



DATA NOTE

A curated transcriptome dataset collection to investigate the immunobiology of HIV infection [version 1; referees: 3 approved]

Jana Blazkova¹, Sabri Boughorbel¹, Scott Presnell², Charlie Quinn², Damien Chaussabel¹

¹Sidra Medical and Research Center, Doha, Qatar

²Benaroya Research Institute, Research Technology, Seattle, WA, USA

v1 First published: 11 Mar 2016, 5:327 (doi: [10.12688/f1000research.8204.1](https://doi.org/10.12688/f1000research.8204.1))
 Latest published: 11 Mar 2016, 5:327 (doi: [10.12688/f1000research.8204.1](https://doi.org/10.12688/f1000research.8204.1))

Abstract

Compendia of large-scale datasets available in public repositories provide an opportunity to identify and fill current gaps in biomedical knowledge. But first, these data need to be readily accessible to research investigators for interpretation. Here, we make available a collection of transcriptome datasets relevant to HIV infection. A total of 2717 unique transcriptional profiles distributed among 34 datasets were identified, retrieved from the NCBI Gene Expression Omnibus (GEO), and loaded in a custom web application, the Gene Expression Browser (GXB), designed for interactive query and visualization of integrated large-scale data. Multiple sample groupings and rank lists were created to facilitate dataset query and interpretation via this interface. Web links to customized graphical views can be generated by users and subsequently inserted in manuscripts reporting novel findings, such as discovery notes. The tool also enables browsing of a single gene across projects, which can provide new perspectives on the role of a given molecule across biological systems. This curated dataset collection is available at:

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



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



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

Open Peer Review

Referee Status:

	Invited Referees		
	1	2	3
version 1 published 11 Mar 2016			

- Amalio Telenti**, J. Craig Venter Institute (JCVI) USA
- Nicolas Chomont**, Université de Montréal Canada
- José Alcami**, Instituto de Salud Carlos III Spain, **Francisco Diez-Fuertes**, Instituto de Salud Carlos III Spain

Discuss this article

Comments (0)

Corresponding author: Jana Blazkova (jblazkova@sidra.org)

How to cite this article: Blazkova J, Boughorbel S, Presnell S *et al.* **A curated transcriptome dataset collection to investigate the immunobiology of HIV infection [version 1; referees: 3 approved]** *F1000Research* 2016, **5**:327 (doi: [10.12688/f1000research.8204.1](https://doi.org/10.12688/f1000research.8204.1))

Copyright: © 2016 Blazkova J *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. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: JB, SB and DC were supported by 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: 11 Mar 2016, **5**:327 (doi: [10.12688/f1000research.8204.1](https://doi.org/10.12688/f1000research.8204.1))

Introduction

Uncovering the gene transcription signature associated with different outcomes of HIV infection is paramount to a deeper understanding of HIV pathogenesis and to identifying potential therapeutic targets for improving immunological response and for eradicating HIV infection¹. HIV has a complex life cycle during which it engages multiple host cellular components, including the immune cells in which it replicates, undermining immune functions. It also hijacks host transcription factors and enzymes to assure viral production and subsequent infections². HIV dysregulates host genes resulting in aberrant immune response, disease progression, and opportunistic infections^{3,4}. The ability to pool and analyze samples across various groups of HIV infected individuals with different disease outcomes and across various cell types or tissues, offers a unique opportunity to define common denominators of the immune control of HIV infection, the regulation of HIV replication, and/or the virus-host interaction. With this in mind, we make available, via an interactive web application, a curated collection of transcriptome datasets relevant to HIV infection.

With over 65,000 studies deposited in the NCBI Gene Expression Omnibus (GEO), a public repository of transcriptome profiles, the identification of datasets relevant to a particular research area is not straightforward. Furthermore, GEO is primarily designed as a repository for storing data, rather than for browsing and interacting with the data. Thus, we used a custom web application, the gene expression browser (GXB), to host a collection of datasets that we identified as particularly relevant to the study of the immunobiology of HIV infection. This tool has been described in detail and the source code released as part of a recent publication⁵. It allows seamless browsing and interactive visualization of large volumes of heterogeneous data. Users can easily customize data plots by adding multiple layers of information, modifying the sample order and generating links that capture these settings and can be inserted in email communications or in publications. Accessing the tool via these links also provides access to rich contextual information essential for data interpretation. This includes for instance access to gene information and relevant literature, study design, and detailed sample information.

