BIOO SCIENTIFIC®

Optimized Library Prep for Cell-Free DNA from Human Plasma

Authors: Jan Risinger, Josh Kinman, Marianna Goldrick

INTRODUCTION

Researchers and clinicians are increasingly interested in using Next Generation Sequencing (NGS) analysis of cell-free DNA (cfDNA) found in plasma (the cell-free fraction of anticoagulated blood) for biomarker discovery and diagnostic applications. Two areas of specific interest are non-invasive prenatal diagnostics (1-6) and monitoring efficacy of treatment in cancer patients (7-20). Many recent studies have shown the feasibility of detecting fetal aneuploidies such as Chromosome 21 trisomy (the cause of Down's Syndrome) by shotgun sequencing of DNA-Seq libraries produced from cell-free DNA isolated from maternal blood. Other aneuploidies including trisomies of Chromosomes 13 and 18 have also been detected (3). This approach is attractive since it avoids the risk of miscarriage associated with invasive tests (amniocentesis and chorionic villi sampling) and may offer cost benefits for prenatal diagnosis. Also, NGS-based assessment has the potential to uncover defects arising from more subtle genetic alterations including point mutations, insertion/deletion mutations, translocations, etc. Although intact fetal cells can be detected in maternal blood, they are vastly outnumbered by the mother's blood cells, and the proportional concentration of informative fetal DNA sequences has been reported to be higher in cell-free circulating DNA in plasma compared to DNA recovered from whole blood. The same findings hold true in the case of cancer diagnostics, where the genetic signal from rare circulating tumor cells is harder to discern against the background of non-malignant cells in whole blood, compared to circulating cell-free DNA. Assessing cancer-related genetic alterations in cell-free DNA can also avoid positional bias inherent in direct sampling of tumors, where the spectrum of mutations observed can differ for different biopsy locations within the malignant tissue. The concept of "liquid biopsy" refers to using readily obtainable body fluids, primarily blood and blood fractions, as surrogate tissue for monitoring levels of malignancy-associated mutations that arise from tumor cells and cell-free DNA shed from tumors (18 - 20). Monitoring levels of cancer-associated genetic alterations in cell-free DNA in blood plasma is a promising new approach to assess the benefits of chemotherapy and other types of cancer treatment. Liquid biopsy of circulating cell-free DNA also has potential for early detection of cancer, and for stratifying cancer patients for treatment decisions. To realize the full potential of these promising opportunities, robust and standardized methods are needed for creating DNA-Seq libraries from cell-free DNA extracted from plasma. To meet this need, Bioo Scientific has developed the NEXTflex™ Cell Free DNA-Seq Kit designed for making NGS libraries from low-input samples. Below, we report results demonstrating the use of this kit to produce high-quality informative libraries from cell-free DNA extracted from plasma.

MAKING CELL FREE DNA-SEQ LIBRARIES

To assess performance of the NEXTflex™ Cell Free DNA-Seq Kit, plasma was obtained from a commercial source (HemaCare Corp) and from healthy consented volunteers. DNA was extracted using a magnetic bead-based protocol from 0.6 mL plasma samples. The cfDNA was eluted from the magnetic beads in 50 µL volume, and 3 – 4 µL was used to verify cfDNA recovery. Then 32 µL of the recovered cfDNA was used as input to make barcoded DNA-Seq libraries using the Bioo Scientific NEXTflex Cell Free DNA-Seq Kit. Libraries were amplified using various numbers of PCR cycles, and analyzed on agarose gels and Agilent BioAnalyzer high-sensitivity DNA chips. Since cell-free plasma DNA is naturally fragmented, no further fragmentation was needed prior to library construction. In some cases, the cell-free DNA was enriched for small fragments (reflecting the smaller size of fetal DNA) before library construction. Control libraries were produced using 1 ng of sheared genomic DNA and analyzed in conjunction with the cell-free DNA libraries. Select libraries were sequenced on an Illumina NextSeq instrument, as 2 x 150 paired-end reads. The entire process of extracting cfDNA and creating libraries from multiple plasma samples can be easily completed in a single day.

