



Investigating Pregnancy and Its Complications Using Circulating Cell-Free RNA in Women's Blood During Gestation

Mira N. Moufarrej¹, Ronald J. Wong², Gary M. Shaw², David K. Stevenson^{2*} and Stephen R. Quake^{1*}

¹ Departments of Bioengineering and Applied Physics, Stanford University, and Chan Zuckerberg Biohub, Stanford, CA, United States, ² Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, United States

OPEN ACCESS

Edited by:

Iliana Bersani,
Bambino Gesù Children Hospital
(IRCCS), Italy

Reviewed by:

Joris Vermeesch,
KU Leuven, Belgium
Jill L. Maron,
Tufts Medical Center, United States

*Correspondence:

David K. Stevenson
dstevenson@stanford.edu
Stephen R. Quake
steve@quake-lab.org

Specialty section:

This article was submitted to
Neonatology,
a section of the journal
Frontiers in Pediatrics

Received: 11 September 2020

Accepted: 12 November 2020

Published: 14 December 2020

Citation:

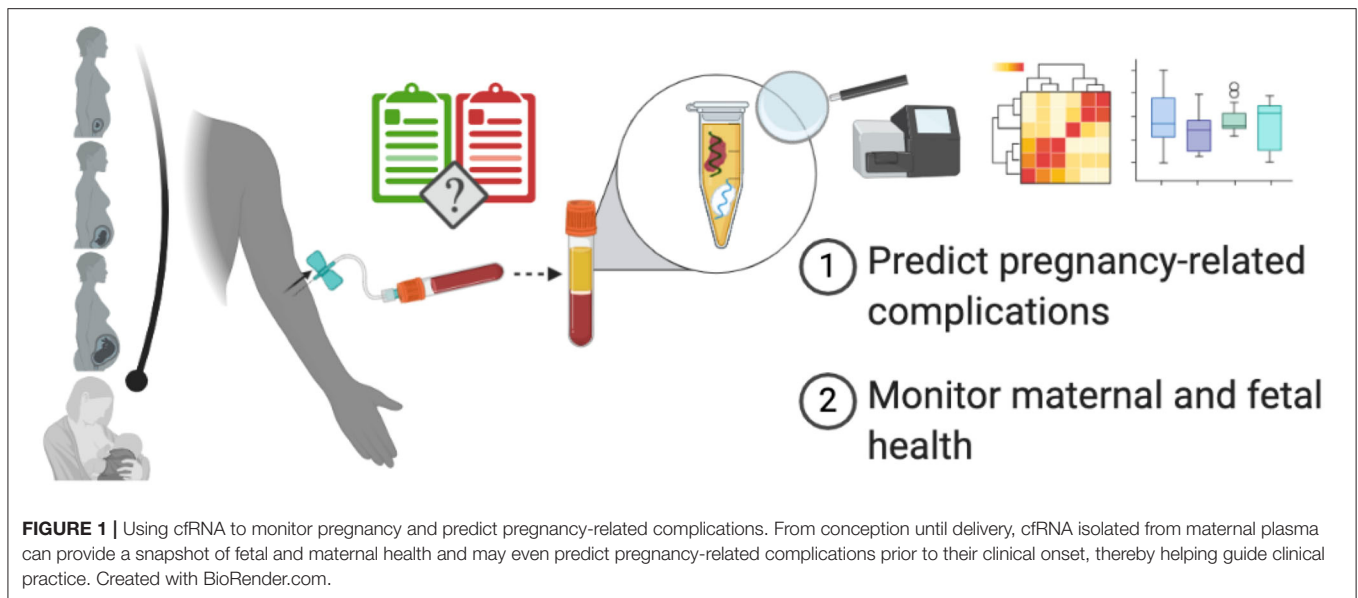
Moufarrej MN, Wong RJ, Shaw GM,
Stevenson DK and Quake SR (2020)
Investigating Pregnancy and Its
Complications Using Circulating
Cell-Free RNA in Women's Blood
During Gestation.
Front. Pediatr. 8:605219.
doi: 10.3389/fped.2020.605219

In recent years, there have been major advances in the application of non-invasive techniques to predict pregnancy-related complications, for example by measuring cell-free RNA (cfRNA) in maternal blood. In contrast to cell-free DNA (cfDNA), which is already in clinical use to diagnose fetal aneuploidy, circulating RNA levels can correspond with tissue-specific gene expression and provide a snapshot of prenatal health across gestation. Here, we review the physiologic origins of cfRNA and its novel applications and corresponding challenges to monitor fetal and maternal health and predict pregnancy-related complications.

Keywords: transcriptome, cell-free RNA, preeclampsia, prediction, preterm birth, IUGR

INTRODUCTION

For years, ultrasound has been the primary tool used to monitor fetal health allowing for the direct measurement of fetal shape, size, and some aspects of fetal physiology, all of which are important for development. It would be very useful to have a companion tool, which enables measurement of the molecular details of development, like the gene expression profile of the fetus over time. Liquid biopsies that rely on a simple maternal blood sample offer a possible avenue to realize this goal. Indeed, the measurement of circulating cell-free DNA (cfDNA) to screen for fetal aneuploidy as first demonstrated 12 years ago (1) has already seen wide-spread clinical adoption (2). More recently, in the past 5 years, we have witnessed unprecedented leaps in the applications of this technology to monitor prenatal health following its extension to measurements of circulating RNA in whole blood or cell-free RNA (cfRNA) in plasma. Today, circulating and cfRNA measurements have been applied to predict and characterize pregnancy-related complications like spontaneous preterm birth (PTB) (3), preeclampsia (4), and intrauterine growth restriction (IUGR) (5). In contrast to cfDNA levels, cfRNA levels of specific genes change during gestation in predictable ways that map closely with their placental and fetal gene expression (3, 4, 6–9). These timed patterns provide a snapshot of prenatal health across gestation and can be leveraged to predict pregnancy-related complications months before clinical diagnoses can be made and perhaps even used to monitor fetal health. Here, we review the physiology of cfRNA, its novel applications, and its challenges as they relate to fetal and maternal health (**Figure 1**).



