




METHOD ARTICLE

Organizing gene literature retrieval, profiling, and visualization training workshops for early career researchers

[version 1; peer review: awaiting peer review]

Fatima Al Ali ^{1*}, Alexandra K Marr^{1*}, Zohreh Tatari-Calderone¹,
Mohamed Alfaki ¹, Mohammed Toufiq ¹, Jessica Roelands ¹,
Basirudeen Syed Ahamed Kabeer ¹, Davide Bedognetti¹⁻³, Nico Marr ^{1,3},
Mathieu Garand¹, Darawan Rinchai¹, Damien Chaussabel ¹

¹Research Branch, Sidra Medicine, Doha, Qatar

²Department of Internal Medicine and Medical Specialties, University of Genoa, Genoa, 16126, Italy

³College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar

* Equal contributors

V1 First published: 06 Apr 2021, 10:275
<https://doi.org/10.12688/f1000research.36395.1>

Latest published: 06 Apr 2021, 10:275
<https://doi.org/10.12688/f1000research.36395.1>

Abstract

Developing the skills needed to effectively search and extract information from biomedical literature is essential for early-career researchers. It is, for instance, on this basis that the novelty of experimental results, and therefore publishing opportunities, can be evaluated. Given the unprecedented volume of publications in the field of biomedical research, new systematic approaches need to be devised and adopted for the retrieval and curation of literature relevant to a specific theme. Here we describe a hands-on training curriculum aimed at retrieval, profiling, and visualization of literature associated with a given topic. This curriculum was implemented in a workshop in January 2021. We provide supporting material and step-by-step implementation guidelines with the ISG15 gene literature serving as an illustrative use case. Through participation in such a workshop, trainees can learn: 1) to build and troubleshoot PubMed queries in order to retrieve the literature associated with a gene of interest; 2) to identify key concepts relevant to given themes (such as cell types, diseases, and biological processes); 3) to measure the prevalence of these concepts in the gene literature; 4) to extract key information from relevant articles, and 5) to develop a background section or summary on the basis of this information. Finally, trainees can learn to consolidate the structured information captured through this process for presentation via an interactive web application.

Open Peer Review

Reviewer Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Literature profiling, Science education, Concept extraction, Data visualization



This article is included in the [Sidra Medicine gateway](#).

Corresponding author: Damien Chaussabel (dchaussabel@sidra.org)

Author roles: **Al Ali F:** Data Curation, Formal Analysis, Methodology, Visualization, Writing – Review & Editing; **Marr AK:** Data Curation, Formal Analysis, Methodology, Visualization, Writing – Review & Editing; **Tatari-Calderone Z:** Methodology, Writing – Review & Editing; **Alfaki M:** Methodology, Writing – Review & Editing; **Toufiq M:** Methodology, Writing – Review & Editing; **Roelands J:** Methodology, Writing – Review & Editing; **Syed Ahamed Kabeer B:** Methodology, Writing – Review & Editing; **Bedognetti D:** Funding Acquisition, Methodology, Writing – Review & Editing; **Marr N:** Funding Acquisition, Methodology, Writing – Review & Editing; **Garand M:** Methodology, Writing – Review & Editing; **Rinchai D:** Conceptualization, Methodology, Writing – Review & Editing; **Chaussabel D:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The work presented here was supported in part by NPRP grant # 10-0205-170348 from the Qatar National Research Fund (a member of Qatar Foundation). The work reported herein is solely the responsibility of the authors.

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

Copyright: © 2021 Al Ali F *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Al Ali F, Marr AK, Tatari-Calderone Z *et al.* **Organizing gene literature retrieval, profiling, and visualization training workshops for early career researchers [version 1; peer review: awaiting peer review]** F1000Research 2021, **10**:275 <https://doi.org/10.12688/f1000research.36395.1>

First published: 06 Apr 2021, **10**:275 <https://doi.org/10.12688/f1000research.36395.1>

Introduction

Peer-reviewed publications constitute the body of knowledge upon which biomedical research is based. This resource is essential for the generation of novel hypotheses and the design of studies, experiments, and trials and is thus key to the discovery process itself. However, the volume of literature on any given specific research topics has grown dramatically in recent years, making it increasingly challenging for a single person to manually survey the relevant literature in its entirety, and consequently to acquire the foundational knowledge needed for discovery research. Hence, developing a solid foundation of skills for literature retrieval and profiling is of critical importance for early-career biomedical scientists. These competencies will, for instance, be needed: 1) to acquire a sound knowledge base through developing the ability to compile and summarize large volumes of literature. 2) to develop data interpretation skills as well as become able to assess novelty and potential impact of a given finding; and 3) to develop scientific writing skills and become able to write background material on specific topics.

Publicly available omics profiling data provides ideal material for training new generations of biomedical researchers.¹⁻³ One of the “collective omics data” training modules that we have developed follows a reductionist analysis and interpretation workflow using publicly available transcriptome profiling data as the source material.⁴ The first activities that form part of this training module involve literature retrieval and profiling for a given candidate gene.

In this article, we describe a detailed stepwise approach that we have developed running literature profiling training workshops. The training program that we have devised takes advantage of publicly available omics data.⁴ The training module presented here focuses on retrieval of gene-centric literature. Supporting material such as sets of slides, templates, and a handout, is also provided along with an illustrative use case. Notably, although the training activity that is presented focuses on gene-centric literature retrieval and profiling, the skills and approaches can be adapted to any other application (e.g., focusing on disease-centric, or pathway/process-centric literature).

