

Nonhematopoietically Derived DNA Is Shorter than Hematopoietically Derived DNA in Plasma: A Transplantation Model

Yama W.L. Zheng,^{1,2†} K.C. Allen Chan,^{1,2†} Hao Sun,^{1,2} Peiyong Jiang,^{1,2} Xiaoxi Su,^{1,2} Eric Z. Chen,^{1,2} Fiona M.F. Lun,^{1,2} Emily C.W. Hung,^{1,2} Vincent Lee,³ John Wong,⁴ Paul B.S. Lai,⁴ Chi-Kong Li,³ Rossa W.K. Chiu,^{1,2} and Y.M. Dennis Lo^{1,2*}

BACKGROUND: Plasma DNA is predominantly hematopoietic in origin. The size difference between maternal and fetal-derived DNA in maternal plasma prompted us to investigate whether there was any discrepancy in molecular size between hematopoietically and nonhematopoietically derived DNA in plasma.

METHODS: Plasma DNA samples from 6 hematopoietic stem cell transplant recipients and 1 liver transplant recipient were analyzed by massively parallel paired-end sequencing. The size of each fragment was deduced from the alignment positions of the paired reads. In sex-mismatched transplant recipients, the reads from chromosome Y were used as markers for the male donor/recipient. For other transplant recipients, the reads of the donor- and recipient-specific alleles were identified from the single-nucleotide polymorphism genotypes.

RESULTS: In male patients receiving female hematopoietic stem cells, more chromosome Y-derived DNA molecules (nonhematopoietically derived) were ≤ 150 bp than the autosome-derived ones (mainly hematopoietically derived) (median difference, 9.9%). In other hematopoietic stem cell transplant recipients, more recipient-specific DNA molecules (nonhematopoietically derived) were ≤ 150 bp than the donor-specific ones (hematopoietically derived) (median difference, 14.8%). In the liver transplant recipient, more donor-derived DNA molecules (liver derived) were ≤ 150 bp than the recipient-derived ones (mainly hematopoietically derived) (difference, 13.4%). The nonhematopoietically derived DNA exhibited a reduction in a 166-bp peak compared with the hematopoietically derived DNA. A 10-bp periodicity in size distribution below approximately 143 bp was observed in both DNA populations.

CONCLUSIONS: Massively parallel sequencing is a powerful tool for studying posttransplantation chimerism. Plasma DNA molecules exhibit a distinct fragmentation pattern, with the nonhematopoietically derived molecules being shorter than the hematopoietically derived ones.

© 2011 American Association for Clinical Chemistry

Circulating nucleic acids isolated from plasma or serum are increasingly being used as biomarkers for prenatal diagnosis (1) and for cancer detection and monitoring (2). Yet, much remains to be learned regarding their cellular origin, release, and clearance mechanisms. Different hypotheses have been proposed concerning the release mechanisms, including DNA release after cell apoptosis, after necrosis (3, 4), and by active liberation (5).

Investigations into the molecular size of circulating DNA have provided valuable information regarding the possible mechanisms of liberation of these molecules. Circulating DNA molecules have been shown to be mainly short fragments of < 200 bp (4, 6), observations consistent with an origin related to DNA fragmentation during apoptosis. The size distributions of circulating DNA are altered in such clinical settings as pregnancy (6) and malignancy (7). In particular, fetal-derived DNA and tumor-specific DNA have been shown to be shorter than maternally derived DNA molecules (6) and DNA originating from nonneoplastic cells (8), respectively. These observations have led to the use of the integrity of circulating DNA for oncology (9, 10) and the development of strategies for enriching fetal DNA in maternal plasma by size selection (11, 12).

¹ Li Ka Shing Institute of Health Sciences and Departments of ² Chemical Pathology, ³ Paediatrics, and ⁴ Surgery, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong.

* Address correspondence to this author at: Department of Chemical Pathology, Prince of Wales Hospital, 30–32 Ngan Shing St., Shatin, New Territories, Hong

Kong SAR, China. Fax +852-2636-5090; e-mail loym@cuhk.edu.hk.

† These authors contributed equally to this work.

Received May 23, 2011; accepted August 12, 2011.

Previously published online at DOI: 10.1373/clinchem.2011.169318

Table 1. Sex and diagnosis of transplantation pairs with cellular and plasma chimerism results.

Case no.	Sex of donor	Sex of recipient	Diagnosis	Chimerism, % ^a	Quantitative contribution in plasma, % ^b		
					Donor (chrY based)	Donor (SNP category 2 based)	Recipient (SNP category 3 based)
HSCT1	F	M	WAS ^c	99.3	86.0	NA	NA
HSCT2	F	M	ALL	99.8	82.8	NA	NA
HSCT3	M	F	ALL	99.8	89.2	NA	NA
HSCT4	M	F	ALL	99.8	75.0	71.5	28.1
HSCT5	M	M	β TM	100	NA	79.9	19.9
HSCT6	M	M	SAA	100	NA	90.7	9.0
LT 1	M	M	Child's C HBV cirrhosis	NA	NA	27.2	72.5

^a Determined by fluorescence in situ hybridization analysis and DNA short tandem repeat analysis.
^b calculated using the % chrY values (only available for the sex-mismatched cases) or the number of donor- and recipient-specific PE reads (available only for cases with genotype data).
^c WAS, Wiskott–Aldrich syndrome; NA, not available; ALL, acute lymphoblastic leukemia; β TM, β -thalassemia; SAA, severe aplastic anemia; HBV, hepatitis B virus.

In contrast to gel electrophoresis or locus-specific PCR assays, which provide only a crude estimate of the amount and size profiles of DNA molecules, massively parallel sequencing of circulating DNA allows us to study these molecules at an unprecedented resolution and in a genome-wide fashion (13). Paired-end (PE)⁵ sequencing of circulating DNA provides size information for each sequenced DNA fragment. Hence, a comprehensive and high-resolution size profile can be obtained. This approach has been used to elucidate the difference in size between fetal and maternal DNA in maternal plasma (13, 14). Recently, PE deep sequencing of maternal plasma DNA has unveiled that plasma DNA molecules possess a fragmentation pattern reminiscent of nuclease-cleaved nucleosomes, with fetal DNA showing a reduction in a 166-bp peak and relative prominence of a 143-bp peak, compared with maternal DNA (13).

