Clinical Chemistry 60:7
954–962 (2014)

Maternal Plasma RNA Sequencing for Genome-Wide Transcriptomic Profiling and Identification of Pregnancy-Associated Transcripts

Nancy B.Y. Tsui,[1,2] Peiyong Jiang,[1,2] Yuen Fei Wong,[1,2] Tak Y. Leung,[3] K.C. Allen Chan,[1,2] Rossa W.K. Chiu,[1,2] Hao Sun,[1,2] and Y.M. Dennis Lo[1,2*]

BACKGROUND: Analysis of circulating RNA in the plasma of pregnant women has the potential to serve as a powerful tool for noninvasive prenatal testing and research. However, detection of circulating RNA in the plasma in an unbiased and high-throughput manner has been technically challenging. Therefore, only a limited number of circulating RNA species in maternal plasma have been validated as pregnancy- and placenta-specific biomarkers.

METHODS: We explored the use of massively parallel sequencing for plasma transcriptome profiling in first-, second-, and third-trimester pregnant women. Genotyping was performed for amniotic fluid, placental tissues, and maternal blood cells, with exome-enriched sequencing.

RESULTS: In the early pregnancy group comprising 1 first- and 1 second-trimester pregnancy cases, the fetal contribution to the RNA pool in maternal plasma was 3.70%. The relative proportion of fetal contribution was increased to 11.28% in the late pregnancy group comprising 2 third-trimester pregnancy cases. The placental biallelic expression pattern of *PAPPA* (pregnancy-associated plasma protein A, pappalysin 1), a known pregnancy-specific gene, and the monoallelic expression pattern of *H19* [H19, imprinted maternally expressed transcript (non-protein coding)], an imprinted maternally expressed gene, were also detected in the maternal plasma. Furthermore, by direct examination of the maternal plasma transcriptomic profiles before and after delivery, we identified a panel of pregnancy-associated genes.

CONCLUSIONS: Plasma RNA sequencing provides a holistic view of the maternal plasma transcriptomic repertoire. This technology is potentially valuable for using circulating plasma nucleic acids for prenatal testing and research.

The presence of cell-free fetal RNA in maternal plasma was reported more than a decade ago *(1)*. Many studies have since been conducted to detect circulating RNA of fetal and placental origin in maternal plasma *(2–4)*. Interestingly, the expression levels of placenta-specific transcripts in the plasma positively correlated with those in placental tissues *(3)*, underscoring the potential clinical utility of plasma RNA analysis as a noninvasive tool to monitor placental development and fetal health. Examination of maternal circulating RNA has found clinical applications for pregnancy- or placenta-related disorders such as preeclampsia *(5–8)*, intrauterine growth retardation *(9)*, and preterm birth *(10)*, as well as for noninvasive testing of fetal chromosomal aneuploidies *(11–13)*. Such developments highlight the potential utilities of RNA biomarkers for the molecular assessment of prenatal disorders.

Despite the promising outlook of plasma RNA analysis in prenatal testing, there remain a limited number of well-validated pregnancy- or placenta-related transcripts in maternal plasma to date. In this regard, examination of RNA biomarkers in the plasma has been conducted using reverse transcriptase-PCR *(2–4, 6, 9, 10)*, an analytically sensitive method that can typically target a relatively small number of RNA species per analysis. Of note, these studies have mainly focused on genes with relatively high levels of expression in the placenta compared with maternal blood cells, on the premise that pregnancy-associated genes are largely derived from the placenta. Such approaches

[1] Centre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, [2] Department of Chemical Pathology, and [3] Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR, China.
* Address correspondence to this author at: Department of Chemical Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, 30-32 Ngan Shing St., Shatin, New Territories, Hong Kong SAR, China. Fax +852-26365090; e-mail loym@cuhk.edu.hk.
Received January 23, 2014; accepted March 25, 2014.
Previously published online at DOI: 10.1373/clinchem.2014.221648

have identified pregnancy-associated RNA targets that are highly expressed in the placenta but might have missed other important targets. Furthermore, given the low concentrations and poor integrity of plasma RNA *(8, 14)*, conventional high-throughput methods such as serial analysis of gene expression and microarray analysis would not be robust for the direct examination of the plasma transcriptome.

The aforesaid technical limitations of plasma RNA analysis could potentially be resolved by employing massively parallel sequencing (MPS)[4] for RNA analysis, namely RNA sequencing (RNA-seq) *(15, 16)*. Given the enhanced analytical sensitivity and wide dynamic range, RNA-seq has been employed to examine gene expression in many human tissues, including the placenta *(17)*. The superiority of MPS has been further demonstrated by its feasibility in direct profiling of plasma microRNAs in healthy individuals *(18)* and in pregnant women *(19, 20)*. Nevertheless, the full spectrum of the plasma transcriptome remains elusive, which might be due to the lower stability of long RNA species compared to short microRNAs in the plasma *(21)*.

In this proof-of-concept study, we have shown that fetal- and maternal-derived transcripts can be detected in maternal plasma and that their relative contributions can be estimated using RNA-seq, by examination of fetal- and maternal-specific single-nucleotide polymorphisms (SNPs). In addition, the allele-specific expression (ASE) patterns of the placenta can be monitored in maternal plasma. We have also demonstrated that pregnancy-associated transcripts can be identified by direct examination of the maternal plasma before and after delivery.

