



DATA NOTE

REVISED A curated transcriptome dataset collection to investigate the development and differentiation of the human placenta and its associated pathologies [version 2; referees: 2 approved]

Alexandra K. Marr¹, Sabri Boughorbel¹, Scott Presnell², Charlie Quinn², Damien Chaussabel¹, Tomoshige Kino¹

¹Sidra Medical and Research Center, Doha, Qatar

²Systems Immunology Division, Benaroya Research Institute, Seattle, WA, USA

v2 First published: 09 Mar 2016, 5:305 (doi: [10.12688/f1000research.8210.1](https://doi.org/10.12688/f1000research.8210.1))
 Latest published: 11 May 2016, 5:305 (doi: [10.12688/f1000research.8210.2](https://doi.org/10.12688/f1000research.8210.2))

Abstract

Compendia of large-scale datasets made available in public repositories provide a precious opportunity to discover new biomedical phenomena and to fill gaps in our current knowledge. In order to foster novel insights it is necessary to ensure that these data are made readily accessible to research investigators in an interpretable format. Here we make a curated, public, collection of transcriptome datasets relevant to human placenta biology available for further analysis and interpretation via an interactive data browsing interface. We identified and retrieved a total of 24 datasets encompassing 759 transcriptome profiles associated with the development of the human placenta and associated pathologies from the NCBI Gene Expression Omnibus (GEO) and present them in a custom web-based application designed for interactive query and visualization of integrated large-scale datasets (<http://placentalendocrinology.gxbsidra.org/dm3/landing.gsp>). We also performed quality control checks using relevant biological markers. Multiple sample groupings and rank lists were subsequently created to facilitate data query and interpretation. Via this interface, users can create web-links to customized graphical views which may be inserted into manuscripts for further dissemination, or e-mailed to collaborators for discussion. The tool also enables users to browse a single gene across different projects, providing a mechanism for developing new perspectives on the role of a molecule of interest across multiple biological states. The dataset collection we created here is available at: <http://placentalendocrinology.gxbsidra.org/dm3>.

Open Peer Review

Referee Status:

	1	2
REVISED version 2 published 11 May 2016	<input checked="" type="checkbox"/> report	
	↑	
version 1 published 09 Mar 2016	<input type="checkbox"/> ? report	<input checked="" type="checkbox"/> report

1 **Brian Cox**, University of Toronto Canada

2 **Maria Belen Rabaglino**, National Scientific and Technical Research Council Argentina

Discuss this article

Comments (0)



This article is included in the **Collective Data Access** channel.



This article is included in the **Sidra Research** channel.



This article is included in the **Molecular Profiling of Pregnancy** channel.

Corresponding author: Alexandra K. Marr (amarr@sidra.org)

How to cite this article: Marr AK, Boughorbel S, Presnell S *et al.* **A curated transcriptome dataset collection to investigate the development and differentiation of the human placenta and its associated pathologies [version 2; referees: 2 approved]** *F1000Research* 2016, **5**:305 (doi: [10.12688/f1000research.8210.2](https://doi.org/10.12688/f1000research.8210.2))

Copyright: © 2016 Marr AK *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: All authors listed on this publication affiliated with the Sidra Medical and Research Center received support from the Qatar Foundation.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 09 Mar 2016, **5**:305 (doi: [10.12688/f1000research.8210.1](https://doi.org/10.12688/f1000research.8210.1))

REVISED Amendments from Version 1

The second version of the manuscript includes additional validation results presented in an additional figure.

See referee reports

Introduction

We aimed to make available via an interactive web-based application a collection of transcriptome datasets curated from the GEO public repository for their relevance to human placental development and pathology.

The placenta is a fetal organ indispensable for the establishment and maintenance of pregnancy. It connects the fetus to the maternal uterine wall via the umbilical cord, supplies nutrients and oxygen to the fetus, allows elimination of fetal waste, protects it from maternal infections and produces various hormones required for maintaining pregnancy¹. The placental trophoblasts play critical roles for all of these activities by acting as an interface between fetus and mother². In early embryogenesis, trophoblasts are the first cells differentiating from the fertilized egg, and eventually fusing with each other: a process transforming the monolayer cytotrophoblasts into syncytiotrophoblasts³. Several morphological and biochemical changes occur during this fusion process throughout pregnancy (known as trophoblast differentiation)⁴ to guarantee the development and appropriate functionality of the placenta¹. Failure in placental formation, differentiation and function, particularly by trophoblast dysfunction, impacts fetal development and is associated with a wide range of pathologies, including gestational diabetes, hypertension, pre-eclampsia and intrauterine growth restriction (IUGR) of the fetus⁵. In addition, exposure of placenta to toxic compounds by a mother's cigarette smoking alters the transcriptome of placental and fetal cells⁶. Indeed, maternal tobacco use causes various pregnancy-associated problems, including spontaneous fetal abortion, placental abruption, preterm birth, stillbirth, fetal growth restriction, and sudden infant death syndrome^{6,7}.