RESULTS

A. Assessment of Libraries Prior to Sequencing

12 µL (~35%) of each library were run on the gel. Outer lanes are 100 bp ladder run as molecular weight marker. Libraries were made using 15 cycles of PCR. Lanes 1 – 4 show libraries from Donors 1 – 4. Donor 2 was from a woman in third trimester of pregnancy. Sample in lane 5 was made from DNA recovered from re-elution of the magnetic beads from Donor 2, in 100 µL additional Elution Solution (optional Step 11 in kit protocol). Lane 6 shows a library made from 1 ng of sheared human genomic DNA. Note the distinct size distribution of libraries made from cell-free DNA, which reflects its origin from apoptotic cells.

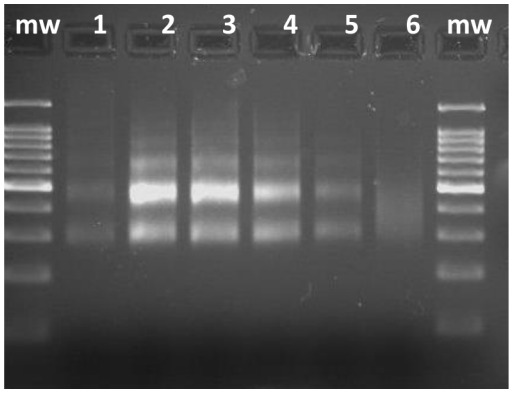


Figure 1. Assessment of library products by ethidium bromide staining.

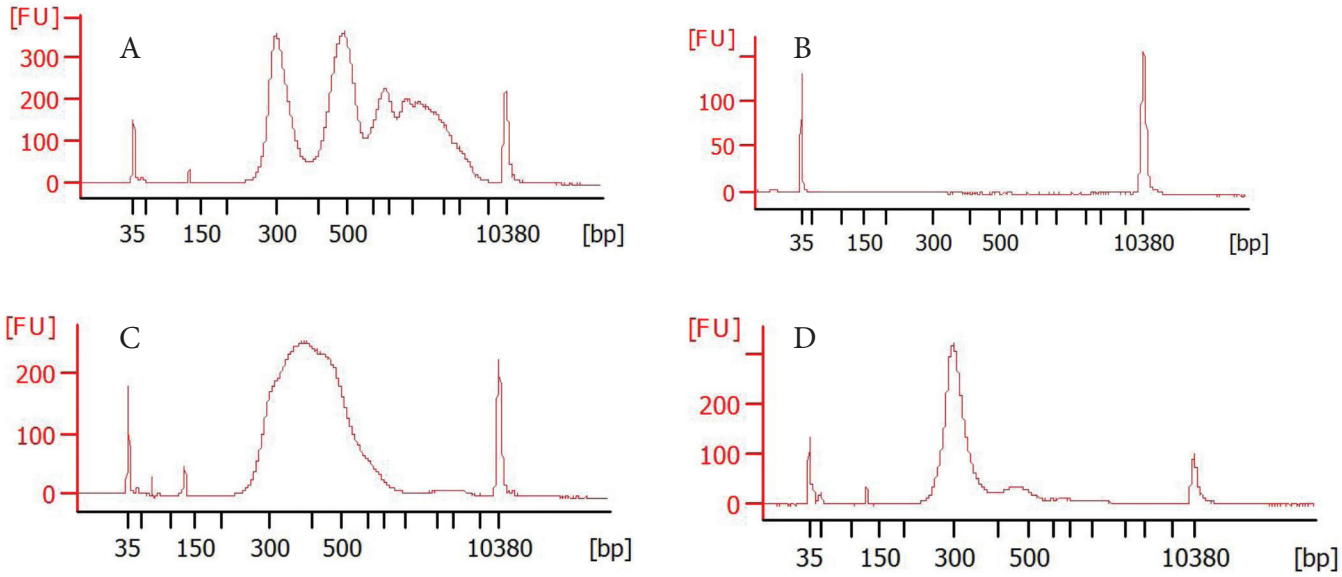


Figure 2. Assessment of libraries on Agilent BioAnalyzer High Sensitivity DNA Chip

Panel A: Library from 32 µL (64% of prep) of cell-free DNA from a male donor amplified for 15 cycles. Note the broad size distribution, which reflects the discrete sizes of cell-free DNA fragments. The cell-free DNA was not fragmented prior to use.
Panel B: Analysis of the corresponding input cell-free DNA used to make the library shown in Panel A. Note, the concentration of cell-free DNA is too low to be detected, which is typical.
Panel C: Library made from 1 ng of sheared human genomic DNA, amplified for 15 cycles. Note the much different size distribution compared to the library made from cell-free DNA.
Panel D: Library made from cell-free DNA size-selected prior to library construction, using Ampure magnetic beads to enrich for small cell-free DNA. Library was amplified for 12 cycles. Sample was from a male donor. After adapter ligation, the desired library products are approximately 300 bp.