Materials and Methods

CASE RECRUITMENT AND SAMPLE COLLECTION

We recruited 1 first-, 1 second-, and 2 third-trimester pregnant women with singleton pregnancies from the Department of Obstetrics and Gynaecology, Prince of Wales Hospital, Hong Kong. For the first- and second-trimester pregnant women, categorized as "early pregnancy cases," we collected peripheral blood, as well as chorionic villi or amniotic fluid; for the third-trimester pregnant women, categorized as "late pregnancy cases," we collected peripheral blood before cesarean delivery and at 24 h after delivery, as well as the delivered placental tissues. Peripheral blood samples from 20 additional third-trimester pregnant women and 2 nonpregnant females were also collected. The study was approved by the Clinical Research Ethics Committee of The Chinese University of Hong Kong. All subjects were recruited with written informed consent. Details for sample collection are described in the Data Supplement that accompanies the online version of this report at http://www.clinchem.org/content/vol60/issue7.

SAMPLE PROCESSING AND LIBRARY SEQUENCING

Total RNA was extracted from the maternal blood cells, placenta, and plasma. A summary of samples analyzed by RNA-seq is shown in online Supplemental Table 1. Placental and blood cell RNA samples were pretreated with a Ribo-Zero Gold Kit (Epicentre) to remove ribosomal RNA (rRNA) before sequencing library preparation. Plasma samples of case 9356 were pretreated with a Ribo-Zero Gold Kit, but owing to a high GC bias, Ribo-Zero Gold pretreatment was omitted for all other plasma samples (see Results in the online Supplemental Data file). Maternal blood cell and placental RNA samples were subjected to chemical fragmentation for 5 min; the fragmentation time was reduced to 2.5 min for plasma RNA samples. cDNA libraries were synthesized using an mRNA-Seq sample preparation kit (Illumina), following the manufacturer's instructions with slight modification, and were sequenced for 50 bp or 75 bp in a paired-end format on a HiSeq 2000 instrument (Illumina). Genomic DNA was extracted from the peripheral blood, chorionic villi, amniotic fluid, and placenta for genotyping. Exome enrichment of genomic DNA libraries was performed using a TruSeq exome enrichment kit (Illumina) following the manufacturer's protocol. The genomic DNA libraries were sequenced for 75 bp in a paired-end format on a HiSeq 2000 instrument. Additional details are described in Materials and Methods in the online Supplemental Data file.

DATA ALIGNMENT AND ANALYSIS

For RNA-seq, the sequenced reads were preprocessed and aligned to the hg19 reference human genome by an in-house bioinformatics pipeline. For exome-enriched genomic DNA library sequencing, the sequenced reads were aligned to the hg19 reference human genome by the Short Oligonucleotide Alignment Program 2 *(22)*. The detailed bioinformatics work flow and data analysis are described in Materials and Methods in the online Supplemental Data file and in online Supplemental Figs. 1 and 2.

IDENTIFICATION OF FETAL- AND MATERNAL-DERIVED TRANSCRIPTS

To identify whether the plasma RNA transcripts were of fetal or maternal origin, we examined approximately one million SNPs in the dbSNP Build 135 database *(23)* that were located in the exons. For the following description, we use "A" and "B" to refer to alleles for a

[4] Nonstandard abbreviations: MPS, massively parallel sequencing; RNA-seq, RNA-sequencing; SNP, single-nucleotide polymorphism; ASE, allele-specific expression; rRNA, ribosomal RNA.

particular SNP site. We first sorted for informative SNPs containing fetal-specific alleles, i.e., the fetus was heterozygous (AB) and the mother was homozygous (AA), and for informative SNPs containing maternal-specific alleles, i.e., the fetus was homozygous (AA) and the mother was heterozygous (AB). We then sought these SNP alleles in the RNA transcripts *(12)* of the placenta, blood cells, and plasma. Because this analysis could potentially be confounded by ASE *(24, 25)*, we filtered out fetal- and maternal-derived transcripts in the plasma, which had evidence of allelic imbalances in the placenta and in the maternal blood cells, respectively. For the early pregnancy cases, RNA-seq was not performed on chorionic villi and amniotic fluid, because these samples had been exhausted for the genotyping analysis. Data sets of the 2 early pregnancy cases and the 2 late pregnancy cases were combined, respectively, to increase the number of informative SNPs for analysis. A detailed description is included in Materials and Methods in the online Supplemental Data file.

Results

RNA-SEQ DATA

After removal of duplicated reads and rRNA reads, a mean of 3 million analyzable reads were obtained for each nonpregnant female plasma sample, and a mean of 12 million analyzable reads were obtained for each plasma sample of pregnant women. For tissue RNA-seq, a mean of 173 million and 41 million analyzable reads per sample were obtained for placenta and blood cells, respectively. The RNA-seq alignment statistics are summarized in online Supplemental Table 2. The GC content of sequenced reads was also examined in all samples (see Results in the online Supplemental Data file, online Supplemental Table 3, and online Supplemental Fig. 4).