With over 65,000 data deposited into the **NCBI Gene Expression Omnibus** (GEO), a public repository for functional genomics data, it can be challenging to identify datasets relevant to a particular research area. Indeed, GEO was primarily designed for the purpose of data archiving rather than data browsing and downstream analysis. Thus, we identified datasets from GEO particularly relevant to our interest in the development and pathology of the human placenta and uploaded them to the custom web-based Gene Expression Browser application (GXB) (<http://placentalendocrinology.gxbsidra.org/dm3/landing.gsp>), which provides seamless access to the data. GXB allows browsing and interactive visualization of large volumes of heterogeneous data. It also provides access to rich contextual information essential for data interpretation, such as: detailed gene information, relevant literature, study design, and sample information. In addition, the user can customize data plots by adding multiple layers of parameters, modify the sample order and generate links that can be shared via e-mail or used in future publications.

Thus we provide a resource enabling browsing of datasets relevant to placental development and pathology that offers a unique opportunity to identify genes that play a role in placental development/differentiation and are commonly modulated in pregnancy-associated diseases.

Material and methods

111 datasets, potentially relevant to the development and pathology of the human placenta, were identified in GEO using the following search query: "Homo sapiens[Organism] AND placenta[DESC]". The majority of retrieved datasets were generated using Illumina or Affymetrix microarrays besides a few other commercial platforms.

The relevance of each one of the entries retrieved with our query pertaining to "development and pathology of the human placenta" was assessed on a case-by-case basis. This process included reading the NCBI's GEO-linked article and examining the list of samples profiled and their annotations as well as the study design. We selected for studies using primary placenta cells, placenta tissue or trophoblast cell lines comparing a) healthy with diseased samples or smoker vs. non-smoker or b) placenta cells in different differentiation/gestational stages. Finally we retrieved 24 curated datasets encompassing 759 transcriptome profiles, including 22 datasets generated using primary placenta cells and 2 datasets generated using trophoblast cell lines. Among the 24 datasets, there are 18 studies comparing control vs. diseased/smoker placenta cells, and 6 investigating transcriptomic changes during placental development and/or trophoblast differentiation. Among the many noteworthy datasets, several stood out, such as an extensive study comparing the diseased placentas with pre-eclampsia or unexplained IUGR vs. normal placentas (GSE24129)⁸. The datasets that comprise our collection are listed in **Table 1**.

Once a final selection had been made, each dataset was downloaded from GEO in the SOFT file format. These files were then uploaded to the "placentaendocrinology" instance of GXB (<http://placentalendocrinology.gxbsidra.org/dm3>), an interactive web-based application developed at the Benaroya Research Institute, hosted on the Amazon Web Services cloud⁹. Information about samples and study design was also uploaded. Samples were then grouped based on relevant study variables and genes were ranked based on specified group comparisons. Details of the GXB software are described in a recent publication⁹. A tutorial for this software is also available online (<https://gxb.benaroyaresearch.org/dm3/tutorials.gsp#gxbtut>). GXB provides the user with a means to navigate and filter the dataset collection available at (<http://placentalendocrinology.gxbsidra.org/dm3>). Briefly, the datasets of interest can be quickly identified either by filtering with the pre-existing criteria listed in the column located in the left side of the dataset navigation window or by entering a query term in the search box located at the top of the window. Clicking one of the studies listed in the dataset navigation window opens a viewer, which is designed to support interactive browsing and graphic representations of the data in an interpretable format. This interface was developed to navigate ranked gene lists and to display expression results in a figure with a context-rich environment. Selecting a gene from the rank-ordered list located on the left side of the navigation window displays its expression values

Table 1. List of all datasets included in our curated collection, also available at <http://placentalendocrinology.gxbsidra.org/dm3>.

For more information, see <http://placentalendocrinology.gxbsidra.org/dm3/geneBrowser/list>: PLAC1, placenta specific 1: CSH1, placental lactogen: XIST, X-inactive specific transcript NP: not published, no PubMed publication for this data set.

