



DATA NOTE

A curated transcriptome dataset collection to investigate the functional programming of human hematopoietic cells in early life [version 1; referees: 2 approved]

Mahbuba Rahman¹, Sabri Boughorbel¹, Scott Presnell², Charlie Quinn², Chiara Cugno¹, Damien Chaussabel¹, Nico Marr¹

¹Sidra Medical and Research Center, Doha, Qatar

²Benaroya Research Institute, Seattle, WA, USA

v1 First published: 30 Mar 2016, 5:414 (doi: [10.12688/f1000research.8375.1](https://doi.org/10.12688/f1000research.8375.1))
 Latest published: 30 Mar 2016, 5:414 (doi: [10.12688/f1000research.8375.1](https://doi.org/10.12688/f1000research.8375.1))

Abstract

Compendia of large-scale datasets made available in public repositories provide an opportunity to identify and fill gaps in biomedical knowledge. But first, these data need to be made readily accessible to research investigators for interpretation. Here we make available a collection of transcriptome datasets to investigate the functional programming of human hematopoietic cells in early life. Thirty two datasets were retrieved from the NCBI Gene Expression Omnibus (GEO) and loaded in a custom web application called the Gene Expression Browser (GXB), which was designed for interactive query and visualization of integrated large-scale data. Quality control checks were performed. Multiple sample groupings and gene rank lists were created allowing users to reveal age-related differences in transcriptome profiles, changes in the gene expression of neonatal hematopoietic cells to a variety of immune stimulators and modulators, as well as during cell differentiation. Available demographic, clinical, and cell phenotypic information can be overlaid with the gene expression data and used to sort samples. Web links to customized graphical views can be generated and subsequently inserted in manuscripts to report novel findings. GXB also enables browsing of a single gene across projects, thereby providing new perspectives on age- and developmental stage-specific expression of a given gene across the human hematopoietic system. This dataset collection is available at:

<http://developmentalimmunology.gxbsidra.org/dm3/geneBrowser/list>.

Open Peer Review

Referee Status:

	Invited Referees	
	1	2
version 1 published 30 Mar 2016	 report	 report

1 Stanislas Goriely, Université Libre de Bruxelles Belgium

2 Peter Ghazal, University of Edinburgh UK

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: Nico Marr (nmarr@sidra.org)

How to cite this article: Rahman M, Boughorbel S, Presnell S *et al.* **A curated transcriptome dataset collection to investigate the functional programming of human hematopoietic cells in early life [version 1; referees: 2 approved]** *F1000Research* 2016, 5:414 (doi: [10.12688/f1000research.8375.1](https://doi.org/10.12688/f1000research.8375.1))

Copyright: © 2016 Rahman M *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: The authors listed on this publication with the exception of CQ and SP 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: 30 Mar 2016, 5:414 (doi: [10.12688/f1000research.8375.1](https://doi.org/10.12688/f1000research.8375.1))

Introduction

Human immune defenses are highly dynamic and vary with age, reflecting the different environmental challenges and needs for adaptation during the fetal, neonatal and postnatal period, and throughout life. Not surprisingly, functional differences of the human immune system are most profound very early in life due to the limited antigen exposure *in utero*, and a variety of developmental, maternal, nutritional, and environmental factors that can act in concert to modulate innate and adaptive effector functions of hematopoietic cells¹⁻³. At the same time, newborns and young infants are particularly vulnerable to infection, with each developmental stage representing a ‘window of vulnerability’ to a very specific subset of pathogenic microbes⁴. In this context, an increasing number of studies have been designed to gain a deeper understanding of immunity in early life, and ultimately, to reveal the underlying immune defense and regulatory mechanisms that determine the clinical outcome of primary infections, and responses to early childhood vaccination¹⁻³. Nonetheless, our understanding of the developing immune system in early life remains very limited, in part because of the difficulty to access biological specimen from human fetuses, neonates, and young children. Most often, *in vitro* studies utilizing umbilical cord and peripheral blood samples were used to assess neonatal immune defenses, and in particular to reveal critical differences in the functional programming of neonatal hematopoietic cells in comparison to that of adults. Aside from the limited repertoire of memory B and T lymphocytes in neonates, such studies have revealed substantial gestational- and postnatal age-dependent differences in the phenotype and function of a variety of hematopoietic cell types upon *in vitro* stimulation of (whole) cord/peripheral blood and isolated blood mononuclear cells with a variety of immune stimulators and modulators, including purified Toll-like receptor (TLR) and RIG-I-like receptor (RLR) agonists, cytokines, and whole pathogens, which engage a variety of pattern recognition receptors (PRRs) and signaling pathways⁵⁻¹⁴. The underlying reasons for the functional differences between hematopoietic cells obtained at different gestational and postnatal ages remain largely unclear. There is little evidence for postnatal age-specific variation in the PRR gene expression at baseline (i.e. in the absence of infection or *in vitro* stimulation)^{8,15,16}, suggesting critical differences in downstream signaling networks and regulatory mechanisms by which hematopoietic cells exert their specific effector functions. These age-specific differences have yet to be revealed.