IDENTIFICATION AND ESTIMATION OF FETAL- AND MATERNAL-DERIVED TRANSCRIPT IN MATERNAL PLASMA

Informative genes, defined as genes with at least 1 informative SNP, were first identified based on the genotyping data. In the 2 early pregnancy cases, a total of 6714 and 6753 informative genes were available to examine the relative proportions of fetal and maternal contributions, respectively (see online Supplemental Table 4). In the 2 late pregnancy cases, a total of 7788 and 7761 informative genes were available to examine the relative proportions of fetal and maternal contributions, respectively. To measure the relative proportion of fetal contribution in the maternal plasma, we sorted for RNA transcripts for which the fetal-specific alleles were covered by at least 1 RNA-seq read in the maternal plasma samples. The relative proportions of such fetal-derived transcripts were 3.70% and 11.28% during early and late gestations, respectively. Using a similar approach, we estimated the relative proportions of maternal contribution in the circulation, examined using the maternal-specific SNP alleles, to be 76.90% and 78.32% during early and late gestations, respectively (Fig. 1A).

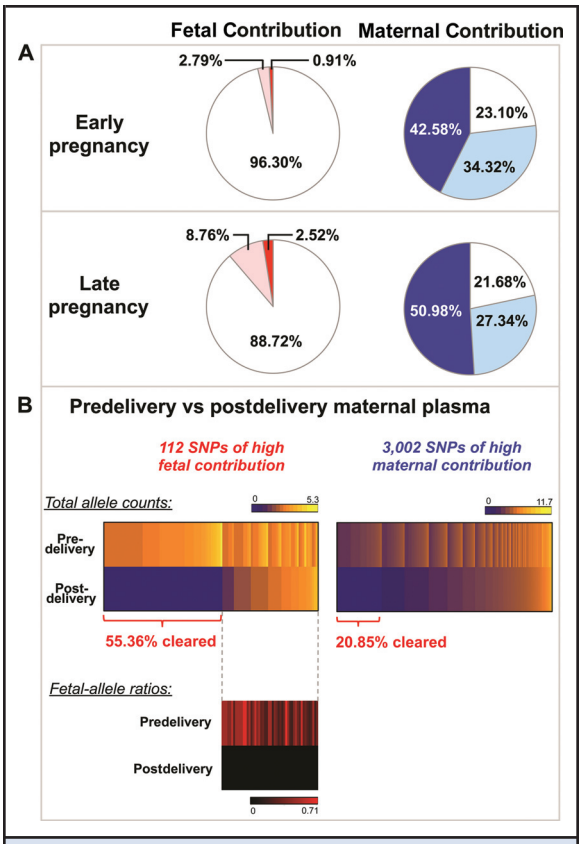

Fig. 1. Fetal and maternal contributions in maternal plasma transcriptome.

(A), Proportions of fetal- and maternal-derived transcripts in the maternal plasma of early and late pregnancies. Red plus pink sectors indicate proportions of all plasma transcripts derived from the fetus; red sector indicates proportion of transcripts predominantly derived from the fetus; pink sector indicates proportion of transcripts derived from both the fetus and the mother. Dark blue plus light blue sectors indicate proportions of all plasma transcripts derived from the mother; dark blue sector indicates proportion of transcripts predominantly derived from the mother; light blue sector indicates proportion of transcripts derived from both the mother and the fetus. (B), Upper panel: total RNA allele counts [$\log_2$ (count+1)] were examined for SNPs on RNA transcripts of high fetal and high maternal contributions, respectively. Lower panel: fetal-allele ratios were calculated for SNPs with residual RNA allele counts in the 24-h postdelivery maternal plasma.

Table 1. RNA-seq read counts on SNP sites in *PAPPA* RNA for case 9415.[a]

	DNA genotype		Placenta		Maternal blood cells		Predelivery maternal plasma		Postdelivery maternal plasma	
PAPPA SNP rsID	Fetus	Mother	A	G	A	G	A	G	A	G
rs386088	GA	AA	2421	2164	0	0	16	21	0	0

[a] Allele-A is the shared allele and allele-G is the fetal-specific allele.

To further identify circulating transcripts that have a high level of fetal or maternal contribution, we first calculated the allele ratio as follows:

$$\text{B-allele ratio} = \frac{\text{B-allele count}}{(\text{A-allele count} + \text{B-allele count})}$$

The B-allele refers to fetal- and maternal-specific alleles and the A-allele refers to the shared allele between the fetus and the mother. Theoretically, for a plasma transcript that is contributed solely by the fetus or by the mother, the B-allele ratio should be 0.5, assuming no ASE. In this study, we defined the B-allele ratio cutoff as ≥0.4 for an RNA transcript with high fetal or maternal contribution. This cutoff also took into consideration the Poisson distribution and random sampling of RNA-seq reads (details are described in Materials and Methods in the online Supplemental Data file). Using these criteria, 0.91% of the circulating transcripts were found to show high contribution by the fetus during early gestation (i.e., the first and second trimesters). This percentage increased to 2.52% during late gestation (i.e., the third trimester). On the other hand, 42.58% and 50.98% showed high contribution by the mother during early and late gestations, respectively (Fig. 1A; also see online Supplemental Table 4).

We further examined the 112 SNPs on RNA transcripts with high fetal contributions (see online Supplemental Table 4A) in the pre- and postdelivery maternal plasma samples. In this subset, 55.36% of the SNPs were devoid of RNA-seq reads in the postdelivery maternal plasma samples. As for the remaining 44.64%, fetal-specific alleles were not detected in the postpartum maternal plasma samples (Fig. 1B). In the subset of 3002 SNPs with high maternal contribution (see online Supplemental Table 4A), 20.85% were undetected in the postpartum maternal plasma (Fig. 1B).

