Annotation of Gene Alteration in Amniotic Fluid Cell Free RNA Transcriptome via Weighted Gene Co-expression Network Analysis

Shufa Yang
Beijing Obstetrics and Gynecology Hospital, Capital Medical University
Chenghong Yin
Beijing Obstetrics and Gynecology Hospital, Capital Medical University
Yan Liu
Beijing Obstetrics and Gynecology Hospital,Capital Medical University
Taifeng Zhuang (ztf19731972@126.com)
Beijing Obstetrics and Gynecology Hospital, Capital Medical University

Research article

Keywords: amniotic fluid, cell-free fetal RNA, co-expression network, fetal development

DOI: https://doi.org/10.21203/rs.3.rs-67755/v1

Abstract

Background

Human amniotic fluid (AF) cells are commonly used in prenatal diagnosis. The AF cell-free mRNA (cfRNA) derived from necrotic or apoptotic cells of fetus may provide feasible marker of organ development and fetal malformation. Analysis of gene co-expression network will help to annotate AF cfRNA gene variations in fetal diseases.

Methods

Datasets of amniotic fluid free RNA were downloaded from the Gene Expression Omnibus database. Co-expressed modules based on normal fetus AF cfRNA transcriptome were established via Weighted Gene Co-expression Network Analysis. The relationship between modules and tissues were set up via tissue-specific gene expression analysis of genes in modules. Differential expressed genes in Down syndrome (DS), Edwards syndrome (ES), and Turner syndrome (TS) were analyzed via linear models implemented in the limma package. Gene Ontology (GO) analysis of modular specific differential expressed genes for these three syndrome fetus was performed separately. Based on relationship of GO terms, GO graph of all enriched GO terms were constructed. The associated enriched GO terms in GO graph were considered as the same subset. Numbers of GO terms of the same subset with diseases and modules were calculated.

Results

A total of 22 co-expressed modules were constructed. Most of co-expressed module eigengenes showed no correlation with gestation weeks. Probe sets with higher expression values showed significant clustering tendencies. Dominant tissues and modules were different for different modules and tissues respectively through analysis of tissue-specific genes distribution in modules. The total numbers of enriched GO terms in six major modules were related to diseases severity for ES, DS, and TS. Disease-specific modules for ES, DS, and TS were detected in relation to biological processes through GO analysis. Differential expressed genes in disease-specific module were identified for DS, ES, and TS.

Conclusions

The gene co-expression network for functional classification provides a potential tool for annotating AF cfRNA variations in fetal diseases.

Introduction

Cells in amniotic fluid (AF) are commonly used for prenatal diagnosis to detect genetic diseases. AF supernatants contain cell-free mRNA (cfRNA) derived from apoptosis or necrosis of fetal cells [1–3]. The sources of AF cfRNA are not only limited to cells in amniotic fluid[4]. The AF cfRNA could be influenced by fetal sex [2], gestation age [2, 5–7], genetic disorders [1, 8–11], developmental disorders [12–15],

maternal obesity [16], and preeclampsia [17]. Therefore analysis of the transcriptome for the AF cfRNA may provide feasible markers for organ development and fetal malformation [18].

Tissue-specific gene expression analysis and gene set enrichment analysis are the major annotation methods for the AF cfRNA research. Tissue-specific gene analysis is based on public gene expression data in human tissues, such as BioGPS [19], Genomics Institute of the Novartis Research Foundation [20], DFLAT [21], and Gene Atlas [20]. According to the results of tissue-specific gene expression analysis [1, 2, 6, 7, 14, 22, 23], the genes derived multiple tissues (brain, liver, kidney, lung, blood, ovary, pancreas, tongue, adrenal gland, thyroid, etc) are detected in AF cfRNA.

Enrichment analyses of different expressed genes in abnormal fetus AF cfRNA include Gene Ontology analysis [13, 23], Kyoto Encyclopedia of Genes and Genomes pathway analysis [17], and QIAGEN's Ingenuity Pathway Analysis (IPA) [23]. Gene Ontology analysis was the most frequent method used. Disease-associated pathways and -biological processes are enriched *via* enrichment analysis. Hematologic/immune category functions and gastrointestinal/metabolic functions are significantly enriched in Turner syndrome (TS) [10]. The neurological disease is identified as the most highly enriched disease pathway in Down syndrome (DS) and Edwards syndrome (ES) [1].

Embryonic development comprises the dynamic complex processes, including multiple coordinated development and gene co-expression intra-tissues or inter-tissues. The above relationships are not included in tissue-specific gene analysis and gene function enrichment analysis. We assumed that gene co-expression networks in AF cfRNA may reflect these coordinated development and gene co-expression. To verify this hypothesis, AF cfRNA array data from GEO database were analyzed for weighted co-expression networks analysis. Enrichment analysis of different expressed genes and tissue-specific gene analysis were also performed.

Materials And Methods

Microarray data

The following datasets of amniotic fluid cell free RNA (AF cfRNA) were downloaded from Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) [24]: GSE16176 [8], GSE46286 [5], GSE47394 [12], GSE48521 [16], GSE25634 [9], and GSE58435 [10] (Affymetrix GPL570 platform, Affymetrix Human Genome U133 Plus 2.0 Array). Sample information including sample record (GSM), series record (GSE), clinical information, pregnancy trimester, sex, gestational age, platform accession number (GPL), and fetus karyotype were summarized in Table S1, Table S2, Table S3, and Table S4. Table S1 containing 26 normal samples with gestation age information were used in co-expression networks analysis. Samples in Table S2 (7 DS versus 7 normal fetus), Table S3 (5 ES versus 6 normal fetus), and Table S4 (5 TS versus 5 normal fetus) were used in differential genes analysis for DS, ES and TS groups. Microarray data were preprocessed (background correction, normalization, and summarization) using the Robust Multi-array Average approach [25] implemented in the oligo package

[26]. Probe sets were converted to gene symbol based on Affymetrix Human Genome U133 Plus 2.0 Array annotation data ("hgu133plus2.db" package).

Co-expression net works analysis

An unsigned co-expression net works was performed with Weighted Gene Co-expression Network Analysis (WGCNA) package [27]. Normal second trimester samples (Table S1) with available gestational ages were included in WGCNA analysis. Before WGCNA analysis, raw expression values of probe sets were preprocessed, and the batch effects were removed with ComBat function in the SVA package [28]. Soft-thresholding power was set to 15, which was calculated by the pickSoftThreshold algorithm. The automatic block-wise network construction and module detection method was employed. Module assignments were made using dynamic tree cutting method by setting mergeCutHeight to 0.2. Module eigengene was calculated via moduleEigengenes algorithm. To generate co-expressed modular genes, probe sets in clustered modules were annotated to genes via "hgu133plus2.db" package.

Tissue-specific gene analysis

Gene expression data in human tissues were downloaded from tissue-based map of the human proteome [29] in the Human Protein Altas database (http://www.proteinatlas.org). The Human Protein Atlas aims to map all the human proteins in cells, tissues and organs using the integration of various omics technologies. Tissue-specific genes in a specific tissue were defined as expression value in particular tissue higher than cut-off value times the median expression value of all other tissues. Cut-off value was set as 5, 10, 15, 20, 25, and 30 respectively. Tissue-specific gene names were transformed to probe sets for further analysis via "hgu133plus2.db" package. Tissue-specific genes that were not tested by array were discarded. Distribution of Tissue-specific genes (probes) at a different cut-off values in co-expression modules were analyzed via R script.