Here we make available, via an interactive web application, a curated collection of transcriptome datasets of either whole blood samples, isolated blood mononuclear cells, or a variety of sort-purified hematopoietic cell populations obtained from human neonates or fetal tissue. In the selected datasets, transcriptional profiles were obtained in the absence or presence of various intrinsic and exogenous immune modulators. For comparison, these datasets contain samples from other age groups (most often from healthy adult volunteers), or cell populations at multiple differentiation stages. The ability to pool and analyze samples across various age and risk groups, and across various hematopoietic cell types, offers a unique opportunity to define common denominators of early life immunity and to reveal critical differences in the functional programming of fetal and neonatal hematopoietic cells.

To this date, over 65,000 high-throughput functional genomics studies have been deposited in the [NCBI Gene Expression Omnibus \(GEO\)](#), a public repository of transcriptome profiles. However, identifying datasets relevant to a particular research area is not straightforward, because GEO is primarily designed as a repository for the storage of data, rather than browsing and interaction with the deposited data. Thus, we used a custom interactive web application, called the Gene Expression Browser (GXB)¹⁷, to host the datasets we identified as particularly relevant to reveal gestational and postnatal age-specific differences in the gene expression pattern of fetal and neonatal hematopoietic cells. GXB allows seamless browsing and interactive visualization of our GEO dataset collection containing large volumes of heterogeneous data, such as transcriptome profiles, demographic information, as well as clinical information. Users can easily customize data plots by adding multiple layers of information (such as postnatal age, weeks of gestation at birth, and gender), modify the ordering of samples and genes, change the plot type, and generate links (mini URLs) capturing the user’s settings, which can then be inserted in email communications or in publications. These user-generated mini URLs provide access not only the transcription data but also to rich contextual information and data interpretation, including gene information, relevant literature, a description of the study design, as well as detailed sample information that was supplied along with the transcriptome data submission to GEO.

Material and methods

Potentially relevant datasets deposited in GEO were identified using two search queries which were designed to retrieve entries where the title and description of the datasets contained the words newborn OR neonate OR neonatal OR fetal OR cord. The search was restricted to datasets that were generated from human whole blood, human blood mononuclear cells, or sort-purified human hematopoietic cells using Illumina or Affymetrix platforms. Studies on cancer patients or cell lines were excluded. First, the following query was used: Homo sapiens[Organism] AND (newborn[DESC] OR neonate[DESC] OR neonatal[DESC] OR fetal[DESC] OR cord[DESC]) AND (blood[DESC] OR PBMC[DESC] OR PBMCs[DESC] OR lymphocyte[DESC] OR lymphocytes[DESC] OR “B cell”[DESC] OR “B cells”[DESC] OR “plasma cells”[DESC] OR “T cell”[DESC] OR “T cells”[DESC] OR Treg[DESC] OR Tregs[DESC] OR monocyte[DESC] OR monocytes[DESC] OR dendritic[DESC] OR DC[DESC] OR DCs[DESC] OR “natural killer”[DESC] OR NK[DESC] OR NKT[DESC] OR neutrophil[DESC] OR neutrophils[DESC] OR erythroblast[DESC] OR erythroid[DESC] OR CD19[DESC] OR CD20[DESC] OR CD3[DESC] OR CD4[DESC] OR CD8[DESC] OR CD71[DESC]) AND (“Expression profiling by array”[gdsType] OR “Expression profiling by high throughput sequencing”[gdsType]) NOT (cancer[DESC] OR leukemia[DESC] OR lymphoma[DESC] OR “cell line”[DESC] OR myeloma[DESC] OR mesenchymal[DESC] OR endothelial[DESC]). In addition, we used the following query to specifically retrieve datasets containing samples from neonatal sepsis patients: sepsis AND (neonate OR newborn). In total, more than 450 datasets were retrieved by the two queries. The list of datasets retrieved from the 2 queries was manually curated and restricted to datasets that: (i) contained transcriptional profiles from primary hematopoietic

