Published in final edited form as:

Am J Obstet Gynecol. 2015 July ; 213(1): 59.e1–59.e172. doi:10.1016/j.ajog.2015.03.023.

The Pathway Not Taken: Understanding ‘Omics Data in the Perinatal Context

Andrea G. EDLOW, MD, MSc^{1,2}, Donna K. SLONIM, PhD³, Ms Heather C. WICK, BS³, Lisa HUI, MBBS, PhD⁴, and Diana W. BIANCHI, MD^{2,5}

¹ Division of Maternal-Fetal Medicine, Department of Ob-Gyn, Tufts Medical Center, Boston, MA, United States.

² Mother Infant Research Institute, Tufts Medical Center, Boston, MA, United States

³ Computer Science, Tufts University, Medford, MA, United States.

⁴ Perinatal Medicine, Mercy Hospital for Woman, Heidelberg, VIC, Australia.

⁵ Division of Genetics, Department of Pediatrics, Tufts Medical Center, Boston, MA, United States.

Abstract

Objective—“Omics” analysis of large datasets has an increasingly important role in perinatal research, but understanding gene expression analyses in the fetal context remains a challenge. We compared the interpretation provided by a widely-used systems biology resource (Ingenuity Pathway Analysis, IPA) to that from Gene Set Enrichment Analysis (GSEA) with functional annotation curated specifically for the fetus (Developmental Functional Annotation at Tufts or DFLAT).

Study Design—Using amniotic fluid supernatant transcriptome datasets previously produced by our group, we analyzed three different developmental perturbations: aneuploidy (Trisomy 21, T21); hemodynamic (twin-twin transfusion syndrome, TTTS); and metabolic (maternal obesity, MAT OB) versus sex and gestational age matched controls. Differentially expressed probe IDs were identified using paired t-tests with the Benjamini-Hochberg correction for multiple testing (BH $p < 0.05$). Functional analyses were performed using IPA and GSEA/DFLAT. Outputs were compared for biological relevance to the fetus.

Results—Compared to controls, there were 414 significantly dysregulated probe IDs in T21 fetuses, 2226 in TTTS recipient twins, and 470 in fetuses of obese women. Each analytic output was unique but complementary. For T21, both IPA and GSEA/DFLAT identified dysregulation of

Corresponding author: Andrea G. Edlow, MD MSc Mother Infant Research Institute Tufts Medical Center 800 Washington Street, Box 394 Boston, MA 02111 aedlow@tuftsmedicalcenter.org Ph (work): 617-636-1468 Ph (cell): 215-913-6611 Fax: 617-636-1469.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The authors report no conflicts of interest.

Presentation at a scientific meeting: This work has been accepted for presentation at the 35th Annual Pregnancy Meeting of the Society for Maternal-Fetal Medicine, San Diego, CA, February 2-7, 2015; abstract #89 (February 5, 2015)

brain, cardiovascular, and integumentary system development. For TTTS, both analytic tools identified dysregulation of cell growth/proliferation, immune and inflammatory signaling, brain, and cardiovascular development. For maternal obesity, both identified dysregulation of immune and inflammatory signaling; brain and musculoskeletal development; and cell death. GSEA/DFLAT identified substantially more dysregulated biological functions in fetuses of obese women (1203 vs. 151). For all three datasets, GSEA/DFLAT provided more comprehensive information about brain development. IPA consistently provided more detailed annotation about cell death. IPA produced many dysregulated terms pertaining to cancer (14 in T21, 109 in TTTS, 26 in MAT OB); GSEA/DFLAT did not.

Conclusion—Interpretation of the fetal AFS transcriptome depends on the analytic program. This suggests that more than one resource should be utilized. Within IPA, physiologic cellular proliferation in the fetus produced many “false positive” annotations pertaining to cancer, reflecting its bias toward adult diseases. This study supports the use of gene annotation resources with a developmental focus, such as DFLAT, for ‘omics studies in perinatal medicine.