PHYSIOLOGIC ORIGINS OF CELL-FREE NUCLEIC ACIDS DURING PREGNANCY

Circulating RNA in pregnant women was first identified by Mandel and Metais in 1948 (10)—long before the invention of reverse-transcription quantitative polymerase chain reaction (RT-qPCR) or next generation sequencing (NGS). Without the necessary means to interrogate the nucleic acids they identified, this discovery laid dormant until the 2000s when researchers leveraged RT-qPCR and NGS to answer five key questions about the physiologic origins of prenatal cfRNA. Specifically, they asked who (mom or fetus), what (RNA types), where (tissues), why (cellular activity), and when (clearance) contribute to prenatal cfRNA levels and in what fractions.

Several groups have now confirmed through orthogonal methods that both the fetus and mother contribute to measured cfRNA levels (3, 4, 6–9, 11, 12). Their fractional cfRNA contributions can be estimated by detecting exomic single-nucleotide variations (SNVs) at sites where either mother or child, but not both, is heterozygous. Although the mother and child may share both alleles across many genes (i.e., both heterozygous or both homozygous), we can still obtain reasonable, informative estimates of fractional contributions. Late in the first trimester, fetal-specific contributions comprise <1% of the total cfRNA measured (6). Later across gestation, fetal-specific contributions conservatively increase from <4% before 18 weeks of gestation to about 15% after 24 weeks (6, 9).

Furthermore, cfRNA from many types of RNA, including circular and non-coding, can be measured in the cf-transcriptome throughout pregnancy (6). As Koh et al. highlight, most fetal cfRNA is coding (i.e., mRNA), whereas about 15% appears to be non-coding (i.e., no open reading frame, long intergenic non-coding, or antisense that survives DNase treatment). The fractional contribution of microRNA across gestation as compared with other RNA types remains

unknown—owing to the unique technical challenges of isolating, amplifying, and sequencing both small and long RNA molecules.

To address where in the body prenatal cfRNA originates from, researchers used available body atlases (13) to identify tissue-specific genes that are also detected prenatally. Defining what constitutes “tissue-specific” expression remains an evolving question as more atlases become available (14–18). In general, it is defined as when expression of a gene in a single tissue can be measured exclusively or several fold (typically 2–20) above that from all other tissues. Using this definition, many genes detected prenatally originate from the placenta (9, 11, 19, 20); however, other tissues and body systems also contribute measurably like fetal tissues and the maternal immune system (3, 4, 7, 21, 22).

What specific cellular process—death, active signaling, or both—gives rise to cfRNA in circulation remains an open question (20, 23–25). In general, cfRNA is both protected from degradation and remarkably stable after blood collection (26). Protection can occur in the form of encapsulation in extracellular vesicles or apoptotic bodies suggesting both active secretion and cell death as possibilities, and most evidence available today remains indirect. Based on prenatal cfDNA and placental research, the majority of cfDNA arises from placental cell turnover, which occurs constantly (27, 28). Therefore, prenatal cfRNA may be released by the same placental cell turnover. Extending this idea to pregnancy-related complications, it is possible that placental-specific complications could result in higher cell death and consequently increased cfRNA release, although this remains an active area of research. Indeed, the connection between apoptosis and proliferation for both cfDNA as reviewed elsewhere (29), and by extension, cfRNA, points to apoptotic nucleic acid release as a means to understand both dying and living, proliferating cells. Recent convincing evidence outside the area of pregnancy also points to the possibility of active secretion (30). In longitudinal studies of cancer patients who underwent bone marrow ablation, Ibarra et al. found that

cf-mRNA captured active transcriptional activity. Specifically, they observed that bone marrow cell type-specific cfRNA level changes preceded parallel changes in the circulating cell count of the same cell type. These findings correspond with expected behavior during active secretion and exactly contrast behavior during cell death, where cfRNA level changes should follow those for cell counts.

Finally, there remains the question of cfRNA clearance or how long do RNA molecules remain in circulation following their release. In general, for healthy pregnancies, we expect all cell-free nucleic acids to clear within hours of delivery (3, 6, 8, 31). More specifically, it is known that certain placental cf-mRNAs have a half-life of about 14 min following delivery for healthy pregnancies (19), and most pregnancy-related cfRNA has been observed to clear from circulation within hours of delivery (6, 8, 11, 32). In contrast, women with preeclampsia (31) or women who underwent elective first-trimester surgical termination (33) both experienced delayed clearance of fetal cfDNA or cfRNA, respectively. Generally, the effect of pregnancy-related complications on cfRNA clearance remains an open question; however, we can hypothesize that placental-specific complications (like placenta accreta) would also result in the delayed clearance of placental-specific cfRNA molecules post-delivery when compared with healthy pregnancies.

