Simultaneously monitoring immune response and microbial infections during pregnancy through plasma cfRNA sequencing

Wenying Pan[1], Thuy Ngo[1], Joan Camunas-Soler[1], Chunxiao Song[1], Mark A. Kowarsky[1], Yair J. Blumenfeld[2], Ronald J. Wong[3], Gary M. Shaw[3], David K. Stevenson[3,4], Stephen R. Quake[1]

[1]Department of Bioengineering, Stanford University, Stanford, CA 94305,

[2]Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Stanford University, Stanford, CA 94305.

[3]March of Dimes Prematurity Research Center, Stanford University School of Medicine, Stanford, CA 94305;

[4]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305

Corresponding Author:

Stephen Quake

James H Clark Center E300

318 Campus Drive

Stanford CA 94305

Tel: 650-721-2195

Email:quake@stanford.edu

Keywords: cell-free RNA, prenatal diagnosis, blood microbiome

Abbreviations:

cfRNA: cell-free RNA

cfDNA: cell-free DNA

NITP: noninvasive prenatal testing

B19V: human parvovirus B19 virus

rRNA: ribosomal RNA

FDR: false discovery rate

Human Genes:

ANXA1: annexin A1

S100A8: S100 calcium binding protein A8

S100A9: S100 calcium binding protein A9

S100P: S100 calcium binding protein P

BPI: bactericidal/permeability-increasing protein

PGLYRP1: peptidoglycan recognition protein 1

CSH1: chorionic somatomammotropin hormone 1

CSH2: chorionic somatomammotropin hormone 2

CGA: glycoprotein hormones, alpha polypeptide

PSG1: pregnancy specific beta-1-glycoprotein 1

PAPPA: pappalysin 1

ALAS2: 5'-aminolevulinate synthase 2

GYPA: glycophorin A

ANK1: ankyrin 1

Abstract

Background:Plasma cell-free RNA (cfRNA) encompasses a broad spectrum of RNA species which can be derived from both human cells and microbes. Because cfRNA is fragmented and of low concentration, it has been challenging to profile its transcriptome using standard RNA-seq methods.

Methods: We assessed several recently developed RNA-seq methods on cfRNA samples. We then analyzed the dynamic changes of both the human transcriptome and the microbiome of plasma during pregnancy from 60 women.

Results: The cfRNA reflects a well-orchestrated immune modulation during pregnancy: a up-regulation of anti-inflammatory genes and an increased abundance of antimicrobial genes. We observed that the plasma microbiome remained relatively stable during pregnancy. The bacteria Ureaplasma shows an increased prevalence and elevated abundance at postpartum, which is likely to be associated with postpartum infection. We demonstrated that cfRNA-seq can be used to monitor viral infections. We detected a number of human pathogens in our patients, including an undiagnosed patient with a high load of human parvovirus B19 virus (B19V), which is known to be a potential cause of complications in pregnancy.

Conclusion: Plasma cfRNA-seq demonstrates the potential to simultaneously monitor immune response and microbial infections during pregnancy.

Introduction

In the last decade, cell-free nucleic acids have drawn significant attention in both academia and industry for their potential in non-invasive diagnosis, especially cancer detection and prenatal testing. Since its invention in 2008, cell-free DNA (cfDNA) based noninvasive prenatal testing (NITP) has been used widely by clinics for the blood screening of fetal aneuploidy(1). While cfDNA can reveal static genetic information about the fetus, cell-free RNA (cfRNA) can reflect the dynamic changes of different tissues during pregnancy in both the mother and the fetus. Our previous work shows that both placental and fetal mRNA are detectable in maternal plasma; demonstrating temporal variation during pregnancy that reflect the underlying developmental pathways of the fetus(2). In addition to the interaction between the fetus and the placenta, pregnancy also involves a complex interplay between the mother's immune system and microbiota. From the perspective of the mother's immune system, the fetus is an allograft that contains foreign antigens from the father. To prevent fetal rejection, the placenta secretes hormones to down-regulate the inflammatory response of the immune system(3). Dysregulation of the complex interplay between the immune system and microbiota might cause serious complications during pregnancy, such as preterm birth and preeclampsia. Nonetheless, it is challenging to monitor the mother's immune response and microbial infection simultaneously. The cfRNA in the maternal plasma contains transcripts from immune cells, placenta, fetal tissues and microbiota. We hypothesize here that global profiling of cfRNA in maternal blood using a next generation sequencing technique could simultaneously monitor immune response, infectious disease, and placental development during pregnancy.

Because cfRNA is derived from apoptotic bodies and exosomes, it is highly degraded and of low quantity. This makes it challenging to prepare sequencing libraries from cfRNA. The standard poly-A-based library preparation method is not applicable because of the low integrity. Instead, random primers are used for the reverse transcription. One downside of random priming-based amplification is that it amplifies the highly abundant, but undesirable, ribosomal RNA (rRNA). As a result, an additional rRNA removal step is usually required for random-priming based library preparation. Several RNA-seq library preparation methods have recently been proposed for low-quantity RNA samples, including: RNase H depletion of rRNA, exome capture of target cDNA, and CRISPR to remove ribosomal cDNA. However, these methods have not been applied to cfRNA samples and their relative merits have not been carefully assessed.

In this study, we first systematically assessed different RNA-seq library preparation methods for cfRNA samples. We then characterized the dynamic changes of the plasma transcriptome and microbiome during pregnancy. Finally, we did a proof-of-principle case study using cfRNA sequencing to monitor infectious disease during pregnancy.

Materials and Methods

Sample collection

Blood samples from pregnant women were collected from Stanford hospital under an Institutional Review Board-approved protocol. The blood samples were collected into EDTA-coated Vacutainer tubes. The blood sample were centrifuged at 1,600 × g for 10 minutes at 4 °C as soon as possible, usually within 30 minutes to 2 hours of collection, and the plasma was centrifuged at 16,000 × g for 10 minutes at 4 °C to remove residual cells.

cfRNA isolation

The cfRNA was extracted from 1 mL of plasma using a Plasma/Serum Circulating RNA and Exosomal Purification kit (Norgen, cat 42800). The DNA residual in the cfRNA was then digested using Baseline-ZERO™ DNase (Epicentre). The cfRNA was then purified using a RNA Clean & Concentrator-5 kit (Zymo), yielding 10 μl of cfRNA per sample.