Keywords

amniotic fluid; bioinformatics; fetus; gene expression; transcriptome

Introduction

The growing awareness of the impact of the *in utero* environment on life-long health¹⁻³ has coincided with recognition of the ability to obtain real-time information about fetal development from cell-free fetal RNA in amniotic fluid.⁴ The amniotic fluid transcriptome has been utilized by our group and others to obtain valuable information about fetal development in a variety of health and disease states.⁴⁻¹⁰ The literature on fetal molecular biology has expanded exponentially in recent years, with an increasing focus on a variety of fetal transcriptomic studies.^{9,11-17} The need to adapt or customize transcriptomic bioinformatics analysis to obtain more relevant and interpretable output has been recognized by a wide variety of other disciplines, ranging from researchers studying breast cancer¹⁸ to chicken models of disease.^{19,20} Obstetrician gynecologists face unique issues in interpreting ‘omics data, as the performance of widely-used systems biology analytic resources has never been specifically evaluated for application in the fetus or placenta. Members of our group previously have addressed the need for more fetal-focused gene expression analytic tools by adding human-specific, developmentally-relevant annotation to the Gene Ontology (GO) database²¹ and maintaining a collection of gene sets tailored for use in studying human development, called “Developmental Functional Annotation at Tufts” or DFLAT (<http://dflat.cs.tufts.edu>).²² Using these gene sets in the Gene Set Enrichment Analysis²³ (GSEA/DFLAT), we sought to compare interpretation provided by this publicly-available fetus-specific functional annotation to that of a commercially-available widely-used functional analytic tool, Ingenuity Pathway Analysis (IPA).

Materials and Methods

In order to compare the functional analytic output of GSEA/DFLAT versus IPA, we performed an *in silico* experiment utilizing three amniotic fluid supernatant (AFS)

transcriptome datasets previously produced by our group and publicly available in the Gene Expression Omnibus [GEO] (GSE16176, GSE47393, GSE48521). These datasets represent three different developmental perturbations in second-trimester fetuses: aneuploidy (Trisomy 21, T21); hemodynamic (twin-twin transfusion syndrome, TTTS); and metabolic (maternal obesity, MAT OB). Each dataset contains information obtained from cell free RNA in AFS from 14-16 fetuses. Within each dataset, cases were matched to controls for gestational age and fetal sex, both of which have been demonstrated to influence fetal gene expression.^{24,25} There was no pooling of samples.

The original amniotic fluid samples for these studies were collected with human subjects approval from the Institutional Review Board at Tufts Medical Center and from each of the participating centers. Subjects signed informed consent for amniocentesis, which was performed for routine clinical indications. Details of subject recruitment and sample collection, as well as RNA extraction, amplification, and microarray hybridization have been previously described.^{5,7,8} All studies utilized the same whole genome expression array, the Affymetrix HGU133 Plus 2.0 (Affymetrix, Santa Clara, CA). The matched case and control gene expression data, experimental conditions, and data normalization methods are publicly available in the associated GEO records. Microarray data for all three datasets were normalized using the three-step command from the affyPLM package in Bioconductor, using ideal-mismatch background-signal adjustment, quantile normalization, and the Tukey biweight summary method.²⁶ This summary method includes a logarithmic transformation to improve the normality of the data. Identification of differentially-regulated probe IDs in cases versus controls was performed via 2-sided paired t-tests, with the Benjamini Hochberg (BH) adjustment for multiple testing. BH- $p < 0.05$ was defined as significant. Three working files containing significantly differentially-regulated probe IDs were generated to perform the IPA analyses on the three datasets. Supplementary Table/File 1 contains the “working files” for the IPA analyses.

Functional genomic analysis

Functional analyses were performed using the IPA “Core Analysis” function (content version 18841524, release date 6/24/13), and via GSEA, using the DFLAT-augmented Gene Ontology Biological Process gene sets. Outputs were compared for biological relevance to the fetus.

Within IPA, both up- and downregulated probe IDs were incorporated into the analysis. We considered pathways and functional annotations to be significantly dysregulated if they were associated with a right-tailed Fisher's exact $p < 0.01$, or bias-corrected absolute Z score 2.^{27,28} Only those functional annotations or terms associated with three or more genes were considered in the IPA and GSEA/DFLAT analyses. Only the “Diseases & Functions” aspect of the IPA analysis could be directly compared to the DFLAT/GSEA analysis, given that there is no direct GSEA correlate for IPA's Canonical Pathways, Upstream Analysis, Regulator Effects, and Networks analysis modes. For this *in silico* experiment, we focused on the “Diseases and Functions” “Canonical Pathways” and “Upstream Analysis” modes within IPA.

In the Canonical Pathways function of IPA, pathways were considered to be significantly dysregulated if they were associated with a right-tailed Fisher's exact p-value < 0.01.²⁷ The Upstream Analysis feature of IPA was used to predict the activation or inhibition of transcriptional regulators based on the direction of gene expression changes in our data set. We defined upstream regulators as significantly activated or inhibited if the activation Z-score was ≥ 2.0 or ≤ -2.0 , in accordance with recommended thresholds.²⁸

The combined DFLAT and Gene Ontology (GO) annotations of human genes can be downloaded as gene sets formatted for use in GSEA. The DFLAT annotation contains 13,344 new terms to use in conjunction with the existing GO annotation. The derivation and validation of DFLAT has been described in detail in a previous publication.²² Briefly, DFLAT was created via manual curation from the literature using the Protein2GO curation tool and the methodology of the Gene Ontology Consortium, as well as Gene Ontology Non-Eligible annotations and mouse-to-human ortholog derived annotations. DFLAT was then validated using both external datasets and ones produced by our own laboratory.