LONGITUDINAL cfRNA CHANGES OVER UNCOMPLICATED, FULL-TERM GESTATION

Intriguingly, several groups have shown that cfRNA levels for specific pregnancy-related genes and circular RNAs change over gestation in predictable patterns (4, 6). Late in gestation, some of these changes map closely with placental gene expression (9) or fetal organ function (7). They may also reflect maternal immune system dynamics (8). Furthermore, prenatal changes in only nine placental genes predict gestational age (GA) for healthy pregnancies with comparable accuracy to ultrasound, providing a possible alternate method to accurately date pregnancies in the second and third trimesters (3). Importantly, most pregnancy-specific genes, typically those of placental or fetal organ origin, clear from circulation within 2 h of delivery (11, 32) with the exception of certain maternal immune genes that eventually return to measurable non-pregnant baseline levels (3, 8).

Longitudinal changes in gene expression follow three main trends: those that change continuously over pregnancy, those that change early in pregnancy and reach a plateau late in the second trimester, or, conversely, those that remain constant early in pregnancy only to change dramatically later in gestation (3, 4, 6, 8). In uncomplicated full-term pregnancies, the cfRNA profiles of placental, maternal immune, and fetal tissue-specific genes correlate with those of other genes specific to the same body system. Placental and fetal tissue-specific cfRNA levels also moderately correlate with each other (3). Despite successful measurement of fetal tissue-specific genes individually, their detection early in gestation can be variable, thereby impeding broad meaningful conclusions about fetal health and organ

development. Overall, their predictive power has yet to be thoroughly explored.

UNDERSTANDING AND PREDICTING PREGNANCY-RELATED COMPLICATIONS USING CIRCULATING RNA LEVELS

Until the last 2 years, research with the goal of predicting pregnancy-related complications has focused on measuring very few genes using RT-qPCR in small, single-hospital cohorts without external validation. These studies have been thoroughly reviewed elsewhere (20) and will not be our focus here. Recently, three groups have systematically identified (using RNA-sequencing, RNAseq) and separately validated (using RT-qPCR) changes in gene levels between matched cases and controls for spontaneous PTB (3), preeclampsia (4), and IUGR (5). The aforementioned studies all used samples from distinct collection sites for discovery and validation to increase generalizability. Nonetheless, each will require larger blinded validation prior to clinical adoption.

Prediction of spontaneous PTB has been reported in Black women at two sites in the USA, where women either presented with symptoms of early labor or experienced a prior spontaneous PTB in the discovery and validation cohorts, respectively. In validation, cfRNA levels for seven genes successfully identified women who delivered preterm up to 2 months in advance of delivery; however, both discovery and validation cohorts were small (3). Larger follow-up studies in ethnically diverse populations that include asymptomatic women will be required to understand the generality of these results.

Another study used cfRNA to characterize preeclampsia. They focused on women with early onset, defined as a diagnosis prior to 34 weeks of gestation and severe features. Blood was sampled at the time of diagnosis between 24 and 34 weeks of gestation. The cfRNA levels of 49 genes, of which 19 were specific to fetal or placental tissue, successfully identified women who experienced early-onset preeclampsia with an accuracy of 85–89% in a small, independent validation cohort (4). The ability of this test to predict preeclampsia prior to symptomatic onset remains unknown and will likely be crucial to understanding its clinical utility. Specifically, prediction of preeclampsia early in gestation may guide the prophylactic use of low-dose aspirin starting at or before 16 weeks of gestation, which has been shown to significantly reduce the risk of preterm preeclampsia (34).

Finally, Hannan et al. (5) focused on identifying pregnancies with severe IUGR using whole blood sampled at the time of diagnosis. In validation, they showed that two (EMP1 and PGM5) of the five genes identified during discovery significantly differed between women carrying significantly growth-restricted fetuses and women with uncomplicated pregnancies at the same GA. Like Munchel et al. (4), the ability of this test to identify pregnancy-related complications prior to symptomatic onset remains unknown. Here, the authors suggest that these genes could be used to monitor IUGR after its identification via ultrasound; however, what further clinical information these

genes will provide if used as a monitoring tool remains to be determined.

Other studies have also reported on differences between early-onset preeclampsia (35), women with adverse pregnancy outcomes in general (36), or women at high risk of experiencing preeclampsia (37) as compared with uncomplicated full-term pregnancies; however, they either did not include results for a separate validation cohort (35, 36) or only included limited results. Tsang et al. (35) identified specific placental cell-type signatures using single-cell sequencing and then measured these signatures in maternal plasma. They found that the signature for extravillous trophoblasts was elevated at diagnosis for women with early-onset preeclampsia as compared with GA-matched controls. Separately, Del Vecchio et al. (36) investigated whether cfRNA measured prior to diagnosis could predict several pregnancy-related complications and identified five cfRNA molecules that could predict preeclampsia in the first term for a very small discovery cohort (five cases of preeclampsia). Without an external validation dataset, it is difficult to interpret the results of both of these small, but encouraging, studies. Finally, Srinivasan et al. (37) explored the utility of bivariate microRNA ratios to predict preeclampsia in high-risk women prior to diagnosis. They reported that certain microRNA ratios could distinguish between women who went on to experience preeclampsia and those who did not as early as 17 weeks into pregnancy; however, limited validation results were provided. Furthermore, as aforementioned, the ability of any test to predict preeclampsia prior to 16 weeks remains untested, but critical to guiding prophylactic aspirin use.