cells; (ii) contained samples of fetal or neonatal origin; (iii) contained a minimum of 3 samples (i.e. biological repeats) for each of the major variables assessed in the respective study; and (iv) allowed within the same dataset, the comparison of transcriptional profiles either between different age groups (e.g. neonate versus adult), between infants born at different gestational ages, between different risk groups (e.g. infants with low birth weight versus those with normal birth weight), or between cell differentiation stages. This process involved reading through the descriptions and examining the list of available samples and their annotations. For the filtering of the dataset list, the Bioconductor package GEOmetadb, version 1.30.0, and its SQLite database was used to capture detailed information on selected GEO datasets in a single table (<https://www.bioconductor.org/packages/release/bioc/html/GEOmetadb.html>)¹⁸. Sometimes, it also required accessing the original published report in which the design of the study and generation of the dataset is described in more detail. Using the stringent criteria detailed above, we reduced the list down to 41 GEO datasets (excluding SuperSeries), of which 32 GEO datasets were uploaded into our interactive web application, GXB, together with corresponding SuperSeries if available (4 additional GEO datasets). For the remaining datasets the platform used to generate the transcriptome profiles was not supported by GXB (9 datasets). Out of the 32 curated datasets, 8 include samples of fetal origin, and 25 datasets include samples of neonatal origin, usually in conjunction with samples of adult subjects (including 3 datasets containing peripheral blood samples from the mothers). The majority of neonatal samples were obtained from healthy subjects, mostly utilizing umbilical cord blood. In these studies, a variety of factors were assessed that may induce and/or reveal differences in the functional programming of neonatal hematopoietic cells, including the effect of active/passive smoking of the mothers during pregnancy (GSE27272, GSE30032)^{19,20}, standards of living and hygiene (GSE53471, GSE53472, GSE53473)²¹, as well as *in vitro* exposure of neonatal and adult cells to purified TLR ligands (GSE67057, GSE3140), and to whole pathogens (GSE24132). In 6 studies, peripheral blood samples were obtained from babies with neonatal sepsis (GSE25504, GSE26440, GSE26378, GSE69686)²²⁻²⁴ bronchopulmonary dysplasia (GSE32472)²⁵, or from babies with low birth weight (GSE29807). The transcriptional profiles were either generated from whole blood (11 datasets), cord and peripheral blood mononuclear cells (1 dataset), or a variety of sort-purified hematopoietic cell populations at different differentiation stages, including cells derived from neonatal and adult hematopoietic stem cells as well as erythroid cells. The latter cells have recently been shown to play an important immunosuppressive role in the context of neonatal infection²⁶. The datasets that make up our collection are listed in Table 1. We also generated a word cloud from the title of published journal articles where the datasets were first reported (or the dataset title if no journal article was available), which provides information on the type of studies that make up our dataset collection (Figure 1).