ALLELIC EXPRESSION PATTERNS OF RNA TRANSCRIPTS IN THE MATERNAL PLASMA

RNA-seq has been employed to examine the allelic expression pattern *(26)*. We postulated that the allelic expression pattern for a given gene would be retained when the RNA transcripts were released from the tissues into the circulation, and hence could be detected in the plasma. In this study, we have examined the allelic counts of 2 RNA transcripts, namely pregnancy-associated plasma protein A, pappalysin 1 (*PAPPA*)[5], a pregnancy-specific gene, and H19, imprinted maternally expressed transcript (non-protein coding) (*H19*), an imprinted maternally expressed gene *(27, 28)*.

For the *PAPPA* gene, we analyzed an SNP, rs386088, which contains a fetal-specific SNP allele (Table 1). The absence of *PAPPA* RNA-seq reads in the postdelivery plasma samples indicated that it was indeed pregnancy specific *(9)*. Of note, there was no statistically significant difference in the proportions of fetal-allele read counts between the predelivery maternal plasma and the placental RNA samples ($P = 0.320$, χ^2 test), indicating that the maternal plasma data reflected the biallelic expression pattern of *PAPPA* in the placenta.

We recently reported that the DNA methylation pattern of the imprinted maternally expressed *H19* gene in the placenta and the maternal blood cells could be detected by bisulfite DNA sequencing of maternal plasma DNA *(29)*. Here, we further examined whether the genomic imprinting status of the *H19* gene could be explored at the RNA level. We first focused on an SNP site in exon 1 of the *H19* gene, rs2839698, which contains a maternal-specific allele, i.e., AA in the fetus and AG in the mother. Only the G-allele was detected in the postdelivery maternal plasma (Table 2). Such a monoallelic pattern was in accordance with its linkage to the unmethylated G-allele on the rs4930098 SNP site in the imprinting control region (see Results in the online Supplemental Data file and online Supplemental Fig. 5). In the predelivery maternal plasma, while the G-allele was present, the A-allele, which was contributed by the placenta, was also detected (Table 2). In 3 other SNP sites, i.e., rs2839701, rs2839702, and

[5] Human genes: *PAPPA,* pregnancy-associated plasma protein A, pappalysin 1; *H19,* H19, imprinted maternally expressed transcript (non-protein coding); *STAT1,* signal transducer and activator of transcription 1, 91kDa; *GBP1,* guanylate binding protein 1, interferon-inducible; *HSD17B1,* hydroxysteroid (17-beta) dehydrogenase 1; *KRT18,* keratin 18; *GADD45G,* growth arrest and DNA-damage-inducible, gamma.

Table 2. RNA-seq read counts on SNP sites in *H19* RNA for case 9415.

H19 SNP rsID	DNA genotype		Placenta		Maternal blood cells		Predelivery maternal plasma		Postdelivery maternal plasma	
	Fetus	Mother	Shared	Maternal-specific	Shared	Maternal-specific	Shared	Maternal-specific	Shared	Maternal-specific
rs2839698[a]	AA	AG	3418 (A)	10 (G)	0 (A)	0 (G)	246 (A)	1679 (G)	0 (A)	808 (G)
rs2839701	GG	CG	4959 (G)	8 (C)	0 (G)	0 (C)	33 (G)	219 (C)	1 (G)	78 (C)
rs2839702	CC	AC	4996 (C)	8 (A)	0 (C)	0 (A)	38 (C)	233 (A)	2 (C)	92 (A)
rs3741219	GG	AG	2255 (G)	15 (A)	0 (G)	0 (A)	17 (G)	158 (A)	0 (G)	96 (A)

[a] The shared A-allele was imprinted while the maternal-specific G-allele was transcribed. Details of the methylation status are shown in online Supplemental Fig. 5.

rs3741219, which bear maternal-specific alleles, a similar allelic pattern was found, i.e., biallelic in the predelivery maternal plasma and monoallelic in the postdelivery maternal plasma (Table 2). It might be that the maternal-specific alleles on these SNP sites were in the same maternal haplotype, which was unmethylated and was therefore transcribed. Notably, *H19* RNA was not expressed in the maternal blood cells (Table 2), suggesting that the *H19* RNA molecules in the plasma were derived from maternal tissues/organs other than the blood cells. Nonplacental and nonfetal tissues that have been reported to show *H19* expression included the adrenal gland, skeletal muscles, uterus, adipocytes, liver, and pancreas *(30)*.