The field of prenatal cfRNA diagnostics has seen many exciting developments in the past 5 years; however, as noted by others previously (20), these results are still awaiting large-scale validation. Nonetheless, the inclusion of small, but independent, validation cohorts in three of these studies increases their likelihood of generalizability (38) and provides further systematic confirmation of the diagnostic value of measuring cfRNA levels.

EXPERIMENTAL CHALLENGES WHEN ISOLATING CIRCULATING RNA

As a short-lived molecule in its naked form, RNA can prove challenging to isolate. Although cfRNA specifically is known to be surprisingly stable in circulation, several experimental decisions may influence research conclusions including the sample type, collection tubes, and volumes as well as processing, extraction, and amplification protocols. Here, we will highlight current best practices to ensure high sample quality and review quality metrics that can be used to identify samples with unusually high levels of RNA degradation, DNA contamination, or ribosomal RNA. Generally, cfRNA isolated from plasma can be extracted and then prepared for RNAseq or RT-qPCR using some of the many commercially available kits as described below.

Circulating RNA can be isolated from whole blood, serum, or plasma—all of which require specific blood collection tubes. Tubes for whole blood collection like PAXgene, Qiagen, or

Tempus RNA contain stabilizing agents that cause maternal cell lysis, thereby increasing background signals and obscuring the specific origin—cellular or cell-free—of any RNA measured. Similarly, coagulation of the blood sample required prior to the isolation of serum also leads to cell lysis resulting in the same obfuscation as first identified with circulating DNA (39). On the other hand, plasma isolation from whole blood requires tubes with reagents that inhibit coagulation and stabilize blood cells, thereby reducing the likelihood of cell lysis. Tubes used for plasma isolation should contain either EDTA, ACD, Streck cfDNA, or Streck cfRNA solutions, but not heparin as the latter has been shown to inhibit reverse transcription (40). In sum, plasma is the ideal sample type to identify pregnancy-specific cfRNA alone and minimize any background from maternal blood cell lysis.

A major concern when using plasma has been the recommendation to isolate plasma immediately after whole blood collection to ensure sample quality. Recent work by Munchel et al. (4) has shown that overnight shipping and storage had no adverse effect on measured pregnancy-associated signals for any common tube used for plasma isolation. Furthermore, they showed a strong cfRNA level agreement broadly (Pearson's $R \geq 0.7$) and, more specifically, for prediction of early-onset preeclampsia across paired samples for which plasma was isolated 0 to 5 days post-collection of blood in Streck cfDNA tubes.

cfRNA can be successfully extracted and amplified from a minimum of 0.5 ml of plasma with strong agreement between technical replicates (30). In a direct comparison, Munchel et al. (4) recommended that 2–4 ml of plasma be used as starting material to minimize variability and maximize library complexity. Since this may not always be possible when using samples from biobanks, which may have limited quantities, findings by Ibarra et al. emphasize that even smaller initial volumes like 0.5 ml can be used.

Finally, extraction and amplification of sufficient cfRNA have persisted as challenges. Recent work has converged on several best practices. Like most RNA-based discovery work, cfRNA analysis can be separated into three parts—extraction, amplification, and downstream analyses.

Published work today relies on commercially available RNA extraction kits from Norgen Biotek Corp., Qiagen, or Thermo Fisher Scientific for example. Importantly, RNA extraction must be followed by DNA digestion as recently emphasized in a re-analysis (41, 42) of data produced by a novel technique, SILVER-seq, to isolate cfRNA from 3 μ l of serum that may have also isolated cfDNA (43). Furthermore, quality control measures should be taken to ensure DNA digestion occurs as expected. Following digestion, cfRNA can be cleaned and concentrated using commercially available kits like those from Zymo Research to a final volume of 15 μ l or less. The amount, size, and integrity of the isolated RNA can be confirmed using Agilent's RNA 6,000 Pico chip.

Following RNA extraction, samples may be prepared for sequencing or RT-qPCR. Library preparation for sequencing has converged on two key methods—total RNA sequencing with ribosomal RNA (rRNA) depletion (8) and exonic-focused (i.e., mRNA) sequencing via whole-exome enrichment (4, 30). The

appropriate method depends on desired downstream analyses. Total RNA sequencing permits simultaneous measurements of many RNA types (circular, non-coding, or coding) and non-human sequences, some of which map to human bacteria and viruses (8); however, even following rRNA depletion, in our and another group's experience, typically 10–20% of all reads will still map to ribosomal sequences (3, 4) and only about 20–85% (mean of 35%) will map to coding regions. On the other hand, about 90% of reads from whole-exome enrichment and sequencing map to coding genes, thereby trading off a high exonic mapping efficiency with loss of potential information from non-coding RNA or non-human sequences (4, 8). Best practices for RT-qPCR have not changed in some time and similarly rely on commercially available kits that use TaqMan probes.

METRICS TO QUANTIFY SAMPLE QUALITY

Even with best practices, sample quality can still vary. Several quality metrics can be used to identify samples with unusually high levels of RNA degradation, DNA contamination, and/or ribosomal RNA (44). Any sample for which at least one of these three metrics deviates from expected values described below should be considered a quality outlier and excluded from further analyses. In our experience, we find that these metrics are well-calibrated and quickly flag water samples, a negative control, as quality outliers. Indeed, when we visualize the principal component analysis (PCA) of cfRNA data from all samples, we find that the samples flagged as outliers by these metrics cluster separately from most samples and typically drive the first or second principal component, thereby introducing unwanted technical bias. Removing these sample outliers from any downstream analysis circumvents the introduction of such bias.