Methods

The hands-on training exercise described here is suitable for implementation as part of a broader course on research methodology, or as a stand-alone workshop. The training is appropriate for undergraduate, graduate and post-doctoral trainees and no prior bioinformatics experience is required in order to participate. The time commitment depends on the level of experience of the attendees, the overall organization of the workflow and the volume of literature associated with the candidate gene(s) selected. For instance, for a gene with $\pm 1,000$ associated articles and five participants each focusing on a different theme, the training could be covered in one introductory session (40 min), and three two-hour hands-on sessions. The format and content of these sessions are described in more detail below. These would cover:

- Introductory session
- Step 1 – Retrieving the relevant literature
- Step 2 – Extracting concepts
- Step 3 – Generating literature profiles
- Step 4 – Developing interactive data visualizations
- Step 5 – Writing a narrative

In addition to covering workshop pre-requisites and time commitment, announcements for such workshops can also list the sets of skills trainees are expected to develop including:

- Literature retrieval (development of advanced PubMed queries).
- Literature profiling (information extraction, determination of keyword frequencies).
- Data visualization (structuring and presenting information in an interactive format).
- Assimilation of biomedical knowledge (synthesis and presentation).
- Scientific publication (optional: manuscript preparation and peer-review).

Trainees are required to undertake some independent preparation between each of the sessions. Alternative formats are possible depending on the needs of the participants and a range of biomedical themes (e.g., diseases, pathways, or cell types) could be selected as the focus of the literature workshop. The endpoint for such workshops may include interactive literature profiling representations generated as part of the hands-on training activities and/or a peer-reviewed publication that, for instance, could build on such a resource.

Introductory session

The introductory session is designed to provide participants with an understanding of basic concepts and present an outline of the training curriculum. In addition, this session should present the overall rationale and teaching objectives of the training program. The introduction should also define the endpoint of the workshop activity as the development of a web resource that permits the visualization and exploration of structured literature profiles for a gene of interest (described in detail below, taking *ISG15* as an illustrative case). This should be followed by an overview of each of the steps that will be undertaken during the hands-on session. These are outlined above and described in more details below. It should also provide participants with an opportunity to ask questions before the start of the hands-on sessions. Ideally, the presentation should last no longer than 40 minutes and an illustrative case could serve to support the presenter's narrative. A generic presentation is provided (**Extended data File 1**).³⁴

Step 1: Retrieving the relevant literature

As a first step, all the literature that is relevant to a given gene of interest must be identified. This forms the body of literature that will be subjected to literature profiling in subsequent steps. Most researchers will already be familiar with PubMed, the search engine hosted by the US National Center for Biotechnology Information (NCBI) (<https://pubmed.ncbi.nlm.nih.gov/>). While this tool is straightforward to use, developing queries that will permit the comprehensive retrieval of literature associated with a given theme can be more complicated.

It is important to design a PubMed query that will permit the retrieval of all the literature that is available for the gene of interest. Challenges when retrieving literature associated with a given gene include capturing all aliases and variations that may exist in addition to the official gene names and symbols. For instance, 14 aliases were used to develop a query for the gene officially known as “*ISG15* Ubiquitin Like Modifier”. Failure to include all the aliases can lead to under-representation of the literature and erroneous judgments that may have major consequences (e.g., when gauging the novelty of a given finding and its suitability for publication). The use of synonyms among gene names and aliases can also result in high proportions of false positive results, particularly when aliases are common language terms (e.g., *CAMEL* genes [official symbol: *CTAG2*] or *WARS* genes [official symbol: *WARS1*]). If the literature retrieved in this step misses relevant records for the candidate gene (= false negatives), the resulting literature profiles will be incomplete. If the literature retrieved contains records that are irrelevant (= false positives), the resulting literature profiles will be inaccurate. For these reasons, it is important to optimize PubMed queries used for gene literature retrieval. Notably, the same principle applies when generating PubMed queries for any given topic. For instance, when querying PubMed to identify literature associated with inflammation, terms such as “inflammatory” or “inflamed” should also be captured.

The endpoint of this step is the development of an optimized PubMed query for a given gene. Any criterion can be used for the selection of a candidate gene to be assigned to a trainee or group of trainees. However, a significant body of literature should be associated with the gene in question. For the purposes of illustration in this article, we have selected *ISG15*.

Practical activities for this step:

- a. The official gene name, symbol and all aliases are retrieved from the GeneCard website (e.g., for *ISG15*: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ISG15>). Although gene names may also be retrieved from different sources, the advantage of using GeneCards is that information has already been compiled from reference databases such as NCBI Entrez Genes, Uniprot or genenames.org.
- b. PubMed queries are built using the official gene name, official symbol, and all aliases for the gene of interest as search terms, and by using the appropriate Boolean operators (AND, OR, NOT), field restriction tags, and suitable syntax. For instance, in a query using multiple search terms, the Boolean operator OR must be capitalized in between each keyword. The field restriction tag [tw] (or alternatively the more restrictive [tiab]) should be added after each term (field restriction tags must be in square brackets), in order to limit the search to titles and abstracts (e.g., *ISG15* [tw] OR *IFI-15K* [tw] OR *IFI15* [tw]), or [pt] to restrict/exclude the search of a particular publication

type (e.g., NOT review [pt]). Quotation marks should be employed when compound words should appear as an exact phrase in the search (e.g., “*ISG15 Ubiquitin Like Modifier*” [tw]). And for reference, a wide range of training material can also be found on the NCBI website:

<https://pubmed.ncbi.nlm.nih.gov/help/>

<https://learn.nlm.nih.gov/documentation/training-packets/T0042010P/>

- c. The PubMed query is run, and quality checks are performed among the publications that were returned. This is to identify search terms that may be too permissive and return false positive results. For instance, short three-character acronyms tend to be more problematic, as are terms that are otherwise used as part of a common language (examples provided above).
- d. If necessary, the query is optimized by addressing problematic terms. Removing the ambiguous or problematic term altogether may be a solution, but this could also lead to false negative results (missing literature that is actually relevant to the gene of interest). As a compromise, the search term could be retained but optimized by adding a keyword that would restrict the search (e.g., provided below for HUCRP and UCRP). One may also find in some instances that the list of aliases known for a given gene is incomplete and needs to be amended (e.g., below).