Material and methods

Identification of relevant datasets

Potentially relevant datasets deposited in GEO were identified using an advanced query based on the Bioconductor package GEOmetadb, version 1.30.0, and on the SQLite database that captures detailed information on GEO data structure (<https://www.bioconductor.org/packages/release/bioc/html/GEOmetadb.html>)⁶. The search query was designed to retrieve entries where the title or summary contained the word HIV, and were generated from human samples using Illumina or Affymetrix commercial platforms.

The relevance of each entry returned by this query was assessed individually. This process involved reading through the descriptions and examining the list of available samples and their annotations. Sometimes it was also necessary to review the original published report in which the design of the study and generation of the dataset are described in more details. We identified 87 datasets meeting the search criteria and containing HIV infected samples (some studies related to HIV problematics contained uninfected samples only). Out of the 87 datasets, 41 were generated from tissues or cells isolated from HIV infected individuals, 46 contained cell lines or primary cells infected *in vitro*. Since molecular, cellular and physiological processes involved in the context of *in vivo* and *in vitro* infections are dramatically different, we decided to create two separate collections. Here we describe the “*in vivo* collection” composed of 34 curated datasets (after filtering out datasets that did not meet quality control criteria, as described in “Dataset Validation” section, or datasets generated using an unsupported array platform). Of the 34 datasets, 7 are from whole blood, 7 from peripheral blood mononuclear cells (PBMCs), 8 from CD4⁺ and/or CD8⁺ T-cells, 4 from monocytes, 1 from dendritic cells (DCs), and 7 from tissues different from blood (Figure 1). Four datasets comprise samples from patients co-infected with tuberculosis (TB)⁷⁻¹⁰, one dataset comprises samples from AIDS related lymphomas¹¹, and four datasets addressed HIV infected patients with neurological disorders, such as HIV related fatigue syndrome¹², major depression disorder (MDD)¹³, or HIV-Associated Neurocognitive Disorder (HAND)^{14,15}. Among the many noteworthy datasets, several stood out, such as the extensive

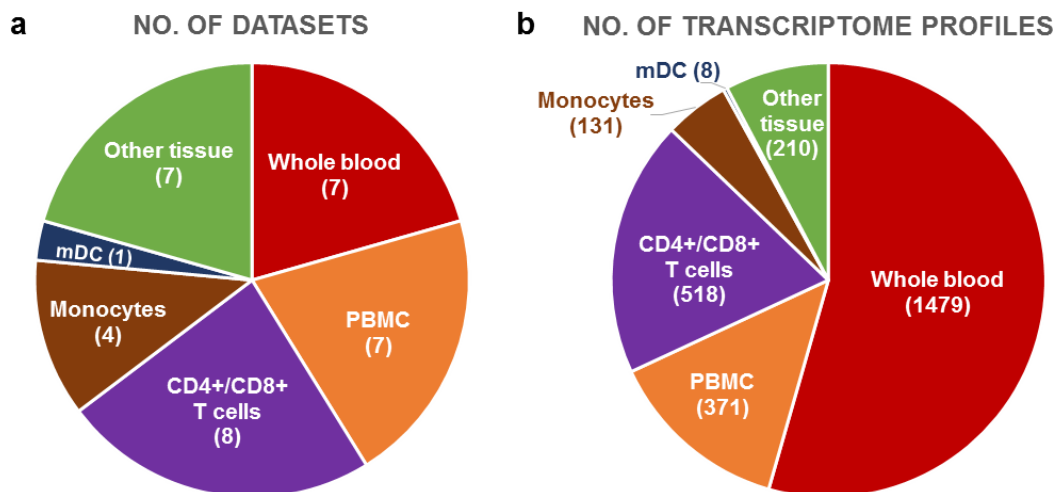


Figure 1. Sample source composition of the dataset collection. Pie charts representing the numbers of datasets (a) or transcriptome profiles (b) for different cell types and tissues.