For degraded input material like cfRNA, the RNA integrity number (RIN) may not be informative. Instead, to assess RNA degradation, researchers can leverage that ribonucleases present in human plasma exhibit 5'-3' catalytic activity owing to poor 3' poly(A) cleavage (45). Specifically, by counting and annotating identified exons from RNAseq data, researchers can determine the fraction of genes for which all assigned reads come from the 3' exon (44). Based on our own work using total RNAseq for almost 700 samples, we expect sample quality such that around 25–40% (with a median of 28%) of all genes will contain reads from only the most 3' exon. Samples for which this metric is above 40%, the 95th percentile, likely contain significantly degraded cfRNA and may have been mishandled prior to plasma isolation.

To assess DNA contamination, one can estimate the intron-to-exon read ratio by counting the number of reads that map exclusively to intronic as compared with exonic sequences (44). Here, we expect most samples to have a ratio of <1 , indicating more exonic than intronic reads. This is not a perfect metric since non-coding RNA will inflate the ratio and can be mistaken for DNA contamination, but it remains useful as a rule of thumb. There are opportunities to develop more sophisticated informatic

subtractions of known transcribed regions of the genome, which will add more accuracy to this metric. In our samples, we generally expect an intron-to-exon ratio of around 0.20–2.85 with a median ratio of 0.60 and consider any values >3 , the 95th percentile estimated from nearly 700 cfRNA samples, to be extremely unusual and typically the result of very few exonic reads (i.e., a small denominator).

Finally, to estimate the rRNA fraction, researchers can count reads that map to known ribosomal sequences as compared with the total number of reads for a given sample. Here, we expect values of 10–20% for total RNAseq as previously noted and $<10\%$ for whole-exome enriched RNAseq (4, 8, 44).

COMPUTATIONAL CHALLENGES WHEN INTERPRETING CIRCULATING RNA MEASUREMENTS

Most computational methods for cfRNA data analyses mirror those used for any RNA-based application. RNAseq data pre-processing relies on tools to remove adapter sequences from reads (i.e., trimomatic or cutadapt), map reads (i.e., STAR), remove PCR duplicates (i.e., Picard or samtools), and finally count the number of reads that map to any given gene (i.e., htseq-count, featureCount, or STAR). RT-qPCR analysis requires far fewer pre-processing steps, namely, just normalization as described in more detail below. Subsequently, to identify genes associated with pregnancy-related complications, researchers frequently use methods like differential expression (i.e., limma-voom, DESeq2, or EdgeR) or machine learning (i.e., sklearn). Despite the numerous pre-processing steps in common between cfRNA and general RNA analyses, there remain several important differences as well.

Unlike RNA measurements from specific tissues, cfRNA levels represent a mixture of RNA from many tissues and both mother and fetus for any given gene (6). Interpreting any specific gene result can be especially difficult and, for some measures, relies on the completeness of relevant atlases. The recent rising popularity of single-cell techniques underscores the importance of contextually defining tissue or cell “specificity.” For cfRNA present in circulation, identifying genes as cell-specific requires that they be unique in expression both for an individual tissue and across adult and fetal tissues. Importantly, special attention should be paid to the specificity of any RNA molecule in both adult and fetal tissues as it is possible that an otherwise “tissue-specific” RNA molecule in atlases that examined adult tissues alone could be expressed elsewhere in early development. Consequently, applying single-cell data from one organ to interpret plasma cfRNA as described previously (35) may be exciting but premature, as the specificity of cell-specific gene signatures in the context of the whole body prenatally remains unknown.

Furthermore, standard practice when analyzing cellular RNA may not apply as expected to cfRNA since it represents a mixture of RNA levels from across the body. One such example is the use of reference genes to normalize RNA levels across samples.

When planning RT-qPCR experiments and choosing an appropriate number of reference genes, we point the reader to the MIQE guidelines (46). In the context of prenatal cfRNA, researchers should be especially cognizant that traditional reference gene levels like those for GAPDH may be variable across people and gestation, and one group has reported on more stable reference genes to use when studying pregnancy (47). Moreover, reference genes can be avoided by converting cycle thresholds to molecular counts with a standard curve (3). In all inquiries, initial plasma volumes should also be normalized for if they differ across samples.

Finally, cfRNA discovery work is also plagued by the same issues as any method that seeks to identify predictive signatures (48, 49). Building a classifier should pull exclusively from discovery data with validation data used only for testing as opposed to further feature selection or model training. In practice, applying such a rigid boundary can be tricky but necessary as validation metrics may otherwise prove overly positive. Concretely, building a model during discovery with a set of genes, but refining it in validation to only include a subset that showed significance again (5), falls into this paradigm. In these cases, although the individual gene changes hold, the aggregate reported statistics for the validation set are likely to be overly optimistic.