Comparison of cfRNA-seq library preparation methods

We first evaluated four methods (ScriptSeq, RNase H, Exome and Clontech) for preparing sequencing libraries from cfRNA. The ScriptSeq method uses random primers to synthesize cDNA from total cfRNA (Illumina ScriptSeq v2 kit). The RNase H method depletes rRNA (NEBNext® rRNA Depletion Kit) from the total cfRNA using RNase H before the RNA library preparation (Illumina ScriptSeq v2 kit). The Clontech method removes ribosomal cDNA after reverse transcription using CRISPR technology (Clontech SMARTer Stranded Total RNA-Seq Kit - Pico). The Exome method captures the exome regions of the cDNA library using probe hybridization (Roche NimbleGen SeqCap EZ Library SR). Multiple biological replicates of plasma samples (1 ml per sample) from healthy donors were used for each method. 8 plasma samples were used for each of the ScriptSeq, Exome, and Clontech methods, and 2 samples were used for the RnaseH method. The input cfRNA for each of these libraries was isolated using the same cfRNA isolation methods described above, using 5 μl of cfRNA each. See more details in **Supplemental Material**.

Quantification of human transcriptome in cfRNA

For each library, more than 10 million sequencing reads (2x75 bp) were generated using a Illumina NextSeq sequencer. Reads were aligned to hg19 human transcriptome using the STAR(4) aligner. The read counts for each gene were calculated using the htseq-count(5) tool.

The number of reads mapped to each gene was normalized with the total number of reads mapped to human transcriptome.

Quantification of microbiome in cfRNA

Reads in the cfRNA libraries that failed to align to human transcriptome were extracted and aligned to the bacteriophage phi X174, which is used as a control sequence in the Illumina platform. The remaining reads were aligned against a custom database(6) of potential pathogens which encompasses viruses, bacteria, and fungi in the NCBI database using BLAST (2.2.28+). Alignments were required to have an identity of at least 90% across 90% of the bases of the query. For high-confidence reads that aligned to multiple microbes, a customized Python script was used to find the lowest common ancestor in the taxonomical tree for each such read.

Detection of pathogenic viruses from cfRNA

The virus species belonging to known human pathogens were included in the analysis (Supplemental Table S**1**. For each sample, the number of reads that was uniquely aligned to these virus species were counted. In this study, each patient has multiple cfRNA samples collected from different time points. The prevalence was calculated at the patient level. For each patient, if there was at least one read aligned to a certain pathogenic species in any of her samples, this virus was counted as detected in this patient.

Quantification of microbiome from cfDNA

Plasma cfDNA libraries were prepared for selected samples with viral infections as a validation for cfRNA microbiome results. The method used for cfDNA sequencing library preparation and microbiome analysis was the same method described previously(6).

Calculation of the QC metrics of RNA-seq libraries

The QC metrics, including the rRNA rate, estimated library size and strand-specificity for each sample was calculated using the RNA-SeQC program(7).

Identify cfRNA genes with variations in abundance during pregnancy

204 plasma samples (**Supplemental Table S2**) were collected from 60 pregnant women at four time points: 1st trimester (T1: week 1 to weeks 13) , 2nd trimester (T2: week 13 to week 23), 3rd trimester (T3: week 23 to week 40), and postpartum (P: within 24 hours after delivery). The cfRNA-seq libraries were made using the ScriptSeq v2 kit. The libraries with poor sample quality (due to RNA degradation, DNA contamination or blood coagulation) were removed from the subsequent statistical analysis (**Supplemental Material**). Samples from women with preterm birth were also removed from the analysis. A generalized linear model (GLM) likelihood ratio test was used to determine the genes with variations in abundance between any pair of time points (T1, T2, T3, P). The GLM test was implemented using the edgeR version 3.12.0 package in R, using the glmFit() and glmLRT() functions.

qPCR validation

To validate the genes with temporal variations identified by RNA-seq, qPCR was performed on the same sets of plasma samples using Fluidigm Biomark system. See more details in **Supplemental Material**.

Results

Assessing different methods for cfRNA-seq

We evaluated the four cfRNA-seq library preparation methods (ScriptSeq, RNase H, Exome and Clontech) by comparing the rRNA rate (the fraction of reads mapped to rRNA among all reads

mapped to human genome), microbiome fraction (the fraction of reads aligned to microbial organisms among all aligned reads), estimated library size (the number of expected fragments based on the total reads and duplication rate assuming a Poisson distribution) and the number of genes detected (**Figure 1** and **Supplemental Table S3**). All of the values discussed in the following section are the average value among the biological replicates.

The rRNA rate is as high as 0.66 when preparing the RNA-seq library directly from total cfRNA (Scriptseq). The other three methods all include a rRNA removal step which results in reduced rRNA rates of different levels (0.24 for RNase H, 0.05 for Exome and 0.10 for Clontech). The Scriptseq, Exome and Clontech methods can recover approximately 9000 protein coding genes, while the RNase H method can only detect approximately 3700 genes. The ScriptSeq and Clontech methods are also able to recover a broad spectrum of noncoding RNAs including lincRNA, snRNA, miscRNA, etc. Besides the dominant reads of human origin, a small fraction of reads were mapped to microbial genomes (0.02 for ScriptSeq, 0.16 for RNase H, 0.18 for Clontech and 0.009 for Exome libraries). For these methods the estimated library size (0.64 million for ScriptSeq, 0. 31 million for Clontech, 0.29 million for Exome and 0.06 million for RNase H) is very small, which correlates to the scarce amount of cfRNA in plasma.

Based on the above analysis, the RNase H method has low sensitivity for cfRNA-seq library preparation. The decrease in the number of genes detected and the estimated library size in the RNase H method implies that a large number of mRNA transcripts were lost during the RNase H depletion step, either due to non-specific hybridization of the ssDNA or RNA degradation. Both the Exome and Clontech methods were effective at the removal of rRNA and a comparable

number of mRNA transcripts was recovered. The Clontech method is a better option if the microbiome information is of interest. However, because of the small number of unique molecules in the cfRNA-seq library, it becomes possible to achieve sequencing saturation at an affordable sequencing depth even without rRNA depletion. After comparing the cost of the reagents for adding the rRNA depletion step with the cost for the extra depth of sequencing, we found that there is not a significant difference between the ScriptSeq and the Clontech methods in terms of the performance and cost. We chose to use the ScriptSeq method for the following pregnant women's study.