study of the transcriptional signature of early acute HIV infection in whole blood samples of both antiretroviral-treated and untreated populations over the course of infection¹⁶ [GXB: [GSE29429-GPL10558](#) and [GSE29429-GPL6947](#)]. Several datasets investigate differences in gene expression between distinct stages of HIV infection (early/acute, chronic)^{17,18} [GXB: [GSE6740](#), [GSE16363](#)], or different host responses to infection (progressors, non-progressors, elite controllers)¹⁹⁻²³ [GXB: [GSE28128](#), [GSE24081](#), [GSE56837](#),

[GSE23879](#), [GSE18233](#)]. Other studies address different stages or responses to antiretroviral therapy²⁴⁻²⁶ [GXB: [GSE44228](#), [GSE19087](#), [GSE52900](#)], or transcriptional changes after therapy interruption²⁷⁻²⁹ [GXB: [GSE10924](#), [GSE28177](#), [GSE5220](#)]. The entirety of the datasets that makes up our collection is listed in [Table 1](#). Thematic composition of our collection is illustrated by a graphical representation of relative occurrences of terms in the list of titles loaded into the GXB tool ([Figure 2](#)).

Table 1. List of datasets constituting the collection, also available at <http://hiv.gxbsidra.org/dm3/geneBrowser/list>.

Title	Platform	Number of samples	Sample source	Validation genes	GEO ID	Ref
Blood Transcriptional Signature of hyperinflammation in HIV-associated Tuberculosis	Illumina HumanHT-12 v4	107	Whole blood	N/A	GSE58411	7
CD4 ⁺ T Cell Decline is Predicted by Differential Expression of Genes in HIV seropositive patients	Affymetrix HG-Focus v1	96	PBMC	N/A	GSE10924	27
CD4 ⁺ T cell gene expression in virologically suppressed HIV-infected patients during Maraviroc intensification therapy	Illumina HumanHT-12 v4	77	CD4 ⁺ T cells	CD3 , CD4	GSE56804	30
Chronic CD4 ⁺ T cell Activation and Depletion in HIV-1 Infection: Type I Interferon-Mediated Disruption of T Cell Dynamic	Affymetrix HG-U133_Plus_2	20	CD4 ⁺ T cells	CD3 , CD4	GSE9927	31
Comparative analysis of genomic features of human HIV-1 infection and primate models of SIV infection	Illumina HumanWG-6 v3	79	CD4 ⁺ CD8 ⁺ T cells	CD4 , CD8	GSE28128	19
Comparison of CD4 ⁺ T cell function between HIV-1 resistant and HIV-1 susceptible individuals (Affymetrix)	Affymetrix HG-U133_Plus_2	18	CD4 ⁺ T cells	CD3 , CD4	GSE14278	32
Comparison of gene expression profiles of HIV-specific CD8 T cells from controllers and progressors	Affymetrix HG-U133A	42	CD8 ⁺ T cells	CD8 , CD4-neg	GSE24081	20
Comparison of transcriptional profiles of CD4 ⁺ and CD8 ⁺ T cells from HIV-infected patients and uninfected control group	Affymetrix HG-U133A	40	CD4 ⁺ CD8 ⁺ T cells	CD4 , CD8	GSE6740	17
Differential Gene Expression in HIV-Infected Individuals Following ART	Illumina HumanWG-6 v3	72	PBMC	XIST	GSE44228	24
Differential Gene Expression of Soluble CD8 ⁺ T-cell mediated suppression of HIV replication in three older children	Affymetrix HG-U133_Plus_2	3	PBMC	XIST	GSE23183	33
Expression data from CD11c ⁺ mDCs in HIV infection	Affymetrix HG-U133_Plus_2	8	mDC	CD11c	GSE42058	34
Expression data from HAART interruption in HIV patients	Affymetrix HG-U133_Plus_2	6	GALT	N/A	GSE28177	28
Expression data from HIV exposed and uninfected women	Affymetrix HG-U133_Plus_2	86	Whole blood	N/A	GSE33580	35
Fatigue-related HIV disease gene-networks identified in CD14 ⁺ cells isolated from HIV-infected patients	Affymetrix FATMITO1a 520158F v1	15	Mono cytes	CD14	GSE18468	12
Gene expression analysis of PBMC from HIV and HIV/TB co-infected patients	Illumina HumanHT-12 v4	44	PBMC	XIST	GSE50834	8
Gene expression before HAART initiation predicts HIV-infected individuals at risk of poor CD4 ⁺ T cell recovery	Illumina HumanWG-6 v3	24	PBMC	XIST	GSE19087	25
Gene Expression in Frontal Cortex in Major Depression and HIV	Affymetrix HG-U133_Plus_2	8	Brain	XIST	GSE17440	13
Gene-expression profiling of HIV-1 infection and perinatal transmission in Botswana	Affymetrix HG-U133A	45	PBMC	N/A	GSE4124	36