IDENTIFICATION OF PREGNANCY-ASSOCIATED TRANSCRIPTS IN MATERNAL PLASMA TRANSCRIPTOMES

A subset of circulating RNA transcripts bearing fetal-specific alleles completely disappeared from the maternal plasma after delivery (Fig. 1B). These transcripts were considered to be fetal-specific in maternal plasma. On the other hand, a portion of maternal-specific alleles was also undetectable after delivery. Thus, we explored genes that showed upregulation during pregnancy, termed pregnancy-associated genes, by directly comparing their representation in the pre- and postdelivery maternal plasma. We defined pregnancy-associated genes as those that were detected in the predelivery plasma of the third-trimester pregnant women and those in which the postpartum plasma level was decreased by ≥2-fold. By using a bioinformatics algorithm for data normalization and differential gene expression analysis (for details see Materials and Methods in the online Supplemental Data file), we compiled a list of 131 pregnancy-associated genes (see online Supplemental Table 7). Among these genes, 15 were previously reported to be pregnancy specific in maternal plasma *(2–5, 9, 12, 14)*. Using one-step real-time reverse transcriptase-PCR, we have further validated the pregnancy association of 5 newly identified transcripts, which were abundant in predelivery maternal plasma, i.e., signal transducer and activator of transcription 1, 91kDa (*STAT1*), guanylate binding protein 1, interferon-inducible (*GBP1*), and hydroxysteroid (17-beta) dehydrogenase 1 (*HSD17B1*) in 10 additional plasma samples from third-trimester pregnant women, as well as keratin 18 (*KRT18*) and growth arrest and DNA-damage-inducible, gamma (*GADD45G*) in 10 plasma samples from another cohort of third-trimester pregnant women (see online Supplemental Fig. 6).

To assess the association of these 131 genes with pregnancy, hierarchical clustering was performed for all the plasma samples. A clear separation was observed between the plasma samples from pregnant women (i.e., early and late pregnancy) and those not associated with ongoing pregnancy (i.e., nonpregnant controls and postdelivery women) (Fig. 2).

Interestingly, when the expression patterns of these 131 genes were compared between the plasma samples of the 2 late-pregnancy cases and their corresponding placenta and maternal blood cells, a closer resemblance was observed between the placenta and the predelivery plasma samples and between the maternal blood cells and the postpartum plasma samples (Fig. 3A). This observation supports the thesis that most pregnancy-associated genes in maternal plasma are preferentially expressed in the placenta rather than in the maternal blood cells. Furthermore, the expression levels of these pregnancy-associated transcripts in the placentas and maternal plasma were positively correlated ($P < 1e{-}16$, Spearman correlation = 0.68) (Fig. 3B).

We further studied the 131 pregnancy-associated genes to determine whether they were derived from the fetus or the mother by analyzing the informative SNPs. The relative proportions of fetal and maternal contributions within these transcripts that contained informative SNPs were then computed. In late pregnancy, a larger proportion of these genes was found to exhibit

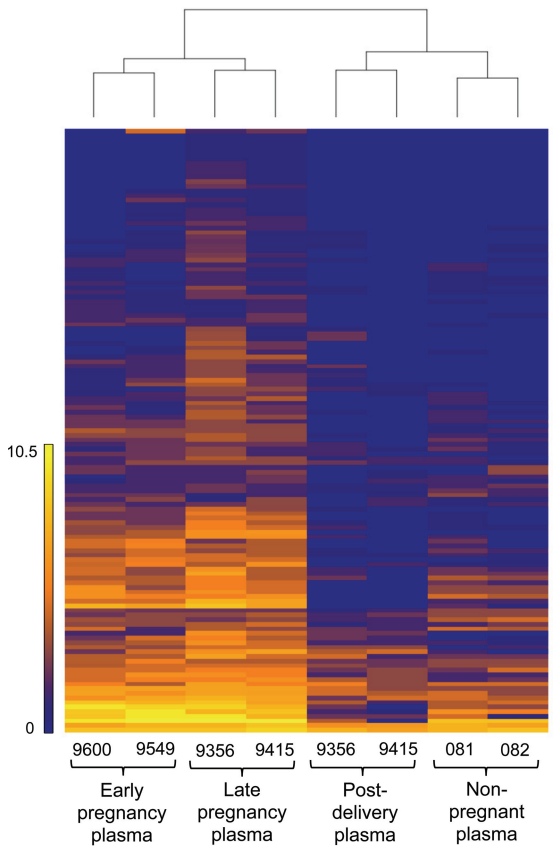

Fig. 2. Hierarchical clustering of plasma samples using the 131 pregnancy-associated genes.

By performing RNA-seq directly on the pre- and postdelivery maternal plasma, we identified 131 pregnancy-associated genes. When hierarchical clustering was performed using these 131 genes, the different plasma samples (identified by case numbers on the *x* axis) were correctly classified. Log_2(transcript levels) are shown.

predominant fetal contribution (Fig. 4A) than when all circulating transcripts in the plasma were considered (Fig. 1A). On the other hand, a smaller proportion of these 131 genes showed predominant maternal contribution than when all circulating transcripts were taken into account (Figs. 1A and 4A). These results suggested that the representations of placenta- or fetal-derived transcripts were enriched among transcripts that were deemed as pregnancy associated.

Although we were able to catalog a panel of pregnancy-associated genes through direct examination of the pre- and postdelivery maternal plasma samples, we also mined the placental and blood cell RNA-seq data for comparison purposes. Assuming that pregnancy-associated genes should be those that were expressed at a high level in the placenta and at a low level in the maternal blood cells *(3, 8)*, we arbitrarily set a 20-fold difference as a minimum cutoff for the tissue-based analysis. This tissue-based analysis yielded a total of 798 candidate genes, in which the proportions of fetal and maternal contributions in the maternal plasma were calculated. A relatively high proportion of genes with predominant fetal contribution was identified (Fig. 4B) compared to that of the full transcriptome (Fig. 1A). However, the plasma-based strategy outperformed the tissue-based strategy in being able to identify a higher proportion of genes with predominant fetal contribution (Fig. 4).