FUTURE APPLICATIONS

Recent years have seen several proof-of-concept studies, which suggest that cfRNA can be a powerful diagnostic tool. However, owing to the study scale required to prospectively collect enough samples from women with pregnancy-related complications, all of the highlighted studies used at least some samples collected at onset of symptoms or at diagnosis. While encouraging, this present emphasis on prediction at diagnosis cannot replace the required work to show that cfRNA diagnostics can also anticipate complication onset and consequently allow for clinical intervention. To address this, large-scale, prospective studies hold promise as they can both confirm prediction at and prior to diagnosis.

In contrast to several studies so far that have focused on cf-mRNA, future applications may explore the predictive power of different types of RNA. Data from total RNA-sequencing like that from Koh et al. (6) have revealed that other RNA types, like circular RNA or non-coding RNA, can also change over gestation in predictable patterns. Furthermore, work like that by Srinivasan et al. (37) points to the potential utility of microRNAs to predict risk of preeclampsia. Altogether, novel methods that capture both small and long RNA molecules could provide an opportunity to directly compare the clinical utility of many RNA types to predict pregnancy-related complications.

Future applications may also benefit from incorporating measurements from other nucleic acids entirely like cfDNA or

other molecules like proteins and metabolites as has recently proven informative for prostate cancer (50) and PTB prediction (51), respectively. One recent report that examined both cfDNA methylation and cfRNA measurements for pregnancy-related complications found that in a small discovery cohort, placental-specific cfDNA increased prior to the subsequent development of gestational diabetes and that other cfRNA changes preceded preeclampsia (36). Withstanding validation and further exploration, these studies point to the possibility of incorporating cfDNA, proteins, metabolites, and other molecules with cfRNA in future prenatal applications as recently highlighted by Paquette et al. (52).

Separate from the prediction of pregnancy-related complications, another goal is to monitor fetal health in real-time. Despite encouraging results, researchers have thus far only identified a small number of fetal genes in circulation and noted the variabilities of such detection early in gestation when the fetus contributes <4% of the total cfRNA measured. Creating signatures, which combine information from groups of co-regulated genes may prove key to developing fetal tissue-specific signatures that can be reliably measured across gestation.

Furthermore, extension of such work from correlative observations to causative inference remains a major challenge not only for cfRNA analyses but also for RNA discovery work broadly. Here, we can imagine applying both cutting-edge computational and experimental methods like the construction of generative models or the application of correlative cfRNA insights to refine experimental hypotheses. Nonetheless, such experimental or computational work must be informed by our knowledge about the physiology of cfRNA. Here, one important question remains unanswered: What biological process—active secretion, cell death, or both—produces prenatal cfRNA?

Step by step, researchers have answered many important questions in a field that is only two decades old. Excitingly, groups over the past few years have clearly shown that cfRNA is not only a rich data source but also one with diagnostic potential. Yet still, key questions must be addressed for cfRNA diagnostics to be clinically used and impact prenatal care.

AUTHOR CONTRIBUTIONS

MNM, RJW, GMS, DKS, and SRQ conceptualized this review. MNM wrote the original manuscript draft and created the overview figure. All authors revised the manuscript and approved it for publication.

FUNDING

This work was supported in part by the Bill and Melinda Gates Foundation, the March of Dimes Prematurity Research Center at Stanford University, the Charles and Marie Robertson Foundation, the Christopher Hess Research Fund, the Charles B. and Ann L. Johnson Endowed Fund, the Chan Zuckerberg Biohub, and the Prematurity Research Fund.

