Plasma nucleic acid analysis by massively parallel sequencing: pathological insights and diagnostic implications

YM Dennis Lo1,2* and Rossa WK Chiu1,2

1 Li Ka Shing Institute of Health Sciences and
2 Department of Chemical Pathology, Chinese University of Hong Kong, Hong Kong SAR, China

*Correspondence to: YM Dennis Lo, Department of Chemical Pathology, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, New Territories, Hong Kong SAR, China. e-mail: loym@cuhk.edu.hk

Abstract

Over the past 15 years there has been increasing interest in the biology and diagnostic applications of circulating DNA in the plasma of human subjects. In particular, DNA from a fetus, a tumour, a transplanted organ and injured tissues has been found in the plasma of pregnant women, cancer patients, transplant recipients and patients suffering from multiple pathologies, respectively. The advent of massively parallel sequencing has given us a quantitative and powerful tool for studying circulating DNA on a genome-wide level. Using this approach, fetal chromosomal aneuploidies can be robustly detected using maternal plasma. Furthermore, a genome-wide genetic map of a fetus can also be constructed using this approach. This method has also allowed one to identify tumour-associated chromosomal translocations, which can then be detected in plasma. The direct application of massively parallel sequencing to the serum of cancer patients has also allowed quantitative aberrations that are associated with malignancy to be detected in serum. The use of massively parallel sequencing on the plasma of transplantation recipients has opened up an approach for detecting rejection. The application of circulating DNA sequencing has also opened up a new method for elucidating the quantitative aberration of circulating DNA in many pathological conditions. Such developments would provide new modalities for molecular diagnostics and would improve our understanding of the biology of circulating nucleic acids.

Copyright © 2011 Pathological Society of Great Britain and Ireland. Published by John Wiley & Sons, Ltd.
opened up new possibilities for non-invasive prenatal diagnosis. Early clinical applications have focused on the detection of fetal-specific targets that are paternally inherited and that are not present in the pregnant mother’s genome. Examples of such targets include markers present on the Y chromosome for detecting a male fetus [9,10] and the *RHD* gene for detecting a RhD-positive fetus carried by a RhD-negative pregnant mother [11,12]. However, the use of maternal plasma DNA for detecting the presence of fetal chromosomal aneuploidies, e.g. trisomy 21, is much more technically challenging. For illustration, if fetal DNA represents 10% of the DNA that is present in a particular maternal plasma sample, then for one to detect the presence of a trisomy 21 fetus, the analytical method has to be able to discriminate a 5% difference [13]. In 2007, Lo et al [13] outlined the principle whereby such a precision could be reached by approaches that allow single DNA molecules to be counted. This group demonstrated the feasibility of the concept using digital PCR [14]. However, the performance of tens of thousands or >100,000 digital PCRs for each case is extremely labour-intensive.

With the development of massively parallel sequencing, a much more efficient way of analysing millions of DNA molecules has become available [8] (Figure 1). In 2008, two groups independently showed that the massively parallel sequencing of maternal plasma DNA would allow a fetus with trisomy 21 to be detected with high accuracy [15,16]. These results have recently been validated in two large-scale studies involving 753 samples [17] and 449 samples [18]. Both Chiu et al and Ehrich et al applied massively parallel sequencing to the plasma of pregnant women who had previously been classified as high risk by conventional screening modalities. Chiu et al [17] reported that fetal trisomy 21 could be detected with a sensitivity of 100% and a specificity of 98%. The data by Ehrich et al [18] were consistent with those of Chiu et al and showed a sensitivity of 100% and a specificity of 99.7%. The precision of massively parallel sequencing for the measurement of the dosage of chromosomes 13 and 18, on the other hand, is lower [16,19]. Special bioinformatics procedures have been demonstrated to improve this precision [20,21]. With further validation in large cohorts, it is likely that massively parallel sequencing of maternal plasma DNA would be used clinically in the near future for the screening of the common fetal trisomies.