Once a final selection has been made, each dataset was downloaded from GEO using the SOFT file format. For GEO datasets generated using multiple platforms (GSE1460, GSE25504), the series matrix file format was used instead, and separate datasets for each platform were downloaded. The retrieved datasets were in turn uploaded on an instance of GXB hosted on the Amazon Web Services cloud

(39 datasets in total, including 4 SuperSeries and 3 additional datasets that were uploaded due to the use of multiple platforms per GEO dataset). The GXB software has been described in detail in a recent publication¹⁷. This custom software interface provides the user with the means to easily navigate and filter the dataset collection, and is available at <http://developmentalimmunology.gxbsidra.org/dm3/geneBrowser/list>. A web tutorial is also available online: <https://gxb.benaroyaresearch.org/dm3/tutorials.gsp#gxbtut>. Annotation and functionality of the web software interface were described previously by our group^{27,28}, and is reproduced here so that readers can use this article as a standalone resource. Available sample and study information were uploaded along with the gene expression data. Samples of each dataset were grouped according to study design and gene rankings were computed for the different group comparisons. Datasets of interest can be quickly identified either by filtering on criteria from pre-defined sections on the left or by entering a query term in the search box at the top of the dataset navigation page. Clicking on one of the studies listed in the dataset navigation page opens a viewer designed to provide interactive browsing and graphic representations of large-scale data in an interpretable format. This interface is designed to present ranked gene lists and display expression results graphically in a context-rich environment. Selecting a gene from the rank ordered list on the left of the data-viewing interface will display its expression values graphically in the screen's central panel. Directly above the graphical display drop down menus give users the ability: a) To change how the gene list is ranked - this allows the user to change the method used to rank the genes, or to only include genes that are selected for specific biological interest; b) To change sample grouping (Group Set button) - in some datasets, a user can switch between groups based on cell type to groups based on disease type, for example; c) To sort individual samples within a group based on associated categorical or continuous variables (e.g. gender or age); d) To toggle between the bar chart view and a box plot view, with expression values represented as a single point for each sample. Samples are split into the same groups whether displayed as a bar chart or box plot; e) To provide a color legend for the sample groups; f) To select categorical information that is to be overlaid at the bottom of the graph - for example, the user can display gender or smoking status in this manner; g) To provide a color legend for the categorical information overlaid at the bottom of the graph; h) To download the graph as a portable network graphics (png) image. Measurements have no intrinsic utility in absence of contextual information. It is this contextual information that makes the results of a study or experiment interpretable. It is therefore important to capture, integrate and display information that will give users the ability to interpret data and gain new insights from it. We have organized this information under different tabs directly above the graphical display. The tabs can be hidden to make more room for displaying the data plots, or revealed by clicking on the blue "show info panel" button on the top right corner of the display. Information about the gene selected from the list on the left side of the display is available under the "Gene" tab. Information about the study is available under the "Study" tab. Rolling the mouse cursor over a bar chart feature while displaying the "Sample" tab lists any clinical, demographic, or laboratory information available for the selected sample. Finally, the "Downloads" tab allows advanced users to retrieve the original dataset for analysis outside this tool. It also provides all available sample annotation

Table 1. List of curated datasets. WB, whole blood; HSC, hematopoietic stem cells; DC, dendritic cells; QC, quality control; NA, not applicable (<http://developmentalimmunology.gxbsidra.org/dm3/geneBrowser/list>).

