

Full-Length Clonal Immunoglobulin Rearrangements in cfDNA: Improved Recovery and Sequencing

Rena R. Xian^{a,b,*}, Prisca Mbonu,^a Lisa M. Haley,^a Troy Billson,^c Kevin He,^a Emily Adams,^a Ming-Tseh Lin,^a Moosa Patel,^d Atul Lakha,^d Garrick Laudin,^d Tanvier Omar,^e Philippa Ashmore,^f Sugeshnee Pather,^e Elizabeth Mayne,^{g,h} Wendy Stevens,^{d,g,i} Vinitha Philip,^d Christopher D. Gocke,^{a,b} Neil A. Martinson,^{j,k,†} Samantha L. Vogt,^{b,j,k,†} and Richard F. Ambinder^{a,b,k,†}

BACKGROUND: Circulating tumor DNA present in cell-free DNA (cfDNA) serves as diagnostic, predictive, and prognostic markers in cancer patients. In lymphoma patients, clonal immunoglobulin (cIg) sequences enable accurate identification of disease states. Despite its importance, there is sparse literature surrounding the preanalytical and analytical conditions that affect cIg detection in cfDNA. This study evaluates the effects of cfDNA isolation and next-generation sequencing (NGS) library preparation on *IGH* sequencing.

METHODS: Pooled plasma obtained from healthy donors and plasma collected from study participants (N = 118) enrolled in cross-sectional studies were analyzed. Two cfDNA isolation methods, paramagnetic bead and spin-column, were evaluated with plasma volume being variable. Two NGS library preparation procedures, single and dual library purification, were evaluated with cfDNA input being variable.

RESULTS: Paramagnetic bead-based cfDNA isolation improved recovery of cfDNA when compared to the column-based method. Higher purity of cfDNA resulted in more reproducible detection of cIg sequences. NGS library conditions affected *IGH* sequencing read depth, read diversity, and enrichment for full-length immunoglobulin sequences. A modified library procedure consisting of repeat library purifications improved the read depth of short and intermediate length targets independent of cfDNA input.

CONCLUSIONS: This is the first study to evaluate preanalytical and analytical factors that impact immunoglobulin sequencing in cfDNA. Our findings show optimal cfDNA isolation and library preparation procedures can improve the detection of clonal immunoglobulin rearrangements.

Introduction

Circulating free nucleic acids were first described in 1948 (1). Since then, the understanding of cell-free DNA (cfDNA) has expanded, enabling applications in virtually all areas of medicine. The study of circulating tumor DNA (ctDNA) has driven the rapidly evolving field of liquid biopsies. Next-generation sequencing (NGS) has emerged as a cornerstone for investigating cfDNA and ctDNA. The discovery and validation of sequencing-based ctDNA markers relies on an in-depth understanding of cfDNA dynamics and fragmentation, technical challenges, and clinical implications. The development of novel cfDNA NGS methods requires optimization of preanalytical and analytical parameters, such as specimen handling, nucleic acid extraction, and sequencing procedures (2–4).

Plasma cfDNA is the byproduct of cell turnover and is mostly derived from normal peripheral blood cells (5, 6). cfDNA is highly fragmented (7), highly labile (8, 9), and present at relatively low abundance. Most healthy individuals have ≤ 10 ng cfDNA per mL of plasma (9, 10). While there may be physiologic spikes in cfDNA (9), only sustained cellular proliferations like

^aDepartment of Pathology, Johns Hopkins School of Medicine, Baltimore, MD, United States; ^bDepartment of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD, United States; ^cClinical Lab Services, Wits Health Consortium, University of the Witwatersrand, Johannesburg, South Africa; ^dClinical Haematology Unit, Department of Medicine, Chris Hani Baragwanath Academic Hospital and Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; ^eDivision of Anatomical Pathology, National Health Laboratory Service, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; ^fClinical Haematology, Netcare Olivedale Hospital, Johannesburg, South Africa; ^gMolecular Medicine and Haematology, University of the Witwatersrand, Johannesburg, South Africa; ^hDivision of Immunology, University of Cape Town, Cape Town, South Africa; ⁱWits Diagnostic Innovation Hub, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; ^jPerinatal HIV Research Unit, University of the Witwatersrand, Johannesburg, South Africa; ^kDepartment of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States.

*Address correspondence to this author at: Department of Pathology, Johns Hopkins Genomics, 1812 Ashland Ave., Suite 200, Baltimore, MD 21205, United States. E-mail rxian1@jhmi.edu.

†These authors contributed equally.

Received December 10, 2024; accepted April 28, 2025.
<https://doi.org/10.1093/clinchem/hvaf064>

cancer lead to continuous and substantial levels of cfDNA (11). Irrespective of the origin, cfDNA mostly occurs as short double-stranded DNA fragments that approximate the length of DNA loops around nucleosomes (12) corresponding to mononucleosomal (mncfDNA) or dinucleosomal cfDNA (dncfDNA). Most sequencing technologies show that mncfDNA is 150 to 170 bp (12–14), while dncfDNA is 350 to 400 bp (13). When comparing ctDNA to cfDNA from nontumor cells, ctDNA is 10 to 40 bp shorter for both mncfDNA (15–17) and dncfDNA (13). Recently, another component of cfDNA, which is mostly single stranded, termed ultrashort cfDNA (50 bp), has been described (18, 19).

In cancer patients, ctDNA comprises a minor component of cfDNA (<10%) (20, 21) with the lowest proportions found in early-stage solid tumor patients (<1%) (20). Higher proportions are found in patients with lymphoma, such as classical Hodgkin lymphoma (5%) (22) and diffuse large B-cell lymphoma (10%) (23). Given the low abundance of cfDNA, even lower ctDNA, and the risks of ex vivo cell lysis leading to contamination with cellular genomic DNA (gDNA), specialized blood collection tubes (24) and processing parameters (25) are necessary to maximize cfDNA recovery. These preanalytical factors have been shown to directly impact the sensitivity of ctDNA detection. Aside from specimen collection and processing, cfDNA isolation methods have been studied as a determinant of cfDNA recovery and ctDNA detection. Approximately 50% of cfDNA studies to date have used column-based isolation methods (4, 26), which show similar cfDNA profiles when compared to bead-based extraction methods, at least for 1 mL of plasma (25). Others have suggested that bead-based methods recover more cfDNA with more abundant ctDNA (27, 28). Not surprisingly, higher plasma volume extracts (>7 mL) (29) and higher input cfDNA (>10–50 ng) (30) are associated with improved ctDNA detection. Additionally, the sequencing methodology has an independent effect on ctDNA detection (30), as well as in silico enrichment for shorter ctDNA molecules (15).

Unlike epithelial tumors, lymphomas derived from B-cells carry unique clonal B-cell receptor rearrangements, encoded by the *IGH*, *IGK*, or *IGL* genes, that can be used as a plasma tumor marker (31). Full-length variable diversity, and joining (VDJ) region sequences reflecting lymphoma clonotypes are detectable in plasma from lymphoma patients (31–36). Clearance of these clones is associated with improved outcomes (31–33). Despite their clinical importance, there are no studies to date that address the unique characteristics of clonal immunoglobulin (cIg) sequencing in cfDNA with respect to preanalytical and analytical variables. Full-length VDJ sequences are typically amplified

using consensus primers, resulting in 3 different amplicon lengths ranging from <150 bp to <350 bp (37). Most ctDNA studies in solid tumors focus on mncfDNA or ultrashort cfDNA molecules, but cIg sequencing requires longer molecules rather than shorter. We previously showed that lymphoma clones are detectable in cfDNA even in the largest amplicon target (36). Given the importance of high quantity, quality, and purity of cfDNA for detecting tumor markers in plasma, the present study evaluates the effects of the isolation method on cfDNA recovery and the effects of the library preparation method on the sequencing of full-length immunoglobulin rearrangements (Fig. 1). To our knowledge, this is the first study to identify cfDNA isolation and sequencing procedures that improve cIg sequencing, a key ctDNA marker in lymphoma patients.