For all analyses, the java implementation of GSEA (version 2-2.07) was run in batch mode. GSEA was run using the preranked option, ranking by paired t-scores, for greater consistency with the IPA input and to preserve the original matching of AFS case and control samples for gestational age and fetal sex. Gene sets were considered to be significantly dysregulated if they were associated with false discovery rate q-values (FDR q) of <0.25, in accordance with recommended stringency thresholds.²³ We extended the analysis to include gene sets with raw p-values < 0.01, given controversy about adjusting for multiple testing between highly-overlapping gene sets.^{22,29,30}

Because the Gene Sets in the DFLAT annotation of GO have different names than the categories and functional annotations within IPA, comparing the two outputs required identifying and categorizing common and unique annotations. All significantly dysregulated gene sets and IPA annotation terms identified in the functional analysis of the three datasets were manually reviewed by the first author, and a list of 23 developmental categories and 443 associated keywords was created (detailed in Supplementary Table 2). These categories and keywords encompassed significantly dysregulated annotations within IPA Molecular and Cellular Functions, Physiological System Development and Function, and Diseases and Disorders categories, in addition to unique Gene Set categories within GSEA/DFLAT that did not correspond to any categories within IPA. We wrote Perl scripts to count the number of terms within each category identified by DFLAT/GSEA versus IPA. Outputs were manually reviewed for accuracy. A category was designated as "common to both" GSEA/DFLAT and IPA if there were at least 5 significantly dysregulated functional annotations or terms in that category for each analysis. A category was determined to be "more frequent" in IPA or GSEA/DFLAT if there were at least 5 significantly dysregulated functional annotations or terms in that category and there was at least a three-fold difference between the two methods for that category.

Results

Compared to controls, paired t-tests with the BH correction identified 414 significantly dysregulated probe IDs in T21 fetuses, 2226 in TTTS recipient twins, and 470 in fetuses of obese women.

Direct comparison of dysregulated annotations

A direct comparison of the Diseases & Functions output on IPA with the gene set terms identified by GSEA/DFLAT identified unique but complementary outputs for each dataset. Table 1 summarizes a broad overview of the results by functional analysis tool. Figure 1 depicts the relative number of selected dysregulated biological functions for each dataset that were detected by IPA versus GSEA/DFLAT. Considering all genes within significantly dysregulated pathways in both resources, Figure 2 depicts the proportion of genes uniquely recognized by IPA or GSEA/DFLAT, as well as the proportion recognized by both resources. This figure demonstrates that there is very little overlap in the genes identified by each resource within significantly dysregulated pathways. Figure 3 contains heatmaps for all genes within significantly dysregulated pathways for each dataset using IPA and GSEA/DFLAT. The expression values are restricted to only the probe IDs that were associated with BH p-values < 0.05. IPA and GSEA/DFLAT show distinct gene expression patterns for all three datasets, but regardless of the systems biology resource used, gene expression clusters by phenotype. A complete list of all developmental categories and number of dysregulated biological functions recognized by each method may be found in Supplementary Table 3.

Trisomy 21—For T21, both IPA and GSEA/DFLAT identified dysregulation of terms related to brain, cardiovascular, and integumentary system development. There were 4-fold more cardiovascular system terms in IPA compared to GSEA/DFLAT (20 vs 5). Dysregulation of immune/inflammatory signaling, cell-to-cell signaling, cancer-related annotations, and cell growth and proliferation figured prominently in the IPA analysis but were not well represented in GSEA/DFLAT. The GSEA/DFLAT analysis was heavily weighted toward dysregulated musculoskeletal development, cell cycle progression, MAP kinase stress-response signaling cascades, and RNA regulation/processing. The GSEA/DFLAT analysis contained more terms and more detail regarding dysregulated brain and nervous system development than did the IPA analysis. Supplementary Tables 4 and 5 contain a detailed description of dysregulated functional annotations identified by the IPA and GSEA/DFLAT analyses for T21.