Title	Platform	Number of Samples	validation	GEO ID	Ref.
Altered Gene Expression Profile of Microvascular Endothelium in Placentas from IUGR/Preeclamptic Pregnancies	Affymetrix Human Genome U133 Plus 2.0 Array	10	PLAC1, CSH1, XIST	GSE25861	12
Chorionic villus sampling (CVS) microarray in preeclampsia	Affymetrix Human Genome U133 Plus 2.0 Array	12	PLAC1, CSH1	GSE12767	13
Comprehensive Study of Tobacco Smoke-Related Transcriptome Alterations in Maternal and Fetal Cells	Illumina HumanRef-8 v3.0 expression beadchip	183	PLAC1, CSH1	GSE27272	6
Culturing Cytotrophoblasts Reverses Gene Dysregulation in Preeclampsia Revealing Possible Causes	Affymetrix Human Genome U133 Plus 2.0 Array	39	PLAC1, CSH1	GSE40182	14
Deregulation of Gene Expression induced by Environmental Tobacco Smoke Exposure in Pregnancy	Illumina HumanRef-8 v3.0 expression beadchip	104	PLAC1, CSH1	GSE30032	15
Differential gene expression in Trophoblast cell cultures	Affymetrix Human Genome U133 Plus 2.0 Array	2	PLAC1, CSH1	GSE4100	NP
Differentially expressed microRNAs revealed by molecular signatures of Preeclampsia and IUGR in human placenta	Illumina human-6 v2.0 expression beadchip	94	PLAC1, CSH1, XIST	GSE35574	16
Dysregulation of the circulating and tissue-based renin-angiotensin system in preeclampsia	Affymetrix Human Genome U133 Plus 2.0 Array	6	PLAC1, CSH1	GSE6573	17
Full-term placenta, smokers and non-smokers	Affymetrix Human Genome U133 Plus 2.0 Array	10	PLAC1, CSH1, XIST	GSE7434	18
Gene expression profiling for placentas from pre-eclamptic, unexplained FGR and normal pregnancies.	Affymetrix Human Gene 1.0 ST Array	24	PLAC1, CSH1	GSE24129	8
Gene expression profiling indicates inflammatory pathways involved in IUGR due to placental insufficiency	ABI Human Genome Survey Microarray Version 2	16	PLAC1	GSE12216	19
Gene expression profiling of trophoblast cells	Affymetrix Human Genome U133A Array	11	PLAC1, CSH1	GSE9773	19
Genome wide analysis of placental malaria	Affymetrix Human Genome U133 Plus 2.0 Array	20	PLAC1, CSH1, XIST	GSE7586	20
Global placental gene expression profiling in the first and third trimesters of normal human pregnancy	ABI Human Genome Survey Microarray Version 2	37	PLAC1	GSE28551	4
Hypoxia induced HIF-1/HIF-2 activity alters trophoblast transcriptional regulation and promotes invasion	Illumina HumanHT-12 V4.0 expression beadchip	18	PLAC1, CSH1	GSE65271	21
Increased placental expression and maternal serum levels of apoptosis-inducing TRAIL in recurrent miscarriage	Affymetrix Human Genome U133 Plus 2.0 Array	10	PLAC1, CSH1	GSE22490	21
Mid-gestational gene expression profile in placenta and link to pregnancy complications	Affymetrix Human Genome U133 Plus 2.0 Array	4	PLAC1, CSH1	GSE37901	22
Placental gene expression in severe preeclampsia.	ABI Human Genome Survey Microarray Version 2	43	PLAC1	GSE10588	22
Profiling Gene Expression in Human Placentae of Different Gestational Ages: an OPRU Network and UW SCOR Study	Affymetrix Human Genome U133 Plus 2.0 Array	12	PLAC1, CSH1	GSE9984	23

Title	Platform	Number of Samples	validation	GEO ID	Ref.
Transcriptomic profiling of human placental trophoblasts in response to <i>Enterococcus faecalis</i> invasion	Illumina HumanHT-12 V4.0 expression beadchip	4	PLAC1, CSH1, XIST	GSE75626	23
Genome-wide analysis of gene expression in placentas derived from patients with preeclampsia	Illumina HumanHT-12 V4.0 expression beadchip	12	PLAC1, CSH1	GSE30186	24
Genomic expression profiles of blood and placenta in Chinese women with gestational diabetes	Aalborg University Illumina human-6 v2.0 expression beadchip	5	PLAC1, CSH1	GSE19649	NP
Severe Preeclampsia-Related Changes in Gene Expression at the Maternal-Fetal Interface Include Siglec-6 and Pappalysin-2	Affymetrix Human Genome U133A Array	23	PLAC1, CSH1	GSE14722	25
Transcriptional Profiling of Human Placentas from Pregnancies Complicated by Preeclampsia Reveals Disregulation of Sialic Acid Acetyltransferase and Immune Signaling Pathways	Illumina human-6 v2.0 expression beadchip	60	PLAC1, CSH1	GSE25906	26