The dynamics of cfRNA reflects the immune response during pregnancy

We systematically examined the temporal variation of human plasma transcriptome during pregnancy. 39 genes (**Supplemental Table S4**) have been identified with variations in the abundance among the four time points (FDR<0.05). We applied hierarchical clustering to these genes and separated them into four major clusters, based on their temporal trends (**Figure 2A**).

For genes in cluster 1, the abundance of genes increased from the first trimester to the third trimester and remained at a higher level than the first trimester after delivery. These genes are specifically expressed in leukocytes and participate in the immune modulation of pregnancy. Two major types of immune response were identified based on gene ontology and pathway analysis: regulation of inflammatory process and increased abundance of endogenous antimicrobials. Genes involved in inflammatory regulation include annexin (*ANXA1*) and S100 calcium binding protein (*S100A8*, *S100A9* and *S100P*). Increased expression of annexin is one of

the mechanisms by which glucocorticoids inhibit inflammation(8). S100A8/A9 is also known to be involved in the migration of neutrophil migration to inflammatory sites(9).

A number of antimicrobial proteins and peptides showed an increased cfRNA abundance during pregnancy, including bactericidal/permeability-increasing protein (BPI) and peptidoglycan recognition protein (PGLYRP1). These are cytotoxic peptides or proteins synthesized mostly in neutrophils which participate in innate immune response to fight against bacteria, fungi and viruses.

Genes in cluster 2 are placenta-specific and only detectable during pregnancy. These genes participate in hormonal regulation (*CSH1*, *CSH2* and *CGA*), immune regulation (*PSG1*) and tissue remodeling (*PAPPA*). Most of these genes were reported in our previous study(2), and these results independently reproduced those findings.

Genes in cluster 3 have a decrease in abundance during the first trimester. Some of these genes (*ANK1* and *ALAS2*) are specific to erythrocytes (red blood cells). It is known that the blood volume increases more than the red blood cell mass during pregnancy, which results in a relative anemia. This might explain the decreased abundance of these erythrocyte-related genes in the plasma at the beginning of pregnancy. There are also several genes that have a lower abundance level in the third trimester. These genes are not enriched for a specific functional group or tissue of origin; their relevance to pregnancy needs investigation.

To prove that these temporal variations in abundance were not caused by RNA-seq technical artifacts, we selected a subset of the genes related to the immune response and pregnancy hormones for qPCR validation. The qPCR was performed on the same pregnant sample set used for cfRNA-seq analysis. The results show consistent temporal trends in qPCR compared to the original RNA-seq analysis (**Figure 2B** and **Supplemental Figure S2**).

The composition and temporal variations of plasma microbiome during pregnancy

After examining the transcriptome, we analyzed the microbiome from plasma cfRNA-seq. Microbial reads in the cfRNA library can be from the blood microbiota or contamination introduced during sample preparation. Negative controls (water) were prepared with the ScriptSeq method from the cfRNA extraction step. The microbial species detected in the negative controls were identified as contaminants and removed from the following microbiome analysis (**Supplemental Material**). The microbial reads from the plasma covered all four superkingdoms. **Figure 3A** displays the microbial read distribution. The percentage of each microbial group is the average value across all the samples.

Overall, the composition of the plasma microbiome remained relatively stable during pregnancy for both bacteria and viruses (**Figure 3B,C**), attesting to the well-orchestrated balance between immune system and microbiome. We performed a statistical test (ANOVA) to find microbial genera that exhibited a change in abundance across the four time points (T1, T2, T3 and P) of pregnancy. Ureaplasma showed an increased average abundance and higher prevalence at postpartum (Benjamini-Hochberg adjusted *P*-value 0.0384, **Figure 3D,E**). Two species of Ureaplasma - Ureaplasma parvum and Ureaplasma urealyticum - were detected in the cfRNA samples. Both species showed higher abundance and prevalence postpartum (**Supplemental**

Figure S3), and the majority of the reads in the Ureaplasma genus were aligned to Ureaplasma parvum. In a previous study, Ureaplasma was found in the blood stream of women with postpartum fever(10). The increased level of Ureaplasma at postpartum might be a reflection of microbial translocation into the maternal blood stream during delivery or an indication of Ureaplasma infection after delivery. Ureaplasma has an important clinical relevance with pregnancy complications. A recent study reported that the abundance of Ureaplasma in the vaginal microbiome is associated with preterm birth(11).

Monitoring viral infection during pregnancy

Viral infections in pregnancy are major causes of maternal and fetal morbidity and mortality. Despite the increased risk to pregnant women, due to their immune suppressed state, from a broad spectrum of viruses, only a small number of species are tested for in current routine pregnancy screening. We hypothesized that plasma cfRNA-seq can be used as an unbiased screening test for viral infection during pregnancy. A number of common pathogenic viruses were detected in our cfRNA samples, including Adenoviridae, Papillomaviridae, Hepesiviridae. Parvoviridae and Polyomaviridae (**Figure 4A**, **Supplemental Table S5**). In particular, Human parvovirus B19 (B19V) was detected from one pregnant plasma sample (patient 10039) using cfRNA-seq. The viral load was very high at the first trimester, and gradually decreased in the later stages of pregnancy. The temporal change of B19V load in this patient was also validated from cfDNA sequencing data (**Figure 4B**). B19V is known to attack erythroid progenitor cells and can trigger an acute cessation of erythrocyte production(12). Based on the cell-free transcriptomic data of the same patient, abundance of erythrocyte-specific genes (ALAS2 and GYPA) plunged at the time of strong B19V infection (**Figure 4C**). Parvovirus B19 virus

infection was not diagnosed for this pregnant woman according to her clinical record. The estimated incidence of parvovirus B19 virus infection in pregnancy ranges from 1-5%. Although B19V infection can be asymptomatic during pregnancy, 3-14% of infections lead to intrauterine fetal deaths(13). Fortunately, this woman recovered from the infection without clinical complications and there were no adverse outcomes for the fetus.