B. NGS Results

Two cell-free plasma DNA libraries, one derived from a healthy male donor and one from a healthy third-trimester pregnant woman carrying a male fetus, were submitted for paired end 150 base shotgun sequencing to the core sequencing facility at University of Texas, Austin. Both libraries were from total cell-free DNA (not size selected) and were amplified for either 12 cycles (Library 14) or 13 cycles (Library 22). The libraries were analyzed at the sequencing facility to verify that they were of adequate concentration and size distribution prior to sequencing, and both libraries passed these quality metrics. The QC results for Library 22 are shown below. Each library was successfully sequenced in two separate multiplexed sequencing runs, on the Illumina HiSeq and on the Illumina NextSeq.

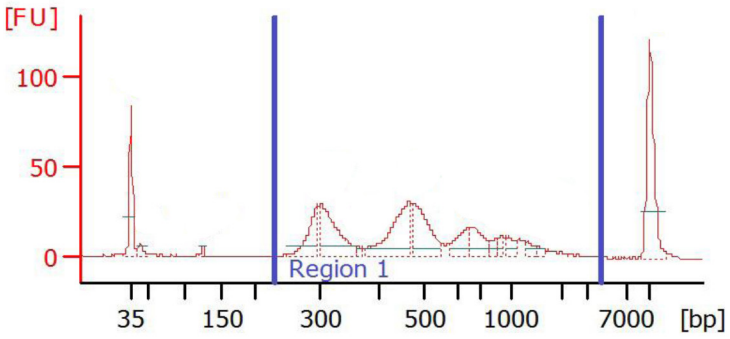
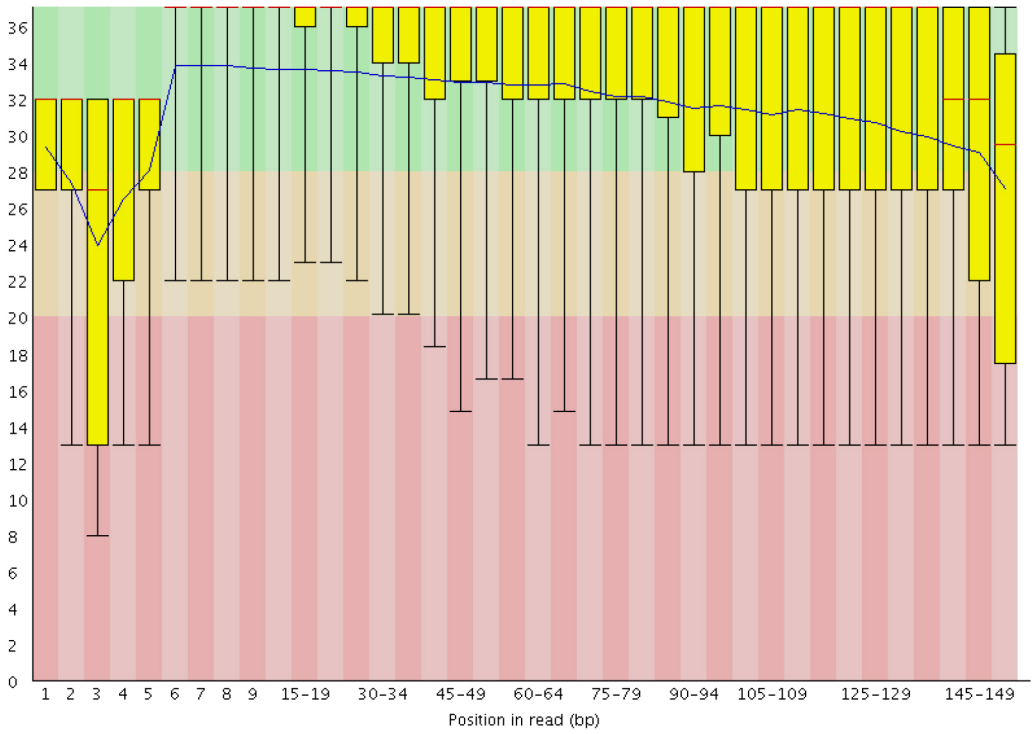