Materials and Methods

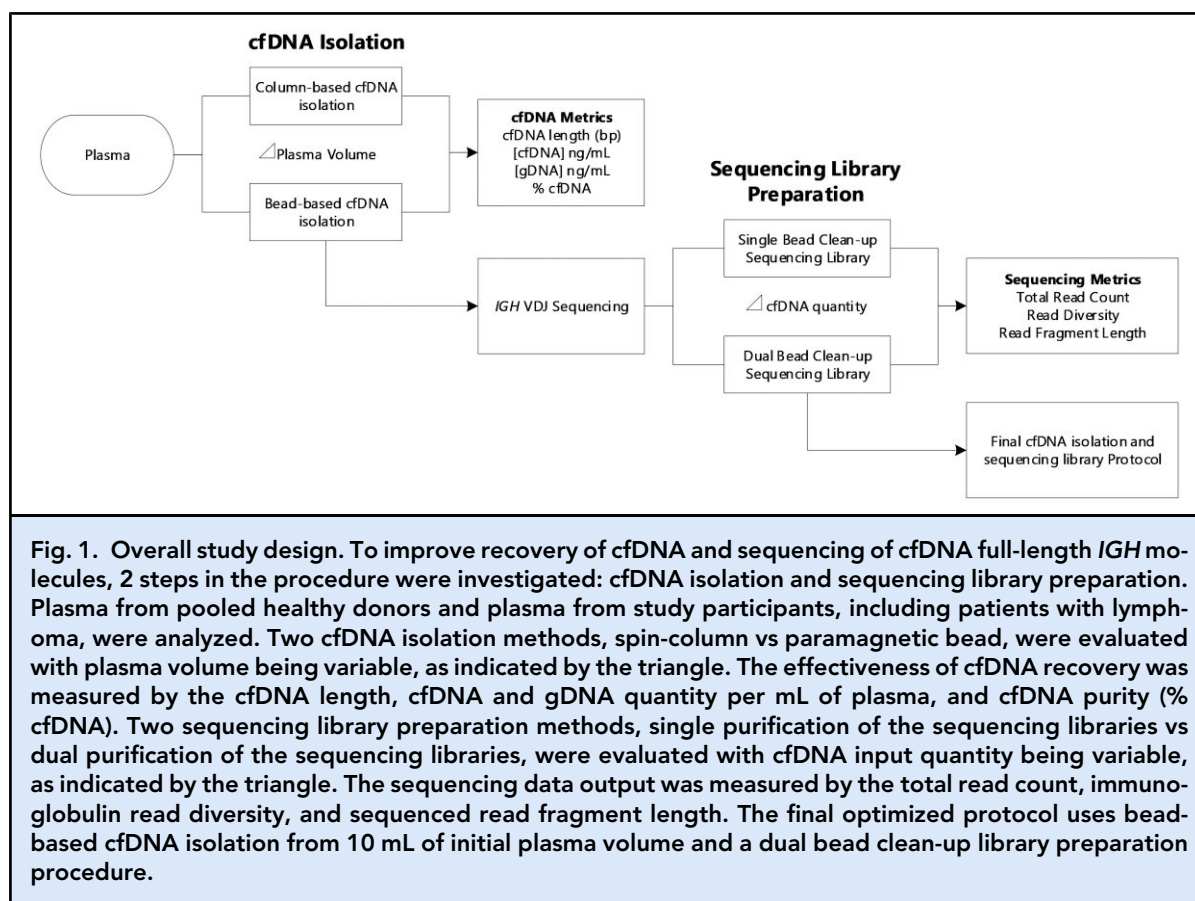
SAMPLE COLLECTION AND STUDY PARTICIPANTS

Pooled plasma from healthy donors and plasma from study participants (N = 118) enrolled in cross-sectional studies were evaluated. Studies were approved by the Johns Hopkins Institutional Review Board (IRB00107027, IRB00165673) and the University of the Witwatersrand Human Research Ethics Committee (160608, 180216, 200208). Pooled frozen plasma was obtained (Innovative Research). Study participants were recruited at Chris Hani Baragwanath Academic Hospital in Soweto, South Africa. Blood was collected in cell-stabilizing tubes (BCT tubes, Streck) from individuals with lymphoma prior to treatment (43/118, 36%) and from individuals without known cancer (75/118, 64%), as previously described (36) (Supplemental Methods).

PLASMA CFDNA ISOLATION AND CFDNA ANALYSIS

Plasma cfDNA was isolated using either a column-based method (QIAamp Circulating Nucleic Acid Kit, Qiagen) or a bead-based method (nRichDx Revolution cfDNA Kit, nRichDx) per vendor protocols (Supplemental Methods). Pooled plasma was extracted in triplicate using both methods. Plasma from study participants was isolated in singlicate using the bead-based method. Four study participants had sufficient plasma to perform parallel extractions using both methods.

cfDNA and gDNA profiles were evaluated using a high-sensitivity electrophoresis system (D1000 DNA ScreenTape, Agilent 4200 TapeStation system, Agilent) and analyzed using the vendor software (version 4.1). DNA fragments smaller than approximately 400 bp that had appropriate cfDNA profiles (36) were designated as cfDNA. Larger fragments were designated as gDNA. Isolation efficiency was assessed using 4 metrics: cfDNA integrity and quality as reflected by the



presence of mncfDNA, dncfDNA, and ultrashort molecules; cfDNA quantity expressed as [cfDNA] ng/mL plasma; gDNA quantity expressed as [gDNA] ng/mL plasma; and cfDNA purity expressed as $\%[\text{cfDNA}] / ([\text{cfDNA}] + [\text{gDNA}])$.

IMMUNOGLOBULIN GENE SEQUENCING AND DATA ANALYSIS

Immunoglobulin heavy chain gene (*IGH*) sequencing libraries (LymphoTrack *IGH* assay, Invivoscribe Inc.) were prepared following vendor protocols as previously described (37), unless otherwise specified. In brief, the assay uses polymerase chain reaction to enrich for rearranged VDJ sequences with consensus framework 1, 2, or 3 (FR1, FR2, or FR3) primers generating median amplicons of <350 bp, <290 bp, and <150 bp and median sequencing read lengths of <300 bp, <240 bp, and <100 bp, respectively. The vendor protocol (condition A) has a single post-polymerase chain reaction bead purification/clean-up step using solid-phase reversible immobilization technology (Agencourt AMPure XP Kit, Beckman Coulter) followed by normalization, pooling,

and sequencing (Supplemental Fig. 1) on the MiSeq (Reagent Kit v2, 500-cycles, Illumina Inc.).

To investigate the effect of the cfDNA isolation method on *IGH* sequencing, plasma from a single participant with lymphoma (L01) was studied. Sequencing was performed in duplicate using 50 ng bead-isolated cfDNA and the standard sequencing procedure. This was compared to sequencing using 6 ng of cfDNA (either bead or column) performed in sextuplicate.

To evaluate the effect of repeat paramagnetic bead clean-up during library preparation, 3 additional library conditions (B, C, D) were devised. All 4 sequencing procedures were applied to the same bead-isolated cfDNA obtained from 20 mL of pooled plasma (Supplemental Fig. 1). This sample was sequenced in sextuplicate using 4 ng of cfDNA (Supplemental Methods). Conditions A and B were further evaluated in study participants using up to 50 ng of cfDNA. Each cfDNA sample was sequenced in duplicate. Based on random sampling, this resulted in a similar proportion of individuals with lymphoma or without known cancer included in each condition (A: 14 and 20; B: 29 and 55; $P = 0.4965$). The skewed samples evaluated by condition B relate to

early improvements leading to a switch to condition B. At each library assessment, library products measuring at the expected size ranges, partial amplicon, and primer/adaptor pairs were annotated. After sequencing, total read counts, read diversity (including the Simpson's diversity index) (38), and read fragment lengths were compared (Supplemental Methods).