Discussion

The immune response during pregnancy from cfRNA is consistent with data from previous measurements

There has been longstanding interest in the role of the maternal immune system during pregnancy, and it is clear from a variety of lines of evidence that pregnancy acts as a complicated modulator of the immune system. From the perspective of cell count, absolute neutrophil counts are increased during pregnancy(14). Since neutrophils are a major site for the synthesis of antimicrobial peptides, this might contribute to the increase of antimicrobial gene transcripts in plasma cfRNA during pregnancy. However, it is also known that placental trophoblasts synthesize antimicrobial peptides; this illustrates the value of measuring cfRNA levels as an approach which integrates the overall contribution of antimicrobial peptides across multiple cell types and tissues. From the perspective of the whole blood transcriptome, the gene expression levels of *MMP8, ANXA1, S100A8* and *S100P* have been previously shown to be up regulated during pregnancy(15,16) and some of these immune related genes have been demonstrated as biomarkers for pregnancy complications. One recent study has shown that women with preeclamptic pregnancy have aberrantly high levels of Annexin protein in their blood(8).

Evidence from another pathophysiological study suggests that inflammatory proteins S100A8 and S100A9 play a role in the recurrent early pregnancy loss(17).

Sample Qualiy control

In the study of temporal change in plasma transcriptome and microbiome, two-thirds of the cfRNA-seq libraries did not pass the quality filering owing to RNA degradation, DNA contamination or blood coagulation. To increase the successful rate of sample processing in clinical applications, the method can be improved from the following aspects: 1) to avoid RNA degradation after the cfRNA extraction, better RNase-free treatment for the reagents and equipments can be implemented; 2) to avoid DNA contamination, the condition for DNA digestion in the cfRNA isolation step can be optimized; 3) to avoid blood coagulation, the blood samples can be processed within 2 hours of collection.

The composition of blood microbiota

Microbial nucleic acids in blood are normally detected using targeted methods such as RT-PCR and pan-viral or pan-microbial microarrays(18). To understand the systematic picture of microbial composition, unbiased deep sequencing is a possible solution. Recently, several efforts have been made toward the characterization of the microbiota in blood using 16S sequencing(19) or cfDNA sequencing(20). Even though there are differences between the genomic and transcriptomic measurements in terms of microbial abundance, the composition of bacteria observed in our study is consistent with the previous studies: the top phyla are Proteobacteira, Firmicutes and Actinobacteria(21).

Compared to our understanding of the bacterial microbiome, our knowledge about human fungal microbiota ("mycobiome") lags behind. Our method also surveyed the mycobiome in plasma. In this cohort of samples, all of the non-human eukaryotic reads in the cfRNA libraries were aligned to fungi. Three fungi species were detected: Candida albicans, Encephalitozoon hellem and Podospora anserina. Candida albicans is a common member of human gut flora(22). Encephalitozoon hellem can cause keratoconjunctivitis in immune compromised patients(23).

Recent studies have investigated temporal variation of the human microbiome during pregnancy at multiple body sites (vagina, distal gut, saliva and tooth gum)(11). Based on these studies, the microbiome at most body sites is stable during pregnancy, except for the vagina. However, the dynamics of plasma microbiome during pregnancy has not been explored before. From the results of our study, we conclude that the plasma microbiome, which integrates signal from many human body sites, is also relatively stable during pregnancy.

Monitoring infection during pregnancy

In the current clinical practice, only the following pathogens are routinely tested for in pregnant women: Rubella, Hepatitis B, Hepatitis C, HIV, Tuberculosis syphilis, Chlamydia and Group B streptococcus. Other pathogenic species that can cause fetal defects are not routinely tested, such as Human Parvovirus B19 virus, Herpes virus, Adenovirus, and Enteroviruses. Unbiased microbiome sequencing could potentially be used as a more comprehensive screening test. Shotgun sequencing of the cfDNA in blood has been used to monitor viral infection in organ transplant patients(6,20). However, this DNA-based method lacks the capability to detect RNA viral pathogens such as Norovirus, Hepatitis C virus (HCV) and Zika virus. In principle,

sequencing of cfRNA in plasma could detect both DNA and RNA viruses. Here, we demonstrate that asymptomatic viral infection could be detected using unbiased plasma cfRNA-seq. Our proof-of-principle study presents the potential use of plasma cfRNA-seq as a screening method for infectious diseases during pregnancy, especially for the asymptomatic and uncommon viruses.

In conclusion, we assessed the relative merits of different RNA-seq methods for cfRNA. We demonstrated that cfRNA-seq is able to monitor the pregnancy-related immune response and the blood-borne microbial infections in pregnant women.

Acknowledgements

We thank Norma Neff, Jennifer Okamoto and Gary Mantalas for assistance with sample preparation and sequencing. We thank Ronald Wong for help in sample collection and processing. We also thank Marina Sirota, Daniel DiGiulio and Brice Gaudilliere for helpful discussions. This work was supported by the March of Dimes.

References

1. 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 U S A 2008;105:16266–71.
2. 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 U S A 2014;111:7361–6.
3. Robinson DP, Klein SL. Pregnancy and pregnancy-associated hormones alter immune responses and disease pathogenesis. Horm Behav. 2012;62:263–71.
4. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
5. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166–9.
6. De Vlaminck I, Martin L, Kertesz M, Patel K, Kowarsky M, Strehl C, et al. Noninvasive monitoring of infection and rejection after lung transplantation. Proc Natl Acad Sci U S A 2015;112:13336–41.
7. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire M-D, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 2012;28:1530–2.
8. Perucci LO, Carneiro FS, Ferreira CN, Sugimoto MA, Soriani FM, Martins GG, et al. Annexin A1 Is Increased in the Plasma of Preeclamptic Women. PLoS One 2015;10:e0138475.

9. Ryckman C, Vandal K, Rouleau P, Talbot M, Tessier P a. Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. J Immunol. 2003;170:3233–42.

10. Chaim W, Horowitz S, David JB, Ingel F, Evinson B, Mazor M. Ureaplasma urealyticum in the development of postpartum endometritis. Eur J Obstet Gynecol Reprod Biol 2003;109:145–8.

11. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al. Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci. 2015;112:201502875.

12. Luo Y, Lou S, Deng X, Liu Z, Li Y, Kleiboeker S, et al. Parvovirus B19 infection of human primary erythroid progenitor cells triggers ATR-Chk1 signaling, which promotes B19 virus replication. J Virol 2011;85:8046–55.

13. de Jong EP, de Haan TR, Kroes ACM, Beersma MFC, Oepkes D, Walther FJ. Parvovirus B19 infection in pregnancy. J Clin Virol 2006;36:1–7.