Fetal-derived DNA in maternal plasma is believed to originate from the placental cells (1). In view of its nonhematopoietic nature, we hypothesized that the size characteristics of placenta-derived fetal DNA could be extrapolated to the other nonhematopoietic tissue-derived DNA molecules in plasma. In this study, we performed massively parallel PE sequencing of plasma DNA from 6 patients who had received hematopoietic stem cell transplants (HSCTs) and 1 patient who had received a liver transplant (LT). We identified the sequenced reads attributable to the do-

nors and/or the recipients to explore the size profile of plasma DNA molecules derived from hematopoietic and nonhematopoietic cells.

Materials and Methods

STUDY PARTICIPANTS AND SAMPLE COLLECTION

Six HSCT recipients and 1 LT recipient were recruited from the Department of Paediatrics and the Department of Surgery, respectively, of the Prince of Wales Hospital, Hong Kong. Informed consent was obtained from each of the patients or their parents. We also recruited 4 adult male and 3 adult female volunteers as controls.

For the HSCT patients, fluorescence in situ hybridization and short tandem repeat DNA analyses for peripheral blood chimerism were performed as previously described (15). From the analyses of their chimerism status, all patients were deduced to be in complete remission with respect to their hematologic conditions (Table 1) (16). For the LT patient, the liver function test profile was normal.

We collected 10 mL of venous blood from each patient. Plasma was harvested after centrifugation as previously described (17). Plasma DNA was extracted according to the blood and body fluid protocol of the QIAamp DSP DNA Blood Mini Kit (Qiagen).

EXTRACTION OF GENOMIC DNA AND MICROARRAY GENOTYPING

Because the hematopoietic system had almost completely been converted to that of the donor after transplantation (Table 1), the genotype of the buffy coat of the HSCT recipients represented the donor's genotype.

⁵ Nonstandard abbreviations: PE, paired-end; HSCT, hematopoietic stem cell transplantation; LT, liver transplantation; SNP, single-nucleotide polymorphism; SOAP2, Short Oligonucleotide Alignment Program 2; chrY, chromosome Y.

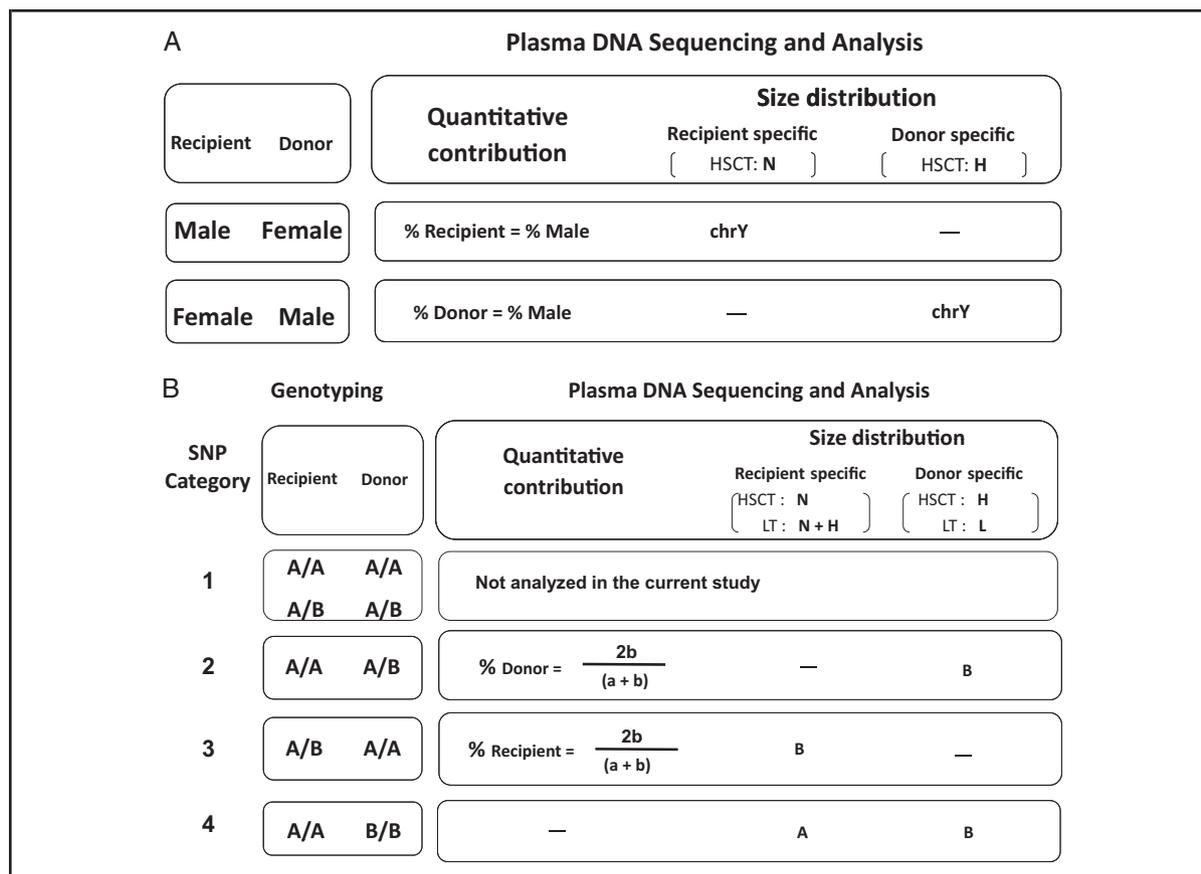


Fig. 1. Schematic diagrams of plasma DNA sequence analysis using sex and SNPs as markers.