Gene differential expression analysis

Before gene differential expression analysis, probe sets expression values were preprocessed (background correction, normalization, and summarization) and log2-transformed. Differential expressed probe sets were analyzed via linear models implemented in the limma [30] package in Bioconductor [31]. Differential expressed probe sets (DEPs) were deemed significant based on p-values < 0.05. Differential expressed genes (DEGs) were obtained after transforming differential expressed probe sets into genes via "hgu133plus2.db" package.

Gene ontology analysis

According to the results of WGCNA analysis, modular distribution of DEPs from ES, DS, and TS groups were analyzed. After transforming probe sets into gene names, modular specific DEGs (DEGs in the certain module) from ES, DS, and TS groups were obtained. Gene ontology analysis of modular specific DEGs from ES, DS, and TS groups was performed separately via the clusterProfiler package [32]. P-values were adjusted using the Benjamini-Hochberg method. Adjusted p-values < 0.05 were considered

statistically significant. Based on the relationship of GO terms (is_a, part_of, regulates, positively_regulates, etc), GO graph of all enriched GO terms from ES, DS and TS groups were constructed. The associated enriched GO terms in the GO graph were considered as a same subset. Numbers GO terms of the same subset concerning diseases and modules were calculated.

Statistical Analysis

All other data statistical analyses and plots were created using R language (version 3.5.3), ggplot2 (version 3.3.1) and VennDiagram (version 1.6.20).

Results

Summary of gene co-expression network analysis

A total of 26 second term normal AF cfRNA chip results (15 males and 11 females) were selected for WGCNA. The mean gestation ages of male and female fetus were 15 and 19 weeks respectively (Table S1). Raw data from each microarray were pre-processed for background correction, normalization and batch correction (Fig 1A). All the 54675 probe sets on the chip were analyzed by WGCNA. For block-wise network construction, a computationally inexpensive and relatively crude clustering method was adopted to classify pre-cluster probe sets into three blocks. A full network analysis was performed in each block separately. A total of 22 distinct probe set modules were generated (Fig 1 B-D). These 22 modules were shown in different colors and module names were labeled as colors, i.e., a color represented a group of co-expressed genes. The size of modules ranged from 20 to 1198 probe sets (Table S6). Probe sets without obvious co-expressions relationship were labeled as grey. After probe sets were converted to gene symbol, co-expressed genes in 22 modules were obtained and summarized in Table S5.

Most of the co-expressed modules showed no correlation with gestation weeks

Module eigengene (ME) is defined as the first principal component of a given module. For all samples employed in WGCNA analysis, MEs of 22 modules were calculated, and compared to modular expression values of the probe sets. As shown in Fig S1, MEs showed the same tendency with expression values of the probe sets and could be representative for expression values of probe sets in the correspondence modules.

The correlation between MEs and gestation age was evaluated by Spearman correlation analysis. As shown in Table S6, most modules showed no significant correlation with gestation age except for green and turquoise modules. Modules were clustered using hierarchical clustering. Green and turquoise modules showing significant correlation with gestation age were classified into different clusters. Modules with similar correlation coefficient were divided into different clusters (Fig 2A). MEs of major modules (turquoise, blue, brown, yellow, green and red) were smoothed by locally weighted regression

and were shown in Fig 2B. A downward trend was observed for MEs of green and turquoise modules with the increase of gestation age. No obvious trend was found for MEs of other modules concerning gestation age. Expression of genes in green and turquoise modules reduced with augment of gestation age in second term.

High expression probe sets showing significant clustering tendencies

Probe sets with higher expression values are more inclined to be used as a marker for fetal organ development. To investigate the expression character of clustered probe sets, mean values and coefficient of variation for all probes in the normal 26 samples employed in WGCNA analysis were calculated. A scatter plot (mean expression value VS coefficient of variation) was drawn to show clustered probe sets distribution in all probe sets, and the clustered probe sets were labeled as modular colors (Fig 3A). Compared to probe sets with low expression values, the higher portions of the probe sets with higher expression values were clustered.

Relationship between modules and tissues via dominant tissues and modules

To establish the relationship between fetal tissues and modules, the distribution of tissue-specific genes in modules were analyzed. Tissue-specific genes were obtained at a different cut-off value conditions. After tissue-specific genes were converted to probe sets, numbers of tissue-specific genes (probe sets) under different cut-off value (5, 10, 15, 20, 25 and 30) were counted (Fig 3B). The number of tissue-specific genes (probe sets) decreased with higher cut-off value. The distribution of tissue-specific probe sets in all 22 modules, was analyzed and summarized in Table S6. As shown in Fig 3C, tissue-specific probe sets were mainly distributed in turquoise, blue, brown, yellow, green, and red modules, which were called major modules.

Considering the difference of gene expression in fetus and adult tissues, the cut-off value was set as 5 to get more gene specific genes. The numbers of tissue-specific probe sets derived from different tissues in major modules were shown in Fig 3D, while cut-off value was set as 5 (detailed data see Table S7). Tissue-specific genes from skeleton, liver, and testis accounted for the largest specific genes in turquoise module. The above three tissues were defined as dominant tissues in turquoise module. Similarly, dominant tissues were calculated in blue module (placenta, skeletal muscle, and testis), brown module (testis, cerebral cortex, and skeletal muscle), yellow module (testis, cerebral cortex, and cerebellum), green module (small intestine, liver, and colon) and red module (esophagus, tongue, and tonsil). Tissue-specific probe sets derived from liver mostly distributed in green and turquoise modules. Green and turquoise modules were defined as dominant modules for liver. Dominant modules for cerebral cortex included

yellow and blue modules. Dominant tissues and dominant modules were different for different modules and tissues respectively.

Blue, brown, and yellow modules included the largest neural-specific genes (cerebral cortex, cerebellum, basal ganglia, etc). Green modules contained more digest system specific genes.

Differential expressed genes in DS, ES, and TS

Samples used for DEGs analysis were listed in Table S2 (DS), Table S3 (ES), and Table S4 (TS). For differential expressed probe sets (DEPs) analysis in each group, the expression values of the abnormal fetus were compared to that of normal fetus via a linear regression model. There were 1049, 1507, and 1448 DEPs detected in DS, ES, and TS groups respectively. Volcano map of DS, ES, and TS group were shown in Fig 4A-C. Relationship of DEPs between DS, ES, and TS group were shown in Venn diagrams (Fig 4D). There were 41 common DEPs between DS ES group, 38 common DEPs between DS and TS group, 43 DEPs between ES and TS group. No common DEPs were found in DEPs of DS, ES, and TS.

Disease-specific modules via GO analysis

Numbers of DEPs of DS, ES, and TS in major clustered modules were counted (Fig 5A). Blue module was the most abundant distributed module in TS and DS groups. However, the brown module had the largest number of DEPs of ES. To analyze the function of clustered genes, DEPs were converted to gene symbols. The modular distribution of DEGs in three groups was summarized in Table S8.

Modular specific DEGs, which mean DEGs of every group (DS, ES, and TS) in a certain module, were extracted from Table S8. Modular specific DEGs in major modules (turquoise, blue, brown, yellow, green, and red) were performed functional enrichment analysis (Table S10). The numbers of enriched GO terms were shown in Fig 5B. The yellow, blue, and red modules were specific modules for DS, ES, and TS groups respectively. DEGs of DS in the yellow module included SORBS1 and FFAR4. DEGs of ES in blue module included AKNAD1, SOX9, ZNF395, PID1, PRPF38A, FAM220A, ATP1B1, NME7, CD9, EPC1, and MAML2. DEGs of TS in the red module included S100A8 and IVL. The total number of enriched GO terms in ES, DS, and TS group ranked the first, the second, and the third respectively. In green and turquoise modules, similar numeric distribution of enriched GO term was shown for DS and ES, rather than TS. In green and turquoise modules, a similar number of GO term numbers were enriched in DS and ES group.