on the interactive plot. The drop-down menus located directly above the graphical display provide the user the following functions: a) Exporting the created graph as a portable network graphics (png) image or a csv file for performing a separate analysis. b) Changing ranking of genes in the list; this function allows the user to manipulate the ways of ranking the genes depending on his/her interest, or to include only the genes selected based on specific biological interest. c) Changing grouping of the samples (by using “Group Set” button); for example, the user can convert the groups created based on “cell type” to those on “disease type”. d) Identifying individual samples within a group using categorical or continuous variables (e.g., mode of delivery, ethnic group and age). e) Toggling between the bar chart view and a box plot view, with demonstration of the values as a single plot for each sample. Samples are split into different categories or groups whatever they are displayed in a bar chart or box plot. f) Providing a color legend for sample groups. g) Providing a color legend for the categorical information overlaid at the bottom of the graph. h) Selection of categorical information that is to be overlaid at the bottom of the graph. Using this function, the user can display, for example, gender or smoking status, at the bottom of the figure. The data without contextual information have no intrinsic utility. It is therefore important to capture and display the context information in the graph, so that the viewer can interpret the data shown in the graph. GXB provides functions for organizing the context information in the tabs located just above the graphical display. The tabs can be hidden to make more room for graphs, or be reappeared by clicking the blue “Show Info Panel” button located on the top right corner of the graphical display. Information on the genes selected from the list located in the left side of the graphical display is shown under the “Gene” tab. Information on the study for the demonstrated dataset is available under the “Study” tab. Detailed information of individual samples is provided under the “Sample” tab. Rolling the mouse cursor over a sample in either the bar chart or box plot view while displaying the “Sample” tab displays any clinical, demographic, or laboratory information available for that sample. Finally, the “Downloads” tab allows the advanced users to retrieve the original data for further analysis using other softwares/tools. It also provides all available sample annotation data along with the expression data. Other functionalities are provided

under the “Tools” drop-down menu located in the top right corner of the user interface. Some of the notable functionalities available through this menu include: a) Annotations, which provides access to all the ancillary information about the study, samples and dataset organized across different tabs; b) Cross project view, which provides the ability to browse through all available studies while focusing on a the currently selected gene; c) Copy link, which generates a mini-URL encapsulating information about the display settings in use and that can be saved and shared with others (clicking on the envelope icon on the toolbar inserts the url in an email message via the local email client); d) Chart options, which gives user the option to customize chart labels.

Dataset validation

We performed quality control checks on the datasets loaded on GXB by validating the expression of placental marker genes. Specifically, we used placenta specific 1 (PLAC1)¹⁰ and human placental lactogen (CSH1)¹¹ (Table 1). As expected, all samples from the datasets we incorporated in our collection expressed these two placental markers (hyperlinked for easy access in Table 1), except some samples in GSE12216, GSE28551 and GSE10588, where CSH1 was not measured.

Gender specific expression of the XIST transcript (in the samples for which sex information was available) was also examined to determine its concordance with demographic information provided with the GEO datasets (Table 1). An overall accordance of a high XIST expression in female samples compared to a low XIST expression in male samples was determined in our entire dataset collection (Table 1). Based on our experience, concordance should be close to 100%. Levels of concordance closer to 50%, which were not observed here, would indicate a potential error with handling of samples during processing that may dramatically affect data analysis and interpretation (e.g. plate inversion).

We also verified and compared the differential expression using the Fold Change values obtained from GXB or the original manuscript. For this verification process, two datasets were selected GSE24129 and GSE30032. As shown in Figure 1, Fold Change