(A), Sex-based analysis. In the sex-mismatched cases, the PE reads aligning to chrY represented the DNA fragments originating from the male side. The percentage and size distribution of male DNA reflected the quantitative contribution and size distribution of the male recipient or the male donor in plasma. (B), SNP-based analysis. SNPs were classified into 4 categories. SNPs in categories 2 and 3 were used to quantify donor- and recipient-contributed DNA in plasma via the equation in the figure, where *b* is the number of sequenced reads of the donor- or recipient-specific allele and *a* is the read count of the other allele, which is shared by the donor and recipient. The PE reads of the donor-specific alleles (e.g., the B allele in categories 2 and 4) and the recipient-specific alleles (e.g., the A allele in category 4) represented the donor- and recipient-specific DNA in plasma, respectively. The size distributions of the recipient- and donor-specific DNA fragments represented those of the nonhematopoietically and hematopoietically derived DNA fragments in the plasma of the HSCT recipients. For the LT case, size distributions of the recipient- and donor-specific DNA fragments represented those of the predominantly hematopoietically derived and liver-derived DNA fragments in the plasma. N, nonhematopoietically derived; H, hematopoietically derived; L, liver derived; N + H, both hematopoietically and nonhematopoietically derived.

Genomic DNA of the recipient was extracted from buccal or hair follicle cells with the QIAamp DNA Mini Kit (Qiagen). For the LT case, the blood and body fluid protocol of the QIAamp DSP DNA Blood Mini Kit was used to extract genomic DNA from the buffy coat of the recipient and the donor.

Two strategies were used to identify DNA molecules of donor and/or recipient origin in plasma. The first was via a sex-mismatch approach (Fig. 1A), and the second was a single-nucleotide polymorphism

(SNP)-based approach (Fig. 1B). For the second approach, we used the Genome-Wide Human SNP Array 6.0 (Affymetrix) for genome-wide SNP genotyping of approximately 900 000 SNPs for the paired donor–recipient genomic DNA. Genotypes were called with the Affymetrix Genotyping Console, version 2.1. The SNPs were then classified into 4 categories (Fig. 1B). Category 1 comprised SNPs in which donor and recipient shared the same genotypes (homozygous for the same allele or heterozygous). Category 2 consisted of

SNPs in which the recipient was homozygous and the donor was heterozygous, category 3 consisted of SNPs in which the recipient was heterozygous and the donor was homozygous, and category 4 consisted of SNPs in which the donor and recipient were both homozygous, but for a different allele.

PLASMA DNA SEQUENCING AND ALIGNMENT

The Paired-End DNA Sample Prep Kit (Illumina) was used as previously described (13) to construct sequencing libraries from the extracted plasma DNA. Details of the library-preparation protocol can be found in the Supplemental Methods file in the Data Supplement that accompanies the online version of this article at <http://www.clinchem.org/content/vol58/issue3>. DNA libraries of plasma samples were loaded onto separate lanes of sequencing flow cells and sequenced with a Genome Analyzer IIx (Illumina) or a HiSeq 2000 (Illumina) that used 36-bp \times 2 (HSCT cases 1 and 2) or 50-bp \times 2 (HSCT cases 3, 4, 5, and 6, and LT case 1) PE formats. Owing to the advances in the sequencing technology, longer sequenced reads per end could be obtained in the later phase of the study. That was particularly useful for the SNP-based analysis, because the longer reads allowed the detection of more informative SNPs with the same number of sequenced reads. All sequenced reads were aligned to the non-repeat-masked human reference genome (Hg18) (<http://genome.ucsc.edu>) with Short Oligonucleotide Alignment Program 2 (SOAP2) (<http://soap.genomics.org.cn/>). A Perl program (<http://www.perl.org/>) was compiled in house to perform the data analysis after sequence alignment. We identified PE reads with individual members sequenced on the same cluster position in a flow cell and uniquely aligned to a single location in the human genome with the correct orientation. Only the PE reads that demonstrated an insert size \leq 600 bp were retrieved for analysis. Duplicated reads (i.e., PE reads in which the insert DNA molecule showed identical start and end locations in the human genome) were likely to be clones of the same original plasma DNA template generated during the 15-cycle enrichment PCR. Hence, we removed all but one of the duplicated PE reads for the subsequent bioinformatics analysis. For the analysis of the sequences aligning to chromosome Y (chrY) and the autosomes, perfect alignment to the reference human genome without any nucleotide mismatch was required in order to minimize the potential for malalignment (18). On the other hand, we allowed up to 2 nucleotide mismatches in either member of the PE reads for the SNP-based analysis so we could identify donor- and recipient-specific alleles. The length of each DNA fragment was inferred from the coordinates of the outermost nucleotides at the 2 ends of the sequenced fragments.

STATISTICAL ANALYSIS

The Perl module WilcoxonRankSum-0.0.6 was downloaded (<http://www.cpan.org/>) and installed according to the manufacturer's instructions. The Mann-Whitney rank sum test inside this module was then invoked in the Perl program for statistical comparisons. A *P* value <0.05 was considered statistically significant. All probabilities were 2-tailed.

Results

PE SEQUENCING OF PLASMA DNA FROM HSCT PATIENTS

On average, 58.6% of the raw sequenced reads of the plasma DNA samples were mapped uniquely and perfectly to the non-repeat-masked human reference genome. HSCT cases 1 and 2 (Table 1) involved male patients who had received female hematopoietic stem cells. The DNA in these 2 cases was sequenced to an extent that the numbers of reads aligning to chrY were comparable with those obtained with SNP-based analysis. The numbers of total PE reads and PE reads aligning to chrY for the HSCT cases are listed in Table 1 in the online Data Supplement.

Paired genomic DNA samples from both the donor and recipient were available for HSCT cases 4, 5, and 6. SNPs were genotyped by microarray analysis and categorized into the 4 groups (see Table 2 in the online Data Supplement). According to the Poisson distribution, a mean coverage of 3-fold for a haploid human genome was required to detect 95% of the informative SNPs (either the A or B allele in SNP categories 2, 3 and 4) at least once. That translates into approximately 90×10^6 aligned PE reads (50 bp sequenced at each of the 2 ends) for each sequenced sample. For HSCT cases 4, 5, and 6, the number of aligned PE reads ranged from 113×10^6 to 121×10^6 . For these 3 cases, 94.4% or more of the informative SNPs were covered by the sequenced reads, with the detection rate for specific alleles (either donor or recipient specific) varying from 20.7% to 86.6% (see Table 3 in the online Data Supplement).