STATISTICAL ANALYSIS

Continuous variables were evaluated using the non-parametric pair-wise Wilcoxon rank sum test. Categorical variables were evaluated using the χ^2 test. Significance was defined as $P \leq 0.05$. The effects of specific variables on a fixed measurement were evaluated using linear regression by standard least squares with crossing of variables to exclude interactions. The transformed P value, Logworth, $-\log_{10}(P \text{ value})$, and native P value were calculated where a Logworth of ≥ 2 or $P \leq 0.01$ was defined as significant. To derive the 95% CI of a linear regression for the sequencing data from study participants, a bootstrap method was applied using an arbitrary sample size of 200. All statistical analyses and graphing were performed using JMP (v19.0.0, SAS Institute Inc.). Additional illustrations were performed using BioRender (BioRender).

Results

PARAMAGNETIC BEAD ISOLATION IMPROVES cfDNA RECOVERY AND IS AFFECTED BY PLASMA VOLUME

Compared to the column-based method, the bead-based method of cfDNA isolation improved the quality and quantity of cfDNA recovered by all metrics using pooled plasma from healthy donors at varying plasma volumes (Fig. 2, Supplemental Fig. 2, Supplemental Table 1). Both methods recovered mncfDNA of similar fragment lengths (median 179 bp bead vs 184 bp column $P = 0.1637$), but only the bead-based method recovered discrete dncfDNA (Fig. 2A). Only the bead-based cfDNA isolates showed discrete ultrashort cfDNA (Supplemental Fig. 2, Supplemental Table 1). The bead-based method recovered 1.6-fold more cfDNA/mL of plasma (Fig. 2B), as well as lower gDNA, although this was not significant (Supplemental Fig. 3A). Nevertheless, the purity of cfDNA was significantly higher using the bead-based method (median 38 vs 9%, $P = 0.0009$) (Supplemental Fig. 3B).

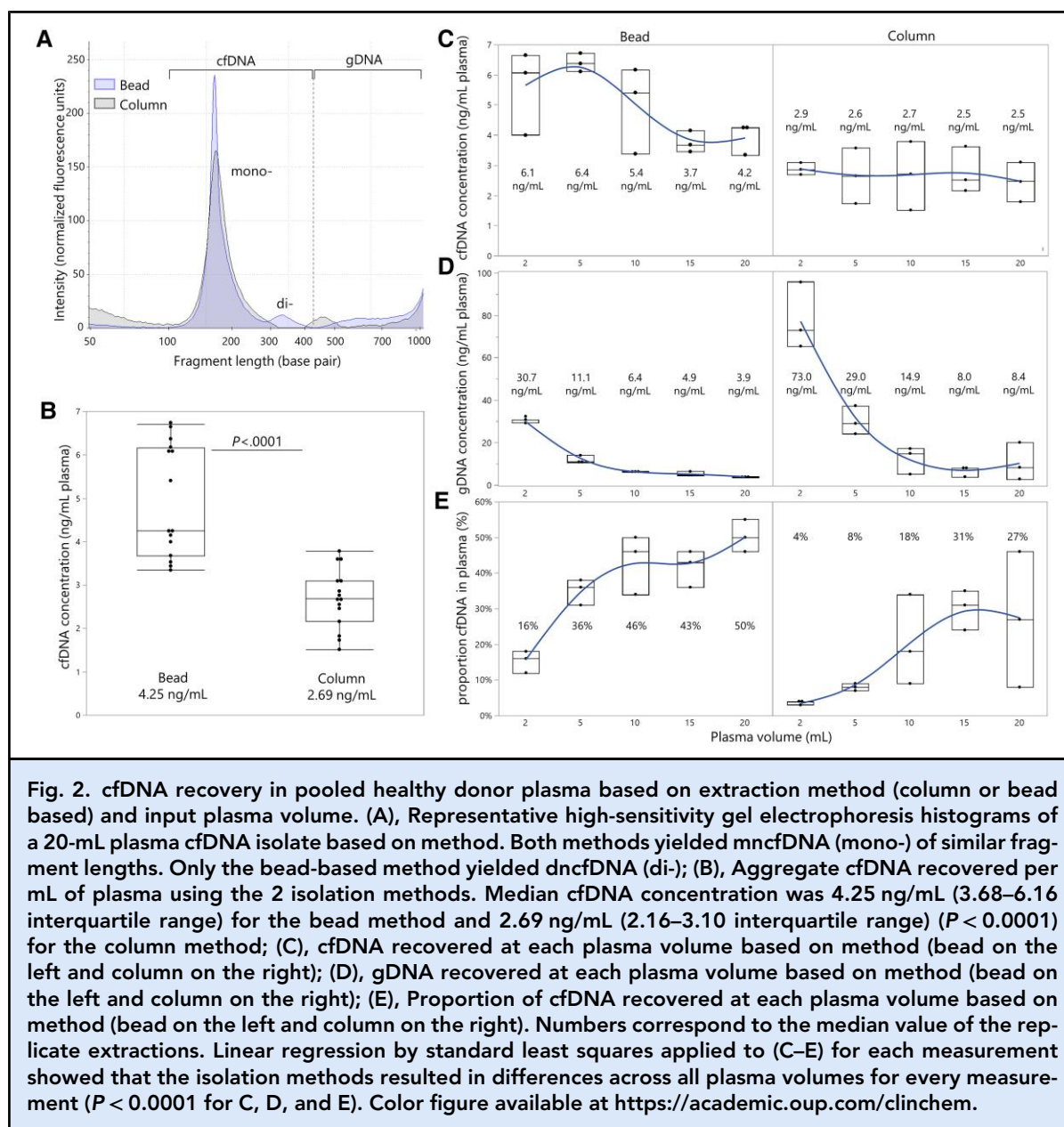
To determine the effect of plasma volume on cfDNA recovery, the same analyses were performed relative to plasma volume and isolation method (Fig. 2C–E), which were independent variables that differentially affected cfDNA recovery (Supplemental Table 2). The isolation method was the major determinant of cfDNA recovery with the bead method showing higher cfDNA recovered per mL of plasma at every volume.

Initial plasma volume only had a minor effect that was mostly attributable to the bead-based method, where cfDNA recovery decreased as plasma volume increased (Fig. 2C). Plasma volume had no effect on cfDNA recovery when isolated by the column method. Regarding gDNA, plasma volume had a major effect on gDNA recovery irrespective of isolation method. As plasma volume increased, the gDNA recovered per mL of plasma decreased (Fig. 2D), particularly at ≥ 10 mL. The isolation method only had a minor effect on gDNA recovery, as the bead method demonstrated lower gDNA recovery than the column method at most volumes. These findings held true when total DNA yields were compared (Supplemental Fig. 4). cfDNA purity increased with increasing volume for both methods (3.1-fold and 6.7-fold, bead and column, respectively, when comparing 20 mL to 2 mL) (Fig. 2E). At every volume, the bead-based method showed higher % cfDNA. The length of mncfDNA molecules decreased as plasma volume increased (Supplemental Fig. 3C), but the length of dncfDNA molecules showed no change. These results indicate that bead-based cfDNA isolation recovered a wider range of cfDNA molecules and at a higher quantity and purity. Initial plasma volume altered cfDNA recovery, especially for the bead method.