Illustrative case:

The optimized query (searching all words and numbers in the title, abstract, other abstract, MeSH terms, MeSH subheadings, publication types, substance names, personal name as subject, corporate author, secondary source, comment/correction notes, and other terms) for ISG15 is as follows:

ISG15 [tw] OR “ISG15 Ubiquitin Like Modifier” [tw] OR “Interferon-stimulated gene 15” [tw] OR “IFN-induced 15-kDa protein” [tw] OR IFI-15K [tw] OR “Ubiquitin Cross-Reactive Protein” [tw] OR “Ubiquitin-Like Protein ISG15” [tw] OR (HUCRP [tw] AND “Cross-Reactive Protein” [tw]) OR G1P2 [tw] OR IP17 [tw] OR (UCRP [tw] AND “Cross-Reactive Protein” [tw]) OR “Interferon-Induced 17-KDa/15-KDa Protein” [tw] OR “Interferon-Stimulated Protein, 15 Kda” [tw] OR “Interferon-Induced 17 Kda Protein” [tw] OR IFI15 [tw] OR IMD38 [tw] NOT review [pt].

This query returned 1,186 results (as of September 1st, 2020). Optimization included the addition of aliases for this gene that were not captured in the GeneCard database but identified by reviewing some of the results (e.g., notably, the search argument Interferon-stimulated gene 15” [tw], which alone retrieves over 230 entries in PubMed). The acronyms HUCRP or UCRP also proved a source of false positive results (CPR standing for cross-reactive protein, rather than C-reactive Protein). This was rectified by adding the AND “cross reactive protein [tw]” argument to each of the ambiguous acronym.

Step 2: Extracting concepts

A large body of literature can be associated with a given gene. It may be useful then to employ a systematic “cataloguing” or indexing approach. For this step, workshop participants may first define themes under which concepts, and keywords associated with these concepts, will be categorized (e.g., themes could be ‘Human diseases and pathogens’, ‘Tissues’, ‘Cell types’, or ‘Cellular processes’). Participants would in turn scan titles of articles associated with *ISG15*, looking for concepts linked to the theme of interest. For example, under ‘Human diseases and pathogens’, a concept could be ‘liver cancer’ and associated keywords could be ‘liver cancer’, ‘hepatic carcinoma’, ‘hepatocellular carcinoma’, ‘HCC’, or ‘liver carcinoma’. The output for this activity would be lists of concepts and keywords associated with the gene literature organized under different themes. It is the prevalence of these concepts that will be measured in the subsequent step (generation of literature profiles).

Notably, this process could be repeated for other themes (or different themes could be assigned to each of the participants). If the literature is extensive and time is limited, it would also be possible to divide the literature into subsets (e.g., by batches of 100 articles, with participants assigned different batches to work on for the same theme). If the literature is sparse, all available articles for the gene may be used rather than just focusing on those articles including search terms in titles (i.e., using [tw], instead of [ti] in Step 2a). However, it may be generally preferable for a workshop to be based on selected genes with a relatively abundant literature (e.g., >100 articles returned when restricting the search to titles).

Practical activities for this step:

- a. A subset of the literature is retrieved, restricting the search to titles (using the optimized query from Step 1, and substituting the field restricting argument [ti] for [tw]).
- b. A given theme is assigned to, or selected by, participants (e.g., 'Human diseases and pathogens', 'Tissues', 'Cell types', 'Biomolecules', 'Pathways', 'Biological processes').
- c. Concepts relevant to the theme in question are identified in the titles of articles retrieved by the query designed in a) (e.g., liver cancer, HIV)
- d. The concepts and associated keywords (e.g., 'hepatocellular carcinoma', 'liver carcinoma', 'hepatic cancer', for the concept 'liver cancer' or 'virus', 'viral', 'human immunodeficiency virus' for the concept HIV) are recorded in a spreadsheet (example and templates are available in **Extended data File 2**).³⁵

Illustrative case:

The subset of the ISG15 literature in which the official gene name, symbol or aliases are present in titles is retrieved. The query below is adapted to retrieve only the literature for which gene name and aliases are present in titles (using [ti]).

ISG15 [ti] OR "ISG15 Ubiquitin Like Modifier" [ti] OR "Interferon-stimulated gene 15" [ti] OR "IFN-induced 15-kDa protein" [ti] OR IFI-15K [ti] OR "Ubiquitin Cross-Reactive Protein" [ti] OR "Ubiquitin-Like Protein ISG15" [ti] OR (HUCRP [ti] AND "Cross-Reactive Protein" [tw]) OR G1P2 [ti] OR IP17 [ti] OR (UCRP [ti] AND "Cross-Reactive Protein" [tw]) OR "Interferon-Induced 17-KDa/15-KDa Protein" [ti] OR "Interferon-Stimulated Protein, 15 Kda" [ti] OR "Interferon-Induced 17 Kda Protein" [ti] OR IFI15 [ti] OR IMD38 [ti] NOT review [pt]. This query returned 312 results (as of September 1st, 2020).

*Titles are then parsed from the ISG15 literature for concepts corresponding to the theme: "human diseases and pathogens". The concepts and associated keywords retrieved for the "human diseases and pathogens" category are listed in **Extended data File 2**³⁵ (Excel File: Extraction of concepts Tab).*

Step 3: Generating literature profiles

Determining the relative prevalence of concepts among the literature associated with a given gene can be useful. For instance, when assessing the novelty of a finding (e.g., change in transcript or protein abundance associated with pathogenesis). Moreover, such an exercise and the information being derived would also be useful in other instances when writing general background/summary about the gene for a report or a manuscript.

With the two previous steps completed, determining the prevalence of concepts in the literature associated with a given gene can be achieved quite simply. For this, the literature query developed in Step 1 is modified to narrow the search and retrieve literature associated with the concepts identified in Step 2. The endpoint for this activity is a table showing the frequency of articles for these concepts in the literature associated with a given gene (e.g., **Extended data File 2**³⁵ [Excel File: Literature Profiling Tab]).

Notes: Participants may also be encouraged to explore different types of visual representations for this type of data. Treemaps or word clouds can for instance show relative prevalence, while other types of graphs, such as 2D bubble graphs, may also show article frequencies (e.g., **Figure 1**). Another exercise could involve visualization of changes in the abundance of the literature associated with the selected concepts over the years. For this purpose, queries can be amended to add a range of publication dates with the field restriction tag [dp] (e.g., adding AND 2000:2010 [dp] to the query).