PERCENTAGE CONTRIBUTION OF HEMATOPOIETICALLY DERIVED DNA IN PLASMA

The hematopoietic systems of all of the analyzed HSCT recipients had almost completely been converted to that of the donor after transplantation (Table 1). Hence, the donor-derived DNA was of hematopoietic origin, whereas the recipient-derived DNA was of non-hematopoietic origin. For sex-mismatched HSCT cases, either the donor or the recipient was male, and the chrY sequences were specific for the male donor or the male recipient, respectively (Fig. 1A). We therefore used the percentage of PE reads aligning to chrY to estimate the male DNA percentage (details are shown

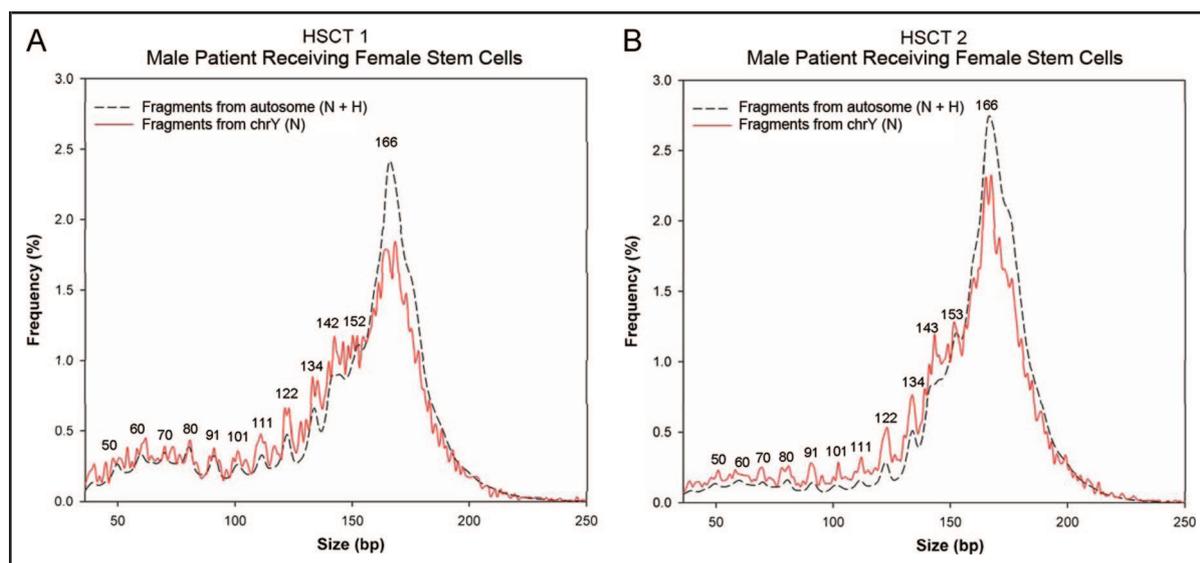


Fig. 2. Size distributions of plasma DNA fragments in sex-mismatched HSCT recipients.

Plots of the size distributions of PE reads that were aligned to chrY (red line) and the autosomes (dashed line) in the plasma from HSCT patients 1 (A) and 2 (B). Numbers denote the DNA sizes at the peaks for the plasma DNA fragments from the autosomes. Autosomal DNA fragments were selected for this purpose because they represented the predominant species in plasma; thus, the sizes of the peaks could be determined with greater precision. For all size plots in this report, only DNA fragments with lengths ≤ 250 bp are shown. The frequencies of PE reads were calculated at 1-bp resolution.

in the Supplemental Methods in the online Data Supplement). In HSCT cases 1 and 2, the percentages of male DNA (i.e., male recipient derived) were 14.0% and 17.2%, respectively. For HSCT cases 3 and 4, the percentages of male DNA in plasma (i.e., male donor derived) were 89.2% and 75.0%, respectively.

The SNP-based quantitative analysis allowed us to determine the proportions of donor- and recipient-derived DNA in plasma, regardless of the sex combination of the donor and recipient (Fig. 1B). For HSCT cases 4, 5, and 6, the percentages of donor-derived DNA determined by the SNPs in category 2 were 71.5%, 79.9%, and 90.7%, respectively. These results are consistent with the percentages of recipient-derived DNA determined by the SNPs in category 3 (Table 1). The values determined for every chromosome were highly consistent (see Table 4 in the online Data Supplement), indicating that the relative proportions of donor and recipient DNA were largely constant across the entire genome.

Taken together, the HSCT recipients exhibited high proportions of donor-derived DNA (i.e., DNA of hematopoietic origin) in their plasma samples, with a median of 84.4% for the 6 cases analyzed.

SIZE DISTRIBUTION OF PLASMA DNA FROM HSCT PATIENTS

We compared the size distributions of DNA from the hematopoietic and nonhematopoietic systems. HSCT

cases 1 and 2 involved male patients who had received female stem cells. Thus, the chrY PE sequences represent the DNA fragments from the nonhematopoietic tissues. In contrast, because the hematopoietic system contributed 86.0% and 82.8% of the plasma DNA for HSCT cases 1 and 2, respectively, the sequences from the autosomes mostly represent the DNA fragments from hematopoietic cells. DNA fragments from chrY and the autosomes exhibited a peak at approximately 166 bp, with peaks occurring at a 10-bp periodicity at sizes of approximately 142 bp and shorter (Fig. 2). The height of the 166-bp peak for chrY, however, was lower than for the autosomes (Fig. 2). To quantify the difference between the size distributions of the hematopoietically and nonhematopoietically derived DNA fragments, we determined the proportion of short DNA fragments for the 2 populations by using 150 bp as a cutoff. In other words, fragments ≤ 150 bp were defined as “short fragments.” A larger proportion of short fragments indicates a distribution with shorter sizes. The difference between the proportions of the 2 DNA populations would reflect the magnitude of the difference in their size distributions. In HSCT cases 1 and 2, the proportions of short DNA fragments were larger for chrY than for the autosomes (Table 2), with a median difference of 9.9%. These results suggested that the nonhematopoietically derived DNA in the plasma was shorter than the hematopoietically derived DNA.

Table 2. Proportions of plasma DNA fragments ≤150 bp.