Discussion

In this work, we aimed to develop a technology for providing a global view of the maternal plasma transcriptome using RNA-seq. We have previously shown that the fractional fetal DNA concentration in maternal plasma can be calculated by targeting one or several fetal-specific loci because the whole fetal genome is evenly represented in the maternal plasma *(31)*. Unlike circulating DNA, measurement of the proportion of fetal-derived RNA transcripts in the maternal plasma is less straightforward because it is complicated by differential gene expression in the fetal and maternal tissues and perhaps their release into the circulation. By performing RNA-seq on maternal plasma and examining the polymorphic differences between the fetus and the mother, we were able to estimate the proportion of plasma transcripts contributed by the fetus. While maternal-derived transcripts dominated the plasma transcriptome, as one would anticipate, 3.70% and 11.28% of the circulating transcripts in the maternal plasma were derived from the fetus during early and late pregnancy, respectively. These fetal-derived transcripts included the RNA molecules contributed by both the fetus and the mother, as well as those contributed solely by the fetus. We found the latter to constitute 0.90% and 2.52% of the maternal circulating transcripts, during early and late pregnancy, respectively. The higher representation of such fetal-specific genes during late pregnancy is perhaps correlated with an increase in the size of the fetus and the placenta as pregnancy progresses.

In this study, we have demonstrated that a balanced RNA allelic expression of the pregnancy-specific *PAPPA* gene in the placenta and the monoallelic expression of the imprinted maternally expressed *H19* gene can be observed in the maternal plasma. These data suggest that the maternal plasma can be used as a noninvasive sample source for the study of allele expression patterns.

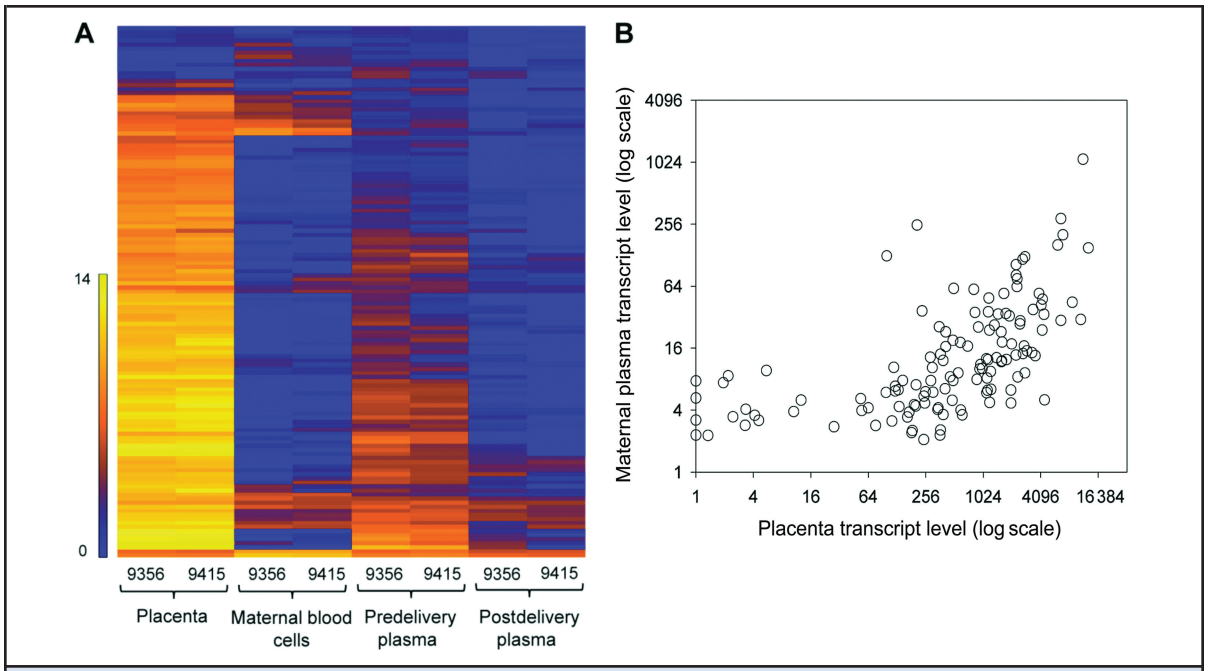

Fig. 3. Expression levels of 131 pregnancy-associated genes in the placenta, maternal blood cells, and maternal plasma.

(A), Heatmap of gene expression [$\log_2$(transcript levels)] of the placenta and maternal blood cells, as well as the pre- and postdelivery maternal plasma of the 2 late pregnancy cases (case numbers 9356 and 9415). The 131 pregnancy-associated genes were preferentially expressed in the placenta. (B), Expression levels of the 131 genes in the placenta and in plasma were positively correlated ($P < 1e\text{-}16$, Spearman correlation = 0.68).