Figure 3. Pre-sequencing QC results for sample Library 22

Number of peaks found: 14 Noise: 0.3 Corr. Area 1: 600.0
Region table for sample 1 From [bp] To [bp] 234 - 4,442 Corr. Area: 600.0 % of Total: 94
Average Size [bp]: 615 Size distribution in CV [%]: 72 Conc. [pg/µl]: 519.50 Molarity [pmol/l]: 1,811.7

Several key metrics of library quality, shown in Table 1, are the percentage of reads that are due to adapter contamination, the % of reads that are attributable to PCR duplicates, and the % of reads that map to the human genome and to each chromosome. The libraries had low levels (1.1% and 4.9%) of uninformative reads representing adapter dimers (unwanted side products consisting of 5' adapter ligated to 3' adapter). The proportion of reads mapping to the human genome was quite high (close to 100%). The percentage of reads attributed to PCR duplicates (defined as reads with identical sequences at the 5' adapter and 3' adapter junctions) was higher in Library 22 (58%) than in Library 14 (16%), possibly reflecting the greater number of PCR cycles used to produce Library 22. Even so, after filtering PCR duplicates, Library 22 still had > 12 million reads, which was sufficient for analysis, as over 97% were mapped to the genome.

The metrics shown in the graphic below for Library 14 depict extremely high-quality sequencing data, as shown by the high Phred scores (Y-axis). The trend line shows average scores above 30, corresponding to >99.9% probability of accuracy, for the average of all reads in the position ranges corresponding to cfDNA sequence (as depicted along the X-axis), with the exception of the longest reads (positions 145-150), which have slightly lower scores (but which are still above 99% probability of being accurate). Results for Library 22 were similar.



% Duplication = (Reads after trimming adapter - Reads after filtering duplicates)/Reads after trimming adapter
% Mapping = (Reads mapping to genome/Reads after filtering duplicates)

Figure 4. Quality scores across all bases. (Sanger/Illumina 1.9 encoding)

Library	Total Reads	Reads after trimming adapter	% reads mapping to adapter	Reads after filtering duplicates	% duplication	Reads mapping to genome	% mapping to genome
Library 14	29,375,523	29,051,363	1.1%	24,338,715	16.22%	24020720	98.69%
Library 22	31,725,510	30,179,252	4.9%	12,651,028	58.08%	12321279	97.39%

Table 1. Low % of adapter contamination and high % of reads mapping to human genome

The results of mapping the reads from each library to each human chromosome are shown in Table 3. There is generally very good agreement between the observed proportion of the total reads that map to each chromosome and the expected proportion based on the size of each chromosome. Of particular interest are the reads mapping to the Y chromosome in Library 14, which was made from cfDNA from a woman carrying a late gestational age (3rd trimester) male fetus; reads in Library 14 that map to Y chromosome are derived from fetal DNA. The number of reads mapped to the Y chromosome in Library 14 (319,835 reads) represent 0.66% of the total number of reads (48,387,865). Since the total size of the human genome is 3,095,677,412 bp and the total size of the Y chromosome is 59,373,566 bp, the Y chromosome represents 1.92 % of the human genome. This value is in good agreement with the observed % of reads mapping to the Y chromosome (1.79 %) in Library 14, which was made from cfDNA from a male donor. Extrapolating from these values, we calculate that 34% of the cfDNA in Library 14 is derived from fetal DNA.