Practical activities for this step:

- a. The literature query from Step 1 is employed (using the field search restriction [tw]).
- b. The Boolean argument AND is added, followed by search terms corresponding to keywords related to one of the concepts identified in Step 2 (e.g., "Liver cancer", "Liver carcinoma", "Hepatic carcinoma", etc.).
- c. Quotation marks, field search restriction and the Boolean OR are added, so that the notation would read as follows: ... AND ("Liver cancer" [tw] OR "Liver carcinoma" [tw] OR "Hepatic carcinoma" [tw], etc.)

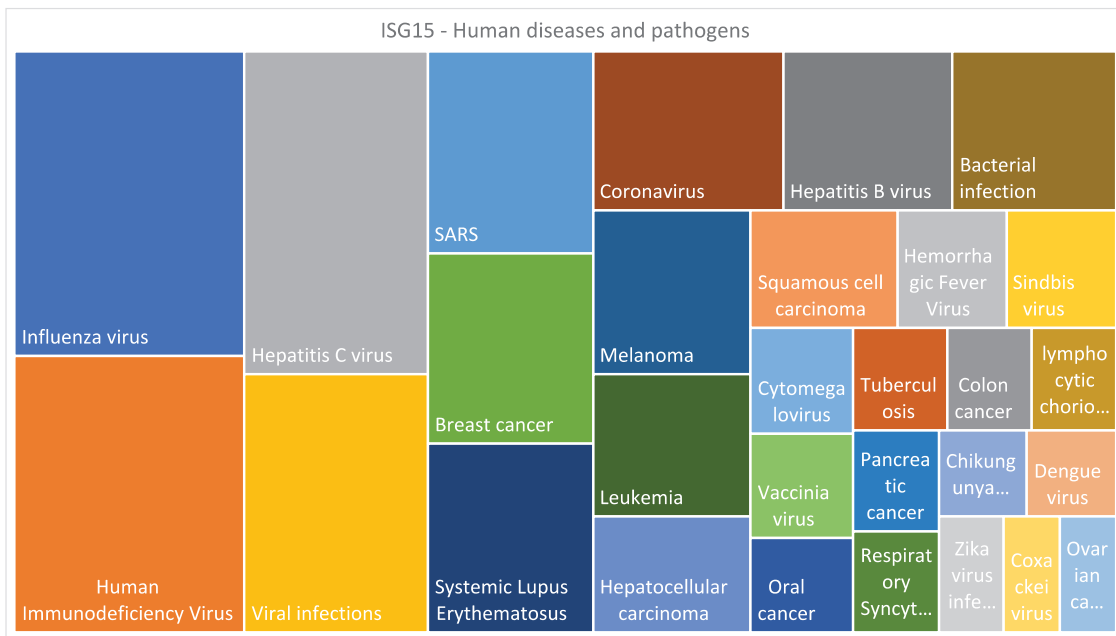


Figure 1. Visualizing ISG15 literature profiles. Treemap representation of the relative prevalence of concepts associated with the “human diseases and pathogens” theme among the ISG15 literature.

- d. The query thus constructed is run, and the number of articles retrieved recorded (for instance in a spreadsheet: **Extended data File 2**).³⁵ Steps b through d are repeated for the rest of the concepts identified in Step 2.
- e. Visual representations of the literature profiles are generated.

Illustrative case:

*The prevalence of concepts among the ISG15 literature associated with the theme “Human diseases and pathogens” is given in **Extended data File 2**³⁵ (Excel File: Literature Profiling Tab) and represented visually in **Figure 1** (also in **Extended data File 2**³⁵ [Excel File: Generating plots Tab]).*

Step 4: Developing interactive data visualizations

Valuable insights can be gained from visual representation of information. And this could be further facilitated when the information underlying these visual representations can be accessed interactively by the end user.

To produce such interactive visual representations in the context of training workshops, we recommend using the Prezi web application (<https://prezi.com>). This tool has been developed to create presentations in which it is possible to zoom in and out between levels of information. This gives users the opportunity to visualize the prevalence of concepts while at the same time allowing them to “drill down” into each individual concept in order to access relevant underlying information. The endpoint for this activity is the production of a “hierarchical circle packing chart” representing the relative abundance of concepts in the literature for a given gene and theme (**Figure 2A**, <https://prezi.com/view/zCedrcYaAEUAON1VeEUi>). Such a resource gives users access to underlying reference information for each of the concepts (**Figure 2B**) and can be made available publicly.

For the practical activity designed for this step, participants will need to register with Prezi and create an account to access and edit the interactive presentations. They can do so free of charge by selecting the basic account at: <https://prezi.com/pricing/basic/>. Ideally, this should be completed ahead of the training session. If instructors have access to a paid subscription, they can create an unlimited number of presentations and invite individual participants to collaborate with editing rights. Otherwise, the participants can create their own presentation and invite the instructors as collaborators. They would also need to copy and paste material from a master template in their own Prezi account. Thus, this alternative solution would be workable, but probably not ideal.

Practical activities for this step:

- a. Participants are given access to a Prezi presentation that contains a template and illustrative example (e.g. <https://prezi.com/view/u65ZqHn9ZBJx1VKHqfZA/>). As indicated above, one such presentation could be created and made available for each participant. It could serve as their own “Sandbox”, to familiarize themselves with the application and use as a starting point to develop an interactive resource. Multiple users can also work simultaneously on the same shared presentation. It would thus also be possible to create one presentation for a group of participants to work on cooperatively.
- b. Starting with the template and following the illustrative example, participants can build a circle packing chart for concepts identified in Step 3.
- c. The size and color of the circles is determined by the frequency of articles in the gene associated literature related to a given concept (as shown in [Figure 2A](#)).
- d. The circles can be arranged manually to create a visually attractive representation.
- e. Underlying information is then added to each of the circles and accessed by zooming in to different levels ([Figure 2B](#)). This information could include, for instance:
 - The gene symbol and concept;
 - The query link;
 - Number of articles retrieved on a specified date;
 - Result link to those articles;
 - Screen capture and links pointing to articles relevant to a specific topic (see Step 5 below).