Recent data have indicated that the massively parallel sequencing of maternal plasma DNA has diagnostic applications outside of trisomy detection. Lo et al [22] have shown that, by combining the maternal plasma DNA sequencing data with genotype information of the father and haplotype information of the mother, one could deduce a genome-wide genetic map of an unborn fetus (Figure 1). These investigators have demonstrated the diagnostic utility of this approach for the prenatal diagnosis of β-thalassaemia. Indeed, in theory, virtually all monogenic diseases could be investigated prenatally using this approach. However, before the technology can be widely used, a number of issues would need to be explored. First, as reported by Lo et al [22], the technology for such whole-genome fetal genetic scanning is expensive. Second, the requirement of maternal haplotype information would potentially reduce the ease with which this technology could be used, as such information is typically generated through pedigree analysis. Third, the social and ethical implications of prenatal fetal whole-genome scanning would need to be discussed by all stakeholders.

A number of the above concerns are already being addressed. With regard to the cost issue, it is unlikely that clinical needs would routinely require the entire fetal genome to be scanned prenatally. A probable scenario would be for the technology to be used for selected genetic diseases prevalent in a particular population. In this regard, Liao et al [23] have recently...
They found that the proportions of different classes of serum of a number of apparently healthy individuals are generally reminiscent of those in the genome, except that certain classes of repeat sequences (eg the Alu repeats) were over-represented in serum DNA. The observation concerning Alu repeats is reminiscent to that made using real-time PCR [32]. This group then applied this approach to study the serum DNA isolated from patients with breast cancer and control subjects [33]. They found that the serum DNA of patients with breast cancer contained significantly more repetitive elements than that from the control subjects. The molecular basis of these observations is unclear at the present time. One possible explanation might be epigenetic aberrations in the cancer genome, which would then lead to an alteration in nucleosome positioning. Such changes might lead to a differential degradation of DNA derived from cancer and normal cells in serum. In this regard, it is important to note that the studies by Beck et al were based on serum, rather than plasma DNA [31,33]. The choice of serum, rather than plasma, might introduce an additional complexity, as it has previously been shown that blood clotting during serum formation would lead to DNA release into the serum [9]. It is currently unknown whether the clotting process in cancer and healthy subjects might have subtle differences that might lead to aberrations in the composition of serum DNA.

The advantage of the method by Beck et al is that it seems to allow the detection of cancer without needing access to tumour tissues. Other investigators, however, have taken a different route. Two groups have developed highly specific PCR-based assays to detect and measure such tumour-specific translocations in the plasma of patients from whom the tumours were originally obtained. This strategy has been found to allow personalized, highly sensitive and specific monitoring of tumour-derived DNA in the circulation. With the further reduction in the cost of DNA sequencing, this approach is expected to be increasingly accessible for clinical care. The disadvantage of this approach is the fact that tumour tissues are necessary for the initial characterization of the translocations and thus this approach, as currently described, is not usable for cancer screening.

Korshunova et al [36] have explored the use of massively parallel sequencing for detecting tumour-associated DNA methylation changes in the serum of breast cancer patients. Following bisulphite conversion, in which methylated and non-methylated cytosine residues would be unchanged and changed to uracil, respectively, Korshunova et al generated multiple amplicons corresponding to four genomic regions in which their previous work had demonstrated differential methylation in breast cancer tissues. They then sequenced these amplicons using massively parallel sequencing. Using this approach, these authors were able to observe a difference in methylation...
levels in breast cancer and normal tissues. However, they were not able to observe a convincing difference between the methylation levels in serum samples from breast cancer patients and control subjects. The most likely explanation for these relatively disappointing results was that the fractional concentrations of tumour DNA in the serum samples were below the sensitivity of the analytical approach. The use of serum, instead of plasma samples, was likely to be a contributory factor, because DNA release from nucleated blood cells into the serum would be expected to further reduce the fractional tumour DNA in the sample [9]. The future analysis of plasma DNA for methylation changes using much deeper sequencing might yield more useful data for the ‘plasma epigenome’.