To get an overall view of all enriched GO terms, the GO graph was established based on the relationship of GO terms. Go maps of all enriched GO terms were list in Table S9. Based on the GO graph, interrelated GO Terms were classified into the same subsets. As shown in Table S10, subsets contained a certain number of enriched GO terms from different modules and abnormal fetus (DE, ES, and TS). A total of 184 subsets were established. The numbers of GO terms in a subset ranged from 1 to 332. The largest nine subsets were summarized in Table 1. Functions of these subsets included basic physiological processes,

absorption and transport of nutrients, response to external stimuli, multi-organ (kidney, lung, and heart) development, protein synthesis process, protein catabolic process and proteolysis, thermoregulation, signal transduction of TGF, and bone development.

Table 1
Subsets of enriched GO terms

ID	No	Fig 6	Function
1	332	A	Basic physiological processes
34	100	B	Absorption and transport of nutrients
20	79	C	Response to external stimuli
48	60	D	Multi-organ (Kidney, lung and heart) development
11	48	E	Protein synthesis process
17	38	F	Protein catabolic process and proteolysis
38	31	G	Thermoregulation
47	15	H	Signal transduction of TGF
106	12	I	Bone development
ID: Subsets id shown in Table S9 subset_id column.			
No: Number of enriched GO terms in subsets, were shown in Table S9 subset_number column.			
Fig 6: Modular distribution of enriched GO term subsets in abnormal fetus as shown in Fig 6.			
Function: summarized function according to in corresponding biological GO term subsets.			

To analyze the relationship between subset function and disease in different modular condition, enriched GO terms numbers in the nine largest subsets (Table 1) in relation to modules and fetus disease were shown in Fig 6. Besides that yellow, blue, and red modules were specific modules for DS, ES, and TS groups, more specific modules were shown in certain subsets. Green, turquoise, and brown modules were specific for ES group in absorption and transport of nutrients subset. The green module was specific to the DS group in multi-organ (kidney, lung, and heart) development subset, protein synthesis process subset, thermoregulation subset and signal transduction of TGF subset.

Discussion

AF cfRNA were considered a biomarker source for abnormal fetal development. And the gene expression level in AF cfRNA could be influenced by fetal status (such as fetal sex [2], gestation age [2, 5–7], genetic disorders [1, 8–11], and developmental disorders [12–15]). Therefore, AF cfRNA was proposed as a mean to monitoring fetal development in real-time. At present, the tissue expression database (Human protein Atlas [29], Gene Atlas [20], and BioGPS [19]) were mostly based on the adult tissues. The lack of fetal-

related tissue expression database hinders the identification of abnormal fetal development. Additionally, AF cfRNAs were multiple organ origins, which increased the interferences for discovering abnormal fetal development biomarkers.

Compared to single genes, the advantages of gene sets contains noise and dimension reduction, as well as greater biological interpretability [33]. A total of 22 co-expressed gene sets (modules) were identified via WGCNA [27]. As shown in Fig. 3A, probe sets with high expression values show obvious clustering tendency. Analogous to dimension reduction, expression values of clustered probes can be simplified as module eigengenes. Correlation between gene expression and gestation ages is easier to analyze via module eigengenes. As shown in Fig. 2B, only genes in green modules showed significant correlation with gestation ages. The above results indicated that correlation between gene expression and gestation ages was not an appropriate way to discover the biomarker of the abnormal fetal development.

To explore the tissue origins of genes in AF cfRNA, different gene expression atlas and cut-off values were used to define tissue-specific genes in previous studies [1, 5, 10, 34, 35]. However, there was no fully comparison to get confirmed standard. Consequently, a series of cut-off values were used to define tissue-specific genes expression in our research. Based on tissue-specific gene expression, the relationships between modules and tissues were constructed. Genes from different organs showed similar co-expression patterns and were classified as the same module. Inter-organ synergetic development was manifested as a co-ordinated expression of genes intra modules. Due to the complexity of gene expression in the same organs, genes from the same organ showed different co-expression patterns. Accordingly, tissue-specific expressed genes from the same organ were divided into different modules. Thus, synergetic developments inter- and intra- organs were shown as gene co-expression networks intra- and inter- modules via gene tissue-specific analysis and weighted correlation network analysis.

DS, ES, and TS were the most frequent chromosomal abnormalities in prenatal diagnosis [36], which were caused by trisomy of chromosome 21, trisomy of chromosome 18, and sex chromosome aneuploidy. The development of multiple tissues and organs was affected in these three symptoms. ES is the most serious defects and most ES cause spontaneous abortion. The most frequent clinical features of ES consist of neurological findings, growth disturbances, malformations of the skull, face, thorax, abdomen, limbs, genitals, skin, skin annexes, and internal organs [37]. Cardiac, airway, pulmonary, hearing, growth, hematologic, oncologic, autoimmune, musculoskeletal, and neurodevelopmental disorders are the most frequent clinical symptoms in DS [38]. Endocrinal, gastrointestinal, hepatic, phenotypic, neurocognitive, and psychosocial disorders [39] were reported in TS patient. Compared with the normal fetus, the differential expressed genes of DS [8], ES [9], and TS [10] were analyzed. Different from common affected organs or closely related clinical symptoms, different expressed genes from these syndrome share few common differential expressed genes [4].

To explain this contradictory phenomenon, DEGs from DS, ES, and TS compared to normal fetuses were extracted. Similar to Zwemer's report, few common misregulated genes were detected in three groups [4]

(Fig. 4D). Interestingly, the rank of the total number of enriched GO terms in ES, DS, and TS group was correlated with disease severity (Fig. 5B). Biological functions of enriched GO terms were simplified via classifying interrelated GO terms into the same subsets. A subset contains function-related GO terms performing associated embryonic development process. In most subsets, the number of enriched GO terms was positively related with disease severity (Fig. 6D). Another tendency, shown in the modular distribution of GO term subsets was modular heterogeneity, which suggested disease-specific modules were detected in functional enrichment analysis of modules related DEPs. Pathogenesis-related different expressed genes were classed into different modules in different diseases. In absorption and transport of nutrients subsets, DEGs in DS and ES were associated with yellow and red modules respectively. Consequently, Preprocessing via co-expression analysis could help constructed the relationship between DEGs and disease phenotype.

Large numbers of GO terms were enriched in most modules with a small amount of DEGs, as shown in Fig. 5B. To explain this phenomenon, DEGs in disease-specific modules were extracted. SORBS1 and FFAR4 were two of DEGs from the yellow module in the DS group. SORBS1 was associated with insulin signaling pathway [40], glucose homeostasis [41, 42], cancer growth and migration [43, 44], etc. FFAR4 have similar functions including glucose and fatty acid metabolism [45, 46], glucose-dependent insulinotropic polypeptide secretion [47], cell cancelation[48, 49], etc. This is accounted for enriched results of yellow modules in DS group. In subsets of a basic physiological process, absorption and transport of nutrients, response to external stimuli, protein synthesis process and thermoregulation were enriched. Taken together, genes without co-expression relationships were filtered via WGCNA. The remaining co-expressed genes were easier to be enriched via functional enrichment analysis.