Twin-Twin Transfusion Syndrome—For recipient twins in TTTS, both IPA and GSEA/DFLAT identified significant dysregulation of immune, brain, cardiovascular, hematologic, endocrine, and musculoskeletal system development, in addition to dysregulated cell growth and proliferation and cell cycle progression. The IPA analysis focused heavily on cancer-related terms, dysregulation of cell growth and proliferation, immune and inflammatory signaling, cell death, and reproductive system development, among others. The GSEA/DFLAT analysis was dominated by terms related to brain and nervous system development, lipid metabolism, phosphate metabolism, and calcium ion regulation. IPA provided more detail about dysregulated CV development than did GSEA/

DFLAT, but the IPA analysis included difficult-to-interpret annotations related to adult CV disease, including arterial occlusion, atherosclerosis, peripheral arterial occlusive disease, and carotid artery disease. Both methods also had many dysregulated terms related to brain development, but GSEA/DFLAT contained more terms (45 vs. 31) and more detail in this regard than did IPA. Supplementary Tables 6 and 7 contain a detailed description of dysregulated functional annotations identified by the IPA and GSEA/DFLAT analyses for TTTS.

Maternal obesity—For MAT OB, both IPA and DFLAT identified dysregulation of terms related to immune and inflammatory signaling; brain development; musculoskeletal, cardiovascular, reproductive, integumentary, and gastrointestinal system development; cell death, and cell growth and proliferation. GSEA/DFLAT identified substantially more (1203 vs. 151) dysregulated biological functions, and in every category except Cell Death provided a wider diversity of dysregulated terms and more detail about development. The GSEA/DFLAT analysis was heavily weighted toward terms related to immune, brain/nervous, musculoskeletal, and cardiovascular system development. GSEA/DFLAT also highlighted dysregulated cell growth and proliferation; insulin signaling, carbohydrate and lipid metabolism; phosphate metabolism, RNA regulation and processing, and response to oxidative stress, while these functional annotations featured minimally, if at all, in the IPA analysis. Supplementary Tables 8 and 9 contain a detailed description of dysregulated functional annotations identified by the IPA and GSEA/DFLAT analyses.

Cancer-Related Annotations—For all three datasets, IPA produced many dysregulated terms pertaining to cancer, for example, “malignant neoplasm of endocrine gland”, “carcinoma in lung”, “breast cancer cell line invasion”, “leukemia”, head and neck cancer” “pancreatic cancer cell lines”, and “gastroesophageal cancer.” IPA identified a total of 14 cancer-related terms in the T21 dataset, 109 in the TTTS dataset, and 26 in MAT OB dataset. GSEA/DFLAT produced no such terms. Keywords used to detect cancer-related annotations are detailed in Supplementary Table 2.

The Canonical Pathways and Upstream Analysis features of IPA do not have a parallel feature in GSEA/DFLAT. A full list of significantly dysregulated Canonical Pathways and Upstream Regulators for each of the three datasets is provided in Supplementary Tables 10 and 11. The gestational age and fetal sex of all cases and controls for the three datasets may be found in Supplementary Table 12.

Comment

The increasing focus on high-throughput techniques for research and clinical applications in perinatology has led to a greater need for meaningful interpretation of large ‘omics datasets. There are scant data regarding the applicability of standard functional analytic tools to fetal/perinatal gene expression datasets. In this comparative analysis of two systems biology resources for fetal gene expression analysis, we demonstrated that IPA and GSEA/DFLAT provided different but complementary interpretations, with IPA providing a wider variety of analytic modes and GSEA/DFLAT giving more focused, fetal-specific information.

While the outputs overlapped in many instances, both provided important and distinct insights for each dataset. GSEA/DFLAT provided more detailed and comprehensive information about brain and nervous system development, while IPA provided more comprehensive annotation about cell death for the three datasets considered. The TTTS analysis serves as a representative example of how the increased detail about nervous system development in GSEA/DFLAT might provide more mechanistic insight. In this analysis, IPA identified fewer, and less specific functional annotations pertaining to brain development compared to GSEA/DFLAT. For example, IPA identified “increased cell death of brain”, “increased cell death of cerebral cortex cells”, and “decreased quantity of CNS cells” in recipient twins based on the patterns of gene expression, while GSEA/DFLAT's significantly dysregulated terms included “regulation of glutamatergic synaptic transmission,” “thalamus development,” “dentate gyrus development,” and “forebrain generation of neurons,” among others.

Key differences between IPA and GSEA/DFLAT for analyses of these three fetal datasets are summarized in Table 2. We identified three potential advantages of GSEA with DFLAT annotation over IPA. First, GSEA/DFLAT's use of Gene Ontology terms is more transparent than IPA's proprietary database, and GSEA/DFLAT is publicly available. Second, GSEA/DFLAT provided more detailed information about nervous system development than did IPA for all three datasets. Finally, perhaps due to its fetal-specific annotation, GSEA/DFLAT did not interpret normal fetal proliferation as cancer or tumor-related terms.