14. Pitkin RM. Platelet and Leukocyte Counts in Pregnancy. JAMA J Am Med Assoc 1979;242:2696.

15. Heung MMS, Jin S, Tsui NBY, Ding C, Leung TY, Lau TK, et al. Placenta-derived fetal specific mRNA is more readily detectable in maternal plasma than in whole blood. PLoS One. 2009;4.

16. Mittal A, Pachter L, Nelson JL, Kjærgaard H, Smed MK, Gildengorin VL, et al. Pregnancy-Induced Changes in Systemic Gene Expression among Healthy Women and Women with Rheumatoid Arthritis. PLoS One 2015;10:e0145204.

17. Nair RR, Khanna A, Singh K. Role of inflammatory proteins S100A8 and S100A9 in

pathophysiology of recurrent early pregnancy loss. Placenta 2013;34:824–7.

18. Leung RK-K, Wu Y-K. Circulating microbial RNA and health. Sci Rep.2015;5:16814.

19. Païssé S, Valle C, Servant F, Courtney M, Burcelin R, Amar J, et al. Comprehensive description of blood microbiome from healthy donors assessed by 16S targeted metagenomic sequencing. Transfusion 2016;5:1138-1147.

20. De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, et al. XTemporal response of the human virome to immunosuppression and antiviral therapy. Cell 2013;155:1178–87.

21. Potgieter M, Bester J, Kell DB, Pretorius E. The dormant blood microbiome in chronic, inflammatory diseases. FEMS Microbiol Rev 2015;39:567–91.

22. Kumamoto CA. Inflammation and gastrointestinal Candida colonization. Curr Opin Microbiol 2011;14:386–91.

23. Didier ES, Didier PJ, Friedberg DN, Stenson SM, Orenstein JM, Vee RW, et al. Isolation and Characterization of a New Human Microsporidian, Encephalitozoon hellem (n. sp.), from Three AIDS Patients with Keratoconjunctivitis. J Infect Dis 1991;163:617–21.

Figure captions:

Figure 1: Comparison of QC metrics from different RNA-seq methods on cfRNA samples. A) rRNA rate: (the fraction of reads mapped to rRNA among all the reads mapped to human genome). B) Estimated library size (the number of expected fragments based on the total reads and duplication rate assuming a Poisson distribution). C) Microbiome fraction (the fraction of reads aligned to microbial organisms among all the aligned reads). D) The number of protein coding genes detected. E) The number of lincRNA genes detected. F) The number of snRNA genes detected. G) The number of miscRNA detected.

Figure 2: Human genes displaying temporal changes during pregnancy in cfRNA. A) The genes with variations in abundance were grouped into four clusters based on their temporal patterns. The "tissue of origin" color bar annotates the origin of each gene based on tissue specificity. The "Gene function" color bar annotates the function of each gene. B) qPCR validation of a subset of the genes with variations in abundance.

Figure 3: The plasma microbiome during pregnancy. A) i: The fraction of reads aligned to different superkindoms of microbiome. ii: The distribution of bacterial reads at phylum level. iii: The distribution of viral reads. B) The composition of bacteria at four time points of pregnancy (1^{st}, 2^{nd}, 3^{rd} and postpartum). C) The composition of viruses at four different time points of pregnancy (1^{st}, 2^{nd}, 3^{rd} and postpartum). The reads count represents the average value across all the samples at that time point. D) The abundance of Ureaplasma over the time course of pregnancy. The reads count represents the average value across all the samples at that time point. E) The prevalence of Ureaplasma at four different time points of pregnancy. All the value in Figure A is the sample average of all the pregnant women's cfRNA samples passed sample quality filtering. The composition in Figure B and C is the sample average after removing the patient (10039) with severe viral infection.

Figure 4: Detection of pathogenic viruses in plasma. A) The prevalence of detectable pathogenic viruses among our pregnant women cohort at patient level. B) The abundance of Human parvovirus B19 virus in one patient (10039) during pregnancy from both cfRNA and cfDNA. c). The abundance of human erythrocytes-specific genes in cfRNA during pregnancy for the same patient with Human parvovirus B19 virus infection.

