hemolyzed plasma. B) Representative Bioanalyzer trace overlay for Donor 1. Genomic DNA contamination is low in the normal plasma (red line), but increases significantly with moderate (blue line) and high hemolysis (green line). Size analysis is a valuable quality control test, as gDNA can interfere with downstream NGS library preparation. (*) Primary cfDNA peak, (**), di-nucleosomal cfDNA, (***) tri-nucleosomal cfDNA, (#) high molecular weight gDNA contamination.

Figure 7 – Effect of hemolysis on MAPD Uniformity and CNVs
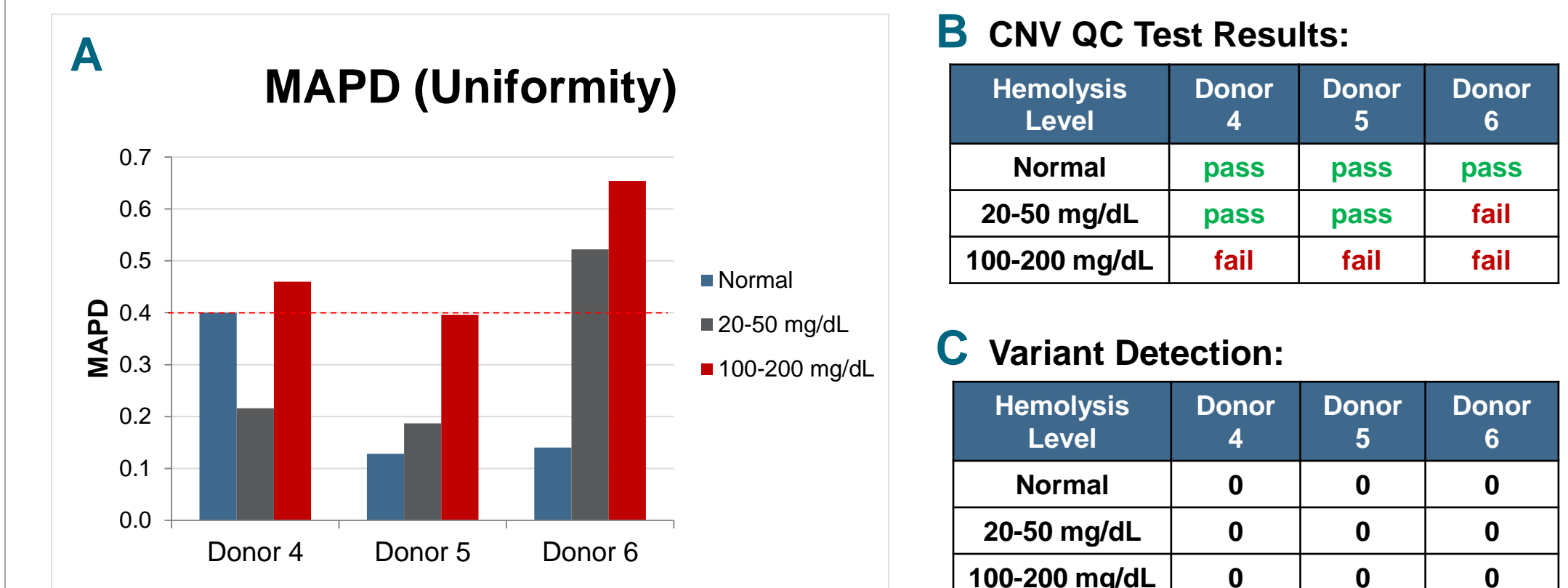
A) NGS libraries were prepared using maximum volume input (10.4 µl) for each sample. The median absolute pairwise difference (MAPD) quality control metric was higher for hemolyzed samples. B) The Ion Reporter QC threshold is MAPD <0.4, which passed for all three donors when plasma was normal. However, this QC test failed for 4 out of 6 samples with either moderate or high hemolysis. The MAPD metric is important for CNV detection, and therefore hemolyzed plasma may negatively impact accurate CNV calling. C) There were no variants detected in any sample, regardless of hemolysis state. As the plasma was collected from healthy donors, this is not unexpected, but suggests increased hemolysis does not cause an increase in false-positive variant detection.

CONCLUSIONS

- Plasma hemolysis in K2EDTA blood collection tubes results in higher yield primarily due to high gDNA contamination.

- Hemolysis can be detected by absorbance at 414 nm.

- Bioanalyzer/Tapestation analysis is a valuable quality control test to visualize cfDNA versus gDNA contamination in a sample.

- Low sample input into NGS libraries reduces mean read length and increases %LOD.

- Library fragment size can be visualized on a Bioanalyzer or Tapestation, as an in-process QC test of library size profile

- Hemolysis impacts NGS uniformity (MAPD), which can adversely affect CNV calls

This study used analytical methods to assess the quality of liquid biopsy NGS samples from initial sample preparation through NGS analysis. The results demonstrate that reduced sample input into NGS libraries increases small byproducts and reduces molecular coverage. In addition, hemolyzed K2-EDTA plasma had higher gDNA contamination, reducing the percentage of cfDNA in the NGS library prep. Hemolysis also negatively impacted sequencing uniformity, evidenced by high MAPD values. Together, these data indicate plasma quality, nucleic acid quantity, and the size profile of isolated nucleic acids are important quality control metrics for liquid biopsy NGS. Highly hemolyzed plasma samples should therefore be avoided for liquid biopsy applications whenever possible.

REFERENCES

1. Shah JS, Soon PS, Marsh DJ (2016) Comparison of Methodologies to Detect Low Levels of Hemolysis in Serum for Accurate Assessment of Serum microRNAs. PLoS ONE 11(4): e0153200. doi:10.1371/journal.pone.01532001.

2. Page K, Guttery DS, Zahra N, Primrose L, Elshaw SR, Pringle JH, et al. (2013) Influence of Plasma Processing on Recovery and Analysis of Circulating Nucleic Acids. PLoS ONE 8(10): e77963. https://doi.org/10.1371/journal.pone.0077963

3. Markus H, et al. (2018). Evaluation of pre-analytical factors affecting plasma DNA analysis. Nature Scientific Reports (8). doi: 10.1038/s41598-018-25810-0.

4. Cheng HH, Yi HS, Kim Y, Kroh EM, Chien JW, Eaton KD, et al. (2013) Plasma Processing Conditions Substantially Influence Circulating microRNA Biomarker Levels. PLoS ONE 8(6): e64795

TRADEMARKS/LICENSING