Illustrative case:

The circle packing chart representation for *ISG15*, focusing on the “Human diseases and pathogens” topic is shown [Figure 2](#). The Prezi can be accessed for interactive exploration via: <https://prezi.com/view/zCedrcYaAEUAON1VeEU/>

A screen recording demonstrating how to add underlying information to a Prezi circle packing chart can be accessed via: <https://soapbox.wistia.com/videos/2tmC1VeQyr>

Step 5: Writing a narrative

Writing material for a report or manuscript can be one of the motivations for profiling the literature associated with a given topic. This is a skill that is worth developing as scholarly output is the measure by which productivity is measured in academic research settings. For instance, it can be at first important to be able to identify the information that is worth capturing and communicating (e.g., types of molecules measure, species, or comparator groups) The last hands-on activity of the workshop consists in capturing such information and using it as a basis for writing up a narrative about the gene of interest.

A specific topic would be selected as the focus of the activity to limit the amount of literature that would need to be covered. One such topic could, for example, be the relevance of a gene (in this case *ISG15*) as a biomarker for diagnostic applications within the overall theme of “Human diseases and pathogens” defined in Step 2. Participants could then focus on the most prevalent disease (e.g. in the case of *ISG15* literature, three diseases have >50 articles and ten diseases have >20 articles). Literature queries corresponding to the diseases that have been selected would then be modified to retrieve only those most likely to contain information about *ISG15* and its relevance as a biomarker.

Practical activities for this step:

- a. Concepts most prevalent in the gene associated literature for the theme of interest are selected (e.g. concept = Hepatitis C infection, within the theme = “Human diseases and pathogens”).
- b. The query developed to retrieve literature for a given concept is modified to restrict the search in order to retrieve articles relevant to a given topic (e.g. relevance of the gene as a biomarker). The Boolean “AND” is appended at the end of the query along with relevant keywords in parenthesis separated by the Boolean “OR” (e.g. “AND (biomarker OR biomarkers OR diagnostic OR diagnosis OR prognostic OR prognosis)”).
- c. Articles are reviewed and those deemed relevant to the topic are selected (e.g., those in which changes in abundance of the gene product in clinical specimens are reported).
- d. Relevant information is captured from the abstract and/or full text and recorded in a spreadsheet (e.g. analyte name, species, data generation methods, comparator groups, etc.). In some cases, the full text of articles may not be accessible (e.g., paywall), and the abstract has only incomplete information. This can lead to the absence of important information in the “capture” spreadsheet, rendering the findings unusable. Participants can then be reminded of the importance of using best practices when reporting findings (e.g., mentioning the comparator group, species or the specific factor being measured – being protein or RNA). It may also be an opportunity to discuss the merits of publication in open access journals.
- e. A function in the spreadsheet (column X) concatenates the information entered in the previous step to automatically generate standard sentences.
- f. The standard sentences are collated and edited to produce a cohesive paragraph or section that is relevant to the topic. For the final step, attendees may choose to “polish” the automatically generated text themselves, with the possibility of having the instructor making final edits. It would also be possible to use the services of a professional copyeditor.

Illustrative case:

Gene of interest: ISG15

Theme: “Human diseases and pathogens”.

Topic: Biomarker/diagnostic relevance.

*Focusing on ISG15 and “human diseases and pathogens” as a theme, with biomarker/diagnostic/prognostic as a topic, the main diseases identified in step 3 were selected (>20 articles; [Figure 2A](#), red circles and orange circles). Starting with the query developed for retrieving ISG15 literature relevant to Hepatitis C virus, the search is restricted by adding the expression “AND (biomarker OR biomarkers OR diagnostic OR diagnosis OR prognostic OR prognosis)”. Three articles were returned. From each article, relevant information available in the abstract, and in the full text where possible/necessary, is recorded in a spreadsheet (**Extended data File 3**).³⁶ The information in question relates specifically to changes in abundance of ISG15 measured in clinical specimens (since it is directly relevant to the topic that was selected). The spreadsheet automatically generates “standard sentences” by concatenating the information that is entered in the spreadsheet. The steps described above are repeated for other diseases (>20 articles were returned within the ISG15 literature). Automatically generated sentences are compiled and edited to produce the final version of a written section describing the potential relevance of ISG15 as a biomarker in clinical settings.*

*Text that is automatically generated by filling out the spreadsheet from **Extended data File 3**³⁶ is provided below (**Part A**). It is broken down by human disease or pathogen. PubMed IDs are provided in lieu of reference. Gaps are apparent in the sentence where input is missing from the spreadsheet (usually information is not available in the abstract and access to full text is not open). It is followed by a narrative that was prepared from this automatically generated text (**Part B**). In this case the narrative was prepared by a professional scientific copyeditor.*

Step 5e: Automatically generated text (Part A):

Hepatitis C Virus:

The abundance of ISG15 proteins measured by is increased in vivo in human liver in patients with HCV infection compared to uninfected controls [26833585].

The abundance of ISG15 Transcripts measured by is increased in vivo in human liver in patients with HCV infection compared to uninfected controls [26833585].

The abundance of ISG15 Transcripts measured by multiplex branched DNA assay is increased in vivo in human PBMCs and liver cells in patients with HCV infection who were unresponsive to IFN treatment compared to patients with HCV infection who were responsive to IFN treatment [23588721].

The abundance of ISG15 measured by is increased in vivo in human Liver in patients with HCV infection associated with an unfavorable HCV genotype 1, a high hepatic HCV load and a low antiviral response to IFN compared to patients who did not present such characteristics [20639253].

HIV:

The abundance of ISG15 IFN stimulated gene, ISG15-mRNA and ISG15-SNPs measured by TaqMan assays is increased in vivo in human PBMC in untreated HIV-1 patients HIV infection compared to healthy donors [26563749].