**Plasma DNA for transplantation monitoring**

In 1998, it was shown that, following solid organ transplantation, the DNA from the donor could be found in the plasma of the recipient [6]. This phenomenon was first demonstrated in female recipients of organs from male donors [6]. Sequences from the Y chromosome could be observed in 100% of such sex-mismatched liver transplantation recipients and 80% of the kidney transplantation recipients. The ease of detecting such donor-derived sequences appears to be related to the size of the donated organ, with the transplanted liver releasing readily detectable amounts of donor-derived DNA when compared to the situation for kidney and heart transplantation [37]. Furthermore, in sex-mismatched bone marrow transplantation cases, it was found that the sex of the plasma of the recipient was converted to the sex of the donor [38]. This latter observation has suggested that haematopoietic cells are the predominant source of plasma DNA.

As it is widely accepted that DNA is released into plasma when cells die, it has been proposed that the measurement of donor-derived DNA in the plasma of transplantation recipients could be used for monitoring graft rejection [6]. This hypothesis has subsequently been shown to be correct by a number of groups in both humans and animal models [39–41]. Thus, there is an elevation in the concentration of circulating donor-derived DNA during rejection episodes. The use of Y-chromosomal sequences as a marker of the donor has limited such an approach for female recipients with male donors. Several workers have attempted to overcome this disadvantage by using markers from the human leukocyte antigen (HLA) region [39]. While this latter method has expanded the population coverage of this approach for monitoring rejection, it requires the development of multiple quantitative assays.

Recently, it was shown that massively parallel sequencing of the plasma DNA obtained from transplantation recipients might provide a universal solution for this problem [42]. Snyder *et al* used heart transplantation as their model system. They employed a bead-based system for carrying out genome-wide genotyping of single nucleotide polymorphisms (SNPs) of constitutional DNA of the donor and recipient. They then analysed the sequencing data from the recipients’ plasma for reads containing such SNP signatures. They were then able to calculate the fractional concentration of donor-derived DNA in the plasma. They observed that rejection episodes were associated with an elevation in the fractional concentration of donor-derived DNA in plasma.

Future work would be needed to test the stability of the concentration threshold for diagnosing rejection episodes. It is likely that such a threshold would need to be adjusted for different types of transplantation and for different immunosuppressive regimens. As prompt treatment of rejection episodes is needed, future work would also be needed to reduce the turnaround time of such an assay, which currently takes days to complete.

**Plasma DNA for tissue damage**

The relationship between plasma DNA and tissue damage, explored above in the context of graft rejection, can be generalized to other types of tissue damage. Hence, it has been shown that the concentration of plasma DNA would increase in a number of acute medical conditions, eg trauma [7], cardiac ischaemia [43], stroke [44] and sepsis [45]. The measurements of plasma DNA in such applications are typically performed by relatively simple quantitative PCR-based techniques. With the development of massively parallel sequencing, researchers have attempted to explore whether abnormalities in plasma DNA can be seen in a number of chronic disorders. In this regard, it is intriguing that one group of workers have reported the aberration representation of selected coding and repeat sequences in the serum of patients with multiple sclerosis [46]. The biological basis of this observation remains unclear at the moment. One possibility may be related to changes in the methylation status of DNA [47]. In cattle, massively parallel sequencing has identified disease-specific patterns in circulating DNA in animals infected with transmissible spongiform encephalopathies [48]. Thus, it is possible that massively parallel sequencing of circulating DNA would provide us with new tools for the monitoring of chronic neurological disorders and shed light on their pathogenesis.

**Conclusion**

The recent availability of massively parallel sequencing has given us unprecedented precision and ability to analyse DNA in plasma. This development is expected to impact on the future development of molecular diagnostics and to provide us a window on molecular processes that are occurring in tissues that have hitherto
Plasma nucleic acid analysis by massively parallel sequencing

been difficult to sample, eg those from the fetus, tumour, transplanted organ or the nervous system.

Author contributions

Both authors contributed to the planning, writing and revision of this review.

References