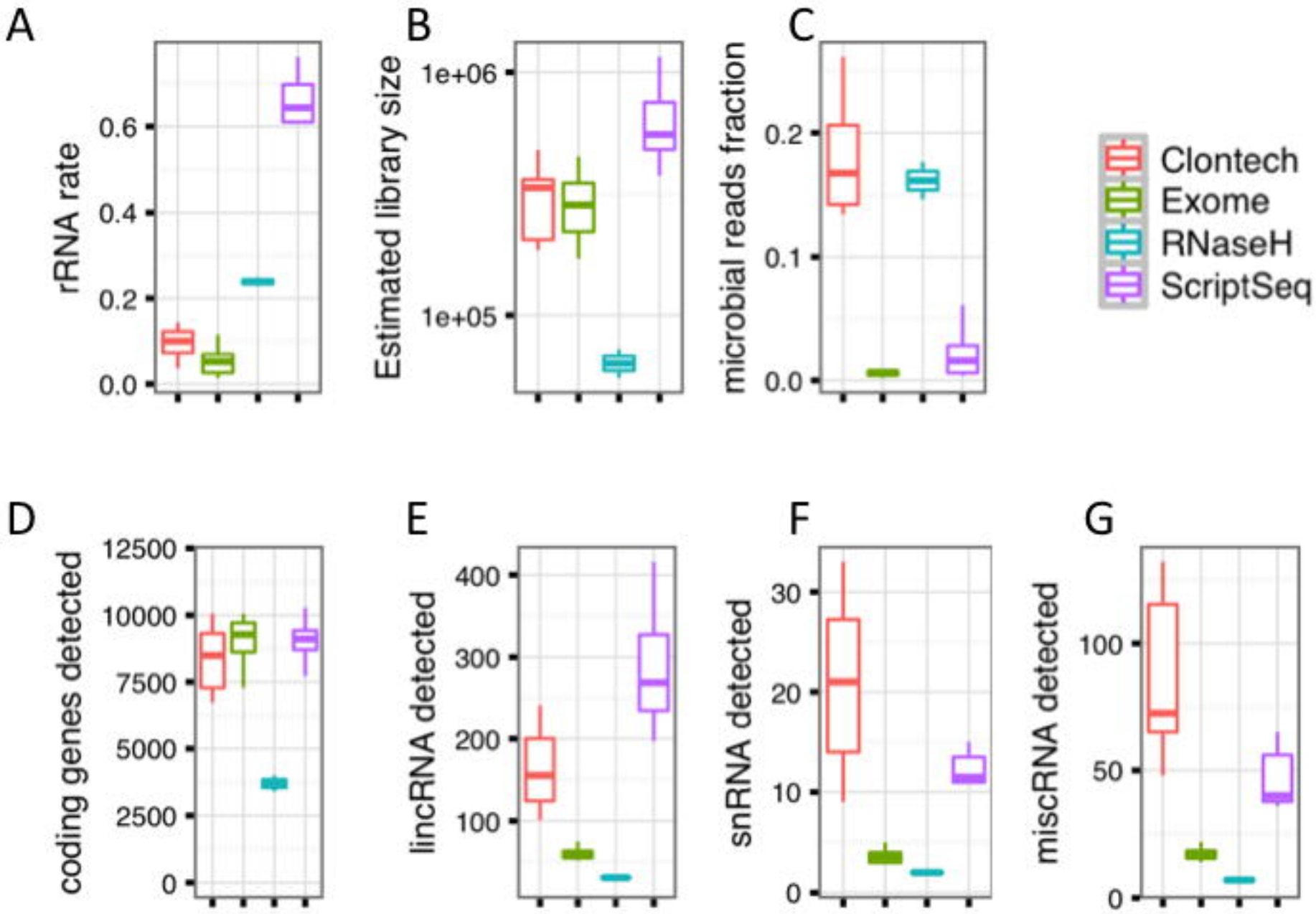
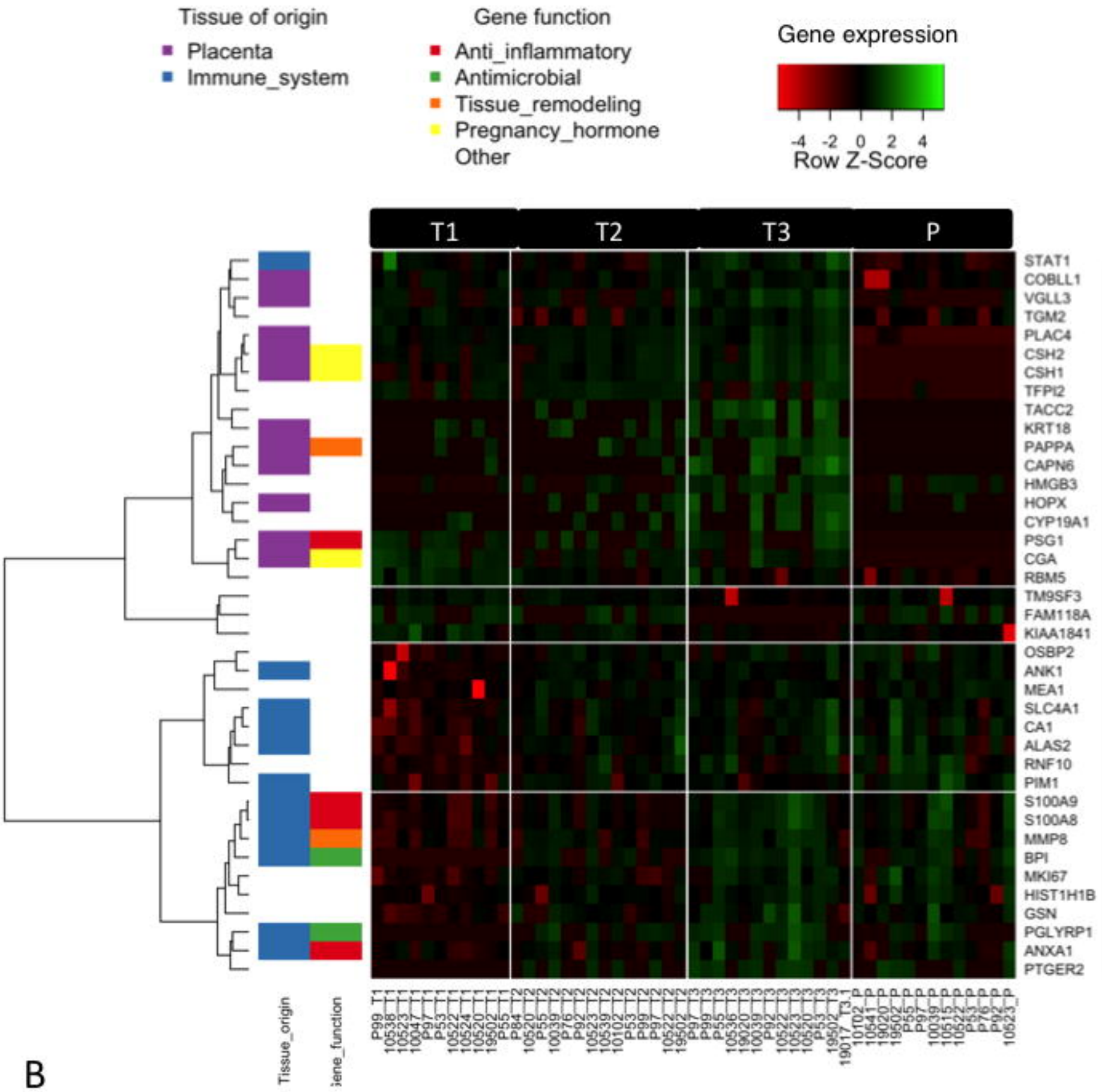

A
rRNA rate
0.0
0.2
0.4
0.6
B
Estimated library size
1e+05
1e+06
C
microbial reads fraction
0.0
0.1
0.2
Clontech
Exome
RNaseH
ScriptSeq
D
coding genes detected
0
2500
5000
7500
10000
12500
E
lincRNA detected
100
200
300
400
F
snRNA detected
0
10
20
30
G
miscRNA detected
0
50
100

Figure 1

A


Tissue of origin
Placenta
Immune_system
Gene function
Anti_inflammatory
Antimicrobial
Tissue_remodeling
Pregnancy_hormone
Other
Gene expression
-4 -2 0 2 4
Row Z-Score
T1
T2
T3
P
cluster
2
4
3
1
STAT1
COBLL1
VGLL3
TGM2
PLAC4
CSH2
CSH1
TFPI2
TACC2
KRT18
PAPPA
CAPN6
HMGB3
HOPX
CYP19A1
PSG1
CGA
RBM5
TM9SF3
FAM118A
KIAA1841
OSBP2
ANK1
MEA1
SLC4A1
CA1
ALAS2
RNF10
PIM1
S100A9
S100A8
MMP8
BPI
MKI67
HIST1H1B
GSN
PGLYRP1
ANXA1
PTGER2
Tissue_origin
Gene_function