Conclusion

Co-expression modules based on samples from the whole second gestation (from 15 to 22 weeks) provided gene classification in this study. Via tissue-specific genes analysis, the relationships between modules and tissues were established. Disease-specific modules can help to annotate AF cfRNA variation in fetal diseases. Pre-grouping via co-expression before functional enrichment analysis could help to build a relationship between DEGs and disease phenotype. With the accumulation of datasets of AF cfRNAs, gestation week specific co-expression modules could be constructed, in which more genes in AF cfRNA would be included. With the finer co-expression modules for different gestation age, comprehensive knowledge for fetal coordinated development and gene co-expression will be elucidated.

Abbreviations

AF: Amniotic fluid; cfRNA: cell-free mRNA; DS: Down syndrome; ES: Edwards syndrome; TS: Turner syndrome; ME: Module eigengene; WGCNA: Weighted Gene Co-expression Network Analysis; DEPs: Differential expressed probe sets; DEGs: Differential expressed genes.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors have no conflicts of interest to report.

Funding

This research was founded by the National Key Research and Development Program of China (No: 2016YFC1000101)

Authors' contributions

SYang, CYin, TZhang designed the project, SYang and YLiu performed statistical analysis, SYang wrote the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgements

Not applicable

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Hui L, Slonim DK, Wick HC, Johnson KL, Koide K, Bianchi DW. Novel neurodevelopmental information revealed in amniotic fluid supernatant transcripts from fetuses with trisomies 18 and 21. Hum Genet. 2012;131(11):1751–9.
2. Larrabee PB, Johnson KL, Peter I, Bianchi DW. Presence of filterable and nonfilterable cell-free mRNA in amniotic fluid. Clin Chem. 2005;51(6):1024–6.

3. Jasinska AJ, Rostamian D, Davis AT, Kavanagh K. Transcriptomic Analysis of Cell-free Fetal RNA in the Amniotic Fluid of Vervet Monkeys (Chlorocebus sabaeus). Comp Med. 2020;70(1):67–74.
4. Zwemer LM, Bianchi DW. **The amniotic fluid transcriptome as a guide to understanding fetal disease**. *Cold Spring Harb Perspect Med* 2015, 5(4).
5. Hui L, Wick HC, Edlow AG, Cowan JM, Bianchi DW. Global gene expression analysis of term amniotic fluid cell-free fetal RNA. Obstet Gynecol. 2013;121(6):1248–54.
6. Tarca AL, Romero R, Pique-Regi R, Pacora P, Done B, Kacerovsky M, Bhatti G, Jaiman S, Hassan SS, Hsu CD, et al. Amniotic fluid cell-free transcriptome: a glimpse into fetal development and placental cellular dynamics during normal pregnancy. BMC Med Genomics. 2020;13(1):25.
7. Jang JH, Jung YW, Shim SH, Sin YJ, Lee KJ, Shim SS, Ahn EH, Cha DH. Global gene expression changes of amniotic fluid cell free RNA according to fetal development. Eur J Obstet Gynecol Reprod Biol. 2017;216:104–10.
8. Slonim DK, Koide K, Johnson KL, Tantravahi U, Cowan JM, Jarrah Z, Bianchi DW. Functional genomic analysis of amniotic fluid cell-free mRNA suggests that oxidative stress is significant in Down syndrome fetuses. Proc Natl Acad Sci U S A. 2009;106(23):9425–9.
9. Koide K, Slonim DK, Johnson KL, Tantravahi U, Cowan JM, Bianchi DW. Transcriptomic analysis of cell-free fetal RNA suggests a specific molecular phenotype in trisomy 18. Hum Genet. 2011;129(3):295–305.
10. Massingham LJ, Johnson KL, Scholl TM, Slonim DK, Wick HC, Bianchi DW. Amniotic fluid RNA gene expression profiling provides insights into the phenotype of Turner syndrome. Hum Genet. 2014;133(9):1075–82.
11. Bianchi DW. **Turner syndrome: New insights from prenatal genomics and transcriptomics**. *Am J Med Genet C Semin Med Genet* 2019.
12. Hui L, Wick HC, Moise KJ Jr, Johnson A, Luks F, Haeri S, Johnson KL, Bianchi DW. Global gene expression analysis of amniotic fluid cell-free RNA from recipient twins with twin-twin transfusion syndrome. Prenat Diagn. 2013;33(9):873–83.
13. Cho HY, Cho Y, Shin YJ, Park J, Shim S, Jung Y, Shim S, Cha D. Functional analysis of cell-free RNA using mid-trimester amniotic fluid supernatant in pregnancy with the fetal growth restriction. Med (Baltim). 2018;97(2):e9572.
14. Xie J, Zhou Y, Gao W, Li Z, Xu Z, Zhou L. The relationship between amniotic fluid miRNAs and congenital obstructive nephropathy. Am J Transl Res. 2017;9(4):1754–63.
15. Tarui T, Kim A, Flake A, McClain L, Stratigis JD, Fried I, Newman R, Slonim DK, Bianchi DW. Amniotic fluid transcriptomics reflects novel disease mechanisms in fetuses with myelomeningocele. Am J Obstet Gynecol. 2017;217(5):587. e581-587 e510.
16. Edlow AG, Vora NL, Hui L, Wick HC, Cowan JM, Bianchi DW. Maternal obesity affects fetal neurodevelopmental and metabolic gene expression: a pilot study. PLoS One. 2014;9(2):e88661.
17. Jung YW, Shim JI, Shim SH, Shin YJ, Shim SH, Chang SW, Cha DH. Global gene expression analysis of cell-free RNA in amniotic fluid from women destined to develop preeclampsia. Med (Baltim).

2019;98(3):e13971.

18. Vora NL, Hui L. Next-generation sequencing and prenatal 'omics: advanced diagnostics and new insights into human development. Genet Med. 2018;20(8):791–9.
19. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW. 3rd et al: BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol 2009, 10(11):R130.
20. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004;101(16):6062–7.
21. Wick HC, Drabkin H, Ngu H, Sackman M, Fournier C, Haggett J, Blake JA, Bianchi DW. Slonim DK: **DFLAT: functional annotation for human development**. BMC Bioinformatics. 2014;15:45.
22. Jung YW, Shim SS, Park JE, Sung SR, Shim SH, Park HR, Cha DH. Analysis of the cell-free amniotic fluid transcriptome expressed during the euploid mid-trimester of pregnancy. Eur J Obstet Gynecol Reprod Biol. 2016;203:94–8.
23. Kang JH, Park HJ, Jung YW, Shim SH, Sung SR, Park JE, Cha DH, Ahn EH. Comparative Transcriptome Analysis of Cell-Free Fetal RNA from Amniotic Fluid and RNA from Amniocytes in Uncomplicated Pregnancies. PLoS One. 2015;10(7):e0132955.
24. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207–10.
25. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249–64.
26. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010;26(19):2363–7.
27. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
28. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882–3.
29. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
30. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
31. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):R80.