Dataset 1. Raw data for Figure 1

<http://dx.doi.org/10.5256/f1000research.8204.d115581>

Gene expression browser (GXB) – dataset upload and annotation

Once a final selection had been made, each dataset was downloaded from GEO as a Simple Omnibus Format in Text (SOFT) file. It was in turn uploaded on a dedicated instance of the GXB, an interactive web application developed at the Benaroya Research Institute, hosted on the Amazon Web Services cloud. Available sample and study information were also uploaded. Samples were grouped according to possible interpretations of study results and gene rankings were computed based on different group comparisons (e.g. comparing samples from HIV negative vs HIV positive patients, with or without antiretroviral therapy, in different stages of disease progression, or with or without co-infection, depending on the focus of respective studies).

GXB – short tutorial

The GXB software has been described in detail in a recent publication⁵. This custom software interface provides users with a means to easily navigate and filter the dataset collection available at <http://hiv.gxbsidra.org/dm3/geneBrowser/list>. A web tutorial is also available online: <https://gxb.benaroyaresearch.org/dm3/tutorials.gsp#gxbtut>. Briefly, datasets of interest can be quickly identified either by filtering on criteria from pre-defined lists on the left side of the dataset navigation page, 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 to 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 the rank list by selecting different comparisons (in cases where the dataset is split in more than two groups), or to only include genes that are selected for specific biological interest. b) To change sample grouping (Group Set button); in some datasets, user can switch between interpretations where samples are grouped based on cell type or disease, 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 a bar plot 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 plot or a box plot. e) To provide a color legend for the sample groups. f) To select categorical information 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 or the table with expression values as a comma separated values (csv) file. 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 “hide/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. Information available about individual samples is provided under the “Sample” tab. Rolling the mouse cursor over a bar plot, 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 data for use alongside the expression data in third party analysis software. Other functionalities are provided under the “Tools” drop-down menu located in the top right corner of the user interface. These functionalities include notably: a) “Annotations”, which provides access to all the ancillary information about the study, samples and the dataset, organized across different tabs; b) “Cross Project View”, which provides the ability to browse across all available studies for a given 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); and d) “Chart Options”, which gives user the option to customize chart labels.

Dataset validation

Quality control checks were performed by examination of profiles of relevant biological markers. Known leukocyte surface markers were used to verify consistency of the information provided by dataset depositors, and to identify instances where contamination of samples by other leukocyte populations may be confounding. The markers that were used include: CD3 (CD3D), a T-cell marker; CD4 and CD8 (CD8A), markers of CD4⁺ and CD8⁺ T cells respectively; CD11c (ITGAX), an mDC marker; CD14, expressed by monocytes and macrophages; or Adiponectin (ADIPOQ), expressed in adipose tissue. Expression of the XIST transcripts, which expression is gender-specific, was also examined in datasets containing relevant information, to determine its concordance with demographic information provided with the GEO submission (respective links in [Table 1](#)).

Data availability

All datasets included in our curated collection are also available publicly via the NCBI GEO website: www.ncbi.gov/geo; and are referenced throughout the manuscript by their GEO accession numbers (e.g. [GSE44228](#)). Signal files and sample description files can also be downloaded from the GXB tool under the “downloads” tab.

F1000Research: Dataset 1. Raw data for Figure 1, 10.5256/f1000research.8204.d115581³⁹

Author contributions

JB and DC conceived the theme for this dataset collection. DC, SP and CQ designed the software. SP, CQ and SB installed and tested the software and programmed portions of the web application. SB uploaded datasets. JB curated and annotated datasets. JB and DC prepared the first draft of 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

JB, SB and DC were supported by 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 publically available by depositing them in GEO.

