Next-Generation Sequencing of Plasma/Serum DNA: An Emerging Research and Molecular Diagnostic Tool

Y.M. Dennis Lo* and Rossa W.K. Chiu

Over the past decade, it has been increasingly realized that cell-free DNA and RNA molecules present in plasma and serum are valuable molecular diagnostic tools (1, 2). For example, tumor-derived (3) and fetal-derived (4) nucleic acids have been found in cancer patients and pregnant women, respectively, thus opening up clinical uses in oncology and prenatal diagnosis. A number of these applications have already been incorporated into clinical practice, such as the prenatal determination of fetal RhD status (5) and the detection and monitoring of nasopharyngeal carcinoma using plasma Epstein-Barr virus DNA measurement (6).

In contrast to its clinical applications, the characterization of circulating nucleic acids has not received as much attention. In this regard, DNA sequencing is a powerful method to address this imbalance. A number of groups have used conventional cloning and DNA sequencing techniques to detect and study circulating nucleic acids (7–9). However, such methods are labor-intensive and can generate sequence information for only a small number of molecules from a few genomic loci. The recent advent of next-generation, massively parallel sequencing technologies (10) has provided an alternative approach for the detection, measurement, and characterization of plasma nucleic acids.

Three groups have reported the use of massively parallel sequencing technologies in the analysis of plasma/serum DNA (11–13). These approaches can be divided into 2 main types: the first type involves the random sequencing of DNA molecules in plasma/serum (11, 12), and the second type involves the deep sequencing of a selected (i.e., nonrandom) subset of circulating DNA molecules (13). As an illustration of the power of the random sequencing approach, Chiu et al. (11) and Fan et al. (12) used the Illumina/Solexa platform to randomly sequence a short (25–36 bp) tag on millions of DNA molecules obtained from the plasma of pregnant women. The short tags allowed such molecules to be mapped back to the reference human genome. The relative representation of each chromosome in plasma could then be calculated. These groups showed that in women carrying fetuses affected by a chromosomal aneuploidy (e.g., trisomy 21) the proportional representation of the affected chromosome (e.g., chromosome 21) would be increased in maternal plasma. This approach thus provides a basis for the development of noninvasive prenatal tests for fetal chromosomal aneuploidies. Korshunova et al. (13), on the other hand, illustrated the use of deep sequencing of selected, nonrandom genomic loci by performing massively parallel bisulfite sequencing of 4 loci using the Roche/454 platform in sera obtained from cancer patients and control subjects. They found that DNA molecules containing every conceivable cytosine-methylation pattern could be found in the sera of both the cancer patients and cancer-free controls, thus highlighting the challenge for developing highly specific serum DNA methylation markers for cancer.

In this issue of Clinical Chemistry, Beck et al. (14) provide yet another illustration of the random sequencing approach by applying the Roche/454 platform to the massively parallel sequencing of serum DNA from a cohort of apparently healthy individuals. Beck et al. obtained a total of 450 000 sequences from 50 apparently healthy individuals. Although not explicitly stated, it appears that the authors had pooled the samples and then subjected the pooled samples to sequencing, obtaining an average of 9100 sequence reads per sample. The authors found that most classes of sequences that they analyzed (e.g., genes and RNA and DNA coding sequences) did not appear to differ between serum DNA and genomic DNA. On the other hand, they found evidence of overrepresentation of Alu sequences in serum DNA. It has been well established that serum contains a higher concentration of DNA than plasma, possibly because of DNA released from the blood cells during the clotting process (15). Thus, one can perhaps argue that it might have been more reflective of the in vivo situation in the circulation if the authors had studied plasma rather than serum DNA, using massively parallel sequencing.

It is particularly interesting that the authors found hepatitis B virus DNA sequences in one serum sample,
thus demonstrating that the approach might have future application in infectious disease research. The authors did not provide data to show whether the hepatitis B virus sequences obtained using the massively parallel sequencing approach fit with those obtained using conventional cloning and sequencing approach from the same case. It is also intriguing that the authors reported that some 0.16% of the sequences might be of possible bacterial origin. It would be of interest to see a more detailed analysis of the types of bacteria that might have contributed to this phenomenon and its potential clinical significance.

An increasing number of platforms are available for massively parallel sequencing (10). For the quantitative random sequencing of plasma/serum DNA, such as that discussed above for the noninvasive prenatal diagnosis of fetal chromosomal aneuploidies (11, 12), one could perhaps argue that platforms that provide short reads but with a higher throughput in terms of the number of molecules analyzed might be more efficient and cost effective than platforms that provide longer reads but with a lower throughput in terms of the number of molecules sequenced. In such applications, as long as the read length is sufficient for mapping back to the genome (e.g., 25–36 bp) or another reference set of sequences, the task would be adequately performed. On the other hand, for applications that are focused on the characterization of plasma/serum DNA, e.g., identification of novel pathogens or mutation detection, platforms with a longer read length might have advantages over platforms with shorter read lengths. With the rapid increases in read lengths and throughput of the various platforms, the gaps between these systems for plasma/serum nucleic acid sequencing might become narrower in the near future.

It will probably still be a number of years before the diagnostic applications using massively parallel sequencing become commonplace, however. Currently, the equipment and reagents are still relatively expensive. Furthermore, the bioinformatic support that is needed to analyze the data is immense and out of reach for most diagnostic laboratories at the present time. Nonetheless, with additional technical advances and cost reduction in the coming years, it is highly likely that massively parallel sequencing approaches will eventually become a routine tool in laboratory medicine.

**Author Contributions:** All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (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; and (c) final approval of the published article.

**Authors’ Disclosures of Potential Conflicts of Interest:** Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

**Employment or Leadership:** None declared.

**Consultant or Advisory Role:** Y.M.D. Lo, Sequenom.

**Stock Ownership:** Y.M.D. Lo, Sequenom.

**Honoraria:** None declared.

**Research Funding:** Y.M.D. Lo, Sequenom.

**Expert Testimony:** None declared.

**Other:** Y.M.D. Lo holds patents or patent applications on aspects of the use of plasma DNA/RNA for molecular diagnosis.

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

**References**