32. Yu GC, Wang LG, Han YY, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. Omics-a Journal of Integrative Biology. 2012;16(5):284–7.
33. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
34. Hui L, Slonim DK, Wick HC, Johnson KL, Bianchi DW. The amniotic fluid transcriptome: a source of novel information about human fetal development. Obstet Gynecol. 2012;119(1):111–8.
35. Kamath-Rayne BD, Du Y, Hughes M, Wagner EA, Muglia LJ, DeFranco EA, Whitsett JA, Salomonis N, Xu Y. Systems biology evaluation of cell-free amniotic fluid transcriptome of term and preterm infants to detect fetal maturity. BMC Med Genomics. 2015;8:67.
36. Yang S, Lv J, Si Y, Du X, Chen Z. Diagnostic differences between patients opting for non-invasive prenatal testing and patients having traditional prenatal diagnosis. International Journal of Clinical Experimental Pathology. 2018;11(5):2831–8.
37. Rosa RF, Rosa RC, Zen PR, Graziadio C, Paskulin GA. Trisomy 18: review of the clinical, etiologic, prognostic, and ethical aspects. Rev Paul Pediatr. 2013;31(1):111–20.
38. Bull MJ. Down Syndrome. N Engl J Med. 2020;382(24):2344–52.
39. Gravholt CH, Viuff MH, Brun S, Stochholm K, Andersen NH. Turner syndrome: mechanisms and management. Nat Rev Endocrinol. 2019;15(10):601–14.
40. Chehin MB, Fraietta R, Lorenzon AR, Bonetti TCS, Motta ELA. The insulin signaling pathway is dysregulated in cumulus cells from obese, infertile women with polycystic ovarian syndrome with an absence of clinical insulin resistance. Ther Adv Reprod Health. 2020;14:2633494120906866.
41. Chang TJ, Wang WC, Hsiung CA, He CT, Lin MW, Sheu WH, Chang YC, Quertermous T, Chen YI, Rotter JI, et al. Genetic variation of SORBS1 gene is associated with glucose homeostasis and age at onset of diabetes: A SAPPHIRe Cohort Study. Sci Rep. 2018;8(1):10574.
42. Germain M, Pezzolesi MG, Sandholm N, McKnight AJ, Susztak K, Lajer M, Forsblom C, Marre M, Parving HH, Rossing P, et al. SORBS1 gene, a new candidate for diabetic nephropathy: results from a multi-stage genome-wide association study in patients with type 1 diabetes. Diabetologia. 2015;58(3):543–8.
43. Cho WC, Jang JE, Kim KH, Yoo BC, Ku JL. SORBS1 serves a metastatic role via suppression of AHNAK in colorectal cancer cell lines. Int J Oncol. 2020;56(5):1140–51.
44. Sultan G, Zubair S, Tayubi IA, Dahms HU, Madar IH. Towards the early detection of ductal carcinoma (a common type of breast cancer) using biomarkers linked to the PPAR(gamma) signaling pathway. Bioinformation. 2019;15(11):799–805.
45. Codoner-Alejos A, Carrasco-Luna J, Codoner-Franch P. The rs11187533 C > T Variant of the FFAR4 Gene Is Associated with Lower Levels of Fasting Glucose and Decreases in Markers of Liver Injury in Children with Obesity. Ann Nutr Metab. 2020;76(2):122–8.
46. Kimura I, Ichimura A, Ohue-Kitano R, Igarashi M. Free Fatty Acid Receptors in Health and Disease. Physiol Rev. 2020;100(1):171–210.

47. Reimann F, Diakogiannaki E, Moss CE, Gribble FM. Cellular mechanisms governing glucose-dependent insulinotropic polypeptide secretion. Peptides. 2020;125:170206.
48. Zhou LZ, Cui YX, Wang WZ, Wu J, Sun Z, Ma SY. FFAR4 promotes cell proliferation and migration and servers as a potential biomarker for clinicopathological characteristics and prognosis in laryngocarcinoma. Eur Rev Med Pharmacol Sci. 2019;23(17):7438–44.
49. Chu X, Zhou Q, Xu Y, Jiang J, Li Q, Zhou Q, Wu Q, Jin M, Wang H, Gu Y, et al. Aberrant fatty acid profile and FFAR4 signaling confer endocrine resistance in breast cancer. J Exp Clin Cancer Res. 2019;38(1):100.

Figures

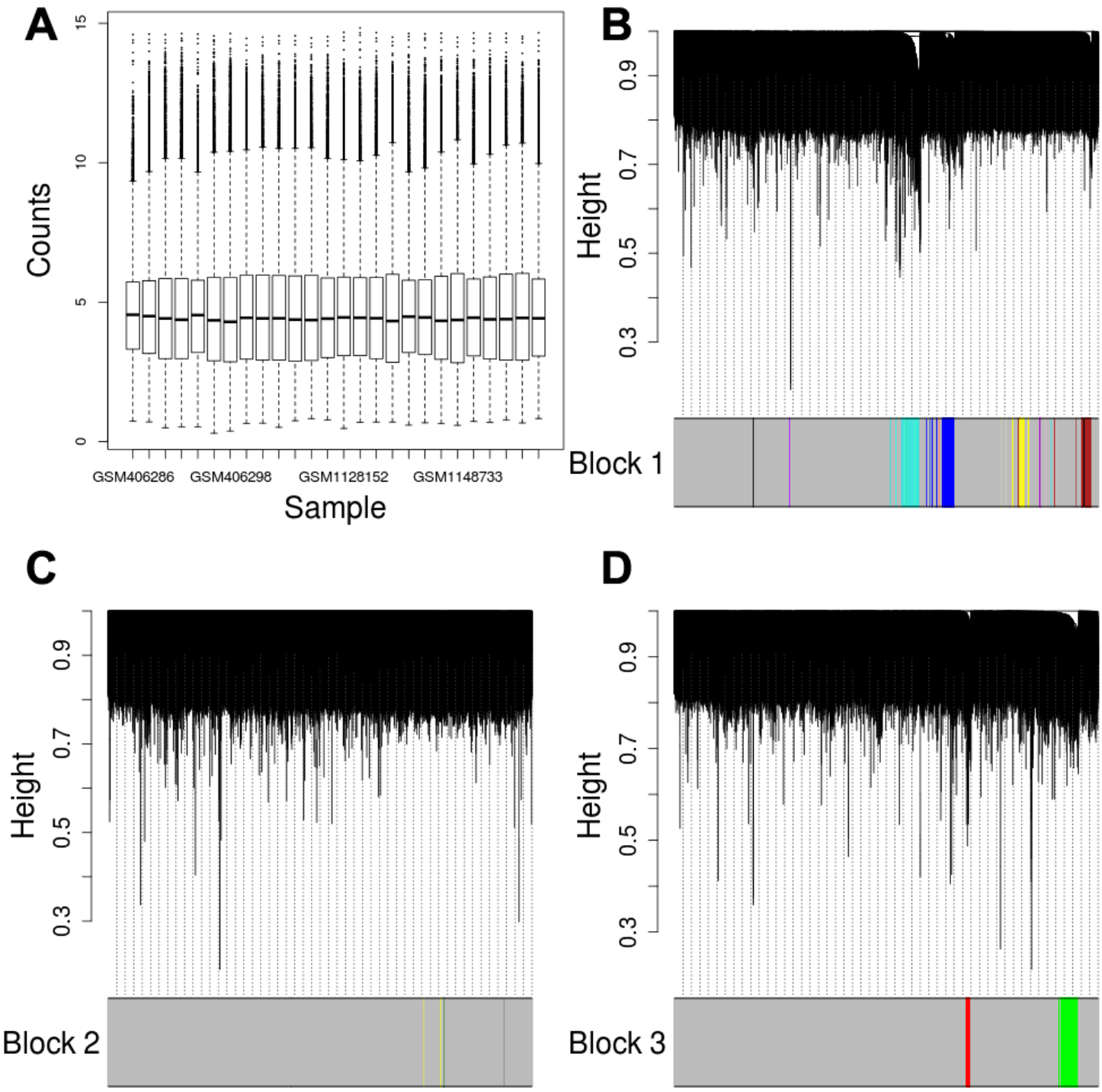

Figure 1

WGCNA in AF cfRNA (A) Pre-procession for background correction, normalization and batch correction. A similar mean and standard deviation were obtained via pre-procession. (B-D) Cluster dendrogram and module assignment for modules from WGCNA. Probe sets on chip were classified into three blocks. A full network analysis was performed in each block separately. A total of 22 distinct probe sets modules were generated.