REFERENCES

- Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci USA*. (2008) 105:16266–71. doi: 10.1073/pnas.0808319105
- Bianchi DW, Parker RL, Wentworth J, Madankumar R, Saffer C, Das AF, et al. DNA sequencing versus standard prenatal aneuploidy screening. *N Engl J Med*. (2014) 370:799–808. doi: 10.1056/NEJMoa1311037
- Ngo TTM, Moufarrej MN, Rasmussen MH, Camunas-Soler J, Pan W, Okamoto J, et al. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science*. (2018) 360:1133–6. doi: 10.1126/science.aar3819
- Munchel S, Rohrbach S, Randise-Hinchliff C, Kinnings S, Deshmukh S, Alla N, et al. Circulating transcripts in maternal blood reflect a molecular signature of early-onset preeclampsia. *Sci Transl Med*. (2020) 12:eaa0131. doi: 10.1126/scitranslmed.aaz0131
- Hannan NJ, Stock O, Spencer R, Whitehead C, David AL, Groom K, et al. Circulating mRNAs are differentially expressed in pregnancies with severe placental insufficiency and at high risk of stillbirth. *BMC Med*. (2020) 18:145. doi: 10.1186/s12916-020-01605-x
- Koh W, Pan W, Gawad C, Fan HC, Kerchner GA, Wyss-Coray T, et al. Noninvasive *in vivo* monitoring of tissue-specific global gene expression in humans. *Proc Natl Acad Sci USA*. (2014) 111:7361–6. doi: 10.1073/pnas.1405528111
- Maron JL, Johnson KL, Slonim D, Lai CQ, Ramoni M, Alterovitz G, et al. Gene expression analysis in pregnant women and their infants identifies unique fetal biomarkers that circulate in maternal blood. *J Clin Invest*. (2007) 117:3007–19. doi: 10.1172/JCI29959
- Pan W, Ngo TTM, Camunas-Soler J, Song CX, Kowarsky M, Blumenfeld YJ, et al. Simultaneously monitoring immune response and microbial infections during pregnancy through plasma cfRNA sequencing. *Clin Chem*. (2017) 63:1695–704. doi: 10.1137/clinchem.2017.273888
- Tsui NB, Jiang P, Wong YF, Leung TY, Chan KC, Chiu RW, et al. Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts. *Clin Chem*. (2014) 60:954–62. doi: 10.1373/clinchem.2014.221648
- Mandel P, Metais P. Les acides nucléiques du plasma sanguin chez l'homme. *C R Seances Soc Biol Fil*. (1948) 142:241–3.
- Ng EK, Tsui NB, Lau TK, Leung TN, Chiu RW, Panesar NS, et al. mRNA of placental origin is readily detectable in maternal plasma. *Proc Natl Acad Sci USA*. (2003) 100:4748–53. doi: 10.1073/pnas.0637450100
- Poon LL, Leung TN, Lau TK, Lo YM. Presence of fetal RNA in maternal plasma. *Clin Chem*. (2000) 46:1832–4. doi: 10.1093/clinchem/46.11.1832
- Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA*. (2004) 101:6062–7. doi: 10.1073/pnas.0400782101
- GTEX Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet*. (2013) 45:580–5. doi: 10.1038/ng.2653
- Sjostedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*. (2020) 367:eay5947. doi: 10.1126/science.aay5947
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. (2015) 347:1260419. doi: 10.1126/science.1260419
- Uhlen M, Karlsson MJ, Hober A, Svensson AS, Scheffel J, Kotol D, et al. The human secretome. *Sci Signal*. (2019) 12:eaa0274. doi: 10.1126/scisignal.aaz0274
- Uhlen M, Karlsson MJ, Zhong W, Tebani A, Pou C, Mikes J, et al. A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science*. (2019) 366:eaax9198. doi: 10.1126/science.aax9198
- Chiu RW, Lui WB, Cheung MC, Kumta N, Farina A, Banzola I, et al. Time profile of appearance and disappearance of circulating placenta-derived mRNA in maternal plasma. *Clin Chem*. (2006) 52:313–6. doi: 10.1373/clinchem.2005.059691
- Whitehead CL, Walker SP, Tong S. Measuring circulating placental RNAs to non-invasively assess the placental transcriptome and to predict pregnancy complications. *Prenat Diagn*. (2016) 36:997–1008. doi: 10.1002/pd.4934
- Tjoo ML, Jani J, Lewi L, Peter I, Wataganara T, Johnson KL, et al. Circulating cell-free fetal messenger RNA levels after fetoscopic interventions of complicated pregnancies. *Am J Obstet Gynecol*. (2006) 195:230–5. doi: 10.1016/j.ajog.2006.01.041
- Wataganara T, LeShane ES, Chen AY, Borgatta L, Peter I, Johnson KL, et al. Plasma gamma-globin gene expression suggests that fetal hematopoietic cells contribute to the pool of circulating cell-free fetal nucleic acids during pregnancy. *Clin Chem*. (2004) 50:689–93. doi: 10.1373/clinchem.2003.030064
- Huppertz B, Kingdom J, Caniggia I, Desoye G, Black S, Korr H, et al. Hypoxia favours necrotic versus apoptotic shedding of placental syncytiotrophoblast into the maternal circulation. *Placenta*. (2003) 24:181–90. doi: 10.1053/plac.2002.0903
- Kosaka N, Yoshioka Y, Hagiwara K, Tominaga N, Katsuda T, Ochiya T. Trash or treasure: extracellular microRNAs and cell-to-cell communication. *Front Genet*. (2013) 4:173. doi: 10.3389/fgene.2013.00173
- Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol*. (2007) 9:654–9. doi: 10.1038/ncb1596
- Tsui NB, Ng EK, Lo YM. Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin Chem*. (2002) 48:1647–53. doi: 10.1093/clinchem/48.10.1647
- Heazell A, Harris L, Forbes K, Crocker I. Placental cell turnover in health and disease. *Rev Gynaecol Perinatal Pract*. (2006) 6:80–6. doi: 10.1016/j.rigapp.2005.12.003
- Taglauer ES, Wilkins-Haug L, Bianchi DW. Review: cell-free fetal DNA in the maternal circulation as an indication of placental health and disease. *Placenta*. (2014) 35:S64–68. doi: 10.1016/j.placenta.2013.11.014
- Heitzer E, Auinger L, Speicher MR. Cell-free DNA and apoptosis: how dead cells inform about the living. *Trends Mol Med*. (2020) 26:519–28. doi: 10.1016/j.molmed.2020.01.012
- Ibarra A, Zhuang J, Zhao Y, Salathia NS, Huang V, Acosta AD, et al. Non-invasive characterization of human bone marrow stimulation and reconstitution by cell-free messenger RNA sequencing. *Nat Commun*. (2020) 11:400. doi: 10.1038/s41467-019-14253-4
- Bianchi DW, Maron JL, Johnson KL. Insights into fetal and neonatal development through analysis of cell-free RNA in body fluids. *Early Hum Dev*. (2010) 86:747–52. doi: 10.1016/j.earlhumdev.2010.08.001
- Morisaki S, Miura K, Higashijima A, Abe S, Miura S, Hasegawa Y, et al. Effect of labor on plasma concentrations and postpartum clearance of cell-free, pregnancy-associated, placenta-specific microRNAs. *Prenat Diagn*. (2015) 35:44–50. doi: 10.1002/pd.4479
- Jeong YJ, Borgatta L, Kapp N, Peter I, Bianchi DW, Johnson KL. Short-term clearance of cell-free nucleic acids after first-trimester termination of pregnancy. *Fertil Steril*. (2007) 88:730–3. doi: 10.1016/j.fertnstert.2006.11.143
- Roberge S, Villa P, Nicolaides K, Giguere Y, Vainio M, Bakthi A, et al. Early administration of low-dose aspirin for the prevention of preterm and term preeclampsia: a systematic review and meta-analysis. *Fetal Diagn Ther*. (2012) 31:141–6. doi: 10.1159/000336662
- Tsang JCH, Vong JSL, Ji L, Poon LCY, Jiang P, Lui KO, et al. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc Natl Acad Sci USA*. (2017) 114:E7786–95. doi: 10.1073/pnas.1710470114
- Del Vecchio G, Li Q, Li W, Thamotharan S, Tosevska A, Morselli M, et al. Cell-free DNA methylation and transcriptomic signature prediction of pregnancies with adverse outcomes. *Epigenetics*. (2020). doi: 10.1080/15592294.2020.1816774. [Epub ahead of print].
- Srinivasan S, Treacy R, Herrero T, Olsen R, Leonardo TR, Zhang X, et al. Discovery and verification of extracellular miRNA biomarkers for non-invasive prediction of pre-eclampsia in asymptomatic women. *Cell Rep Med*. (2020) 1:100013. doi: 10.1016/j.xcrm.2020.100013
- Fang Z, Cui X. Design and validation issues in RNA-seq experiments. *Brief Bioinform*. (2011) 12:280–7. doi: 10.1093/bib/bbr004
- Teare MD, Woll PJ. Genomic tests: unreliable for cancer? A focus on circulating DNA and lung cancer. *Expert Rev Mol Diagn*. (2007) 7:699–702. doi: 10.1586/14737159.7.6.699
- Willems M, Moshage H, Nevens F, Fevery J, Yap SH. Plasma collected from heparinized blood is not suitable for HCV-RNA detection