Case no.	chrY-based analysis						SNP-based analysis					
	From chrY		From the autosomes		chrY vs autosomes		Recipient derived		Donor derived		Recipient vs donor	
	Proportion	Source	Proportion	Source	Difference	Interpretation	Proportion	Source	Proportion	Source	Difference	Interpretation
H SCT1	47.9%	N ^a	38.8%	N + H	9.1%	Shorter	NA	NA	NA	NA	NA	NA
H SCT2	36.4%	N	25.7%	N + H	10.7%	Shorter	NA	NA	NA	NA	NA	NA
H SCT3	27.4%	H	30.7%	N + H	-3.3%	Longer	NA	NA	NA	NA	NA	NA
H SCT4	32.9%	H	39.2%	N + H	-6.3%	Longer	52.2%	N	30.5%	H	21.7%	Shorter
H SCT5	31.0%	N + H	31.6%	N + H	-0.6%	Similar	40.5%	N	26.8%	H	13.7%	Shorter
H SCT6	29.2%	N + H	29.8%	N + H	-0.6%	Similar	40.9%	N	26.1%	H	14.8%	Shorter
LT 1	31.2%	N + H	31.5%	N + H	-0.3%	Similar	25.3%	N + H	38.7%	L	-13.4%	Longer

^a N, nonhematopoietically derived; H, hematopoietically derived; N + H, both nonhematopoietically and hematopoietically derived; NA not available; L, liver derived.

HSCT cases 3 and 4 involved female patients who had received male stem cells. In these 2 cases, the chrY sequences represented DNA fragments from the hematopoietic cells. On the other hand, the sequences from the autosomes represented a mixture of hematopoietically and nonhematopoietically derived DNA, with the hematopoietic component being predominant (89.2% and 75.0% for HSCT cases 3 and 4, respectively). The size distributions of DNA fragments from chrY and the autosomes for these 2 cases are shown in Fig. 1 in the online Data Supplement. The proportions of short DNA fragments were smaller for chrY than for the autosomes, indicating that the hematopoietically derived DNA had a distribution of larger sizes. The difference was 3.3% and 6.3% for HSCT cases 3 and 4, respectively (Table 2). For HSCT cases 5 and 6, which involved male patients who had received male stem cells, DNA fragments from chrY and the autosomes were both predominantly of hematopoietic origin. Hence, the 2 size distributions were almost identical (see Fig. 2 in the online Data Supplement), with similar proportions of short DNA fragments (Table 2).

In contrast to the comparison of DNA fragments from chrY and the autosomes, the analysis of DNA fragments of the donor- and recipient-specific alleles allows size profiling of hematopoietically and nonhematopoietically derived DNA, regardless of the donor's and recipient's sex. PE reads of the donor-specific alleles for the SNPs in categories 2 and 4 were combined to represent the donor-derived DNA fragments, and those of the recipient-specific alleles for the SNPs in categories 3 and 4 were combined to represent the recipient-derived ones. The total numbers of the recipient- and donor-specific PE reads obtained for HSCT cases 4, 5, and 6 were 93 562 vs 220 138, 57 268 vs 231 293, and 27 243 vs 270 086, respectively (see Table 3 in the online Data Supplement). Fig. 3 shows the size distributions of the recipient- and donor-specific DNA fragments for these 3 cases. Irrespective of the origin, >98% of the plasma DNA fragments were shorter than 250 bp, with a 166-bp peak and peaks occurring with a 10-bp periodicity at approximately 143 bp and smaller (Fig. 3). For each case, the size distribution of the recipient-specific DNA fragments consisted of significantly shorter fragments than for the donor-specific DNA fragments ($P < 0.001$, Mann-Whitney rank sum test), indicating that the nonhematopoietically derived DNA fragments were shorter than the hematopoietically derived ones. The height of the 166-bp peak for the recipient-derived DNA was reduced (Fig. 3) relative to that of the donor-derived DNA for all 3 cases. Moreover, the recipient-derived DNA had a higher proportion of short DNA fragments than the donor-derived DNA (Table 2), with a median difference of 14.8%.