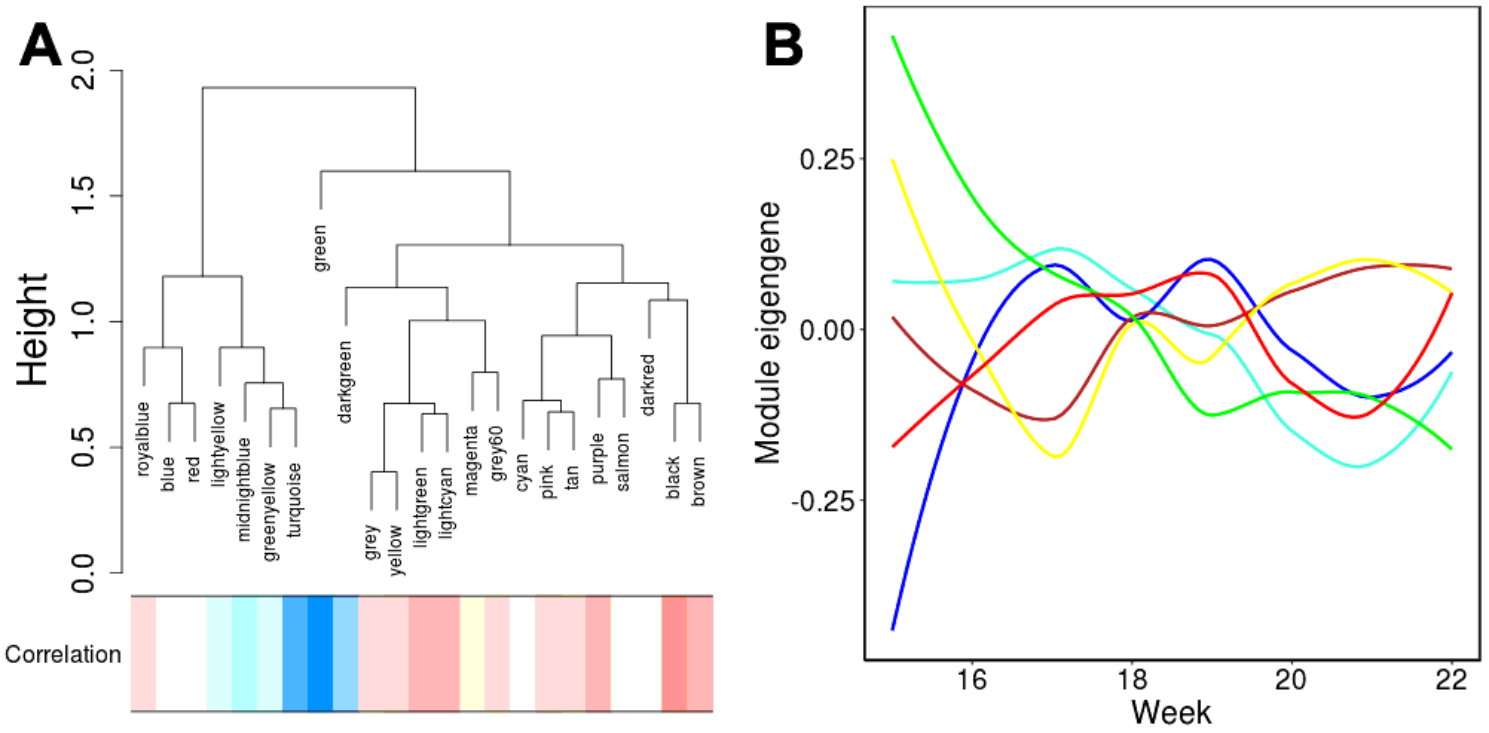

Figure 2

Correlation analysis between modules and gestation weeks (A) Modules were classified in different clusters (upper clustering dengrogram). Correlation coefficient between MEs and gestation ages were shown in lower color gradient. Blue color means higher correlation coefficient. Modules with similar correlation coefficient (lower heat map) with gestation age were divided into different clusters. (B) MEs of major modules (turquoise, blue, brown, yellow, green and red) were smoothed by locally weighted regression. There was a downward trend for MEs of green and turquoise modules with the increase of gestation age. No obvious trend was observed for MEs of other modules in relation to gestation age.