The abundance of ISG15 expression measured by is decreased in vivo in human PBMC in patients with long-term antiretroviral therapy HIV infection compared to uninfected controls [26563749].

The abundance of ISG15 Transcripts measured by multiplex branched DNA assay is increased in vivo in human PBMCs and liver cells in patients with therapeutic responses in 20 HIV/HCV genotype-1 subjects undergoing HCV treatment compared to patients with HCV infection who were responsive to IFN treatment [23588721].

The abundance of ISG15 Protein measured by is increased in vivo in mouse models Brain in mouse models of acute and chronic neuronal injuries HIV infection compared to global ischemia and traumatic brain injury, and in transgenic mice overexpressing HIV gp120 protein [23090498].

Melanoma:

The abundance of ISG15 Proteins measured by immunohistochemistry is increased in vivo in human melanoma cell clones and DC “in patients with” tumor compared to [12067988].

Breast cancer:

The abundance of ISG15 Proteins measured by immunohistochemistry is increased in vivo in human immune cell markers “in patients with” invasive BC patients with long-term follow-up compared to [33073304].

The abundance of ISG15 Proteins measured by is increased in vivo in human mammary tumors “in patients with” mammary tumors compared to normal mammary tissue [30500379].

The abundance of ISG15 Immunoregulatory and antitumor measured by shRNAs is ISG15 inhibits breast tumor growth and increases NK cell infiltration in vivo in nude mice tumorigenesis in patients with Cancer compared to [25749047].

The abundance of ISG15 molecule measured by is highly expressed in vivo in human numerous malignancies “in patients with” breast cancer immunotherapy compared to [22057675].

The abundance of ISG15 Proteins measured by semiquantitative real-time PCR, cDNA dot-blot hybridization and immunohistochemistry is overexpressed in vivo in human invasive breast carcinomas and normal breast tissues “in patients with” breast carcinoma cells compared to normal breast tissue [18627608].

The abundance of ISG15 Proteins measured by short hairpin RNA-mediated knockdown is increased in vivo in human tumors “in patients with” breast cancer ZR-75-1 cells decreased CPT sensitivity compared to [18566215].

The abundance of ISG15 Proteins measured by short hairpin RNA-mediated knockdown is reduced in vivo in human tumor cells selected for resistance to CPT “in patients with” breast cancer compared to ... [18566215].

Systemic Lupus

The abundance of ISG15 mRNA level measured by quantitative polymerase chain reaction is high in vivo in human whole blood cells “in patients with” 28 patients newly diagnosed with SLE compared to with 10 patients with undifferentiated connective tissue disease, and 22 healthy volunteers [28204879].

The abundance of ISG15 measured by ... is induced in vivo in human plasmablasts/Plasma cells “in patients with” active systemic lupus erythematosus compared to ... [27357150].

HBV

The abundance of ISG15 Proteins measured by ... is high in vivo in human HBV-related HCC tissues “in patients with” hepatitis B virus (HBV) compared to non-tumor tissues [26770308].

Step 5f: Edited narrative, based on the standard text (Part B):

“Interferon-stimulated gene 15 (ISG15) is a member of the ubiquitin family, which includes ubiquitin and ubiquitin-like modifiers (Ubls). Ubiquitin and Ubls are involved in the regulation of a variety of cellular activities, including protein stability, intracellular trafficking, cell cycle control and immune modulation. ISG15 has been implicated as a biomarker with diagnostic relevance for a number of human disorders, including cancer (melanoma and breast cancer) and autoimmune diseases (SLE), as well as infection with pathogens such as HBC, HCV and HIV.

The expression of ISG15 has been implicated in a wide range of human cancers, although the roles of ISG15 in tumorigenesis and responses to anticancer treatments remain largely unknown. In patients with breast cancer, ISG15 is overexpressed at both the mRNA and protein levels in mammary tumor tissue compared to that in normal mammary tissue.^{5,6} Increased ISG15 protein expression has also been detected in human melanoma cell lines.⁷ These findings indicate the potential of ISG15 as a tumor biomarker. ISG15 protein expression is upregulated in immune cell markers from patients with invasive breast cancer⁸ as well as dendritic cells from melanoma patients.⁷ ISG15 is also highly expressed in breast cancer patients undergoing immunotherapy.⁹ Short-hairpin RNA-mediated silencing of ISG15 expression has been shown to inhibit breast tumor growth and increase NK cell infiltration a nude mouse model of tumorigenesis.¹⁰ Furthermore, shRNA-mediated knockdown of ISG15 increased the resistance of human tumor cells to CPT.¹¹ These findings indicate that ISG15 is also a candidate biomarker of the responsiveness to immunotherapy among patients with cancer.

ISG15 plays a key role in the host antiviral response. As such, ISG15 is implicated as a diagnostic biomarker of viral infection. Studies in patients with HCC have shown high levels of ISG15 protein in human HBV-related HCC tissues compared to the levels detected in non-tumor tissues.¹² Furthermore, upregulated expression of ISG15 was observed at both the mRNA and protein levels in liver tissue samples from patients with HCV infection compared to the levels detected in uninfected controls.¹³ Similarly, the abundance of ISG15 transcripts was found to be increased in human PBMCs and liver cells in patients with HCV infection who were unresponsive to IFN treatment compared to the levels in corresponding IFN-responsive patients.¹⁴ In addition, high levels of ISG15 in the liver of patients with HCV infection were associated with an unfavourable HCV genotype 1, a high hepatic HCV load and a low antiviral response to IFN compared to patients who did not present such characteristics.¹³ These findings indicate the potential of ISG15 as a biomarker of IFN treatment response in patients with HCV infection.