Library 14 (3rd trimester pregnant female)

	Size Chr bp; % of genome	# reads; % reads
chr1	249250621; 8.0%	3957289; 8.2%
chr2	243199373; 7.9%	4174405; 8.6%
chr3	198022430; 6.4%	3097157; 6.4%
chr4	191154276; 6.2%	3104733; 6.4%
chr5	180915260; 5.8%	2819347; 5.8%
chr6	171115067; 5.5%	2669787; 5.5%
chr7	159138663; 5.1%	2590026; 5.3%
chr8	146364022; 4.7%	2347256; 4.9%
chr9	141213431; 4.6%	2090387; 4.3%
chr10	135534747; 4.4%	2547947; 5.3%
chr11	135006516; 4.4%	2237365; 4.6%
chr12	133851895; 4.3%	2137615; 4.4%
chr13	115169878; 3.7%	1425445; 2.9%
chr14	107349540; 3.5%	1447501; 3.0%
chr15	102531392; 3.3%	1413026; 2.9%
chr16	90354753; 2.9%	1594624; 3.3%
chr17	81195210; 2.6%	1441168; 3.0%
chr18	78077248; 2.5%	1319911; 2.7%
chr19	59128983; 1.9%	1027852; 2.1%
chr20	63025520; 2.0%	1086005; 2.2%
chr21	48129895; 1.6%	686368; 1.4%
chr22	51304566; 1.6%	660874; 1.4%
chrX	155270560; 5%	2191942; 4.5%
chrY	59373566; 1.9%	319835 (0.66%; fetal DNA)
Total:	3095677412	48387865

Library 22 (Hispanic male)

	Size Chr bp; % of genome	# reads; % reads
chr1	249250621;8.0%	2037608; 8.2%
chr2	243199373; 7.9%	2197291; 8.9%
chr3	198022430; 6.4%	1495105; 6.0%
chr4	191154276; 6.2%	1694858; 6.8%
chr5	180915260; 5.8%	1367047; 5.5%
chr6	171115067; 5.5%	1289128; 5.2%
chr7	159138663; 5.1%	1306538; 5.3%
chr8	146364022; 4.7%	1166654; 4.7%
chr9	141213431; 4.6%	1076104; 4.3%
chr10	135534747; 4.4%	1448807; 5.8%
chr11	135006516; 4.4%	1120768; 4.5%
chr12	133851895; 4.3%	1063070; 4.3%
chr13	115169878; 3.7%	671893; 2.7%
chr14	107349540; 3.5%	711182; 2.9%
chr15	102531392; 3.3%	726261; 2.9%
chr16	90354753; 2.9%	883902; 3.6%
chr17	81195210; 2.6%	8072023; 3.0%
chr18	78077248; 2.5%	653564; 2.6%
chr19	59128983; 1.9%	592323; 2.4%
chr20	63025520 2.0%	578161; 2.3%
chr21	48129895 1.6%	376356; 1.5%
chr22	51304566; 1.6%	378480; 1.5%
chrX	155270560; 5%	655529; 2.6%
chrY	59373566; 1.9%	442239; 1.8%
Total:	3095677412	24740070

Table 2. Reads mapping to each chromosome

DISCUSSION/CONCLUSIONS

Results presented herein demonstrate the performance of the NEXTflex Cell Free DNA-Seq Kit for processing cell-free DNA for NGS applications. Recovery of a high percentage of fetal DNA in the maternal plasma sample was shown by the analysis of the proportion of reads mapping to the Y chromosome in the library derived from late-gestational stage plasma; these reads represent DNA from the male fetus that is present in the maternal circulation. The value we report for % fetal DNA (34%) is on the upper end of values reported in other studies. Some of these reports were based on analyses using different techniques, but one study that used NGS reported a similar high level of fetal DNA (40%) (Fan et al 2010). The authors speculated that the relatively high % of fetal DNA in their studies was due to the higher efficiency of amplification of the shorter fetal cfDNA fragments, compared to the longer maternal cfDNA fragments, during the library amplification step. For non-invasive prenatal diagnostics, magnetic bead-based size selection is sometimes used prior to library construction to enrich for the shorter fragments found in maternal plasma. In this case, the size distribution of the predominant library product reflects that expected for fetal DNA. The distribution of the library products we observed using unfractionated cfDNA shows several peaks, which probably reflects the fact that most cfDNA in healthy donors is thought to originate from the "ladder" of DNA fragments derived from apoptotic cells. It is interesting that a significant fraction of the library products in Library 22, from a healthy male donor, are quite large, ranging up to several kilobases. The proportion of larger library products may be increased in plasma from cancer patients.