Plasma from four study participants also underwent the same cfDNA isolation method comparison. The integrity of cfDNA remained higher in the bead-based isolates given the consistent presence of dncfDNA (Fig. 3A and B). None of these isolates showed discrete ultrashort cfDNA molecules. The bead-based isolates contained more cfDNA and less gDNA with higher %cfDNA (Fig. 3C). The smaller volumes used in the column-based isolates may increase gDNA and lower %cfDNA but should not affect cfDNA quantity since there was constant recovery at all volumes using pooled plasma (Fig. 2C). The lengths of mncfDNA were similar in bead isolates compared to the smaller volume column isolates (median 179 bp vs 172 bp, respectively, $P = 0.0833$).

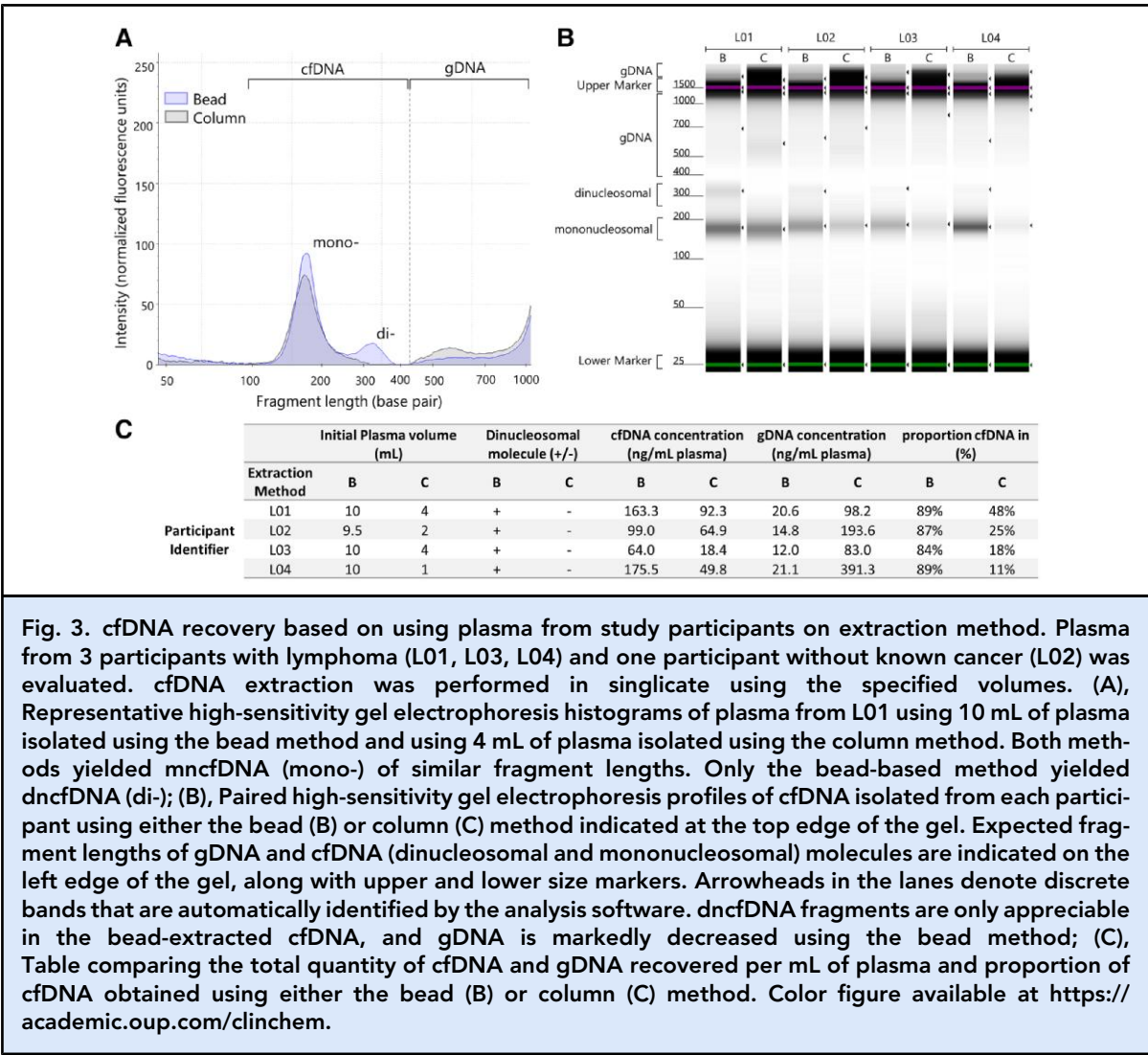
CFDNA ISOLATED USING PARAMAGNETIC BEADS YIELDS MORE REPRODUCIBLE DETECTION OF CLONAL IMMUNOGLOBULIN

To investigate the potential effects of cfDNA isolation method on the detection of cIg rearrangements, plasma from a single study participant (L01 from Fig. 3C) with lymphoma and a cIg rearrangement in FR1 was sequenced at both high and low cfDNA input (Fig. 4). With the reduction in input cfDNA, the total read count and clone proportion showed a nonsignificant reduction irrespective of the isolation method (Fig. 4A and B). Although the median clonal percentage was not different between isolation methods, the interquartile range was smaller for the bead-isolated cfDNA (86.5%–88.2% vs 69.2%–87.5%).



When comparing the cumulative reads from the top 200 aligned *IGH* sequences detected in all 6 replicates of the low-input cfDNA sequencing, the column-cfDNA showed additional *IGH* sequences not detected in any of the bead-cfDNA sequencing studies (Fig. 4C). The lymphoma clone was 278 bases with a IGHV3-11_01;IGHJ4_03 gene usage and was the only sequence detected in the low-input bead cfDNA. The column cfDNA showed unrelated full-length sequences (269-base-IGHV3-11_04; IGHJ4_03, 291-base-IGHV4-b_01;IGHJ3_02) and 2 unrelated partial V-gene sequences (246-base-IGHV3-49_04, 234-base-IGHV3-d_01). The absence of

these sequences from the bead-cfDNA sequencing suggests that they may be derived from gDNA, which comprised 52% of the column cfDNA. These unrelated sequences likely accounted for the highly variable and reduced clonal fractions detected using the column-extracted cfDNA. Since only the top 200 aligned sequences were analyzed, there may be additional low-frequency *IGH* sequences that reduced clonal percentage in all sequencing reactions. Sequencing of the nonclonal frameworks, FR2 and FR3, using the same parameters showed no significant differences in total read count based on the extraction method when input cfDNA was similarly reduced.



IMMUNOGLOBULIN SEQUENCING LIBRARY PREPARATION
PROCEDURES AFFECT SEQUENCING DATA OUTPUT

After determining that the bead method of cfDNA isolation yielded higher integrity, quantity, and purity of cfDNA with more reproducible clg sequencing, we next investigated the possibility of improving sequencing data output through optimization of the sequencing protocol. The vendor-recommended procedure (condition A) is intended for gDNA and yields low sequencing data output for cfDNA (36). This may be related to the retention of non-specific primer/adaptor dimers and partial amplicons present in the cfDNA sequencing library. In an attempt to decrease nonspecific library products and improve sequencing of full-length *IGH* sequences, cfDNA isolated from 20 mL of pooled healthy donor plasma using the bead method was subjected to 4 different library preparation procedures (Supplemental Fig. 1, Supplemental Methods).

Primer/adaptor pairs and partial amplicons were markedly reduced in condition B (Supplemental Fig. 5). Based on the total read count, condition B showed higher read depths in all 3 frameworks (Fig. 5A). *IGH* sequence diversity, as visualized by the density and abundance of polyclonal *IGH* sequences, was highest in condition B for FR1 and FR2 (Fig. 5B). Conditions A and B showed similarly improved profiles for FR3. In terms of the range of fragment lengths obtained, condition B yielded the highest median fragment length for FR1, while all conditions yielded similar fragment lengths for FR2 and FR3 (Fig. 5C). Despite the fact there was no major increase in fragment lengths based on sequencing procedure, the proportion of full-length VDJ sequences, rather than partial V and/or J sequences, was higher in conditions A and B, and condition B showed the highest Simpson's Diversity Index (Supplemental Table 3).