Increased ISG15 expression has also been reported in mouse models related to HIV infection. ISG15 is upregulated in transgenic mice overexpressing the HIV gp120 protein.¹⁵ Moreover, in mouse models of acute and chronic neuronal injuries with HIV infection, ISG15 protein in the brain was increased compared to the levels detected in mice with global ischemia and traumatic brain injuries.¹⁵ The importance of ISG15 as a biomarker of HIV infection and treatment response has also been revealed in human studies. Compared to healthy donors, TaqMan assays revealed high ISG15 expression (mRNA and protein) as well as an increased frequency of ISG15-SNPs in human PBMC from untreated

patients with HIV infection.¹⁶ Moreover, *ISG15* expression decreased in human PBMC in patients with HIV infection after long-term antiretroviral therapy.¹⁶

The key importance of *ISG15* in the human immune system is also reflected in its role in autoimmune diseases, such as SLE. High levels of *ISG15* transcripts were detected in human whole blood cell samples from 28 patients newly diagnosed with SLE compared to with 10 patients with undifferentiated connective tissue disease, and 22 healthy volunteers,¹⁷ indicating the potential of *ISG15* as a specific biomarker of SLE. Moreover, the upregulation of *ISG15* in human plasmablasts/plasma cells from patients with active SLE indicate that *ISG15* can be used as a marker of disease activation in patients in remission.¹⁸

Thus, an increasing body of evidence supports a role of *ISG15* as a biomarker with diagnostic relevance for a number of human disorders.”

Implementation

As proof of principle, a workshop was implemented using the step-by-step guide and supporting information provided with this paper. It was led by FAA and AKM, who participated in the development of the training curriculum, but had no prior experience running such training activities. The workshop took place on January 26, 2021. In total, 29 graduate students from Hamad Bin Khalifa University took part. It was offered as one three-hour class as part of an “Introduction to Data Science” course. Due to Covid-19 restrictions the workshop was run remotely using WebEx. As no information was collected from workshop participants (i.e., no surveys, or questionnaires), the activity was not considered to constitute human research and therefore no ethical approval was required. The generic introductory presentation in **Extended data File 1**³⁴ was adapted to provide more specific context, and notably explain how, *CCR1*, the target gene for this workshop was selected (for illustrative purposes this introductory presentation is also made available here: **Extended data File 4**).³⁷ FAA and AKM, the two co-instructors, also prepared a handout to guide participants through the different steps (**Extended data File 5**),³⁸ specifically for the *CCR1* use case. Several days prior to the workshop, participants had been asked to create a Prezi “Basic” account and a template was created and shared with each one of them ahead of time. Following the introductory presentation (10 minutes) participants carried out hands-on activities following each one of the tasks described in the handout, with each step corresponding to one working session. Given time constraints, material prepared ahead of time by FAA and AKM was provided to the participants at the end of each working session so that they could move on to the next. Ideally sessions should span multiple days in order for participants to complete the assignments. However, the more intensive schedule implemented here had the advantage of introducing literature profiling concepts and approaches to a large group of trainees in a time-effective manner. The opportunity to continue work later on, on their own time being also available to them.

Conclusions

Effectively harnessing biomedical literature is one of the most fundamental skills required by biomedical researchers. Given the current rate at which articles are published, developing more systematic approaches to literature profiling based on defined principles may prove especially beneficial to early-career biomedical researchers.

Here, we present a training workflow as well as supporting material that may be re-used/adapted for the organization of ‘Gene literature retrieval, profiling and visualization’ training workshops. Hands-on activities range from literature retrieval and optimization of PubMed queries, to the development of interactive resources and authoring original material focusing on a specific topic.

A number of the steps described could easily be automated. For instance, the concatenation of official gene names and aliases for retrieving gene-specific literature using PubMed (Step 1). Indeed, such tools are under development by our group and will be made available in the future. These tools would not only save time, but also minimize user error. Nevertheless, some of the steps, such as optimization of search queries or extraction of information from abstracts will always require some critical evaluation and decision making that cannot be automated. Refraining from the use of such automated tools in a training workshop may also be a deliberate choice on the part of the instructor if the emphasis is initially on the development of competency in completing the steps involved throughout the process.

Here, we focused on gene-associated literature, simply because the training curriculum that is currently under development is based on the reuse of publicly available omics data. A similar approach could be employed to profile the literature associated with any given disease, pathway, molecular process, or drug, for instance.

We aim to further develop an illustrative case that would lead to the publication of a review of the *ISG15* literature employing the literature profiling and visualization approaches described herein. We also plan in the future to develop

and make available guides and supporting material for the implementation of hands-on training workshops focusing on other topics, which include: 1) retrieval, visualization and interpretation of gene-specific transcriptional profiles^{19–21} 2) constitution and curation of themed public dataset collections^{22–30} 3) global analysis of large-scale omics data.^{31–33} These hands-on training activities would build in part on the skills developed through the implementation of the literature profiling workshop that we have presented here.

Data availability

Underlying data

No newly generated data are associated with this article. Information was retrieved via the literature search engine PubMed (<https://pubmed.ncbi.nlm.nih.gov/>).

Extended data

This project contains the following extended data:

Figshare: Literature profiling workshop, introduction session slides (S File 1), <https://doi.org/10.6084/m9.figshare.13669070.v2>.³⁴

Figshare: Literature profiling workshop: steps 1-3 (S File 2), <https://doi.org/10.6084/m9.figshare.14160329.v3>.³⁵

Figshare: Literature Profiling Workshop: Step 5 (S File 3), <https://doi.org/10.6084/m9.figshare.14161484.v1>.³⁶

Figshare: Literature profiling workshop: HBKU handout (S File 4), <https://doi.org/10.6084/m9.figshare.14166395.v1>.³⁷

Figshare: Literature profiling workshop: HBKU intro presentation (S File 5), <https://doi.org/10.6084/m9.figshare.14166500.v1>.³⁸

Data are available under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Acknowledgements

We wish to thank all the students from Hamad Bin Khalifa University (HBKU) for participating in the workshop that served as a proof-of-principle for the use of the implementation guide described in this paper. We also thank the Division of Genomics and Translational Biomedicine at the College of Health and Life Sciences at HBKU for hosting this workshop.