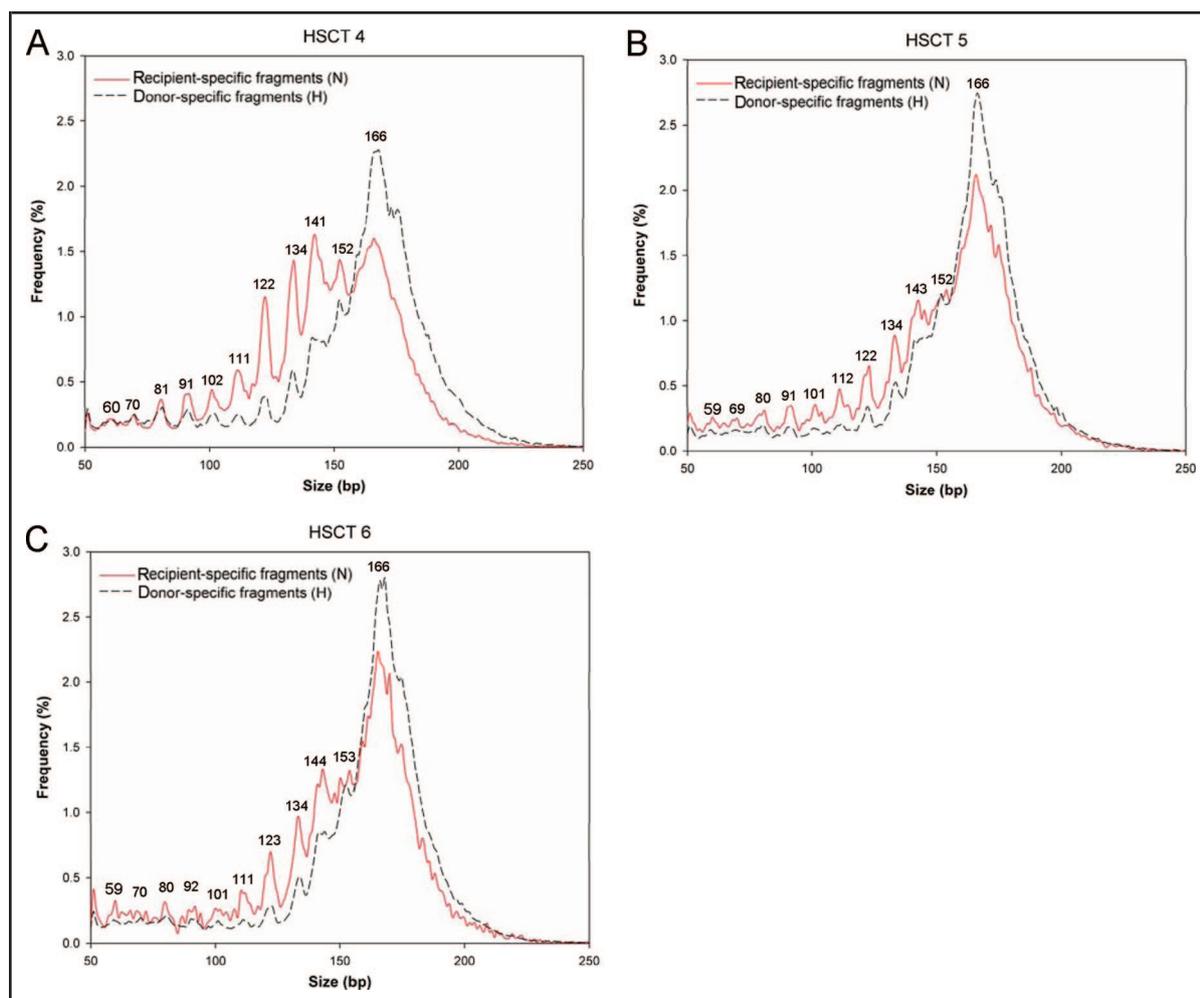


Fig. 3. Size distributions of plasma DNA fragments in HSCT recipients.

Plots of the size distributions of PE reads of the recipient-specific alleles (red line) and the donor-specific alleles (dashed line) in the plasma from HSCT patients 4 (A), 5 (B), and 6 (C). Numbers denote the DNA sizes at the peaks for the donor-specific DNA fragments in plasma. Donor-specific DNA fragments were selected for this purpose because they represented the predominant species in plasma; thus, the sizes of the peaks could be determined with greater precision.

SEQUENCING OF PLASMA DNA FROM THE LT PATIENT

Because the liver is a nonhematopoietic organ in the adult, LT can be used as a model to further characterize the nonhematopoietically derived DNA in plasma (Table 1). In this case, the donor-specific DNA fragments were released from the transplanted liver, whereas the recipient-specific DNA fragments originated from the other systems (including the hematopoietic system and other nonhematopoietic organs exclusive of the liver). With 110×10^6 aligned PE reads, $>95.7\%$ of the informative SNPs were covered (see Table 3 in the online Data Supplement). A donor-specific allele was observed at least once for 48.4% of the SNPs in category 2. A recipient-specific allele was observed at least once for

79.9% of the SNPs in category 3 and for 91.7% of the SNPs in category 4. Using the equation in Fig. 1B, we estimated that 27.2% of the plasma DNA in this patient was donor derived, a result consistent with a previous finding that the liver accounted for only a minority of the plasma DNA in solid-organ transplant recipients (19). The donor's contribution to the plasma DNA was consistent across chromosomes (see Table 4 in the online Data Supplement).

From the above quantitative analysis of plasma DNA in HSCT patients, we postulated that the recipient-derived plasma DNA in this patient (accounting for 72.5% of the total plasma DNA) was predominantly from hematopoietic cells. We therefore

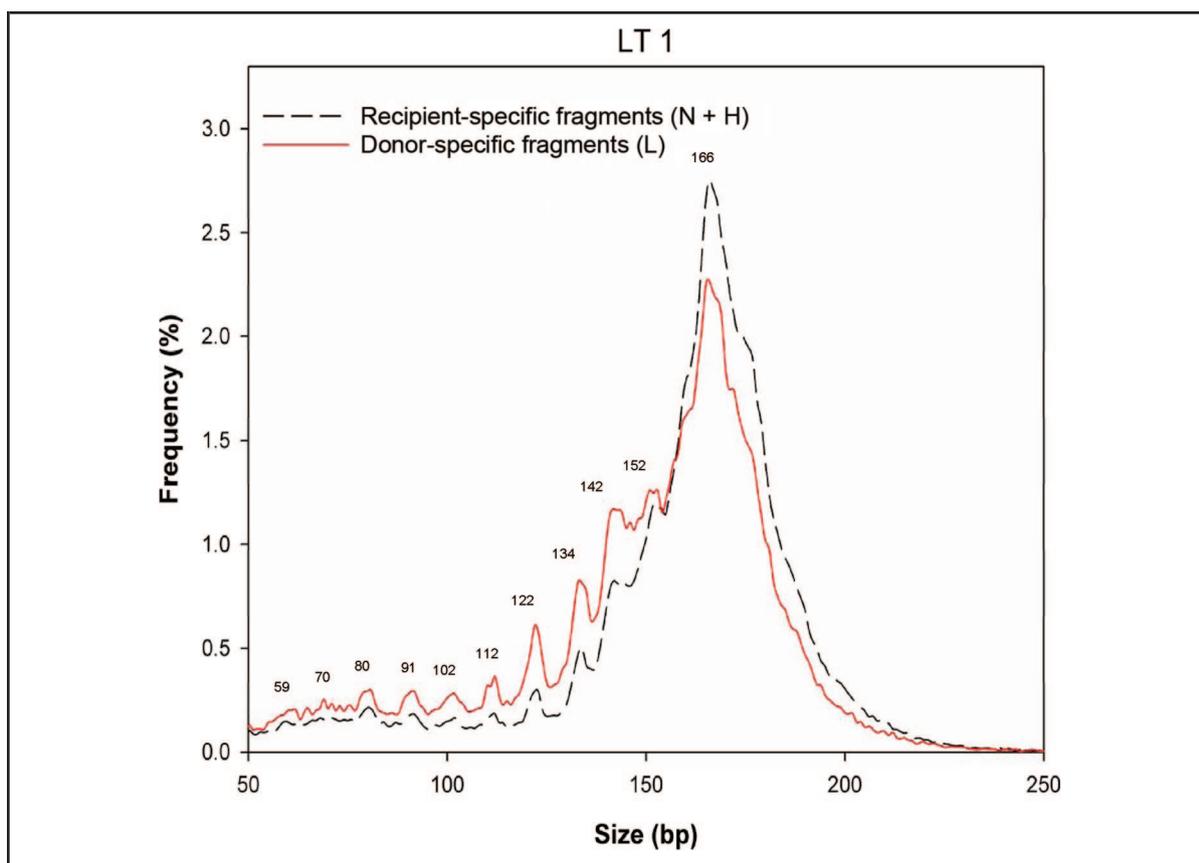


Fig. 4. Size distributions of plasma DNA fragments in the LT recipient.