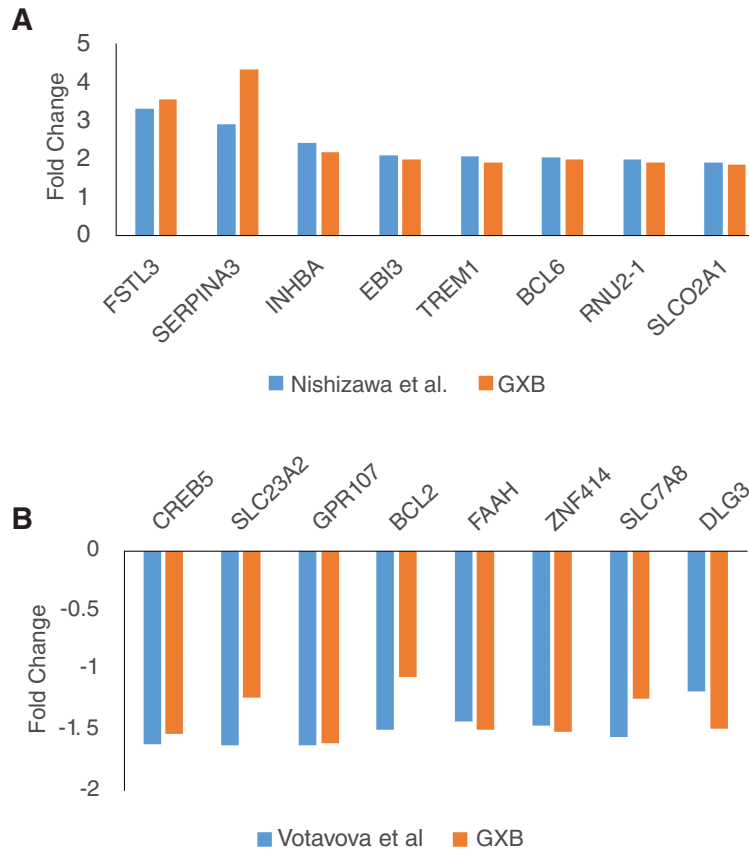


Figure 1. Verification of differential gene expression (Fold Changes) in GXB compared to the values published in the originating manuscript. Two datasets were selected for validation of differential expression: GSE24129 (**A**) and GSE 30032 (**B**). The first 8 genes were selected from Table 2 in Nishizawa *et al.* (for GSE24129) and from supplemental material Table in Votavova *et al.* (for GSE30032).

values obtained from GXB correlate with those obtained from the original published article.

Data availability

All datasets included in our curated collection are available publicly at the NCBI GEO website: <http://www.ncbi.nlm.nih.gov/gds/>. They were cited in our manuscript along with their GEO accession numbers (e.g. GSE24129). Signal files and sample description files for each uploaded GEO dataset can also be downloaded from the GXB tool under the “downloads” tab after accessing our dataset collection.

Author contributions

AM and TK conceived the theme for this dataset collection. AM, SB, SP and CQ contributed to the query, selection, loading and curation of datasets. DC and AM prepared the first draft of the manuscript.

TK edited the manuscript. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

Grant information

All authors listed on this publication affiliated with the Sidra Medical and Research Center received support from the Qatar Foundation.

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We would like to thank all the investigators who decided to make their datasets publicly available by depositing them into GEO.