Title	Platform	Sample source		QC Markers	Number of Samples	GEO ID	Ref.
		Origin	RNA				
Whole blood mRNA expression profiling of host molecular networks in neonatal sepsis (platforms GPL6947)	Illumina	neonatal	WB	XIST	63	GSE25504	22
Whole blood mRNA expression profiling of host molecular networks in neonatal sepsis (platform GPL13667)	Affymetrix	neonatal	WB	XIST	20	GSE25504	22
Whole blood mRNA expression profiling of host molecular networks in neonatal sepsis (platform GPL570)	Affymetrix	neonatal	WB	XIST	5	GSE25504	22
Expression data for derivation of septic shock subgroups	Affymetrix	neonatal, pediatric	WB	NA	130	GSE26440	23
Expression data from validation cohort of children with septic shock	Affymetrix	neonatal, pediatric	WB	NA	103	GSE26378	23
Post-natal age is a critical determinant of the neonatal host response to sepsis	Affymetrix	neonatal	WB	NA	150	GSE69686	24
Maternal influences on the transmission of leukocyte gene expression profiles in population samples (mother and child)	Illumina	neonatal, adult	WB	NA	56	GSE21342	29
Standard of hygiene and immune adaptation in newborn infants [113 cord blood RNA samples] (This SubSeries is part of SuperSeries GSE53473: Standard of hygiene and immune adaptation in newborn infants)	Affymetrix	neonatal	WB	XIST	113	GSE53471	21
Standard of hygiene and immune adaptation in newborn infants [15 rehybridized/batch correction samples] (This SubSeries is part of SuperSeries GSE53473: Standard of hygiene and immune adaptation in newborn infants)	Affymetrix	neonatal	WB	XIST	15	GSE53472	21
Genome-wide analysis of gene expression levels in placenta and cord blood samples from newborns babies	Illumina	neonatal	WB, placenta	NA	96	GSE36828	NA
Oxygen induced complication of prematurity: from experimental data to prevention strategy	Affymetrix	neonatal	WB	NA	299	GSE32472	25
Gene expression study reveals compromised Pattern Recognition Receptors and Interferon Signaling in fullterm Low birth Weight newborns	Affymetrix	neonatal	WB	NA	12	GSE29807	NA
Deregulation of Gene Expression induced by Environmental Tobacco Smoke Exposure in Pregnancy	Illumina	neonatal	WB, placenta	NA	104	GSE30032	20
Comprehensive Study of Tobacco Smoke-Related Transcriptome Alterations in Maternal and Fetal Cells	Illumina	neonatal, adult	WB, placenta	NA	183	GSE27272	19
Gene expression profiles of adult peripheral and cord blood mononuclear cells altered by lipopolysaccharide	Affymetrix	neonatal, adult	PBMCs, CBMCs	NA	12	GSE3140	30
The human reticulocyte transcriptome	Affymetrix	fetal, adult	erythroid cells	TFRC	12	GSE17639	31
Expression Profiling of Primary Human Fetal and Adult Hematopoietic Stem/Progenitor Cells (HSPCs) and Differentiating Proerythroblasts (ProEs); (This SubSeries is part of SuperSeries GSE36994: Comparative profiling of human fetal and adult erythropoiesis)	Affymetrix	fetal, adult	HSCs, erythroid cells	CD34, TFRC	22	GSE36984	32