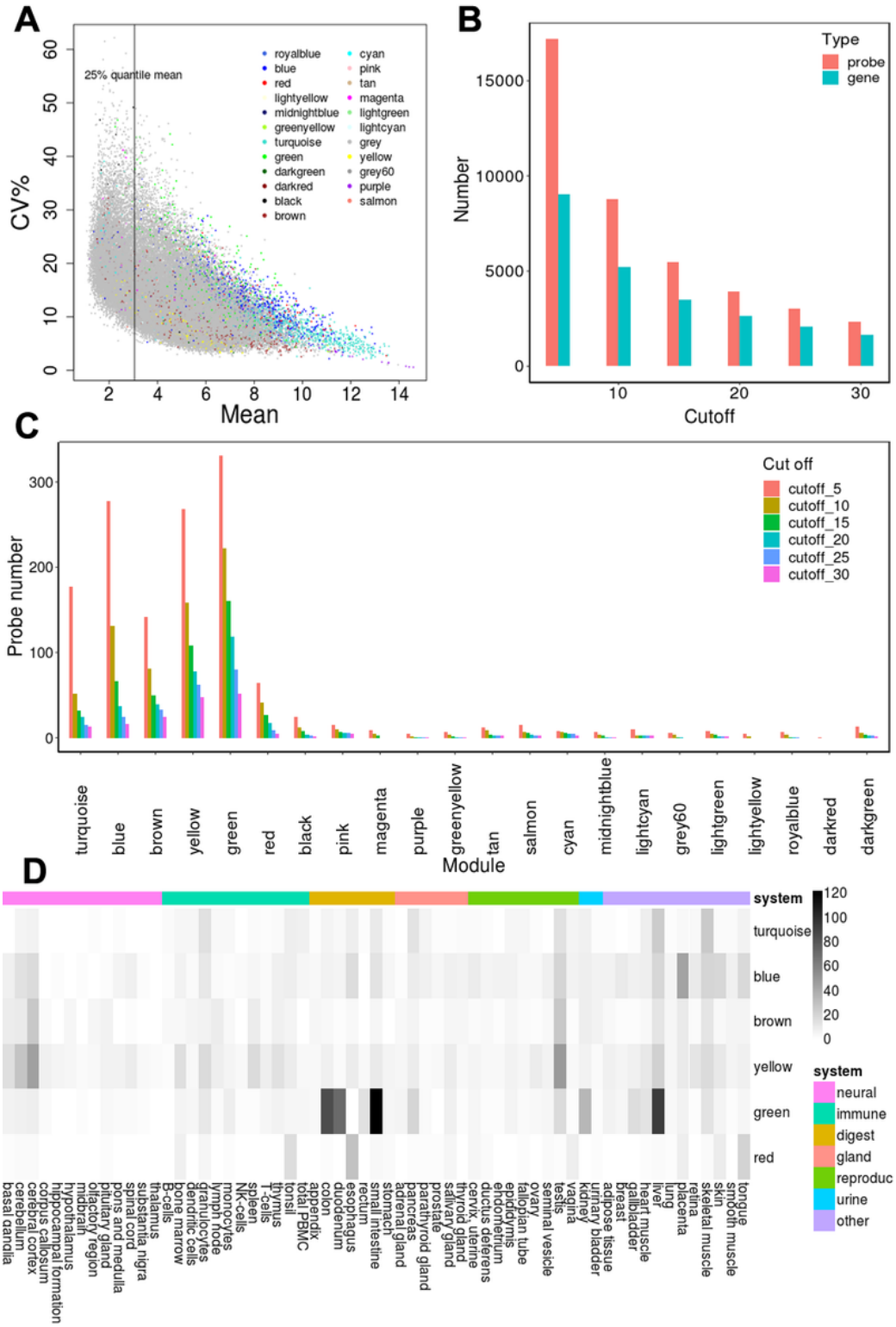

Figure 3

Expression features of genes in clustered modules (A) Distribution of clustered probe sets in all probe sets based on mean expression value and coefficient of variation. Clustered probe sets were labeled as module colors. Probe sets with higher expression values were more likely classified into co-expression modules. (B) Numbers of tissue-specific genes (probes) under different cut-off value. Number of tissue-specific genes (probes) decreased with higher cut-off value. (C) Distribution of tissue-specific probe sets

in all modules. Tissue-specific probe sets were mainly distributed in turquoise, blue, brown, yellow, green and red modules. (D) Numbers of tissue-specific probe sets in major modules and tissues. Tissues (shown in X axis) were classified into digest, gland, immune, neural, reproductive, urine and other systems. Dominant tissues and dominant modules were different for different modules and for different tissues respectively.

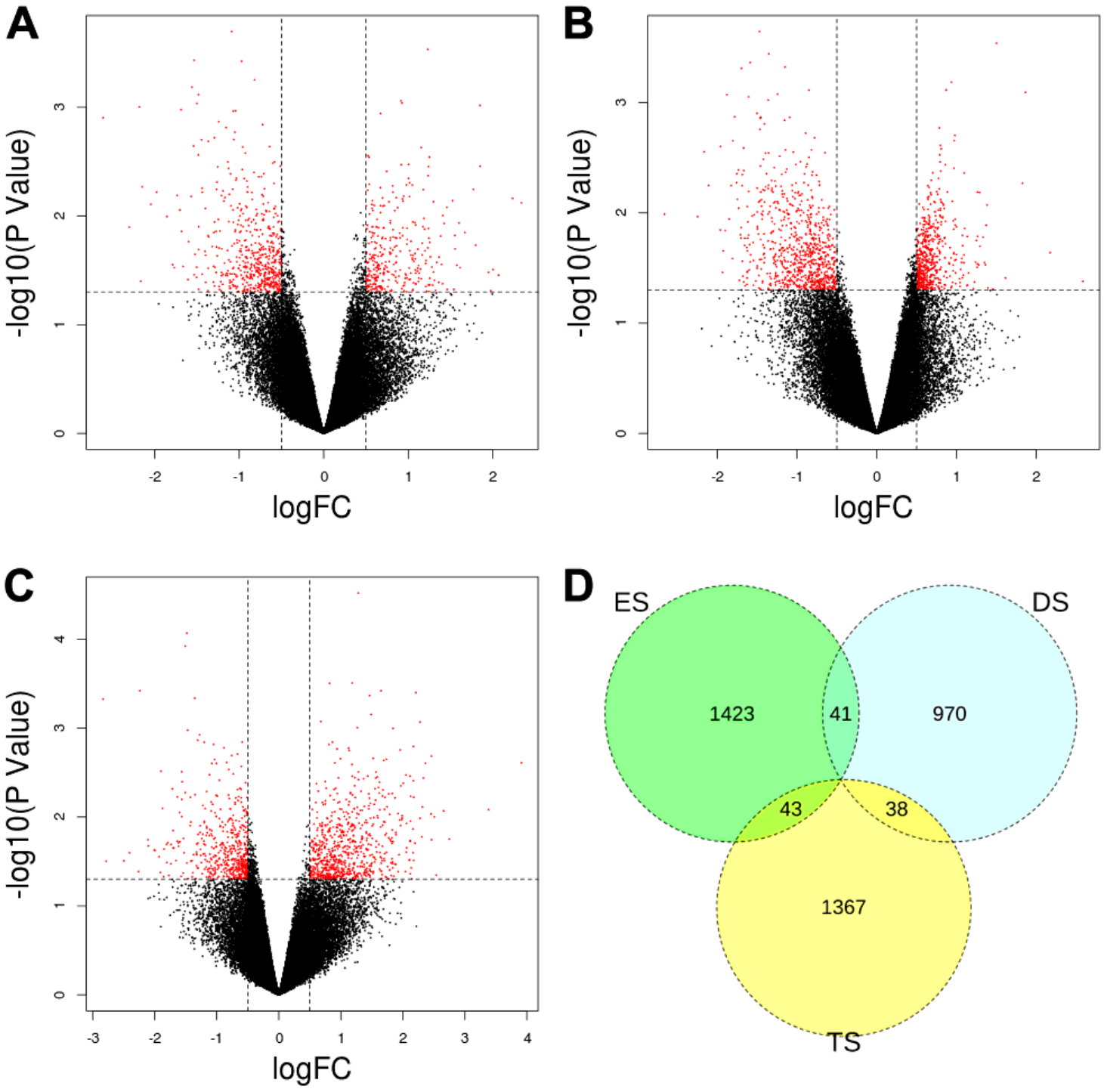

Figure 4

Differential expressed probe sets analysis in DS, ES and TS (A-C) Volcano map of differential expressed probes (DEPs) in DS (A), ES (B) and TS (C). (D) Relationship of DEPs in three groups. We found 1049, 1507 and 1448 different expressed probes (DEPs) in DS, ES and TS respectively. There are 41 common DEPs in DS intersecting ES, 38 common DEPs in DS intersecting TS and 43 DEPs in ES intersecting TS respectively. No common DEPs were found DS, ES and TS.