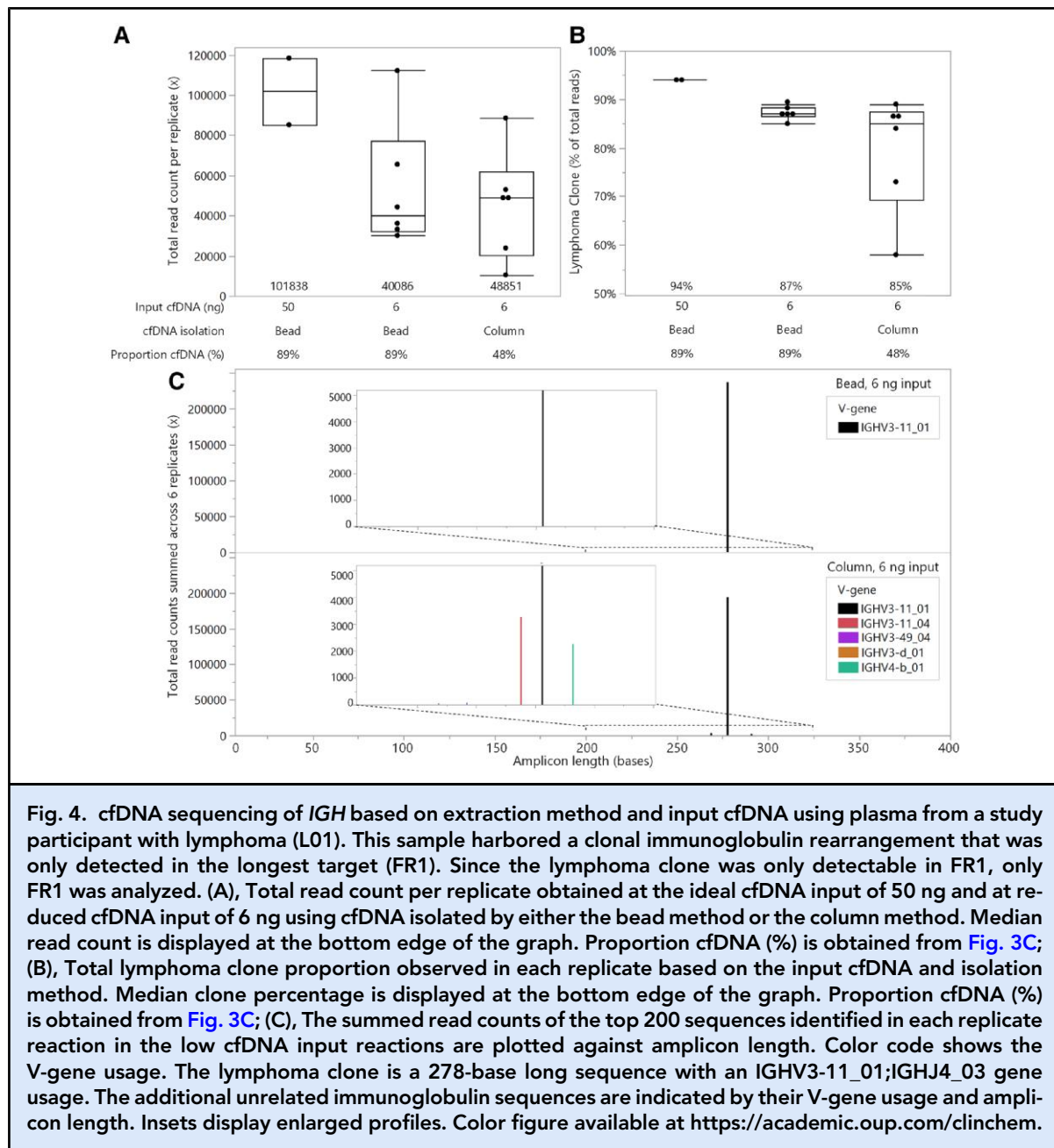


Fig. 4. cfDNA sequencing of *IGH* based on extraction method and input cfDNA using plasma from a study participant with lymphoma (L01). This sample harbored a clonal immunoglobulin rearrangement that was only detected in the longest target (FR1). Since the lymphoma clone was only detectable in FR1, only FR1 was analyzed. (A), Total read count per replicate obtained at the ideal cfDNA input of 50 ng and at reduced cfDNA input of 6 ng using cfDNA isolated by either the bead method or the column method. Median read count is displayed at the bottom edge of the graph. Proportion cfDNA (%) is obtained from Fig. 3C; (B), Total lymphoma clone proportion observed in each replicate based on the input cfDNA and isolation method. Median clone percentage is displayed at the bottom edge of the graph. Proportion cfDNA (%) is obtained from Fig. 3C; (C), The summed read counts of the top 200 sequences identified in each replicate reaction in the low cfDNA input reactions are plotted against amplicon length. Color code shows the V-gene usage. The lymphoma clone is a 278-base long sequence with an IGHV3-11_01;IGHJ4_03 gene usage. The additional unrelated immunoglobulin sequences are indicated by their V-gene usage and amplicon length. Insets display enlarged profiles. Color figure available at <https://academic.oup.com/clinchem>.

REPEAT PURIFICATION OF IMMUNOGLOBULIN LIBRARIES IMPROVES SEQUENCING READ DEPTH

Since conditions A and B performed the best using cfDNA from pooled healthy donors, these 2 protocols were further evaluated in 118 study participants. The median plasma volume used for isolation was 10 mL for all participants, with lymphoma patients showing higher quantities of cfDNA leading to higher input cfDNA (median 34.7 ng/mL vs 18.2 ng/mL, $P=0.0043$). Despite this, the total read counts obtained were similar irrespective of

lymphoma status (Supplemental Fig. 6), and subsequent analyses were performed together. The median cfDNA input quantity was lower in condition A samples, leading to lower total read counts in all 3 frameworks (Supplemental Table 4). Acknowledging that cfDNA input is a confounder, analysis was restricted to 50 ng input samples. In this case, condition B continued to show higher total read counts in all frameworks. Median cfDNA needed to achieve a minimum 10 000x reads was 50 ng by condition A and 32 ng by condition B.