The NEXTflex Cell Free DNA-Seq Kit is optimized for library construction from low input amounts of ctDNA or cfDNA isolated from cell free fluids. The NEXTflex Cell Free DNA-Seq Kit can be used to construct Illumina-compatible libraries from 1 ng of DNA in about two hours. This kit delivers high coverage and reduced bias, along with flexible multiplexing options. The high quality of the DNA-Seq libraries produced using this kit is demonstrated by several metrics of sequencing quality including a high % of reads mapping to the human reference genome, low levels of adapter contamination, Phred scores indicative of extremely accurate base-calling, and excellent correspondence between observed and expected % of reads mapping to each human chromosome.

References

1. 1. Chan KCA, Zhang J, Hui ABY, Wong N, Lau TK, Leung TN, et al. Size distributions of maternal and fetal DNA in maternal plasma. Clin Chem 2004;50:88–92.
2. Dennis Lo YM, Chiu RW. Prenatal diagnosis: progress through plasma nucleic acids. Nat Rev Genet 8:71–7 (2007).
3. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Nat Acad Sci U S A;105:16266–71(2008).
4. Chiu RW, Chan KC, Gao Y, Lau VY, Zheng W,Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Nat Acad Sci U S A 105: 20458–63 (2008).
5. Chiu RW, Sun H, Akolekar R, Clouser C, Lee C, McKernan K, et al. Maternal plasma DNA analysis with massively parallel sequencing by ligation for noninvasive prenatal diagnosis of trisomy 21. Clin Chem 56:459–63 (2010).
6. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, and Quake SR. Analysis of the Size Distributions of Fetal and Maternal Cell-Free DNA by Paired-End Sequencing Clinical Chemistry 56:8 (2010)
7. Wang BG, Huang H-Y, Chen Y-C, Bristow RE, Kassauei K, Cheng C-C, Roden R, Sokoll LJ, Chan DW, and Shih I-M. Increased Plasma DNA Integrity in Cancer Patients. CANCER RESEARCH 63, 3966–3968, July 15, 2003
8. Gautschi, O. et al. Origin and prognostic value of circulating KRAS mutations in lung cancer patients. Cancer Lett. 254, 265–273 (2007).
9. Kuang, Y. et al. Noninvasive detection of EGFR T790M in gefitinib or erlotinib resistant non-small cell lung cancer. Clin. Cancer Res. 15, 2630–2636 (2009).
10. Leary, R.J. et al. Development of personalized tumor biomarkers using massively parallel sequencing. Sci. Transl. Med. 2, 20ra14 (2010).
11. McBride, D.J. et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. Genes Chromosom. Cancer 49, 1062–1069 (2010).
12. Taniguchi, K. et al. Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. Clin. Cancer Res. 17, 7808–7815 (2011).
13. Forshew, T. et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci. Transl. Med. 4, 136ra168 (2012).
14. Leary, R.J. et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci. Transl. Med. 4, 162ra154 (2012).
15. Dawson, S.J. et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N. Engl. J. Med. 368, 1199–1209 (2013).
16. Diehl F, and Smergeliene E. BEAMing for Cancer: Detecting Tumor Mutations in Peripheral Blood Using Digital PCR. Genetic Engineering and Biotechnology News Vol 33 No 15 (2013).
17. Murtaza, M. et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 497, 108–112 (2013).
18. Crowley, E., Di Nicolantonio, F., Loupakis, F. & Bardelli, A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat. Rev. Clin. Oncol. 10, 472–484 (2013).
19. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE, Shrager JB, Loo BW Jr, Alizadeh AA, and Diehn M. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nature Medicine, Received 15 August 2013; accepted 6 November 2013; published online 6 April 2014; doi:10.1038/nm.3519
20. Diaz LA and Bardelli A. Liquid Biopsies: Genotyping Circulating Tumor DNA. Journal of Clinical Oncology/American Society of Clinical Oncology 32:6 Feb 20 2014.

Bioo Scientific Corporation
7050 Burlseon Road
Austin, Texas 78744, USA

www.biooscientific.com
info@biooscientific.com
1.512.707.8993