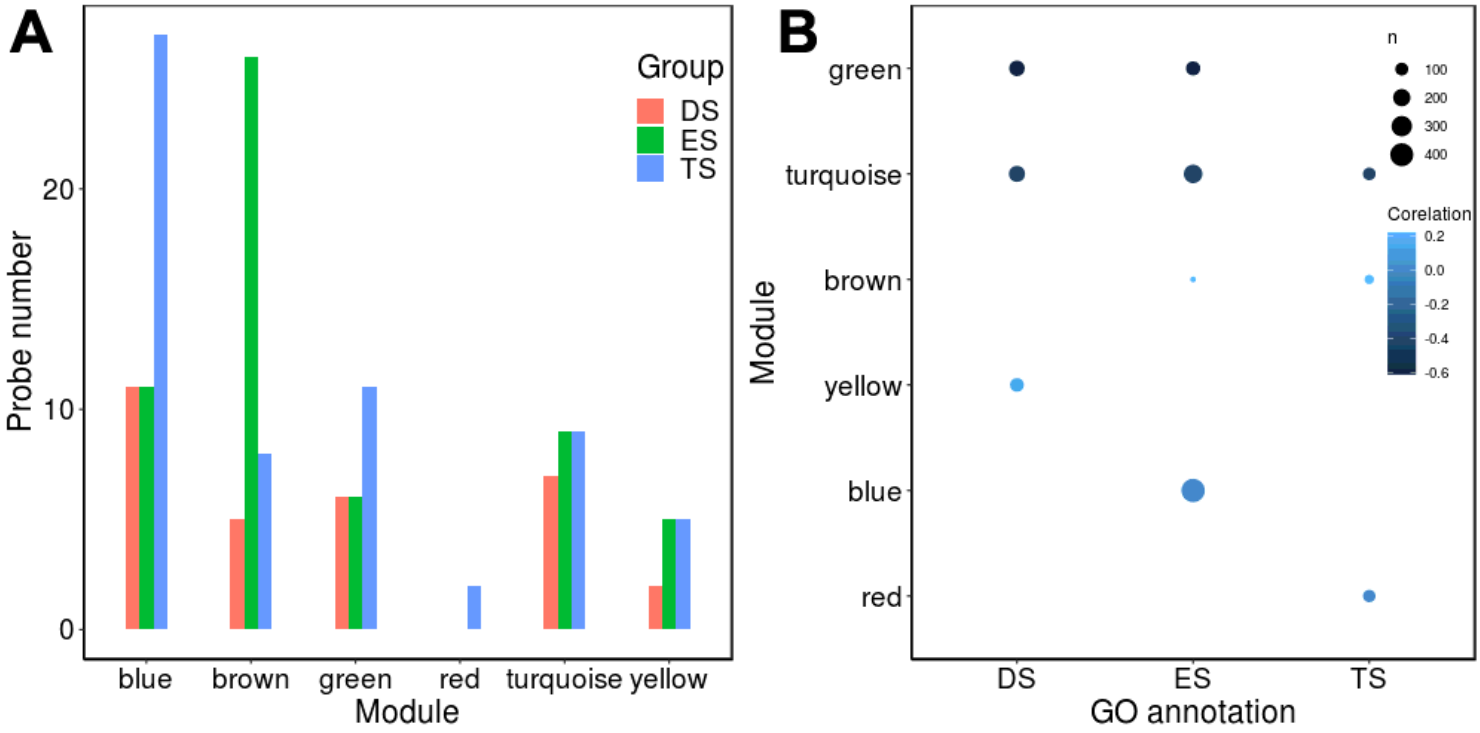

Figure 5

Functional enrichment analysis of modules related DEPs (A) Numbers of DEPs of DS, ES and TS in major clustered modules were counted. Blue module was the most abundant distributed modules in TS and DS group. However, brown module had the largest number of DEPs of ES. (B) Numbers of enriched GO terms for modular DEGs in DS, ES and TS. ES enriched the largest number of GO terms. Yellow, blue and red modules were specific modules for DS, ES and TS modules for GO analysis.

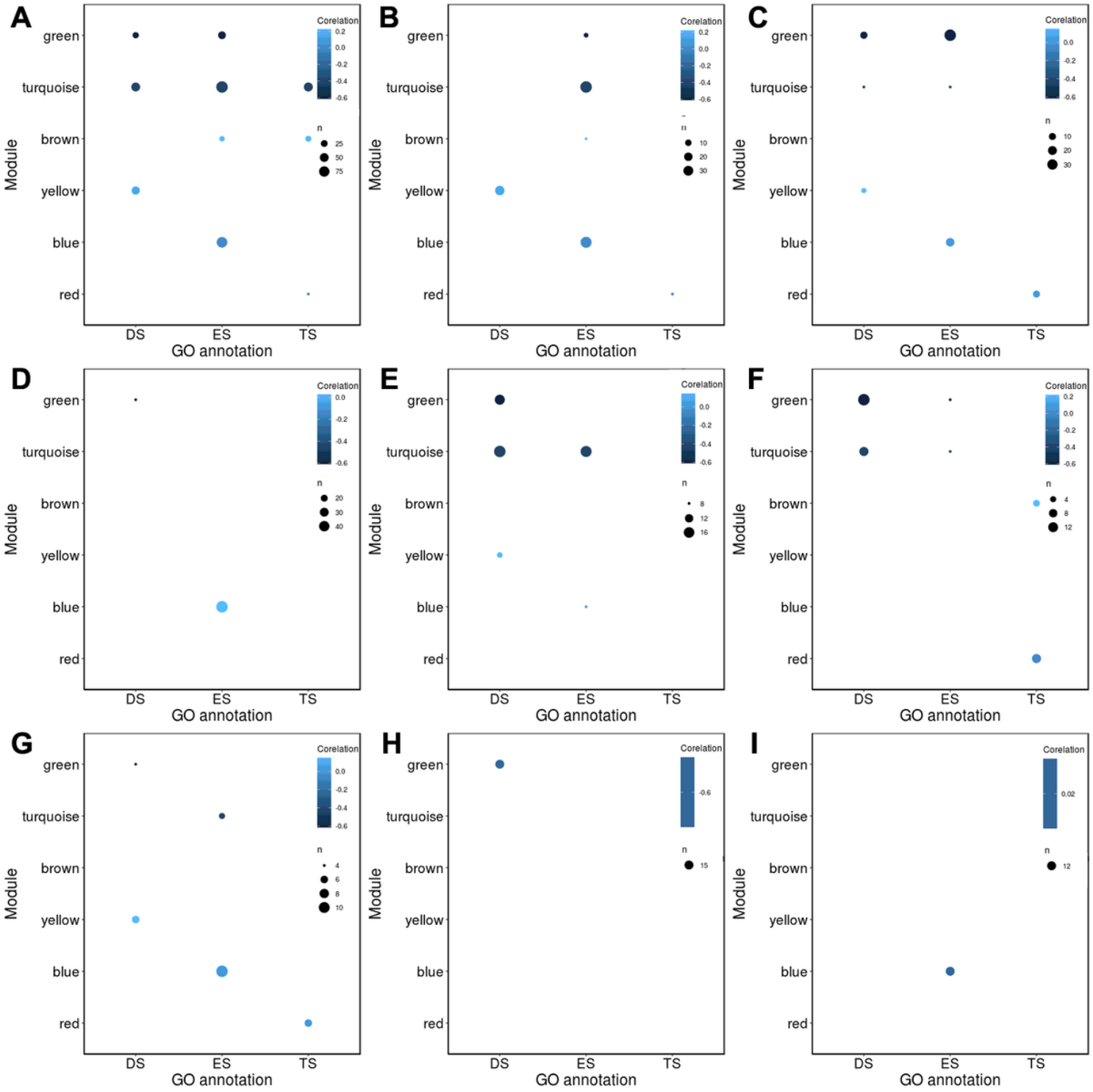

Figure 6

Modular distribution of GO term subsets in abnormal fetus (A) Basic physiological processes. (B) Absorption and transport of nutrients. (C) Response to external stimuli. (D) Multi-organ (Kidney, lung and heart) development. (E) Protein synthesis process. (F) Protein catabolic process and proteolysis. (G) Thermoregulation. (H) Signal transduction of TGF. (I) Bone development. In addition to yellow, blue and red modules were specific modules for DS, ES and TS groups, more specific modules were shown in certain subsets. Green, turquoise and brown module was specific for ES group in absorption and

transport of nutrients subset. Green module was specific to DS group in multi-organ (Kidney, lung and heart) development subset, protein synthesis process subset, thermoregulation subset and signal transduction of TGF subset.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.csv
- TableS2.csv
- TableS3.csv
- TableS4.csv
- TableS5.csv
- TableS6.csv
- TableS7.csv
- TableS8.csv
- TableS9.csv
- TableS10.csv
- Fig1S.Png