References

1. Burton GJ, Fowden AL: **The placenta: a multifaceted, transient organ.** *Philos Trans R Soc Lond B Biol Sci.* 2015; **370**(1663): 20140066.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Silva JF, Serakides R: **Intrauterine trophoblast migration: A comparative view of humans and rodents.** *Cell Adh Migr.* 2016; 1–23.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Ji L, Brkic J, Liu M, *et al.*: **Placental trophoblast cell differentiation: physiological regulation and pathological relevance to preeclampsia.** *Mol Aspects Med.* 2013; **34**(5): 981–1023.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Sitras V, Fenton C, Paulssen R, *et al.*: **Differences in gene expression between first and third trimester human placenta: a microarray study.** *PLoS One.* 2012; **7**(3): e33294.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Zhang S, Regnault TR, Barker PL, *et al.*: **Placental adaptations in growth restriction.** *Nutrients.* 2015; **7**(1): 360–89.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Votavova H, Dostalova Merkerova M, Fejglova K, *et al.*: **Transcriptome alterations in maternal and fetal cells induced by tobacco smoke.** *Placenta.* 2011; **32**(10): 763–70.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Bruin JE, Gerstein HC, Holloway AC: **Long-term consequences of fetal and neonatal nicotine exposure: a critical review.** *Toxicol Sci.* 2010; **116**(2): 364–74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Nishizawa H, Ota S, Suzuki M, *et al.*: **Comparative gene expression profiling of placentas from patients with severe pre-eclampsia and unexplained fetal growth restriction.** *Reprod Biol Endocrinol.* 2011; **9**: 107.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Speake C, Presnell S, Domico K, *et al.*: **An interactive web application for the dissemination of human systems immunology data.** *J Transl Med.* 2015; **13**: 196.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Cocchia M, Huber R, Pantano S, *et al.*: **PLAC1, an Xq26 gene with placenta-specific expression.** *Genomics.* 2000; **68**(3): 305–12.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Samaan N, Yen SC, Friesen H, *et al.*: **Serum placental lactogen levels during pregnancy and in trophoblastic disease.** *J Clin Endocrinol Metab.* 1966; **26**(12): 1303–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Dunk CE, Roggensack AM, Cox B, *et al.*: **A distinct microvascular endothelial gene expression profile in severe IUGR placentas.** *Placenta.* 2012; **33**(4): 285–93.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Founds SA, Conley YP, Lyons-Weiler JF, *et al.*: **Altered global gene expression in first trimester placentas of women destined to develop preeclampsia.** *Placenta.* 2009; **30**(1): 15–24.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Zhou Y, Gormley MJ, Hunkapiller NM, *et al.*: **Reversal of gene dysregulation in cultured cytotrophoblasts reveals possible causes of preeclampsia.** *J Clin Invest.* 2013; **123**(7): 2862–72.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Votavova H, Dostalova Merkerova M, Krejci Z, *et al.*: **Deregulation of gene expression induced by environmental tobacco smoke exposure in pregnancy.** *Nicotine Tob Res.* 2012; **14**(9): 1073–82.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Guo L, Tsai SQ, Hardison NE, *et al.*: **Differentially expressed microRNAs and affected biological pathways revealed by modulated modularity clustering (MMC) analysis of human preeclamptic and IUGR placentas.** *Placenta.* 2013; **34**(7): 599–605.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Herse F, Dechend R, Harsem NK, *et al.*: **Dysregulation of the circulating and tissue-based renin-angiotensin system in preeclampsia.** *Hypertension.* 2007; **49**(3): 604–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Huuskonen P, Storvik M, Reinisalo M, *et al.*: **Microarray analysis of the global alterations in the gene expression in the placentas from cigarette-smoking mothers.** *Clin Pharmacol Ther.* 2008; **83**(4): 542–50.
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Sitras V, Paulssen R, Leirvik J, *et al.*: **Placental gene expression profile in intrauterine growth restriction due to placental insufficiency.** *Reprod Sci.* 2009; **16**(7): 701–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Muehlenbachs A, Fried M, Lachowitz J, *et al.*: **Genome-wide expression analysis of placental malaria reveals features of lymphoid neogenesis during chronic infection.** *J Immunol.* 2007; **179**(1): 557–65.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Highet AR, Khoda SM, Buckberry S, *et al.*: **Hypoxia induced HIF-1/HIF-2 activity alters trophoblast transcriptional regulation and promotes invasion.** *Eur J Cell Biol.* 2015; **94**(12): 589–602.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Uusküla L, Männik J, Rull K, *et al.*: **Mid-gestational gene expression profile in placenta and link to pregnancy complications.** *PLoS One.* 2012; **7**(11): e49248.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Mikheev AM, Nabekura T, Kaddoumi A, *et al.*: **Profiling gene expression in human placentae of different gestational ages: an OPRU Network and UW SCOR Study.** *Reprod Sci.* 2008; **15**(9): 866–77.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Meng T, Chen H, Sun M, *et al.*: **Identification of differential gene expression profiles in placentas from preeclamptic pregnancies versus normal pregnancies by DNA microarrays.** *OMICS.* 2012; **16**(6): 301–11.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Winn VD, Gormley M, Paquet AC, *et al.*: **Severe preeclampsia-related changes in gene expression at the maternal-fetal interface include sialic acid-binding immunoglobulin-like lectin-6 and pappalysin-2.** *Endocrinology.* 2009; **150**(1): 452–62.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Tsai S, Hardison NE, James AH, *et al.*: **Transcriptional profiling of human placentas from pregnancies complicated by preeclampsia reveals dysregulation of sialic acid acetyltransferase and immune signalling pathways.** *Placenta.* 2011; **32**(2): 175–82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Referee Status:



Version 2

Referee Report 07 June 2016

doi:[10.5256/f1000research.9407.r14210](https://doi.org/10.5256/f1000research.9407.r14210)



Brian Cox

Department of Physiology, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

I find the authors responses thoughtful and encouraging that this may become a stable resource that will expand.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 06 April 2016

doi:[10.5256/f1000research.8830.r12946](https://doi.org/10.5256/f1000research.8830.r12946)



Maria Belen Rabaglino

Center of Excellence in Products and Processes of Cordoba, National Scientific and Technical Research Council, Cordoba, Argentina

The authors of this article have collected several gene expression datasets, related with the development of the human placenta and associated pathologies, from the public database Gene Expression Omnibus (GEO) to analyze them in the context of a web application known as Gene Expression Browser (GXB) platform.

This web tool provides a graphical interface to search and visualize gene status across different datasets. The collection of datasets presented on this article makes this platform particularly useful to gather information about transcriptome profiles related to human placenta and associated pathologies. Undoubtedly, this tool allows fast searching of a gene of interest on those datasets, which could facilitate visualization of results from the corresponding studies and helps in manuscripts or grants writing.

There is no information on this manuscript about the statistical analysis followed to rank the genes on each dataset but it refers to the publication that details the GXB software¹. Here, it is specified that genes are ranked according their differential expression using the limma package, and the links to the R scripts

are showed. However, the application should include the information about the normalization/background correction methods used on each dataset. Different methods could yield different expression measures² and so modify the gene ranking.