References

- Martin AR, Siliciano RF: **Progress Toward HIV Eradication: Case Reports, Current Efforts, and the Challenges Associated with Cure.** *Annu Rev Med.* 2016; **67**: 215–28.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moir S, Chun TW, Fauci AS: **Pathogenic mechanisms of HIV disease.** *Annu Rev Pathol.* 2011; **6**: 223–48.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sauter D, Kirchhoff F: **HIV replication: a game of hide and sense.** *Curr Opin HIV AIDS.* 2016; **11**(2): 173–81.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mohan T, Bhatnagar S, Gupta DL, et al.: **Current understanding of HIV-1 and T-cell adaptive immunity: progress to date.** *Microb Pathog.* 2014; **73**: 60–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- 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](#)
- 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](#)
- Lai RP, Meintjes G, Wilkinson KA, et al.: **HIV-tuberculosis-associated immune reconstitution inflammatory syndrome is characterized by Toll-like receptor and inflammasome signalling.** *Nat Commun.* 2015; **6**: 8451.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dawany N, Showe LC, Kossenkov AV, et al.: **Identification of a 251 gene expression signature that can accurately detect *M. tuberculosis* in patients with and without HIV co-infection.** *PLoS One.* 2014; **9**(2): e89925.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Anderson ST, Kaforou M, Brent AJ, et al.: **Diagnosis of childhood tuberculosis and host RNA expression in Africa.** *N Engl J Med.* 2014; **370**(18): 1712–23.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kaforou M, Wright VJ, Oni T, et al.: **Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study.** *PLoS Med.* 2013; **10**(10): e1001538.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Deffenbacher KE, Iqbal J, Liu Z, et al.: **Recurrent chromosomal alterations in molecularly classified AIDS-related lymphomas: an integrated analysis of DNA copy number and gene expression.** *J Acquir Immune Defic Syndr.* 2010; **54**(1): 18–26.
[PubMed Abstract](#)
- Voss JG, Dobra A, Morse C, et al.: **Fatigue-related gene networks identified in CD14⁺ cells isolated from HIV-infected patients: part II: statistical analysis.** *Biol Res Nurs.* 2013; **15**(2): 152–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tatro ET, Scott ER, Nguyen TB, et al.: **Evidence for Alteration of Gene Regulatory Networks through MicroRNAs of the HIV-infected brain: novel analysis of retrospective cases.** *PLoS One.* 2010; **5**(4): e10337.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gelman BB, Chen T, Lisinicchia JG, et al.: **The National NeuroAIDS Tissue Consortium brain gene array: two types of HIV-associated neurocognitive impairment.** *PLoS One.* 2012; **7**(9): e46178.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Levine AJ, Horvath S, Miller EN, et al.: **Transcriptome analysis of HIV-infected peripheral blood monocytes: gene transcripts and networks associated with neurocognitive functioning.** *J Neuroimmunol.* 2013; **265**(1–2): 96–105.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chang HH, Soderberg K, Skinner JA, et al.: **Transcriptional network predicts viral set point during acute HIV-1 infection.** *J Am Med Assoc.* 2012; **19**(6): 1103–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hyrca MD, Kovacs C, Loutfy M, et al.: **Distinct transcriptional profiles in ex vivo CD4⁺ and CD8⁺ T cells are established early in human immunodeficiency virus type 1 infection and are characterized by a chronic interferon response as well as extensive transcriptional changes in CD8⁺ T cells.** *J Virol.* 2007; **81**(7): 3477–86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li Q, Smith AJ, Schacker TW, et al.: **Microarray analysis of lymphatic tissue reveals stage-specific, gene expression signatures in HIV-1 infection.** *J Immunol.* 2009; **183**(3): 1975–82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rotger M, Dalmau J, Rauch A, et al.: **Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque.** *J Clin Invest.* 2011; **121**(6): 2391–400.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quigley M, Pereyra F, Nilsson B, et al.: **Transcriptional analysis of HIV-specific CD8⁺ T cells shows that PD-1 inhibits T cell function by upregulating BATF.** *Nat Med.* 2010; **16**(10): 1147–51.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Xu X, Qiu C, Zhu L, et al.: **IFN-stimulated gene LY6E in monocytes regulates the CD14/TLR4 pathway but inadequately restrains the hyperactivation of monocytes during chronic HIV-1 infection.** *J Immunol.* 2014; **193**(8): 4125–36.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Vigneault F, Woods M, Buzon MJ, et al.: **Transcriptional profiling of CD4 T cells identifies distinct subgroups of HIV-1 elite controllers.** *J Virol.* 2011; **85**(6): 3015–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rotger M, Dang KK, Fellay J, et al.: **Genome-wide mRNA expression correlates of viral control in CD4⁺ T-cells from HIV-1-infected individuals.** *PLoS Pathog.* 2010; **6**(2): e1000781.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Massanella M, Singhanian A, Beliakova-Bethell N, et al.: **Differential gene expression in HIV-infected individuals following ART.** *Antiviral Res.* 2013; **100**(2): 420–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Woelk CH, Beliakova-Bethell N, Goicoechea M, et al.: **Gene expression before HAART initiation predicts HIV-infected individuals at risk of poor CD4⁺ T-cell recovery.** *AIDS.* 2010; **24**(2): 217–22.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wu JQ, Sasse TR, Saksena MM, et al.: **Transcriptome analysis of primary monocytes from HIV-positive patients with differential responses to antiretroviral therapy.** *Virol J.* 2013; **10**: 361.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vahey MT, Wang Z, Su Z, et al.: **CD4⁺ T-cell decline after the interruption of antiretroviral therapy in ACTG A5170 is predicted by differential expression of genes in the ras signaling pathway.** *AIDS Res Hum Retroviruses.* 2008; **24**(8): 1047–66.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