Plots of the size distributions of PE reads of the donor-specific alleles (red line) and the recipient-specific alleles (dashed line) in the plasma from LT patient 1. Numbers denote the DNA sizes at the peaks for the recipient-specific DNA fragments in plasma. Recipient-specific DNA fragments were selected for this purpose because they were mainly of hematopoietic origin and represented the predominant species in plasma; thus, the sizes of the peaks could be determined with greater precision.

used the PE reads of the donor- and recipient-specific alleles to determine the size distribution of plasma DNA molecules derived from the liver and the hematopoietic system, respectively. We obtained 155 084 donor-specific and 404 567 recipient-specific PE reads (see Table 3 in the online Data Supplement). Fig. 4 shows the size distributions of the recipient- and donor-specific plasma DNA fragments. Both size distributions exhibited a peak at 166 bp and peaks occurring at a 10-bp periodicity at approximately 142 bp and smaller (Fig. 4). The size distribution of the donor-derived (i.e., liver-derived) DNA fragments consisted of significantly shorter fragments than that of the recipient-specific ones (mainly hematopoietically derived) ($P < 0.001$, Mann-Whitney rank sum test). Again, this shortening was mainly due to the reduction in the 166-bp peak (Fig. 4) and the relative increase in the proportion of short

DNA fragments (Table 2). The proportion of short fragments for the donor-derived DNA was 13.4% higher than that for the recipient-derived DNA (Table 2). The HSCT and the LT data were thus consistent and indicated that the nonhematopoietically derived plasma DNA molecules were shorter than the hematopoietically derived ones.

Discussion

In this study, we performed massively parallel PE sequencing of plasma DNA from patients who had undergone transplantation in order to characterize the nature of the plasma DNA with respect to the cellular contribution and the fragment sizes. The quantitative analyses of 6 HSCT cases and 1 LT case confirm our previous findings that the DNA in plasma is predominantly hematopoietic in origin

(20). In contrast to the previous studies that used locus-specific quantitative PCR (19, 20), sequencing-based analysis characterizes plasma DNA molecules of different cellular origin in a locus-independent manner, provides these data at single-molecule resolution, and yields exact measurements of the sizes of DNA molecules (21).

By revealing the detailed size-distribution pattern of the plasma DNA from transplant recipients, our study helps to enhance the understanding of the biology of plasma DNA. Previous studies suggested that cell apoptosis is an important mechanism for releasing cell-free DNA into the circulation (3, 4, 6). During apoptosis, DNA is degraded into nucleosomal units (22). In the present study, the distributions of both hematopoietically and nonhematopoietically derived DNA demonstrated a 166-bp peak and peaks occurring at a 10-bp periodicity at approximately 143 bp and smaller. The 166-bp peak probably represents a monochromatosome consisting of nucleosomal DNA wrapped around a histone core (146 bp) and a linker (20 bp) (23). The 143-bp peak probably corresponds to the remnant of a monochromatosome that has been trimmed of its linker segment. The 10-bp periodicity at approximately 143 bp and smaller is probably related to the structural periodicity of the DNA double helix, in that every 360° turn of the helix corresponds to 10 bp, where nuclease-sensitive sites lie (24). These characteristics of overall size distribution are consistent with those of studies of maternal plasma (13), suggesting that the fragmentation into nucleosomal DNA fragments during cell apoptosis might be a universal mechanism involved in the liberation of cell-free DNA into the plasma. Although other methods, such as gel electrophoresis and real-time PCR analysis (6, 11), seem to detect long DNA fragments in the circulation of some healthy individuals and pregnant women, the importance of these plasma DNA molecules remains unclear.

A number of factors might contribute toward the relatively shorter distribution of sizes for nonhematopoietically derived plasma DNA fragments, compared with hematopoietically derived plasma DNA. Possible factors include those affecting the intracellular degradation of DNA, the release of DNA, and those operating during DNA transit in plasma (e.g., duration or path of transit). Remarkably, the size characteristics of the hematopoietically and nonhematopoietically derived DNA in plasma detected in the current study are virtually identical to those of maternal and fetal DNA in maternal plasma (13). In this regard, circulating fetal DNA, which is placentally derived and thus nonhematopoietic in origin, exhibits size characteristics that are the same as the nonhematopoietically derived DNA in the

transplantation models investigated in this study. Such consistency implies similarities in some of the above-mentioned factors operating on such nonhematopoietically derived DNA.

Plasma DNA can serve as an alternative sample type for the determination of HSCT chimerism (20) and as a potential marker for monitoring graft rejection or other sources of tissue damage after solid-organ transplantation (19, 25, 26). Our SNP-based sequence analysis of plasma DNA in transplant recipients provides a universal approach for such applications. Recently, Snyder et al. have applied this approach for the noninvasive detection of rejection episodes in heart transplant recipients (27, 28). On the other hand, the difference in sizes between hematopoietically and nonhematopoietically derived plasma DNA might allow us to selectively target specific populations of such nucleic acids, thereby facilitating diagnostic applications of nonhematopoietic tissue-derived circulating DNA in the relevant clinical setting, e.g., detection of tissue damage associated with rejection, ischemia, or trauma (29).

Massively parallel sequencing of plasma DNA has become an emerging promising research and molecular-diagnostic tool for a number of clinical settings (27, 30). The observations we have reported may have general implications for the field of circulating nucleic acids and may help us push forward the clinical utility of this new technology. Such developments may have impacts in multiple diagnostic areas, ranging from transplantation monitoring to noninvasive prenatal diagnosis, to cancer detection and monitoring.