Also, comparisons of gene expression across studies could be a powerful feature of this program, but for that is necessary to apply a method to remove batch effects.

Another suggestion is to extend the verification of the datasets including genes that are known to be deregulated in placental pathologies --such as LEP up-regulation in preeclamptic placenta^{3,4}--. An explanation about these genes along with graphs obtained from the web application could be useful for the reader as an example of its utility.

Finally, the application should include the date of the last update. Obviously, new placenta-related datasets might be uploaded into the GEO database and so they should be added to the collection.

References

1. Speake C, Presnell S, Domico K, Zeitner B, Bjork A, Anderson D, Mason MJ, Whalen E, Vargas O, Popov D, Rinchai D, Jourde-Chiche N, Chiche L, Quinn C, Chaussabel D: An interactive web application for the dissemination of human systems immunology data. *J Transl Med*. 2015; **13**: 196 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Irizarry RA, Wu Z, Jaffee HA: Comparison of Affymetrix GeneChip expression measures. *Bioinformatics* . 2006; **22** (7): 789-94 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Tejera E, Bernardes J, Rebelo I: Co-expression network analysis and genetic algorithms for gene prioritization in preeclampsia. *BMC Med Genomics*. 2013; **6**: 51 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Kaartokallio T, Cervera A, Kyllönen A, Laivuori K: Gene expression profiling of pre-eclamptic placentae by RNA sequencing. *Sci Rep*. 2015; **5**: 14107 [PubMed Abstract](#) | [Publisher Full Text](#)

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 23 March 2016

doi:[10.5256/f1000research.8830.r12824](https://doi.org/10.5256/f1000research.8830.r12824)



Brian Cox

Department of Physiology, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

The web based utility is well assembled and easy to navigate and have very nice layout and graphic capabilities. A large spectrum of RNA microarray human placental sample sets have been assembled.

Reservations are:

The verification of the datasets using CSH1 and PLAC1 is rather minimal. Can you verify the differential expression observed in the originating manuscripts? At least in a couple of examples, such as one smoking and one preeclampsia/IUGR. Figures with examples graphics from these validations could be helpful.

In many sample sets there exist batch effects. In some cases the depositors have identified the batches, but have not corrected for these, in other cases batches can be observed even when not formally identified by the depositor. For maximum utility these effects should be removed or a tool supplied to identify and remove them.

Further to the need for batch correction, viewing a gene across multiple uncorrected/non-normalized data sets is not very informative.

The web utility was frequently unavailable during the review process. A more stable platform or redundant server system should be established to limit the down time of this application. What user load can the server handle, how many concurrent users can the server handle?

Only transcription arrays are used in the assembled data sets, there are numerous data sets involving DNA methylation, microRNAs and protein. Incorporation of these data sets would create a true systems level database of the human placenta.

Improved utility would come from enabling merging of multiple datasets to perform meta analysis of aggregate analysis. This would require the installation of a batch corrections scheme or method/s.

There is also a growing resource of sequencing based data sets that have been neglected in this application/manuscript. Once converted into a table of gene expression/counts, it should not be too challenging to incorporate these datasets as well.

User training could be improved with additional figures to support descriptions of the utility and data sets. Better would be a vignette containing worked examples of the different types of analyses to generate the table and figures discussed in the paper. Similar worked vignette examples are frequently found in R libraries.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 02 May 2016

Alexandra Marr, Sidra Medical and Research Center, Qatar

The web based utility is well assembled and easy to navigate and have very nice layout and graphic capabilities. A large spectrum of RNA microarray human placental sample sets have been assembled.

Reservations are:

1. The verification of the datasets using CSH1 and PLAC1 is rather minimal. Can you verify the differential expression observed in the originating manuscripts? At least in a couple of examples, such as one smoking and one preeclampsia/IUGR. Figures with examples graphics from these validations could be helpful.

We included a new figure (new Figure 1) showing the verification of differential gene expression in

GXB compared to the values published in the originating manuscript. We also included the following paragraph in the “Dataset Validation” section of the manuscript: “We also verified and compared the differential expression using the Fold Change values obtained from GXB or the original manuscript. For this verification process, two datasets were selected GSE24129 and GSE30032. As shown in Figure 1, Fold Change values obtained from GXB correlate with those obtained from the original published article.”

2. In many sample sets there exist batch effects. In some cases the depositors have identified the batches, but have not corrected for these, in other cases batches can be observed even when not formally identified by the depositor. For maximum utility these effects should be removed or a tool supplied to identify and remove them.