We identified three potential advantages of IPA over GSEA/DFLAT. First, the IPA software interface provided a more accessible visual presentation of the results and easier identification of genes implicated in a particular functional annotation. In addition, IPA includes a direct link to their bibliographic database entry for each gene within a pathway, for easier identification of putative gene function. These links are available for the DFLAT annotation, but locating them is not as straightforward. Second, IPA integrated the information of the differentially-regulated functional annotations into a single output for greater ease of use, while GSEA/DFLAT generated separate lists of up- and downregulated terms. Finally, the additional analysis tools available on IPA, in particular the Canonical Pathways and Upstream Regulator analysis, allowed for the identification of biologically-relevant findings that had the potential to provide novel insights into fetal development in the maternal and fetal conditions we examined.

One representative example of the utility of the IPA Canonical Pathways function was demonstrated by the maternal obesity dataset. In that analysis, IPA identified dysregulation of two canonical pathways related to folate metabolism, Folate Transformations I and Folate Polyglutamylation. This finding has the potential to provide additional mechanistic insight into the known increased risk of neural tube defects in fetuses of obese women.³¹⁻³³ Highlighting the synergistic nature of these resources, GSEA/DFLAT identified “neural tube development/formation/closure” and “dorsal/ventral neural tube patterning” as enriched terms in fetuses of obese women. These findings demonstrate how using the two resources together suggests clinically relevant information about the underlying mechanism of neural tube defects, providing hypotheses for future research.

One major disadvantage of IPA compared to GSEA/DFLAT is IPA's bias toward adult disease states. This was apparent in the many significantly dysregulated terms pertaining to cancer and degenerative adult-onset diseases. Our results show that performing a complementary analysis with a tool specifically annotated for development allows for the most biologically relevant interpretation of gene expression data.

Very few studies have compared IPA to GSEA/GO in any context^{34,35}, and we know of none that have directly compared the utility of different systems biology resources for interpretation of fetal datasets. Of the studies comparing IPA to GSEA, one compared the two analytic methods for accuracy of detection of Molecular Signatures DataBase (MSigDB) oncogenic signatures. This study found that GSEA accurately identified more MSigDB oncogenic signatures than did IPA, and that IPA and GSEA had equal performance for predicting treatment of samples with TGF β and identifying progesterone receptor status in different patient populations.³⁵ Similar to our observations, these authors recognized the utility of the Upstream Regulator analysis function of IPA in generating novel insights into activated and inhibited signaling pathways. The second paper identified slightly better performance of IPA compared to GSEA for identification of gene signatures reflecting deranged lipid metabolism, and found that within GSEA, a single gene within a pathway can strongly drive statistical significance, citing this as a limitation of GSEA.³⁴ These two studies are difficult to directly compare to ours, given that they focused primarily on the performance of IPA and GSEA for adult disease gene expression signatures (cancer and lipid metabolic derangements), rather than fetal gene expression. There is a small body of literature on bioinformatics approaches to embryo and single-blastomere transcriptomics, but these studies focus largely on non-human embryos rather than human fetuses, and did not utilize functional analysis tools such as GSEA or IPA.^{36,37}

The primary strength of this study is the novel comparison of standard systems biology resources to a new, developmentally-relevant annotation to the GO database. Additionally, the study includes only human samples and utilizes three fetal gene expression datasets produced in our own laboratory, allowing for the greatest uniformity of sample processing, bench workflow, and data normalization procedures. The maternal obesity dataset represents a more physiologically heterogeneous group than the TTTS and T21 datasets, with the potential for misclassification bias in the setting of possible maternal comorbid conditions. However, performing a comparative analysis of all three datasets allowed us to look for consistent biases across the different analysis tools independent of specific fetal or maternal disease process. For example, IPA's bias toward cancer annotations was evident in all three analyses, suggesting that this misclassification of fetal cell proliferation by IPA is a systematic issue.

Samples from the Trisomy 21 dataset (but not TTTS or MATOB) may have had a minor influence on the prioritization of genes for annotation in the early stages of the DFLAT project, which has the potential to create slight bias in this comparison. Of note, of the 13,344 annotations in DFLAT, less than 2% had the potential to be influenced by the T21 dataset. The DFLAT annotation was then validated using both external and internal datasets, so any bias toward internally-produced datasets should be minimal.²² Other limitations include the difficulties inherent in comparing two resources that have different functional

analytic modes and therefore different outputs. We attempted to overcome this limitation by manually categorizing the outputs of the two methods and writing Perl scripts to enable a more direct comparison of the two. With the recognition that more annotation terms does not necessarily translate to a “better” analysis, this script gave us the ability to quantitatively compare the two methods, while the manual categorization allowed a qualitative review.