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 or 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: R.W.K. Chiu, University Grants Committee of the Government of the Hong Kong Special Administrative Region under the Areas of Excellence scheme (AoE/M-04/06); Y.M.D. Lo, Sequenom, University Grants Committee of the Government of the Hong Kong Special Administration Region, China, under the Areas of Excellence Scheme (AoE/M-04/06), and Endowed Chair from the Li Ka Shing Foundation.

Expert Testimony: None declared.

Other Remuneration: Several authors have filed patent applications on the diagnostic applications of plasma nucleic acids. Part of this patent portfolio has been licensed to Sequenom. Y.M.D. Lo, R.W.K. Chiu, K.C.A. Chan, Y.W.L. Zheng, and F.M.F. Lun, patent applications filed on the detection of fetal nucleic acids in maternal plasma for noninvasive prenatal diagnosis; F.M.F. Lun, patent applications filed on the detection of fetal nucleic acids in maternal plasma for noninvasive prenatal diagnosis; Y.M.D. Lo and R.W.K. Chiu, travel support from Illumina, Life Technologies, and Sequenom.

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.

Acknowledgments: We thank Lisa Y.S. Chan, Yongjie Jin, Katherine C.K. Chow, and Kiwi Y.W. Chan for performing the sequencing and R.W.Y. Chan for extraction of genomic DNA from buccal and hair follicle cells.

References

1. Lo YMD, Chiu RWK. Prenatal diagnosis: progress through plasma nucleic acids. *Nat Rev Genet* 2007;8:71–7.
2. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer—a survey. *Biochim Biophys Acta* 2007;1775:181–232.
3. Giacona MB, Ruben GC, Iczkowski KA, Roos TB, Porter DM, Sorenson GD. Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas* 1998;17:89–97.
4. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayr FO, Hesch RD, 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.
5. van der Vaart M, Pretorius PJ. The origin of circulating free DNA. *Clin Chem* 2007;53:2215.
6. 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.
7. Wang BG, Huang HY, Chen YC, Bristow RE, Kasasuei K, Cheng CC, et al. Increased plasma DNA integrity in cancer patients. *Cancer Res* 2003;63:3966–8.
8. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* 2005;102:16368–73.
9. Umetani N, Kim J, Hiramatsu S, Reber HA, Hines OJ, Bilchik AJ, Hoon DS. Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. *Clin Chem* 2006;52:1062–9.
10. Chan KCA, Leung SF, Yeung SW, Chan ATC, Lo YMD. Persistent aberrations in circulating DNA integrity after radiotherapy are associated with poor prognosis in nasopharyngeal carcinoma patients. *Clin Cancer Res* 2008;14:4141–5.
11. Li Y, Zimmermann B, Rusterholz C, Kang A, Holzgreve W, Hahn S. Size separation of circulatory DNA in maternal plasma permits ready detection of fetal DNA polymorphisms. *Clin Chem* 2004;50:1002–11.
12. Lun FMF, Tsui NBY, Chan KCA, Leung TY, Lau TK, Charoenkwan P, et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc Natl Acad Sci U S A* 2008;105:19920–5.
13. 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.
14. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing. *Clin Chem* 2010;56:1279–86.
15. Hung ECW, Shing TKF, Chim SSC, Yeung PC, Chan RWY, Chik KW, et al. Presence of donor-derived DNA and cells in the urine of sex-mismatched hematopoietic stem cell transplant recipients: implication for the transrenal hypothesis. *Clin Chem* 2009;55:715–22.
16. Antin JH, Childs R, Filipovich AH, Giral S, Mackinnon S, Spitzer T, Weisdorf D. Establishment of complete and mixed donor chimerism after allogeneic lymphohematopoietic transplantation: recommendations from a workshop at the 2001 Tandem Meetings of the International Bone Marrow Transplant Registry and the American Society of Blood and Marrow Transplantation. *Biol Blood Marrow Transplant* 2001;7:473–85.
17. Chan KCA, Yeung SW, Lui WB, Rainer TH, Lo YMD. Effects of preanalytical factors on the molecular size of cell-free DNA in blood. *Clin Chem* 2005;51:781–4.
18. Chiu RWK, Chan KCA, Gao Y, Lau VYM, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci U S A* 2008;105:20458–63.
19. Lui YYN, Woo KS, Wang AYM, Yeung CK, Li PKT, Chau E, et al. Origin of plasma cell-free DNA after solid organ transplantation. *Clin Chem* 2003;49:495–6.
20. Lui YYN, Chik KW, Chiu RWK, Ho CY, Lam CWK, Lo YMD. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem* 2002;48:421–7.
21. Chiu RWK, Cantor CR, Lo YMD. Non-invasive prenatal diagnosis by single molecule counting technologies. *Trends Genet* 2009;25:324–31.
22. Wyllie AH. Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation. *Nature* 1980;284:555–6.
23. Jiang C, Pugh BF. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 2009;10:161–72.
24. Krebs JC, Goldstein ES, Kilpatrick ST. *Lewin's genes X*. 10th ed. Sudbury (MA): Jones and Bartlett Publishers; 2011. p 220–61.
25. Gadi VK, Nelson JL, Boespflug ND, Guthrie KA, Kuhr CS. Soluble donor DNA concentrations in recipient serum correlate with pancreas-kidney rejection. *Clin Chem* 2006;52:379–82.
26. Garcia Moreira V, Prieto Garcia B, Baltar Martín JM, Ortega Suárez F, Alvarez FV. Cell-free DNA as a noninvasive acute rejection marker in renal transplantation. *Clin Chem* 2009;55:1958–66.
27. Lo YMD. Transplantation monitoring by plasma DNA sequencing. *Clin Chem* 2011;57:941–2.
28. Snyder TM, Khush KK, Valentine HA, Quake SR. Universal noninvasive detection of solid organ transplant rejection. *Proc Natl Acad Sci U S A* 2011;108:6229–34.
29. Lo YMD, Rainer TH, Chan LY, Hjelm NM, Cocks RA. Plasma DNA as a prognostic marker in trauma patients. *Clin Chem* 2000;46:319–23.
30. Lo YMD, Chiu RWK. Next-generation sequencing of plasma/serum DNA: an emerging research and molecular diagnostic tool. *Clin Chem* 2009;55:607–8.