28. Lerner P, Guadalupe M, Donovan R, *et al.*: **The gut mucosal viral reservoir in HIV-infected patients is not the major source of rebound plasma viremia following interruption of highly active antiretroviral therapy.** *J Virol.* 2011; **85**(10): 4772–82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Tilton JC, Johnson AJ, Luskin MR, *et al.*: **Diminished production of monocyte proinflammatory cytokines during human immunodeficiency virus viremia is mediated by type I interferons.** *J Virol.* 2006; **80**(23): 11486–97.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Beliakova-Bethell N, Jain S, Woelk CH, *et al.*: **Maraviroc intensification in patients with suppressed HIV viremia has limited effects on CD4+ T cell recovery and gene expression.** *Antiviral Res.* 2014; **107**: 42–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Sedaghat AR, German J, Teslovich TM, *et al.*: **Chronic CD4+ T-cell activation and depletion in human immunodeficiency virus type 1 infection: type I interferon-mediated disruption of T-cell dynamics.** *J Virol.* 2008; **82**(4): 1870–83.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. McLaren PJ, Ball TB, Wachihhi C, *et al.*: **HIV-exposed seronegative commercial sex workers show a quiescent phenotype in the CD4+ T cell compartment and reduced expression of HIV-dependent host factors.** *J Infect Dis.* 2010; **202**(Suppl 3): S339–44.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Katz BZ, Salimi B, Gadd SL, *et al.*: **Differential gene expression of soluble CD8+ T-cell mediated suppression of HIV replication in three older children.** *J Med Virol.* 2011; **83**(1): 24–32.
[PubMed Abstract](#) | [Publisher Full Text](#)
34. Nagy LH, Grishina I, Macal M, *et al.*: **Chronic HIV infection enhances the responsiveness of antigen presenting cells to commensal *Lactobacillus*.** *PLoS One.* 2013; **8**(8): e72789.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Songok EM, Luo M, Liang B, *et al.*: **Microarray analysis of HIV resistant female sex workers reveal a gene expression signature pattern reminiscent of a lowered immune activation state.** *PLoS One.* 2012; **7**(1): e30048.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Montano M, Rarick M, Sebastiani P, *et al.*: **Gene-expression profiling of HIV-1 infection and perinatal transmission in Botswana.** *Genes Immun.* 2006; **7**(4): 298–309.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Ockenhouse CF, Bernstein WB, Wang Z, *et al.*: **Functional genomic relationships in HIV-1 disease revealed by gene-expression profiling of primary human peripheral blood mononuclear cells.** *J Infect Dis.* 2005; **191**(12): 2064–74.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Smith AJ, Li Q, Wietgreffe SW, *et al.*: **Host genes associated with HIV-1 replication in lymphatic tissue.** *J Immunol.* 2010; **185**(9): 5417–24.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Blazkova J, Boughorbel S, Presnell S, *et al.*: **Dataset 1 in: A curated transcriptome dataset collection to investigate the immunobiology of HIV infection.** *F1000Research.* 2016.
[Data Source](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 20 April 2016