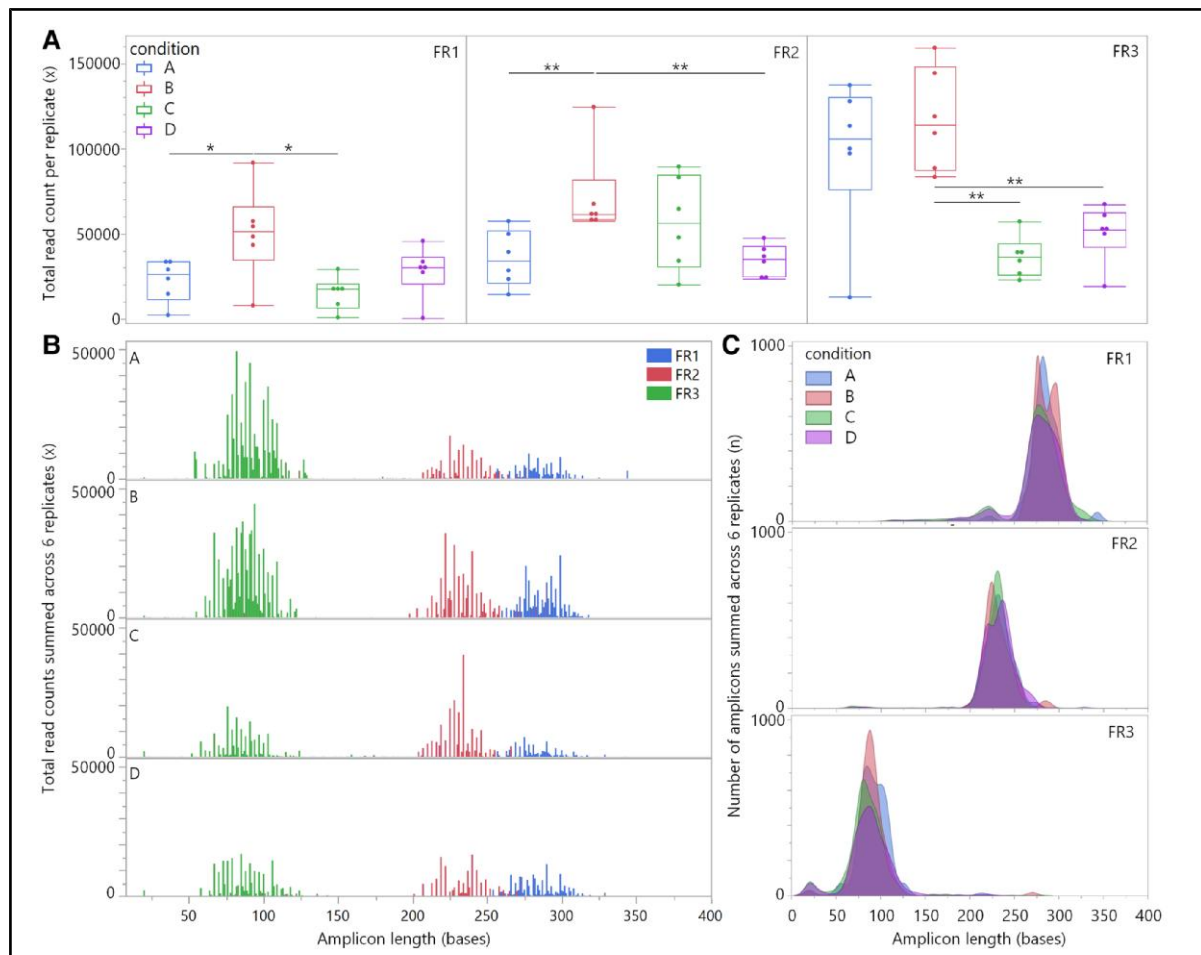


Fig. 5. cfDNA sequencing of *IGH* based on a sequencing library procedure using bead-isolated cfDNA obtained from 20 mL of pooled plasma from healthy donors. Four nanograms of cfDNA were used for each sequencing reaction. Sequencing for each condition was performed in sextuplicate, encompassing 3 different *IGH* targets and amplicon lengths (FR1 > FR2 > FR3). (A), Total read count per replicate obtained for each condition and target. Condition B showed higher read depths in all 3 frameworks. The other conditions showed no differences between one another. In FR1, condition B yielded median total read counts of 51 348x, as compared to 26 332x (A, $P = 0.0453$), 17 523x (C, $P = 0.0453$), and 30 322x (D, ns). In FR2, condition B yielded median total read counts of 61 632x, as compared to 33 853x (A, $P = 0.0051$), 56 191x (C, ns), and 35 232x (D, $P = 0.0051$). In FR3, condition B yielded median total read counts of 114 010x, as compared to 105 702x (A, ns), 36 295x (C, $P = 0.0051$), and 52 439x (D, $P = 0.0051$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; (B), Summed read counts of unique immunoglobulin molecules are aggregated across all replicates for each condition (A–D) and target (FR1, FR2, and FR3) and are plotted against sequence length. The density of sequences separated by amplicon length and the summed total read counts of each of these sequences is an indicator of read diversity; (C), Number of unique immunoglobulin molecules are aggregated across all replicates for each condition (A–D) and target (FR1, FR2, and FR3). The median fragment lengths obtained for FR1 by condition were 284, 287, 280, and 281, respectively. The median fragment lengths obtained for FR2 by condition were 232, 228, 231, and 234, respectively. The median fragment lengths obtained for FR3 by condition were 91, 88.5, 83, and 88, respectively. Abbreviation: ns, nonsignificant. Color figure available at <https://academic.oup.com/clinchem>.

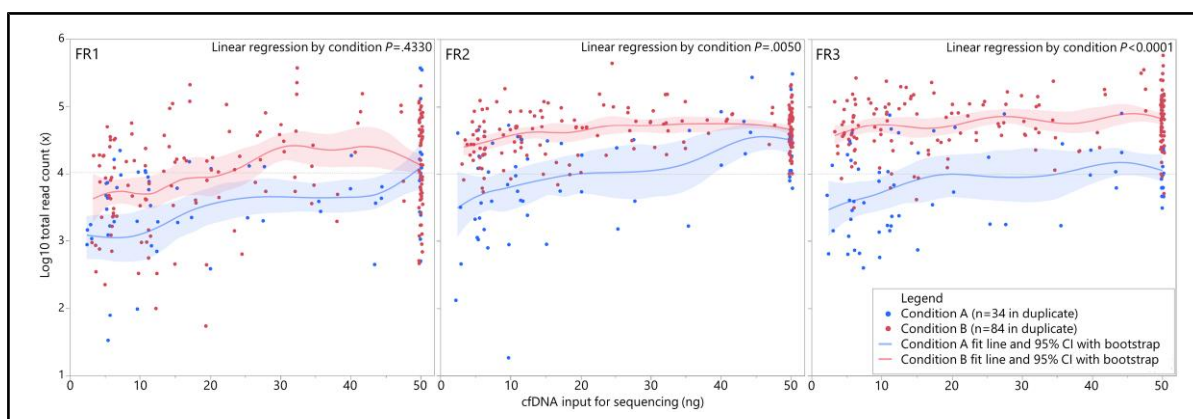


Fig. 6. cfDNA sequencing of *IGH* in 118 study participants based on sequencing library condition using bead-isolated cfDNA. Sequencing was performed in duplicate for each sample (236 reactions) at the indicated input cfDNA per replicate ranging from less than 10 to 50 ng. The dotted line denotes a minimum total read count of 10 000x (Log4), which may be considered a minimum optimal sequencing depth. For each amplicon (lengths FR1 > FR2 > FR3), condition B yielded higher total read counts for FR1 and FR2, particularly at lower cfDNA input. For FR3, condition B maintained higher read counts along all ranges of input cfDNA. The fit lines and regression significances are shown based on sequencing condition and indicate that condition B outperforms condition A in terms of read depth for FR2 and FR3. The 95% confidence interval (shaded zone) is bootstrapped using an arbitrary sample size of 200, which approximates the number of sequencing reactions for each framework. Increasing input cfDNA independently correlated with improved read count for all 3 targets ($P < 0.0001$ for FR1 and FR2, $P = 0.0388$ for FR3). Color figure available at [https://academic.oup.com/clinchem](https://academic.oup.com/clinchem/article/71/8/896/8148924).

When cfDNA input was treated as a continuous variable and the library procedure was treated as a categorical variable (Fig. 6), condition B showed higher total read depths at all cfDNA inputs for FR2 and FR3. cfDNA input and library preparation were independent variables that differentially affected the total read counts (Supplemental Table 5). Higher cfDNA input was the only factor in achieving higher FR1 read count, whereas condition B was the only factor in achieving higher FR3 read count. Both factors had major effects on FR2 read count.