Title	Platform	Sample source		QC Markers	Number of Samples	GEO ID	Ref.
		Origin	RNA				
Expression Profiling of Primary Human Proerythroblasts (ProEs) After IRF2, IRF6, and MYB shRNA Knockdown (This SubSeries is part of SuperSeries GSE36994: Comparative profiling of human fetal and adult erythropoiesis)	Affymetrix	fetal, adult	erythroid cells	TFRC	20	GSE36988	32
Characterization of Transcription Factor Networks Involved in Umbilical Cord Blood CD34+ Stem Cells-Derived Erythropoiesis	Illumina	neonatal	erythroid cells	CD34, TFRC	12	GSE49438	33
Densely interconnected transcriptional circuits control cell states in human hematopoiesis	Affymetrix	neonatal, adult	various cell types	CD34, TFRC, CD19, CD4, CD3D, CD14, CD8A	211	GSE24759	34
Expression data from human CD34+ HPC subpopulations isolated from umbilical cord blood	Affymetrix	neonatal	T/NK and B-lymphoid progenitor cells	CD34, CD19, CD3D	8	GSE29522	NA
Distinct functional programming of human fetal and adult monocytes	Agilent Technologies	fetal, adult	monocytes	CD14	8	GSE54668	35
Restricted Dendritic Cell and Monocyte Progenitors in Human Cord Blood and Bone Marrow	Illumina	neonatal, adult	monocytes, DCs	CD14, CD11C, IL3RA	36	GSE65128	36
DC response to Respiratory syncytial virus from adult peripheral and cord blood	Affymetrix	neonatal, adult	DCs	CD11C, IL3RA	12	GSE24132	37
Differences in the transcriptomic response of human adult and neonatal dendritic cell subsets to TLR7/8 stimulation	Illumina	neonatal, adult	DCs	XIST, IL3RA, CD11C	72	GSE67057	NA
Genome-wide analysis of B lymphocytes derived from human pluripotent stem cells, neonatal and adult sources	Illumina	neonatal, adult	B cells	CD19	19	GSE53572	NA
Functional Analysis and Gene Expression Profile of Umbilical Cord Blood Regulatory T Cells	Affymetrix	neonatal, adult	Tregs	CD4, CD3D, FOXP3	10	GSE22501	NA
Regulatory T cells in human pregnancy	Illumina	fetal, adult	CD4+ T cells	CD4, CD3D	23	GSE31976	38
Comparison of gene expression profiles by CD3+CD4+ thymocytes derived from fetal and adult hematopoietic stem cells (This SubSeries is part of SuperSeries GSE25119: Comparison of CD4+ T cells from human fetal and adult donors)	Affymetrix	fetal, adult	CD4+ T cells	CD4, CD3D, FOXP3	9	GSE25085	39
Human Fetal and Adult Peripheral Naïve CD4+ T cells and CD4+CD25+ Treg cells (This SubSeries is part of SuperSeries GSE25119: Comparison of CD4+ T cells from human fetal and adult donors)	Affymetrix	fetal, adult	CD4+ T cells	CD4, CD3D, FOXP3	12	GSE25087	39
Gene Expression Profile during human CD4+ T cell differentiation (platform GLP96)	Affymetrix	fetal, neonatal, adult	CD4+ T cells	CD4, CD3D	15	GSE1460	40
Gene Expression Profile during human CD4+ T cell differentiation (platform GLP97)	Affymetrix	fetal, neonatal, adult	CD4+ T cells	NA	15	GSE1460	40
Gene expression profile of activated CD4 T cells from adults and newborns	Affymetrix	neonatal, adult	CD4+ T cells	NA	12	GSE52129	41
Expression data from healthy human CD161++CD8aa and CD161++CD8ab T cells (This SubSeries is part of SuperSeries GSE33425: Human MAIT and CD8++ cell development)	Affymetrix	neonatal, adult	CD8+ T cells	CD8A	8	GSE33374	42
Expression data from human cord blood CD161++/CD161-/CD161-/CD8+ T cell subsets (This SubSeries is part of SuperSeries GSE33425: Human MAIT and CD8++ cell development)	Affymetrix	neonatal, adult	CD8+ T cells	CD8A	9	GSE33424	42