doi:10.5256/f1000research.8824.r12871



José Alcamí¹, Francisco Díez-Fuertes²

¹ Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain

² Instituto de Salud Carlos III, Majadahonda, Madrid, Spain

Blazkova *et al.* describe an interactive web application that includes 34 different transcriptome datasets. This open tool facilitates access to transcriptome analysis in the HIV field allowing meta-analyses on transcriptomic changes in HIV infection.

As strengths of the article I will highlight:

- The application is friendly and easy to use and allowed us to compare our results with a large collection of databases in a comprehensive way.
- The software allows searches related with a particular gene and how its expression is modified in different scenarios (infected vs non-infected, long term non-progressors vs typical progressors, treated vs untreated).
- The cellular types in which dataset have been obtained are indicated.
- Datasets included have been selected according to their interest and high methodological standards. For example, when contamination with cell types different from those initially targeted are detected the studies are not considered for the final dataset thus enhancing the quality of the results.

I would propose some suggestions to improve this interesting tool:

- All the studies were performed with microarrays. It would be important to discuss if the inclusion of data using RNA-seq approaches and the current units used in these studies (FPKMs, RPKMs, TPMs) could be incorporated in the future.
- It should be clarified if the results among the different studies are normalized or just described with the units used in each study. If data normalization has been performed it would be important to describe how it was done.

Overall it represents an important effort that can be useful for many researchers working in the field of HIV genetics and pathogenesis.

We have read this submission. We believe that we 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 15 April 2016

doi:10.5256/f1000research.8824.r13356



Nicolas Chomont

Department of Microbiology, Infectiology and Immunology, Université de Montréal, Montreal, QC, Canada

In this interesting article, Blazkova and colleagues describe the development of an interactive web application that allows HIV researchers to access a collection of transcriptome datasets relevant to HIV infection. The collection includes 34 datasets generated with human samples that have been carefully selected based on their relevance and quality control checks.

This is a very useful tool that can be easily used by non-experts in transcriptomics analyses. I have used it and I am convinced that it potentially represents an important contribution to the work performed by HIV researchers. I tested the accuracy of the tool (not in a formal way) by examining differences in the expression for several genes that are well-known to be modulated by HIV infection. The results are clearly presented and can be easily exported to be included in presentations/publications.

A few suggestions: I anticipate that the database will be updated on a regular basis. Therefore, it would be great to specify the date of the last update of the data set available online. Also, the possibility of adding RNAseq data would be important in the future. Maybe a brief description of each dataset would be useful too.

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 12 April 2016

doi:10.5256/f1000research.8824.r12874



Amalio Telenti

J. Craig Venter Institute (JCVI), La Jolla, CA, USA

The article by Blazkova and colleagues constitutes an important contribution to the HIV field. It crystallizes the efforts of multiple groups that characterized the host transcriptional response to infection by providing a viewer of data that are not immediately accessible in a structured interface. I have assessed the performance of the tool, and found it intuitive and user-friendly.

It extends efforts of my group to provide facilitated access to genomic data in HIV disease (<http://www.guavah.org/>).

I would bring two aspects up for discussion. First, that this tool should evolve to display RNAseq data = new generation sequencing data are increasingly available, effectively displacing microarrays. RNAseq is

also easier for standardization across studies. Second, that users should be attentive to the subtleties of analysis: covariates such as gender, age, cellularity, analytical platforms and batch effects can influence expression profiles significantly. In-depth analysis may thus require downloading of original expression data.

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.