Discussion

In patients with B-cell lymphoma, cIg detection in plasma is a key ctDNA marker that tracks with treatment response and clinical outcomes (31–33). Despite its importance, there is sparse literature surrounding the preanalytical and analytical conditions that affect the detection of cIg in cfDNA. In this study, we investigated laboratory parameters that improve the recovery and sequencing of full-length immunoglobulin rearrangements in plasma. We showed that a bead-based cfDNA isolation method improves the quantity and purity of cfDNA recovery. Enriching cfDNA in this manner resulted in more reproducible cIg detection. We also demonstrated that

optimized NGS library preparation procedures improve cfDNA immunoglobulin data quality and sequencing read depth independent of cfDNA input.

The current study highlights observations important for the larger field of cfDNA/ctDNA biomarker discovery and development. While previous studies exploring the effects of cfDNA isolation methodology on cfDNA profiles and ctDNA detection suggested higher proportions of cfDNA isolated by bead-based methods (25, 27, 28), the current study addresses a number of additional cfDNA characteristics. These characteristics include cfDNA integrity, as reflected by the presence of discrete dncfDNA and ultrashort molecules; cfDNA and gDNA concentration (ng/mL plasma); and %cfDNA. Bead-based cfDNA isolates outperformed column-isolates by all measures. The plasma volume experiments also demonstrated notable effects on gDNA recovery and cfDNA purity. Our results suggest that differences in cfDNA characteristics reported in the literature may be attributable, at least partly, to isolation techniques. Therefore, it would be prudent to establish a defined cfDNA isolation procedure and standardized plasma volume prior to initiating large-scale clinical studies or beginning clinical testing. This strategy would ensure internal consistency and validity, if not external generalizability.

The analytical portion of this study showed that NGS library preparation procedures intended for

sequencing of gDNA *IGH* can be adapted for cfDNA. The optimized library preparation procedure consisting of dual library purification steps improved *IGH* diversity and increased read depths, particularly for short FR3 and intermediate FR2 amplicons. This new procedure even overcomes low cfDNA input for both of these targets. In the longest FR1 amplicon, input cfDNA was the only way to increase read depth, speaking to the importance of effective cfDNA isolation. Despite the observation that FR1 targets may be larger than most cfDNA molecules, FR1 continues to be diagnostically important, as cIg sequences are detectable in FR1 and FR1 sequences appear to be more abundant in lymphoma patients (36). The sequencing improvements observed in our study suggest that future cfDNA *IGH* sequencing endeavors would benefit from similar library modifications, particularly when cfDNA quantity is limited.

The current investigation has some limitations. Only a single column-based and a single bead-based method were compared. A single column method was evaluated because this was the predominant method cited in the literature (4, 26), and prior studies show similar cfDNA performance among different spin-column kits and bead-based isolation techniques (25, 27, 28). The design of the current study proceeded in a 2-step manner, where only the selected isolation procedure was paired with cIg sequencing. There were no large-scale comparisons of column cfDNA sequenced with the optimized library protocol. Given the available plasma volume from study participants, sample splitting was not feasible without compromising the individual cfDNA input for sequencing. Comparing *IGH* sequencing of column cfDNA to bead cfDNA is also a complex endeavor. Our observation that column methods enrich gDNA, rather than cfDNA, may result in improved sequencing of full-length *IGH* sequences de facto since gDNA contains cell-derived DNA from normal B-cells. Sequencing of isolates with excess gDNA may artificially enhance *IGH* read diversity and read depth and improve sequencing of larger amplicons, without any relevance to cfDNA. This phenomenon may be especially apparent in samples lacking a detectable tumor-derived cIg where only polyclonal *IGH* sequences are present. Nevertheless, this question may be addressed in future studies where larger plasma volume collections could be obtained and paired leukocyte sequencing could be performed to subtract sequences derived from gDNA. Lastly, this study focused on preanalytical and analytical improvements without addressing diagnostic yield.

The present analyses demonstrate robust immunoglobulin sequencing using plasma cfDNA, especially when pairing high-fraction cfDNA isolates obtained by bead-based extractions with a cfDNA-optimized sequencing procedure. Many future directions may be pursued as a result of these findings. Evaluating the

sensitivity of detecting lymphoma clonotypes in cfDNA is of paramount importance, and these studies are underway. We had previously observed that intermediate and long cfDNA molecules harbor lymphoma clones (36). These observations suggest that enriching for dncfDNA with bead-based isolation and improving sequencing of longer fragment *IGH* sequences may improve cIg detection. To our knowledge, this line of investigation has not been explored in other cancer types where the focus has been on mncfDNA and ultrashort cfDNA. In addition to emerging methods of enriching for ctDNA, such as recovery of exosomal DNA (39) and in vivo agents that enhance ctDNA stability (40), future studies dedicated to the study of ctDNA fractions in dncfDNA may reveal additional characteristics of longer cfDNA molecules not yet explored.

In summary, our study is the first to evaluate preanalytical and analytical factors that improve immunoglobulin sequencing in cfDNA. We anticipate that cIg detection in cfDNA will become even more important for the clinical management of lymphoma patients. Future investigations in this area will likely be facilitated by our findings.

Supplemental Material

Supplemental material is available at *Clinical Chemistry* online.

Nonstandard Abbreviations: cfDNA, cell-free DNA; ctDNA, circulating tumor DNA; NGS, next-generation sequencing; mncfDNA, mononucleosomal DNA; dncfDNA, dinucleosomal DNA; gDNA, genomic DNA; VDJ, variable diversity and joining; cIg, clonal immunoglobulin.

Human Genes: *IGH*, immunoglobulin heavy locus; *IGK*, immunoglobulin kappa locus; *IGL*, immunoglobulin lambda locus.

Author Contributions: The corresponding author takes full responsibility that all authors on this publication have met the following required criteria of eligibility for authorship: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved. Nobody who qualifies for authorship has been omitted from the list.

Ren R. Xian (Conceptualization-Equal, Data curation-Equal, Formal analysis-Equal, Funding acquisition-Equal, Investigation-Equal, Methodology-Equal, Project administration-Equal, Resources-Equal, Software-Equal, Supervision-Equal, Validation-Equal, Visualization-Equal, Writing—original draft-Equal, Writing—review & editing-Equal), Prisca Mbonu (Investigation-Equal, Methodology-Equal, Writing—original draft-Equal), Lisa M. Haley (Conceptualization-Equal, Investigation-Equal, Methodology-Equal, Writing—original draft-Equal), Troy Billson (Investigation-Equal), Kevin He (Data curation-Equal, Investigation-Equal, Visualization-Equal), Emily Adams

(Investigation-Equal, Methodology-Equal), Moosa Patel (Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Atul Lakha (Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Garrick Laudin (Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Tanvier Omar (Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Philippa Ashmore (Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Sugeshnee Pather (Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Elizabeth Mayne (Funding acquisition-Equal, Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Vinitha Philip (Investigation-Equal, Writing—review & editing-Equal), Ming-Tseh Lin (Conceptualization-Equal, Writing—review & editing-Equal), Wendy Stevens (Funding acquisition-Equal, Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Christopher Gocke (Conceptualization-Equal, Methodology-Equal, Writing—review & editing-Equal), Neil A. Martinson (Funding acquisition-Equal, Investigation-Equal, Supervision-Equal, Writing—review & editing-Equal), Samantha L. Vogt (Funding acquisition-Equal, Investigation-Equal, Project administration-Equal, Supervision-Equal, Writing—review & editing-Equal), and Richard Ambinder (Conceptualization-Equal, Formal analysis-Equal, Funding acquisition-Equal, Investigation-Equal,

Project administration-Equal, Resources-Equal, Supervision-Equal, Writing—review & editing-Equal)

Authors' Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form.

Research Funding: The authors would also like to acknowledge the following funding sources: National Cancer Institute (R01CA250069, R21CA232891, U01CA284811, U01CA27125, UM1CA121947, P30AI094189, P30CA006973) and the Limas Research award.

Disclosures: R.R. Xian has received an honorarium and travel expenses from Invivoscribe Inc. N.A. Martinson has received institutional research funding from Pfizer.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of the manuscript, or final approval of the manuscript.