Further to the need for batch correction, viewing a gene across multiple uncorrected/non-normalized data sets is not very informative.

The reviewer raises an important point. Batch effects resulting in non-biological experimental variation across multiple batches of microarray experiments have been well documented. Several methods are available to correct for batch effects, such as Combat in ‘R’ (<http://www.r-project.org/>). In the context of a dataset compendium batch effects should be considered at two different levels:

a) Within datasets: Unfortunately batch-processing information is available for only a small minority of datasets deposited in GEO (6% in the case of our “placental endocrinology” dataset collection), which makes systematically correcting for such effects impossible. This limitation of course affects any investigator inclined to use GEO as a resource. The practice of randomizing groups across runs, which avoids having batch effects confounding the analysis, is fortunately widely accepted. But it cannot be altogether discounted.

However, this is where one of the distinct advantage of working with a compendium of dataset comes into play: a conclusion regarding differential expression of a given gene can be based on observation of such differences in not just one but multiple datasets, obtained independently by different investigators, in different parts of the world, using different array platforms. The conclusions that can be thus derived are especially robust. This point is illustrated by one of our recent contribution to F1000Research: <http://f1000research.com/articles/4-89/v1>

b) Across datasets: We have previously implemented strategies for meta-analyses per-se, focusing on blood transcriptome data, and found it necessary to reduce the information from each dataset to p-values (comparison within each dataset of cases to a control group; 17724127, 16461797) or co-clustering information (module repertoire analyses 24662387). While GXB is primarily designed as a data browser it is possible to employ it similarly (but with added flexibility) for “meta-interpretation” (rather than meta-analysis). This consists in interpreting findings for a given gene across tens or hundreds of datasets, thus affording a unique perspective to the investigators and an effective means to identify and address knowledge gaps. This again is illustrated in the article mentioned above: <http://f1000research.com/articles/4-89/v1>. Notably, the cross-project view functionality also provides an independent search for datasets with genes that have a certain fold-changes above a chosen threshold.

3. The web utility was frequently unavailable during the review process. A more stable platform or redundant server system should be established to limit the down time of this application. What user load can the server handle, how many concurrent users can the server handle?

We thank the reviewer for this critical feedback. Our instance was created on the Amazon Elastic Compute Cloud (Amazon Web services). The downtime was not due to the GXB application per se but rather to some missing packages related to the tomcat server. These packages have been fixed and the downtime issue is now resolved. The number of concurrent users is configurable and it is related to tomcat web server. The current configuration is 200 concurrent users. This can be reconfigured as needed.

4. Only transcription arrays are used in the assembled data sets, there are numerous data sets involving DNA methylation, microRNAs and protein. Incorporation of these data sets would create a true systems level database of the human placenta.

These are great ideas. We are currently working on the interactive web application (GXB) to be able to expand to other platforms such as working on SomaLogic data etc. We do see the establishment of our instance as an ongoing project, where additional datasets can be uploaded on demand (see "Request to load a Sample Set" function on the upper right corner of the web application page <<http://placentalendocrinology.gxbsidra.org/dm3/geneBrowser/list>>), for example when new datasets are published and deposited, and according to advances in our GXB technology.

5. Improved utility would come from enabling merging of multiple datasets to perform meta analysis of aggregate analysis. This would require the installation of a batch corrections scheme or method/s.

We agree with the reviewer's point of view. However, it is a difficult problem to address at the gene level. We are currently working on approaches and tools involving those modular repertoires (please see Chaussabel D. *et al.* 2008 *Immunology* 29:150-64 and Chaussabel D. & Baldwin N. 2014 *Nat Rev Immunol.* 14:271-80). See also our comments above pertaining to the point raised earlier regarding batch correction.

6. There is also a growing resource of sequencing based data sets that have been neglected in this application/manuscript. Once converted into a table of gene expression/counts, it should not be too challenging to incorporate these datasets as well.

We agree that it would be very valuable to include RNAseq datasets in the future. We are not at this time in measure to accommodate this data type routinely but as a proof of principle a trial RNA-seq dataset have been setup to display RNA-seq profiles of immune cell subsets isolated from subjects with a wide range of diseases (GSE60424, <https://gxb.benaroyaresearch.org/dm3/geneBrowser/show/396>, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0109760>).

7. User training could be improved with additional figures to support descriptions of the utility and data sets. Better would be a vignette containing worked examples of the different types of analyses to generate the table and figures discussed in the paper. Similar worked vignette examples are frequently found in R libraries.

Indeed, thank you for suggesting this. We started working on a 'User Training' document with

detailed, step-by-step descriptions. This document will be published in the Collective Data Access channel of F1000Research very soon.

Competing Interests: No competing interests were disclosed.