- by conventional RT-PCR assay. *J Virol Methods*. (1993) 42:127–30. doi: 10.1016/0166-0934(93)90184-S
41. Verwilt J, Trypsteen W, Van Paemel R, De Preter K, Giraldez MD, Mestdagh P, et al. When DNA gets in the way: a cautionary note for DNA contamination in extracellular RNA-seq studies. *Proc Natl Acad Sci USA*. (2020) 117:18934–6. doi: 10.1073/pnas.2001675117
 42. Zhou Z, Wu Q, Yan Z, Zheng H, Chen CJ, Liu Y, et al. Reply to Verwilt et al. experimental evidence against DNA contamination in SILVER-seq. *Proc Natl Acad Sci USA*. (2020) 117:18937–8. doi: 10.1073/pnas.2008585117
 43. Zhou Z, Wu Q, Yan Z, Zheng H, Chen CJ, Liu Y, et al. Extracellular RNA in a single droplet of human serum reflects physiologic and disease states. *Proc Natl Acad Sci USA*. (2019) 116:19200–08. doi: 10.1073/pnas.1908252116
 44. Pan W. *Development of Diagnostic Methods Using Cell-Free Nucleic Acids*. Stanford University (2016).
 45. Sorrentino S. The eight human “canonical” ribonucleases: molecular diversity, catalytic properties, and special biological actions of the enzyme proteins. *FEBS Lett*. (2010) 584:2194–00. doi: 10.1016/j.febslet.2010.04.018
 46. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. (2009) 55:611–22. doi: 10.1373/clinchem.2008.112797
 47. Chim SSC, Wong KKW, Chung CYL, Lam SKW, Kwok JSL, Lai CY, et al. Systematic selection of reference genes for the normalization of circulating RNA transcripts in pregnant women based on RNA-Seq data. *Int J Mol Sci*. (2017) 18:1709. doi: 10.1101/165654
 48. Ambrose C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci USA*. (2002) 99:6562–6. doi: 10.1073/pnas.102102699
 49. Smialowski P, Frishman D, Kramer S. Pitfalls of supervised feature selection. *Bioinformatics*. (2010) 26:440–3. doi: 10.1093/bioinformatics/btp621
 50. Fettke H, Kwan EM, Docanto MM, Bukczynska P, Ng N, Graham LK, et al. Combined cell-free DNA and RNA profiling of the androgen receptor: clinical utility of a novel multianalyte liquid biopsy assay for metastatic prostate cancer. *Eur Urol*. (2020) 78:173–80. doi: 10.1016/j.eururo.2020.03.044
 51. Peterson LS, Stelzer IA, Tsai AS, Ghaemi MS, Han X, Ando K, et al. Multiomic immune clockworks of pregnancy. *Semin Immunopathol*. (2020) 42:397–412. doi: 10.1007/s00281-019-00772-1
 52. Paquette AG, Hood L, Price ND, Sadovsky Y. Deep phenotyping during pregnancy for predictive and preventive medicine. *Sci Transl Med*. (2020) 12:eaay1059. doi: 10.1126/scitranslmed.aay1059
- Conflict of Interest:** SQ is a founder and shareholder of Mirvie, and Stanford University and the Chan Zuckerberg Biohub have filed patents based on the work of SQ and MM on the use of cfRNA in maternal and fetal health.
- The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2020 Moufarrej, Wong, Shaw, Stevenson and Quake. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.