The long-term goal of this work is the development of targeted treatments that improve pregnancy outcomes. The field of perinatal medicine now recognizes that a systems biology approach is a powerful method for understanding fetal diseases and identifying future candidate fetal therapies. The unbiased nature of whole transcriptome studies overcomes the substantial difficulty in selecting specific genes or pathways to study in the living human fetus. The importance of analyzing whole genome expression data lies in its potential to assist in the identification of novel therapies via resources such as the Connectivity Map.³⁸ Current work from our laboratory on a prenatal therapy for Trisomy 21 using specific antioxidants is a direct example of this approach.^{39,40}

In conclusion, the finding that IPA together with GSEA/DFLAT provided more complete information than either tool alone suggests that more than one functional analytic resource should be utilized in interpretation of fetal transcriptomic data. Relevant and developmental-specific findings provide the best opportunity to understand the molecular underpinnings of fetal programming in a variety of disease states, and therefore to design and implement effective interventions and therapies. The finding of “false positive” annotations within IPA pertaining to cancer and adult-onset diseases highlights the importance of gene annotation resources with a developmental focus in interpreting perinatal gene expression datasets. Future research should focus on creating a developmental/fetal-specific annotation to other widely used systems biology resources, such as IPA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Source of financial support: This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) grants R01HD42053-10 (D.W.B.), R01HD076140 (D.K.S.)

Role of the funding source: The funding source had no involvement in study design, collection, analysis and interpretation of data, the writing of the report, or the decision to submit the article for publication.

References

1. Barker DJ. Fetal origins of coronary heart disease. *BMJ*. 1995; 311:171–4. [PubMed: 7613432]
2. Barker DJ, Osmond C, Simmonds SJ, Wield GA. The relation of small head circumference and thinness at birth to death from cardiovascular disease in adult life. *BMJ*. 1993; 306:422–6. [PubMed: 8461722]
3. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med*. 2008; 359:61–73. [PubMed: 18596274]
4. 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*. 119:111–8. [PubMed: 22183218]

5. Slonim DK, Koide K, Johnson KL, et al. 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; 106:9425–9. [PubMed: 19474297]
6. 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; 121:1248–54. [PubMed: 23812459]
7. Hui L, Wick HC, Moise KJ Jr. et al. Global gene expression analysis of amniotic fluid cell-free RNA from recipient twins with twin-twin transfusion syndrome. *Prenatal diagnosis*. 2013; 33:873–83. [PubMed: 23640821]
8. 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:e88661. [PubMed: 24558408]
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. *Human genetics*. 2011; 129:295–305. [PubMed: 21152935]
10. Romero R, Espinoza J, Goncalves LF, Kusanovic JP, Friel LA, Nien JK. Inflammation in preterm and term labour and delivery. *Semin Fetal Neonatal Med*. 2006; 11:317–26. [PubMed: 16839830]
11. Madsen-Bouterse SA, Romero R, Tarca AL, et al. The transcriptome of the fetal inflammatory response syndrome. *American journal of reproductive immunology*. 2010; 63:73–92. [PubMed: 20059468]
12. Choe HK, Son GH, Chung S, et al. Maternal stress retards fetal development in mice with transcriptome-wide impact on gene expression profiles of the limb. *Stress*. 2011; 14:194–204. [PubMed: 21291317]
13. Votavova H, Dostalova Merkerova M, Fejglova K, et al. Transcriptome alterations in maternal and fetal cells induced by tobacco smoke. *Placenta*. 2011; 32:763–70. [PubMed: 21803418]
14. Stunkel W, Pan H, Chew SB, et al. Transcriptome changes affecting Hedgehog and cytokine signalling in the umbilical cord: implications for disease risk. *PLoS One*. 2012; 7:e39744. [PubMed: 22808055]
15. 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. *Human genetics*. 2014; 133:1075–82. [PubMed: 24850140]
16. Maron JL, Johnson KL, Slonim D, 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. [PubMed: 17885688]
17. Thakali KM, Saben J, Faske JB, et al. Maternal pregravid obesity changes gene expression profiles toward greater inflammation and reduced insulin sensitivity in umbilical cord. *Pediatr Res*. 2014; 76:202–10. [PubMed: 24819376]
18. Huang S, Yee C, Ching T, Yu H, Garmire LX. A novel model to combine clinical and pathway-based transcriptomic information for the prognosis prediction of breast cancer. *PLoS computational biology*. 2014; 10:e1003851. [PubMed: 25233347]
19. Jimenez-Marin A, Collado-Romero M, Ramirez-Boo M, Arce C, Garrido JJ. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC proceedings*. 2009; 3(Suppl 4):S6. [PubMed: 19615119]
20. Hedegaard J, Arce C, Bicciato S, et al. Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. *BMC proceedings*. 2009; 3(Suppl 4):S5. [PubMed: 19615118]
21. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000; 25:25–9. [PubMed: 10802651]
22. Wick HC, Drabkin H, Ngu H, et al. DFLAT: functional annotation for human development. *BMC bioinformatics*. 2014; 15:45. [PubMed: 24507166]
23. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50. [PubMed: 16199517]
24. Larrabee PB, Johnson KL, Lai C, et al. Global gene expression analysis of the living human fetus using cell-free messenger RNA in amniotic fluid. *JAMA*. 2005; 293:836–42. [PubMed: 15713773]