Acknowledgments: The authors would like to thank the patients and their families for their participation. The authors would like to acknowledge Mr. Kamran Syed (nRichDx) for his assistance with the technical aspects of their product. nRichDx did not provide other support and did not contribute to or review the design of the study or the conclusions drawn.

References

- Mandel P, Metais P. Les acides nucleiques du plasma sanguine chez l'homme. C R Seances Soc Biol Fil 1948;142:241–3.
- Meddeb R, Pisareva E, Thierry AR. Guidelines for the preanalytical conditions for analyzing circulating cell-free DNA. Clin Chem 2019;65:623–33.
- El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: preanalytical considerations. Clin Chim Acta 2013;424:222–30.
- Lockwood CM, Borsu L, Cankovic M, Earle JS, Gocke CD, Hameed M, et al. Recommendations for cell-free DNA assay validations: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagn 2023;25:876–97.
- Moss J, Magenheimer J, Neiman D, Zemmour H, Loyfer N, Korach A. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018;9:5068.
- Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell 2016;164:57–68.
- Volik S, Alcaide M, Morin RD, Collins C. Cell-free DNA (cfDNA): clinical significance and utility in cancer shaped by emerging technologies. Mol Cancer Res 2016;14:898–908.
- Lo YD, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM. Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet 1999;64:218–24.
- Fatouros IG, Jamurtas AZ, Nikolaidis MG, Destouni A, Michailidis Y, Vrettou C, et al. Time of sampling is crucial for measurement of cell-free plasma DNA following acute aseptic inflammation induced by exercise. Clin Biochem 2010;43:1368–70.
- Chen E, Cario CL, Leong L, Lopez K, Márquez CP, Chu C, et al. Cell-free DNA concentration and fragment size as a biomarker for prostate cancer. Sci Rep 2021;11:5040.
- Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch R-D, Knippers R. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 2001;61:1659–65.
- Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2010;2:61ra91.
- Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature 2019;570:385–9.
- Sanchez C, Roch B, Mazard T, Blache P, Dache ZAA, Pastor B, et al. Circulating nuclear DNA structural features, origins, and complete size profile revealed by fragmentomics. JCI Insight 2021;6:e144561.
- Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med 2018;10:eaat4921.
- Leal A, van Grieken NC, Palsgrove DN, Phallen J, Medina JE, Hruban C, et al. White blood cell and cell-free DNA analyses for detection of residual disease in gastric cancer. Nat Commun 2020;11:525.
- Lam WKJ, Jiang P, Chan KCA, Cheng SH, Zhang H, Peng W, et al. Sequencing-based counting and size profiling of plasma Epstein-Barr virus DNA enhance population screening of nasopharyngeal carcinoma. Proc Natl Acad Sci U S A 2018;115:E5115–24.
- Cheng J, Morselli M, Huang W-L, Heo YJ, Pinheiro-Ferreira T, Li F, et al. Plasma contains ultrashort single-stranded DNA in addition to nucleosomal cell-free DNA. iScience 2022;25:104554.
- Cheng LY, Dai P, Wu LR, Patel AA, Zhang DY. Direct capture and sequencing reveal ultra-short single-stranded DNA in biofluids. iScience 2022;25:105046.
- Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 2018;359:926–30.
- Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med 2017;9:eaan2415.
- Spina V, Bruscajgin A, Cuccaro A, Martini M, Di Trani M, Forestieri G, et al. Circulating tumor DNA reveals genetics, clonal evolution and residual disease in classical Hodgkin lymphoma. Blood 2018;131:2413–25.
- Rossi D, Diop F, Spaccarotella E, Monti S, Zanni M, Rasi S, et al. Diffuse large B-cell lymphoma genotyping on the liquid biopsy. Blood 2017;129:1947–57.
- Parpart-Li S, Bartlett B, Popoli M, Adleff V, Tucker L, Steinberg R, et al. The effect of

- preservative and temperature on the analysis of circulating tumor DNA. *Clin Cancer Res* 2017;23:24717.
25. Markus H, Contente-Cuomo T, Farooq M, Liang WS, Borad MJ, Sivakumar S, et al. Evaluation of pre-analytical factors affecting plasma DNA analysis. *Sci Rep* 2018;8:7375.
26. Devereaux KA, Souers RJ, Merker JD, Lindeman NI, Graham RP, Hameed MR, et al. Clinical testing for tumor cell-free DNA: College of American Pathologists proficiency programs reveal practice trends. *Arch Pathol Lab Med* 2023;147:425–33.
27. van der Leest P, Boonstra PA, Ter Elst A, van Kempen LC, Tibbesma M, Koopmans J, et al. Comparison of circulating cell-free DNA extraction methods for downstream analysis in cancer patients. *Cancers (Basel)* 2020;12:1222.
28. Bronkhorst AJ, Ungerer V, Holdenrieder S. Comparison of methods for the isolation of cell-free DNA from cell culture supernatant. *Tumour Biol* 2020;42:1010428320916314.
29. Cohen JD, Douville C, Dudley JC, Mog BJ, Popoli M, Ptak J, et al. Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. *Nat Biotechnol* 2021;39:1220–7.
30. Deveson IW, Gong B, Lai K, LoCoco JS, Richmond TA, Schageman J, et al. Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. *Nat Biotechnol* 2021;39:1115–28.
31. Roschewski M, Dunleavy K, Pittaluga S, Moorhead M, Pepin F, Kong K, et al. Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. *Lancet Oncol* 2015;16:541–9.
32. Frank MJ, Hossain NM, Bukhari A, Dean E, Spiegel JY, Claire GK, et al. Monitoring of circulating tumor DNA improves early relapse detection after axicabtagene ciloleucel infusion in large B-cell lymphoma: results of a prospective multi-institutional trial. *J Clin Oncol* 2021;39:3034–43.
33. Lakhotia R, Melani C, Dunleavy K, Pittaluga S, Saba N, Lindenberg L, et al. Circulating tumor DNA predicts therapeutic outcome in mantle cell lymphoma. *Blood Adv* 2022;6:2667–80.
34. Kurtz DM, Green MR, Bratman SV, Scherer F, Liu CL, Kunder CA, et al. Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. *Blood* 2015;125:3679–87.
35. Wagner-Johnston ND, Lensing S, Noy A, Ratner L, Henry D, Lee JY, et al. High frequency of identical clonal immunoglobulin DNA in pre-treatment tumor and plasma from untreated patients with HIV-associated lymphoma: prospective multicenter trial of the AIDS malignancies consortium (AMC 064). *Leuk Lymphoma* 2017;58:2939–42.
36. Vogt SL, Patel M, Lakha A, Philip V, Omar T, Ashmore P, et al. Feasibility of cell-free DNA collection and clonal immunoglobulin sequencing in South African patients with HIV-associated lymphoma. *JCO Glob Oncol* 2021;7:611–21.
37. Halper-Stromberg E, McCall CM, Haley LM, Lin M-T, Vogt S, Gocke CD, et al. CloneRetriever: an automated algorithm to identify clonal B and T cell gene rearrangements by next-generation sequencing for the diagnosis of lymphoid malignancies. *Clin Chem* 2021;67:1524–33.
38. Simpson E. Measurement of diversity. *Nature* 1949;163:688.
39. Li F, Wei F, Huang W-L, Lin C-C, Li L, Shen MM, et al. Ultra-short circulating tumor DNA (usctDNA) in plasma and saliva of non-small cell lung cancer (NSCLC) patients. *Cancers (Basel)* 2020;12:2041.
40. Martin-Alonso C, Tabrizi S, Xiong K, Blewett T, Sridhar S, Crnjac A, et al. Priming agents transiently reduce the clearance of cell-free DNA to improve liquid biopsies. *Science* 2024;383:eadf2341.