B

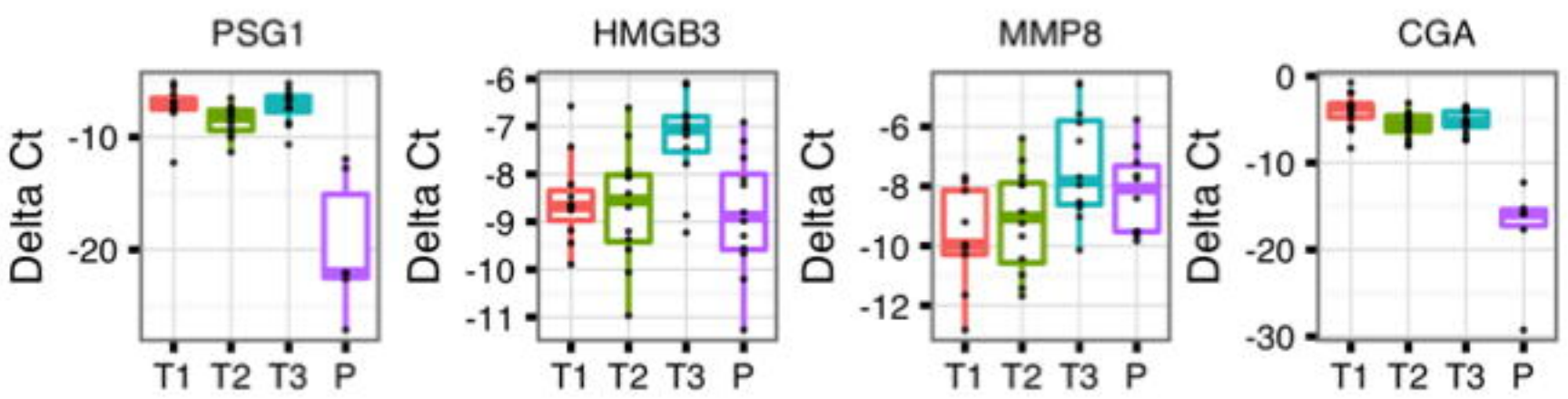

PSG1
HMGB3
MMP8
CGA
Delta Ct
-10
-20
-6
-7
-8
-9
-10
-11
-6
-8
-10
-12
0
-10
-20
-30
T1 T2 T3 P