25. Massingham LJ, Johnson KL, Bianchi DW, et al. Proof of concept study to assess fetal gene expression in amniotic fluid by nanoarray PCR. *The Journal of molecular diagnostics : JMD*. 2011; 13:565–70. [PubMed: 21827969]
26. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*. 2004; 5:R80. [PubMed: 15461798]
27. Website I. Calculating and interpreting the p-values for functions, pathways, and lists in IPA. 2013
28. Website I. Ingenuity Upstream Regulator Analysis Whitepaper. 2014
29. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006; 22:1600–7. [PubMed: 16606683]
30. Goeman JJ, Mansmann U. Multiple testing on the directed acyclic graph of gene ontology. *Bioinformatics*. 2008; 24:537–44. [PubMed: 18203773]
31. Haddow JE, Palomaki GE. Is maternal obesity a risk factor for open neural tube defects? *Am J Obstet Gynecol*. 1995; 172:245–7. [PubMed: 7847555]
32. Shaw GM, Todoroff K, Finnell RH, Lammer EJ. Spina bifida phenotypes in infants or fetuses of obese mothers. *Teratology*. 2000; 61:376–81. [PubMed: 10777833]
33. Parker SE, Yazdy MM, Tinker SC, Mitchell AA, Werler MM. The impact of folic acid intake on the association among diabetes mellitus, obesity, and spina bifida. *Am J Obstet Gynecol*. 2013; 209:239, e1–8. [PubMed: 23711668]
34. Hong MG, Pawitan Y, Magnusson PK, Prince JA. Strategies and issues in the detection of pathway enrichment in genome-wide association studies. *Human genetics*. 2009; 126:289–301. [PubMed: 19408013]
35. Warden CD, Kanaya N, Chen S, Yuan YC. BD-Func: a streamlined algorithm for predicting activation and inhibition of pathways. *PeerJ*. 2013; 1:e159. [PubMed: 24058887]
36. El-Sayed A, Hoelker M, Rings F, et al. Large-scale transcriptional analysis of bovine embryo biopsies in relation to pregnancy success after transfer to recipients. *Physiological genomics*. 2006; 28:84–96. [PubMed: 17018689]
37. Rodriguez-Zas SL, Ko Y, Adams HA, Southey BR. Advancing the understanding of the embryo transcriptome co-regulation using meta-, functional, and gene network analysis tools. *Reproduction*. 2008; 135:213–24. [PubMed: 18239050]
38. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006; 313:1929–35. [PubMed: 17008526]
39. Guedj F, Pennings JL, Wick HC, Bianchi DW. Analysis of Adult Cerebral Cortex and Hippocampus Transcriptomes Reveals Unique Molecular Changes in the Ts1Cje Mouse Model of Down Syndrome. *Brain pathology*. 2015; 25:11–23. [PubMed: 24916381]
40. Guedj F, Bianchi DW, Delabar JM. Prenatal treatment of Down syndrome: a reality? Current opinion in obstetrics & gynecology. 2014; 26:92–103. [PubMed: 24573065]

Condensation

Functional analysis of fetal transcriptomic datasets varies according to whether or not the resource has been specifically annotated for developmental processes.

Figure 1. Dysregulated biological and molecular functions by dataset
Selected biological and molecular functions detected by IPA and GSEA/DFLAT for trisomy 21 (A), twin-twin transfusion syndrome (B), and maternal obesity (C) datasets. Number of annotations is depicted on the y-axis.
IPA, Ingenuity Pathway Analysis; *GSEA/DFLAT*, Gene Set Enrichment Analysis with fetal-specific annotation (Developmental Functional Annotation at Tufts or DFLAT); *TTTS*, twin-twin transfusion syndrome; *Immune*, immune and inflammatory signaling; *Neuro*, brain/nervous system; *CV*, cardiovascular system; *Repro*, reproductive system; *Musculoskel*, musculoskeletal system; *Heme*, hematological system; *Ox Stress*, oxidative stress; *Lipid Metab*, lipid metabolism; *RNA Regulat*, RNA processing/transcriptional regulation; *Phos Metab*, phosphate metabolism; *MAPK Signal*, MAP Kinase signaling cascade; *Calcium Reg*, calcium ion regulation/transport.