By quantitative comparison of the RNA transcripts in the pre- and postdelivery maternal plasma samples, we have compiled a list of 131 genes that were upregulated during pregnancy, as evident by their reduced representation in the postpartum plasma samples. As expected, the profiles of these genes could be used to differentiate plasma samples of pregnant women from those of nonpregnant women. Such direct comparison of the pre- and postdelivery maternal plasma samples has allowed us to, in a high-throughput manner, sort out pregnancy-associated genes, which may not necessarily be expressed at a much higher level in the placenta than in the maternal blood cells, as demonstrated in previous work *(3, 8)*. In essence, this direct plasma examination method presents a powerful approach for the discovery of circulating pregnancy-associated RNA transcripts, without a priori knowledge of the transcriptomic profiles of the placental tissues and the blood cells.

Although we have shown that RNA-seq is a feasible method to profile the plasma transcriptome, several technical issues can be further improved. First, the information yield for plasma RNA-seq could be increased by further optimization of the sequencing protocol, particularly in depletion of the highly transcribed rRNA and globin genes from the plasma. Second, we have focused only on the reference transcripts and have not yet explored individual isoforms. Future studies could include detection of novel transcripts and differential analysis of the splicing variants and their isoforms *(32–34)* by increasing the sequencing read depth. Third, we have omitted ASE filtering in the analysis of proportions of fetal- and maternal-derived transcripts for the early pregnancy samples, because the chorionic villi and amniotic fluid had been exhausted for the genotyping analysis. Nonetheless, we have shown in late pregnancy that ASE filtering had no pronounced impact on the identification of genes with predominant fetal and maternal contributions in the maternal plasma (see online Supplemental Table 4, A and B). It is also worth mentioning that systematic errors associated with the MPS platforms, including but not limited to the Illumina system, have been reported *(35, 36)*. Thus, one would need to be mindful of the limitations of the adopted RNA-seq protocol and bioinformatics interpretation. Validation of the results using independent methodologies is of value.

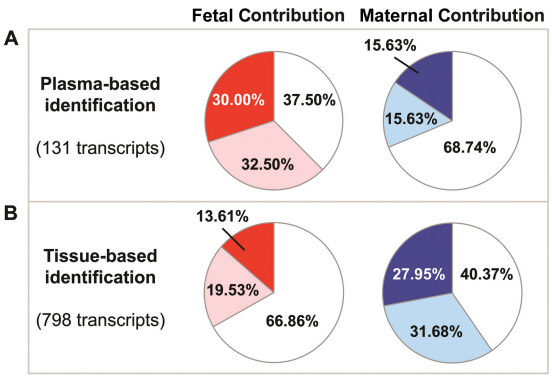

Fig. 4. Comparison between proportions of fetal and maternal contributions in the plasma.

Proportions of fetal and maternal contributions based on candidate transcripts identified by the plasma-based approach (A) and the tissue-based approach (B). The plasma-based approach had a higher fold-enrichment for proportion of fetal contribution. Red plus pink sectors indicate proportions of all plasma transcripts derived from the fetus; red sector indicates proportion of transcripts predominantly derived from the fetus; pink sector indicates proportion of transcripts derived from both the fetus and the mother. Dark blue plus light blue sectors indicate proportions of all plasma transcripts derived from the mother; dark blue sector indicates proportion of transcripts predominantly derived from the mother; light blue sector indicates proportion of transcripts derived from both the mother and the fetus.

In conclusion, we have demonstrated that RNA-seq technology can be used to measure the proportion of fetal-derived transcripts and to identify circulating pregnancy-associated genes in the maternal plasma. This study has paved the path toward better comprehension of the transcriptomic landscape of maternal plasma, hence facilitating the identification of biomarker candidates involved in pregnancy-related diseases. We envision that this technology will lead to new avenues for molecular diagnostics for pregnancy- or placenta-related diseases and also for other diseases such as cancer *(37)*.

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 author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: Y.M.D. Lo, *Clinical Chemistry*, AACC.
Consultant or Advisory Role: R.W.K. Chiu, Sequenom; Y.M.D. Lo, Sequenom.
Stock Ownership: R.W.K. Chiu, Sequenom; Y.M.D. Lo, Sequenom.
Honoraria: None declared.
Research Funding: R.W.K. Chiu and Y.M.D. Lo, University Grants Committee of the Government of the Hong Kong Special Administrative Region, China, under the Areas of Excellence Scheme (AoE/M-04/06) and Sequenom, Inc.; R.W.K. Chiu and Y.M.D. Lo, Sequenom Sponsored Research Agreement.
Expert Testimony: None declared.
Patents: N.B.Y. Tsui, P. Jiang, K.C.A. Chan, R.W.K. Chiu, and Y.M.D. Lo, US patent application 14/194,294.

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 Chan, Yoyo Jin, Coral Lee, Katherine Chow, and Xiaoxi Su for technical and bioinformatics assistance.