Figure 2

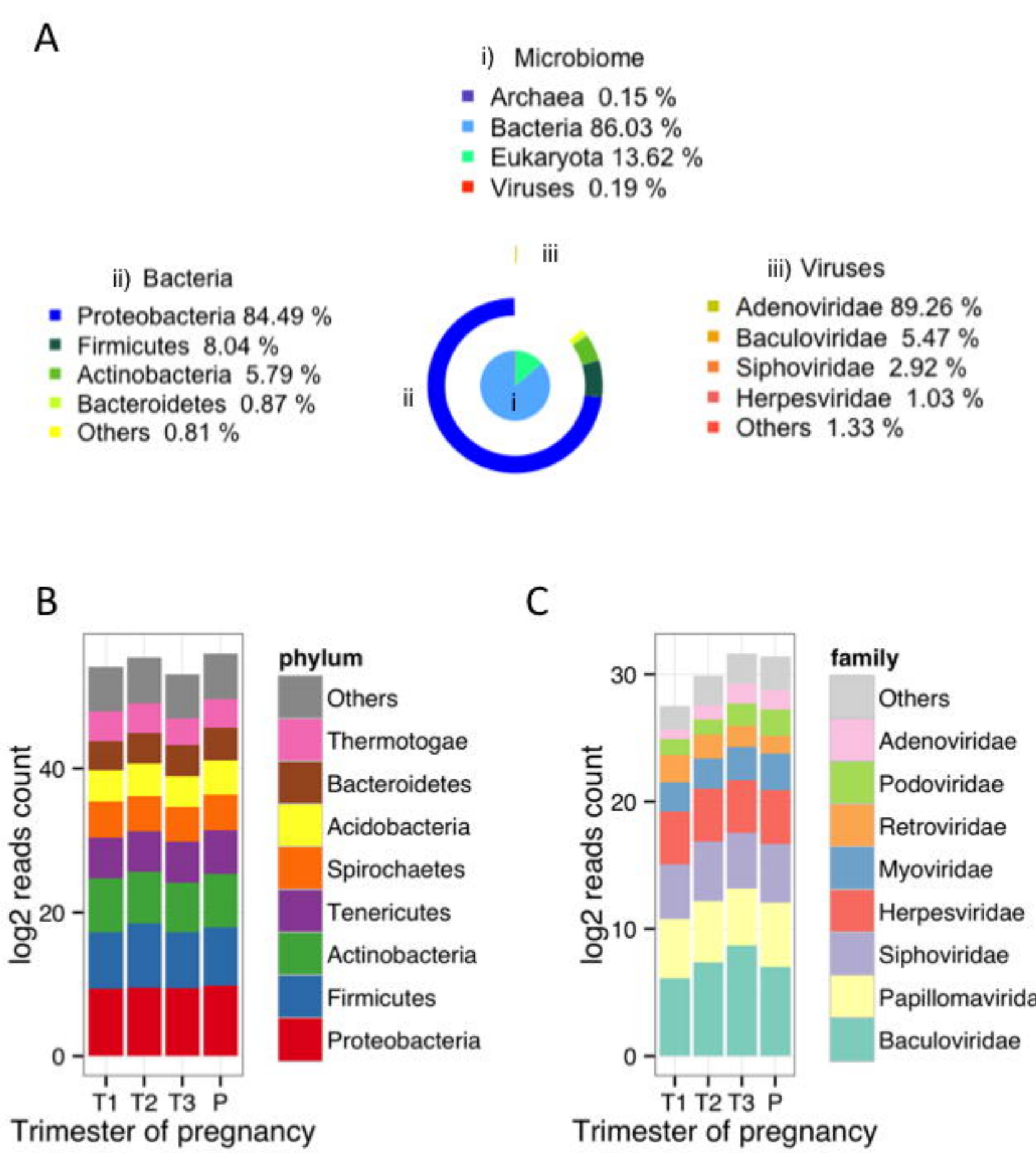

A
i) Microbiome
Archaea 0.15 %
Bacteria 86.03 %
Eukaryota 13.62 %
Viruses 0.19 %
ii) Bacteria
Proteobacteria 84.49 %
Firmicutes 8.04 %
Actinobacteria 5.79 %
Bacteroidetes 0.87 %
Others 0.81 %
iii) Viruses
Adenoviridae 89.26 %
Baculoviridae 5.47 %
Siphoviridae 2.92 %
Herpesviridae 1.03 %
Others 1.33 %
i
ii
iii
B
phylum
Others
Thermotogae
Bacteroidetes
Acidobacteria
Spirochaetes
Tenericutes
Actinobacteria
Firmicutes
Proteobacteria
log2 reads count
0
20
40
T1 T2 T3 P
Trimester of pregnancy
C
family
Others
Adenoviridae
Podoviridae
Retroviridae
Myoviridae
Herpesviridae
Siphoviridae
Papillomaviridae
Baculoviridae
log2 reads count
0
10
20
30
T1 T2 T3 P
Trimester of pregnancy
D
Ureaplasma
Normalized reads count
0
10
20
30
40
10
20
30
40
Gestational age (week)
E
Ureaplasma
Prevalence of infection (%)
0
20
40
60
T1
T2
T3
P
Trimester of pregnancy

Figure 3

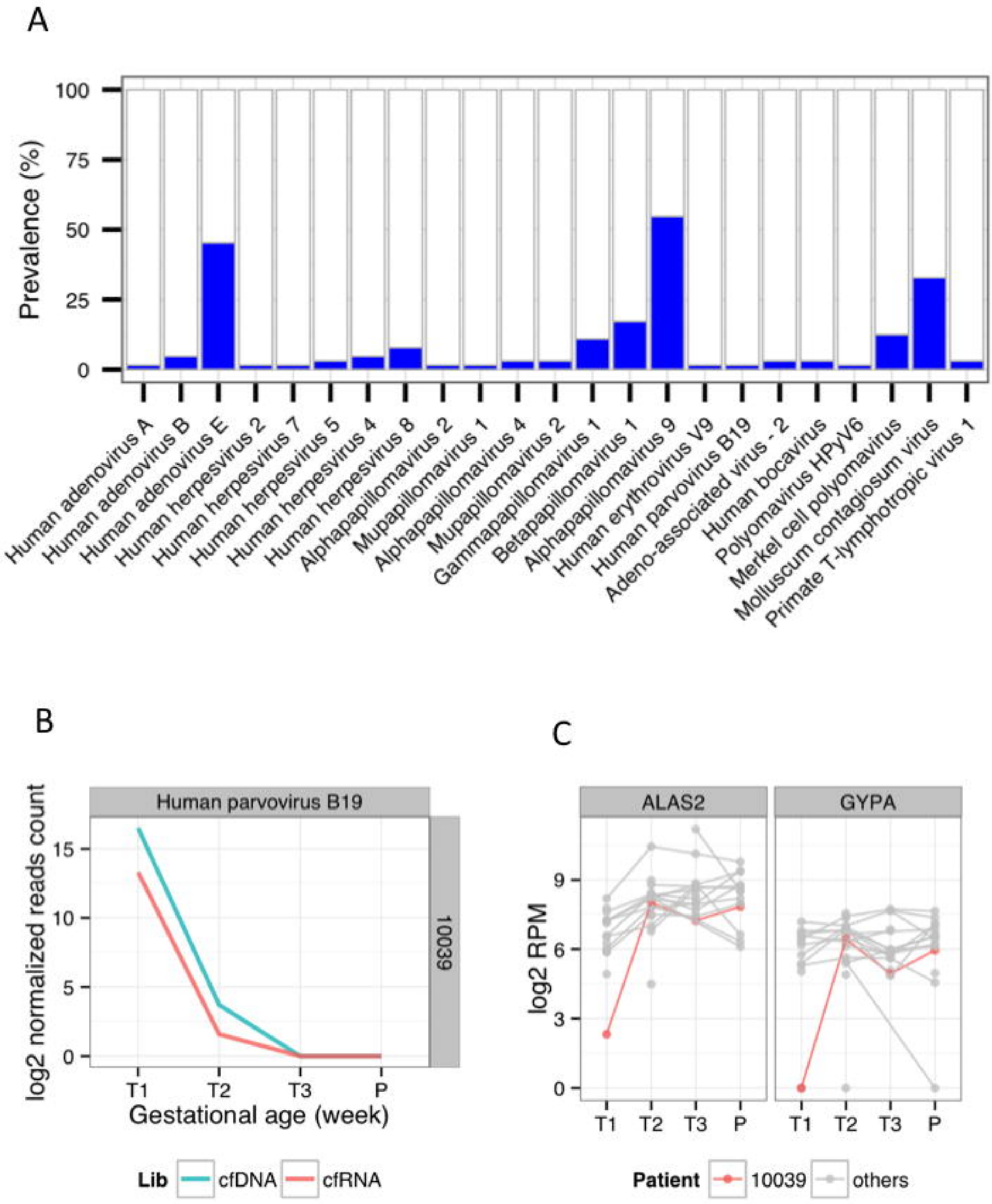

A
Prevalence (%)
100
75
50
25
0
Human adenovirus A
Human adenovirus B
Human adenovirus E
Human herpesvirus 2
Human herpesvirus 7
Human herpesvirus 5
Human herpesvirus 4
Human herpesvirus 8
Alphapapillomavirus 2
Mupapillomavirus 1
Alphapapillomavirus 4
Mupapillomavirus 2
Gammapapillomavirus 1
Betapapillomavirus 1
Alphapapillomavirus 9
Human erythrovirus V9
Human parvovirus B19
Adeno-associated virus - 2
Human bocavirus
Polyomavirus HPyV6
Merkel cell polyomavirus
Molluscum contagiosum virus
Primate T-lymphotropic virus 1
B
Human parvovirus B19
log2 normalized reads count
15
10
5
0
10039
T1
T2
T3
P
Gestational age (week)
Lib
cfDNA
cfRNA
C
ALAS2
GYPA
log2 RPM
9
6
3
0
T1
T2
T3
P
Patient
10039
others

Figure 4