Figure 2. Unique and common genes detected by IPA and GSEA/DFLAT, according to biological category and dataset

The proportion of unique genes in significantly dysregulated pathways detected by IPA is depicted in red, by GSEA/DFLAT in blue, and common genes detected by both in yellow. *IPA*, Ingenuity Pathway Analysis; *GSEA/DFLAT*, Gene Set Enrichment Analysis with fetal-specific annotation (Developmental Functional Annotation at Tufts or DFLAT); *T21*, Trisomy 21; *TTTS*, twin-twin transfusion syndrome; *OB*, maternal obesity; *Immune*, immune and inflammatory signaling; *Neuro*, brain/nervous system; *CV*, cardiovascular system; *Repro*, reproductive system; *Musc*, musculoskeletal system; *Heme*, hematological system

Figure 3. Heatmap of genes in significantly dysregulated pathways within IPA and GSEA/DFLAT

Gene expression patterns are depicted by systems biology resource and dataset. Yellow indicates higher expression (Z-scores > 2), red indicates lower expression (Z-scores < -2). *IPA*, Ingenuity Pathway Analysis; *GSEA/DFLAT*, Gene Set Enrichment Analysis with fetal-specific annotation (Developmental Functional Annotation at Tufts or DFLAT); *T21*, Trisomy 21; *TTTS*, twin-twin transfusion syndrome; *OB*, maternal obesity

Table 1

Dataset characteristics and summary of results by dataset

Dataset	Significantly dysregulated probe IDs ^a	Significantly dysregulated IPA annotations ^b	Significantly dysregulated GSEA/DFLAT terms ^c	Top 5 dysregulated developmental categories		
				Common to IPA and GSEA/DFLAT	More frequent in IPA	More frequent in GSEA/DFLAT
T21-euploid (7 matched pairs) ^d GSE16176 ^e	414	120	143	Brain/Nervous Cardiovascular Integumentary Cancer Cell Growth/Proliferation	Immune/Inflammatory Cell-to-cell Signaling Cardiovascular Cancer Cell Growth/Proliferation	Musculoskeletal Cell Cycle Regulation MAP Kinase Signaling Cascades RNA Processing/Regulation Phosphate Metabolism
TTTS-control (8 matched pairs) ^d GSE47393 ^e	2226	500	207	Cell Growth/Proliferation Immune/Inflammatory Brain/Nervous Musculoskeletal Cardiovascular Immune/Inflammatory Brain/Nervous Musculoskeletal Cardiovascular Cell Death Reproductive	Cancer Cell growth/Proliferation Immune/Inflammatory Cell Death Cancer	Lipid Metabolism Phosphate Metabolism Calcium Regulation/Transport Immune/Inflammatory
Obese-lean (8 matched pairs) ^d GSE4852 ^e	470	151	1203	Immune/Inflammatory Brain/Nervous Musculoskeletal Cardiovascular Cell Death	Cancer Cell Death Cancer	Brain/Nervous Musculoskeletal Cardiovascular Cell Growth/Proliferation

IPA, Ingenuity Pathway Analysis; GSEA/DFLAT, Gene Set Enrichment Analysis with fetal-specific annotation (Developmental Functional Annotation at Tufts or DFLAT); T21, Trisomy 21; TTTS, twin-twin transfusion syndrome.

^a BH-corrected $p < 0.05$

^b Fisher's exact $p < 0.01$ or absolute Z-score 2.0

^c Nominal/raw $p < 0.01$ or false-discovery rate $q < 0.25$

^d Pairs matched for gestational age and fetal sex

^e Gene Expression Omnibus datasets can be retrieved at www.ncbi.nlm.nih.gov/geo/, accession numbers GSE16176, GSE47393, and GSE48521.

Table 2

Advantages and disadvantages of systems biology resources

	GSEA/DPLAT	IPA
Methodology/ability to recreate results	Transparent, code available	Proprietary
Cost	Publicly available, free	Commercially-available, \$2,995 - \$8,820/seat/year
Fetal/developmental-focused annotation	Yes	No
Ease of use	Fewer visual interface options for the user; defaults to providing results in separate files of up- and downregulated annotations	More accessible visual presentation of results with links to the Ingenuity Knowledge Base; incorporates up- and downregulated annotations into a single output
Default analysis modes	Functional enrichment of Gene Ontology terms ^a	Networks, canonical pathways, upstream analysis, regulator effects, diseases and biofunctions, tox functions

IPA, Ingenuity Pathway Analysis; *GSEA/DPLAT*, Gene Set Enrichment Analysis with fetal-specific annotation (Developmental Functional Annotation at Tufts or DPLAT)

^a Most analogous to IPA's diseases and biofunctions analysis. Analyses comparable to many other IPA modes can be performed using other freely-available bioinformatics tools not evaluated here.