References

1. Poon LL, Leung TN, Lau TK, Lo YMD. Presence of fetal RNA in maternal plasma. Clin Chem 2000; 46:1832–4.
2. Ng EKO, Tsui NBY, Lau TK, Leung TN, Chiu RWK, Panesar NS, et al. mRNA of placental origin is readily detectable in maternal plasma. Proc Natl Acad Sci U S A 2003;100:4748–53.
3. Tsui NBY, Chim SSC, Chiu RWK, Lau TK, Ng EKO, Leung TN, et al. Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling. J Med Genet 2004;41:461–7.
4. Go AT, Visser A, Mulders MA, Blankenstein MA, Van Vugt JM, Oudejans CB. Detection of placental transcription factor mRNA in maternal plasma. Clin Chem 2004;50:1413–4.
5. Ng EKO, Leung TN, Tsui NBY, Lau TK, Panesar NS, Chiu RWK, Lo YMD. The concentration of circulating corticotropin-releasing hormone mRNA in maternal plasma is increased in preeclampsia. Clin Chem 2003;49:727–31.
6. Smets EM, Visser A, Go AT, van Vugt JM, Oudejans CB. Novel biomarkers in preeclampsia. Clin Chim Acta 2006;364:22–32.
7. Purwosunu Y, Sekizawa A, Koide K, Farina A, Wibowo N, Wiknjosastro GH, et al. Cell-free mRNA concentrations of plasminogen activator inhibitor-1 and tissue-type plasminogen activator are increased in the plasma of pregnant women with preeclampsia. Clin Chem 2007;53:399–404.
8. Miura K, Miura S, Yamasaki K, Shimada T, Kinoshita A, Niikawa N, et al. The possibility of microarray-based analysis using cell-free placental mRNA in maternal plasma. Prenat Diagn 2010;30:849–61.
9. Pang WWI, Tsui MHY, Sahota D, Leung TY, Lau TK, Lo YMD, Chiu RWK. A strategy for identifying circulating placental RNA markers for fetal growth assessment. Prenat Diagn 2009;29:495–504.
10. Chim SSC, Lee WS, Ting YH, Chan OK, Lee SW, Leung TY. Systematic identification of spontaneous preterm birth-associated RNA transcripts in maternal plasma. PLoS One 2012;7:e34328.
11. Ng EKO, El-Sheikhah A, Chiu RWK, Chan KCA, Hogg M, Bindra R, et al. Evaluation of human chorionic gonadotropin beta-subunit mRNA concentrations in maternal serum in aneuploid pregnancies: a feasibility study. Clin Chem 2004; 50:1055–7.
12. Lo YMD, Tsui NBY, Chiu RWK, Lau TK, Leung TN, Heung MMS, et al. Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. Nat Med 2007;13:218–23.
13. Tsui NBY, Wong BCK, Leung TY, Lau TK, Chiu RWK, Lo YMD. Non-invasive prenatal detection of fetal trisomy 18 by RNA-SNP allelic ratio analysis using maternal plasma SERPINB2 mRNA: a feasibility study. Prenat Diagn 2009;29:1031–7.
14. Wong BCK, Chiu RWK, Tsui NBY, Chan KCA,

Chan LW, Lau TK, et al. Circulating placental RNA in maternal plasma is associated with a preponderance of 5′ mRNA fragments: implications for noninvasive prenatal diagnosis and monitoring. Clin Chem 2005;51:1786–95.
15. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008; 5:621–8.
16. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 2008; 321:956–60.
17. Kim J, Zhao K, Jiang P, Lu ZX, Wang J, Murray JC, Xing Y. Transcriptome landscape of the human placenta. BMC Genomics 2012;13:115.
18. Wang K, Li H, Yuan Y, Etheridge A, Zhou Y, Huang D, et al. The complex exogenous RNA spectra in human plasma: an interface with human gut biota? PLoS One 2012;7:e51009.
19. Li H, Guo L, Wu Q, Lu J, Ge Q, Lu Z. A comprehensive survey of maternal plasma miRNAs expression profiles using high-throughput sequencing. Clin Chim Acta 2012;413:568–76.
20. Williams Z, Ben-Dov IZ, Elias R, Mihailovic A, Brown M, Rosenwaks Z, Tuschl T. Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. Proc Natl Acad Sci U S A 2013;110:4255–60.
21. Chim SSC, Shing TK, Hung EC, Leung TY, Lau TK, Chiu RWK, Lo YMD. Detection and characterization of placental microRNAs in maternal plasma. Clin Chem 2008;54:482–90.
22. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 2009;25: 1966–7.
23. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.
24. Pant PV, Tao H, Beilharz EJ, Ballinger DG, Cox DR, Frazer KA. Analysis of allelic differential expression in human white blood cells. Genome Res 2006;16:331–9.
25. Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. Genome Res 2011;21:1728–37.
26. Smith RM, Webb A, Papp AC, Newman LC, Handelman SK, Suhy A, et al. Whole transcriptome RNA-Seq allelic expression in human brain. BMC Genomics 2013;14:571.
27. Frost JM, Monk D, Stojilkovic-Mikic T, Woodfine K, Chitty LS, Murrell A, et al. Evaluation of allelic expression of imprinted genes in adult human blood. PLoS One 2010;5:e13556.
28. Daelemans C, Ritchie ME, Smits G, Abu-Amero S, Sudbery IM, Forrest MS, et al. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. BMC Genet 2010;11:25.
29. Lun FMF, Chiu RWK, Sun K, Leung TY, Jiang P, Chan KCA, et al. Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. Clin Chem 2013;59: 1583–94.
30. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol 2009; 10:R130.
31. 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.
32. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011;25:1915–27.
33. St Laurent G, Shtokalo D, Tackett MR, Yang Z, Eremina T, Wahlestedt C, et al. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. BMC Genomics 2012; 13:504.
34. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res 2012;22:2008–17.
35. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, et al. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res 2011;39:e90.
36. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. Genome Biol 2013;14: R51.
37. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer–a survey. Biochim Biophys Acta 2007;1775:181–232.