References

1. Dowling DJ, Levy O: **Ontogeny of early life immunity.** *Trends Immunol.* 2014; **35**(7): 299–310.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Basha S, Surendran N, Pichichero M: **Immune responses in neonates.** *Expert Rev Clin Immunol.* 2014; **10**(9): 1171–84.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Levy O: **Innate immunity of the newborn: basic mechanisms and clinical correlates.** *Nat Rev Immunol.* 2007; **7**(5): 379–90.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Lozano R, Naghavi M, Foreman K, *et al.*: **Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010.** *Lancet.* 2012; **380**(9859): 2095–128.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Kollmann TR, Crabtree J, Rein-Weston A, *et al.*: **Neonatal innate TLR-mediated responses are distinct from those of adults.** *J Immunol.* 2009; **183**(11): 7150–60.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Corbett NP, Blimkie D, Ho KC, *et al.*: **Ontogeny of Toll-like receptor mediated cytokine responses of human blood mononuclear cells.** *PLoS One.* 2010; **5**(11): e15041.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Liao SL, Yeh KW, Lai SH, *et al.*: **Maturation of Toll-like receptor 1-4 responsiveness during early life.** *Early Hum Dev.* 2013; **89**(7): 473–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Marr N, Wang TI, Kam SH, *et al.*: **Attenuation of respiratory syncytial virus-induced and RIG-I-dependent type I IFN responses in human neonates and very young children.** *J Immunol.* 2014; **192**(3): 948–57.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Sharma AA, Jen R, Brant R, *et al.*: **Hierarchical maturation of innate immune defences in very preterm neonates.** *Neonatology.* 2014; **106**(1): 1–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Kollmann TR, Levy O, Montgomery RR, *et al.*: **Innate immune function by Toll-like receptors: distinct responses in newborns and the elderly.** *Immunity.* 2012; **37**(5): 771–83.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Marchant EA, Kan B, Sharma AA, *et al.*: **Attenuated innate immune defenses in very premature neonates during the neonatal period.** *Pediatr Res.* 2015; **78**(5): 492–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Lavoie PM, Huang Q, Jollette E, *et al.*: **Profound lack of interleukin (IL)-12/IL-23p40 in neonates born early in gestation is associated with an increased risk of sepsis.** *J Infect Dis.* 2010; **202**(11): 1754–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. You D, Marr N, Saravia J, *et al.*: **IL-4R α on CD4 $^{+}$ T cells plays a pathogenic role in respiratory syncytial virus reinfection in mice infected initially as neonates.** *J Leukoc Biol.* 2013; **93**(6): 933–42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Guilmot A, Carlier Y, Truyens C: **Differential IFN- γ production by adult and neonatal blood CD56 $^{+}$ natural killer (NK) and NK-like-T cells in response to *Trypanosoma cruzi* and IL-15.** *Parasite Immunol.* 2014; **36**(1): 43–52.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Dasari P, Zola H, Nicholson IC: **Expression of Toll-like receptors by neonatal leukocytes.** *Pediatr Allergy Immunol.* 2011; **22**(2): 221–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Viemann D, Dubbel G, Schleifenbaum S, *et al.*: **Expression of toll-like receptors in neonatal sepsis.** *Pediatr Res.* 2005; **58**(4): 654–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. 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](#)
18. Zhu Y, Davis S, Stephens R, *et al.*: **GEOMETADB: powerful alternative search engine for the Gene Expression Omnibus.** *Bioinformatics.* 2008; **24**(23): 2798–800.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. 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](#)
20. Votavova H, Dostalova Merkerova M, Krejciik 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](#)
21. Kallionpää H, Laajala E, Öling V, *et al.*: **Standard of hygiene and immune adaptation in newborn infants.** *Clin Immunol.* 2014; **155**(1): 136–47.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Dickinson P, Smith CL, Forster T, *et al.*: **Whole blood gene expression profiling of neonates with confirmed bacterial sepsis.** *Genom Data.* 2014; **3**: 41–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Wynn JL, Cvijanovich NZ, Allen GL, *et al.*: **The influence of developmental age on the early transcriptomic response of children with septic shock.** *Mol Med.* 2011; **17**(11–12): 1146–56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Wynn JL, Guthrie SO, Wong HR, *et al.*: **Postnatal Age Is a Critical Determinant of the Neonatal Host Response to Sepsis.** *Mol Med.* 2015; **21**: 496–504.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Pietrzyk JJ, Kwinta P, Wollen EJ, *et al.*: **Gene expression profiling in preterm infants: new aspects of bronchopulmonary dysplasia development.** *PLoS One.* 2013; **8**(10): e78585.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Elahi S, Ertelt JM, Kinder JM, *et al.*: **Immunosuppressive CD71 $^{+}$ erythroid cells compromise neonatal host defence against infection.** *Nature.* 2013; **504**(7478): 158–62.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. 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 1; referees: awaiting peer review].** *F1000Res.* 2016; **5**: 305.
[Publisher Full Text](#)
28. Rinchai D, Boughorbel S, Presnell S, *et al.*: **A compendium of monocyte transcriptome datasets to foster biomedical knowledge discovery [version 1; referees: 1 approved].** *F1000Res.* 2016; **5**: 291.
[Publisher Full Text](#)
29. Mason E, Tronc G, Nones K, *et al.*: **Maternal influences on the transmission of leukocyte gene expression profiles in population samples from Brisbane, Australia.** *PLoS One.* 2010; **5**(12): e14479.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Koch L, Linderkamp O, Iltich C, *et al.*: **Gene expression profiles of adult peripheral and cord blood mononuclear cells altered by lipopolysaccharide.** *Neonatology.* 2008; **93**(2): 87–100.
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Noh SJ, Miller SH, Lee YT, *et al.*: **Let-7 microRNAs are developmentally regulated in circulating human erythroid cells.** *J Transl Med.* 2009; **7**: 98.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Xu J, Shao Z, Glass K, *et al.*: **Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis.** *Dev Cell.* 2012; **23**(4): 796–811.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Li B, Ding L, Yang C, *et al.*: **Characterization of transcription factor networks involved in umbilical cord blood CD34 $^{+}$ stem cells-derived erythropoiesis.** *PLoS One.* 2014; **9**(9): e107133.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Novershtern N, Subramanian A, Lawton LN, *et al.*: **Densely interconnected transcriptional circuits control cell states in human hematopoiesis.** *Cell.* 2011; **144**(2): 296–309.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Krow-Lucal ER, Kim CC, Burt TD, *et al.*: **Distinct functional programming of human fetal and adult monocytes.** *Blood.* 2014; **123**(12): 1897–904.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Lee J, Breton G, Oliveira TY, *et al.*: **Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow.** *J Exp Med.* 2015; **212**(3): 385–99.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Thornburg NJ, Shepherd B, Crowe JE Jr: **Transforming growth factor beta is a major regulator of human neonatal immune responses following respiratory syncytial virus infection.** *J Virol.* 2010; **84**(24): 12895–902.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Santner-Nanan B, Straubinger K, Hsu P, *et al.*: **Fetal-maternal alignment of regulatory T cells correlates with IL-10 and Bcl-2 upregulation in pregnancy.** *J Immunol.* 2013; **191**(1): 145–53.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Mold JE, Venkatasubrahmanyam S, Burt TD, *et al.*: **Fetal and adult hematopoietic stem cells give rise to distinct T cell lineages in humans.** *Science.* 2010; **330**(6011): 1695–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Lee MS, Hanspers K, Barker CS, *et al.*: **Gene expression profiles during human CD4 $^{+}$ T cell differentiation.** *Int Immunol.* 2004; **16**(8): 1109–24.
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Gibbons D, Fleming P, Virasami A, *et al.*: **Interleukin-8 (CXCL8) production is a signatory T cell effector function of human newborn infants.** *Nat Med.* 2014; **20**(10): 1206–10.
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Walker LJ, Kang YH, Smith MO, *et al.*: **Human MAIT and CD8 $\alpha\alpha$ cells develop from a pool of type-17 precommitted CD8 $^{+}$ T cells.** *Blood.* 2012; **119**(2): 422–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 17 June 2016

doi:[10.5256/f1000research.9008.r14443](https://doi.org/10.5256/f1000research.9008.r14443)



Peter Ghazal

Division of Infection and Pathway Medicine, Medical School, Edinburgh Infectious Diseases, University of Edinburgh, Edinburgh, UK

The report by Marr and colleagues compiles a valuable set of human early-life publically available expression datasets from the Gene Expression Omnibus (GEO) resource. Obtaining consent and sufficient amounts of sample for this population group is problematical and the limited number of datasets presented reflects the scarcity of studies in this area. The authors have made these datasets web accessible through the Gene Expression Browser (GXB). Interrogating these datasets using GXB application is straightforward but is quite limited providing a restrictive gene analytic view. Incorporating pathway-querying and visualization functions could enhance the overall utility of GXB. Further this would benefit with a note regarding the update frequency for new relevant datasets – during the review process new datasets have already become available. The usefulness and research value of this resource will only be met with continued effort to routinely curate new datasets.

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 13 April 2016

doi:[10.5256/f1000research.9008.r13168](https://doi.org/10.5256/f1000research.9008.r13168)



Stanislas Goriely

WELBIO and Institute for Medical Immunology, Université Libre de Bruxelles, Brussels, Belgium

This report by Nico Marr and colleagues puts together publicly available expression datasets pertinent to the function/development of human immune and hematological cells in early life. The authors use a Web-based application (Gene Expression Browser¹) to facilitate exploration and visualization of the data. This tool is useful for the field and user-friendly. However, the current collection is rather heterogenous (clinical samples from septic shock patients, expression profile in progenitor cells after shRNA knockdown for specific transcription factors, general data on different hematopoietic subpopulations without special emphasis on infant/adult comparison...) so it might not be that easy for researchers or clinicians to navigate between the datasets with a specific question in mind. Furthermore, is not clear whether it will be updated on a regular basis. Will other genome-wide datasets (ChIP-Seq, Methylation Arrays..) be

incorporated?

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](#)

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.