References

- Margolis R, Derr L, Dunn M, *et al.*: **The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data.** *J Am Med Inform Assoc.* 2014 Dec; **21**(6): 957–8. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Van Horn JD, Fierro L, Kamdar J, *et al.*: **Democratizing data science through data science training.** *Pac Symp Biocomput.* 2018; **23**: 292–303. [PubMed Abstract](#) | [Free Full Text](#)
- Garmire LX, Gliske S, Nguyen QC, *et al.*: **The training of next generation data scientists in biomedicine.** *Pac Symp Biocomput.* 2017; **22**: 640–5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chaussabel D, Rinchai D: **Using "collective omics data" for biomedical research training.** *Immunology.* 2018; **155**(1): 18–23. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bektas N, Noetzel E, Veeck J, *et al.*: **The ubiquitin-like molecule interferon-stimulated gene 15 (ISG15) is a potential prognostic marker in human breast cancer.** *Breast Cancer Res.* 2008; **10**(4): R58. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tealco-Cruz AC, Cortés-González CC, Cruz-Ramos E, *et al.*: **Interplay between interferon-stimulated gene 15/ISGylation and interferon gamma signaling in breast cancer cells.** *Cell Signal.* 2019 Feb; **54**: 91–101. [PubMed Abstract](#) | [Publisher Full Text](#)
- Padovan E, Terracciano L, Certa U, *et al.*: **Interferon stimulated gene 15 constitutively produced by melanoma cells induces e-cadherin expression on human dendritic cells.** *Cancer Res.* 2002 Jun 15; **62**(12): 3453–8. [PubMed Abstract](#)
- Kariri YA, Alsalem M, Joseph C, *et al.*: **The prognostic significance of interferon-stimulated gene 15 (ISG15) in invasive breast cancer.** *Breast Cancer Res Treat.* 2020 Oct; **19**. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wood LM, Pan Z-K, Seavey MM, *et al.*: **The ubiquitin-like protein, ISG15, is a novel tumor-associated antigen for cancer immunotherapy.** *Cancer Immunol Immunother.* 2012 May; **61**(5): 689–700. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Burks J, Reed RE, Desai SD: **Free ISG15 triggers an antitumor immune response against breast cancer: a new perspective.** *Oncotarget.* 2015 Mar 30; **6**(9): 7221–31. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Desai SD, Wood LM, Tsai Y-C, *et al.*: **ISG15 as a novel tumor biomarker for drug sensitivity.** *Mol Cancer Ther.* 2008 Jun; **7**(6): 1430–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

12. Qiu X, Hong Y, Yang D, *et al.*: **ISG15 as a novel prognostic biomarker for hepatitis B virus-related hepatocellular carcinoma.** *Int J Clin Exp Med.* 2015; **8**(10): 17140–50.
[PubMed Abstract](#) | [Free Full Text](#)
13. Broering R, Trippler M, Werner M, *et al.*: **Hepatic expression of proteasome subunit alpha type-6 is upregulated during viral hepatitis and putatively regulates the expression of ISG15 ubiquitin-like modifier, a proviral host gene in hepatitis C virus infection.** *J Viral Hepat.* 2016 May; **23**(5): 375–86.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Katsounas A, Hubbard JJ, Wang CH, *et al.*: **High interferon-stimulated gene ISG-15 expression affects HCV treatment outcome in patients co-infected with HIV and HCV.** *J Med Virol.* 2013 Jun; **85**(6): 959–63.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Wang R-G, Kaul M, Zhang D-X: **Interferon-stimulated gene 15 as a general marker for acute and chronic neuronal injuries.** *Sheng Li Xue Bao.* 2012 Oct 25; **64**(5): 577–83.
[PubMed Abstract](#) | [Free Full Text](#)
16. Scagnolari C, Monteleone K, Selvaggi C, *et al.*: **ISG15 expression correlates with HIV-1 viral load and with factors regulating T cell response.** *Immunobiology.* 2016 Feb; **221**(2): 282–90.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Yuan Y, Ma H, Ye Z, *et al.*: **Interferon-stimulated gene 15 expression in systemic lupus erythematosus: Diagnostic value and association with lymphocytopenia.** *Z Rheumatol.* 2018 Apr; **77**(3): 256–62.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Care MA, Stephenson SJ, Barnes NA, *et al.*: **Network Analysis Identifies Proinflammatory Plasma Cell Polarization for Secretion of ISG15 in Human Autoimmunity.** *J Immunol.* 2016 Aug 15; **197**(4): 1447–59.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Toufiq M, Roelands J, Alfaki M, *et al.*: **Annexin A3 in sepsis: novel perspectives from an exploration of public transcriptome data.** *Immunology.* 2020 Jul; **18**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Rinchai D, Kewcharoenwong C, Kessler B, *et al.*: **Increased abundance of ADAM9 transcripts in the blood is associated with tissue damage.** *F1000Res.* 2015; **4**: 89.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Roelands J, Garand M, Hinchcliff E, *et al.*: **Long-Chain Acyl-CoA Synthetase 1 Role in Sepsis and Immunity: Perspectives From a Parallel Review of Public Transcriptome Datasets and of the Literature.** *Front Immunol.* 2019; **10**: 2410.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Rinchai D, Boughorbel S, Presnell S, *et al.*: **A curated compendium of monocyte transcriptome datasets of relevance to human monocyte immunobiology research.** *F1000Res.* 2016; **5**: 291.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Roelands J, Decock J, Boughorbel S, *et al.*: **A collection of annotated and harmonized human breast cancer transcriptome datasets, including immunologic classification.** *F1000Res.* 2017; **6**: 296.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Huang SSY, Al Ali F, Boughorbel S, *et al.*: **A curated collection of transcriptome datasets to investigate the molecular mechanisms of immunoglobulin E-mediated atopic diseases.** *Database (Oxford).* 2019 01; **2019**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Bougarn S, Boughorbel S, Chaussabel D, *et al.*: **A curated transcriptome dataset collection to investigate the blood transcriptional response to viral respiratory tract infection and vaccination.** *F1000Res.* 2019; **8**: 284.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. 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.** *F1000Res.* 2016; **5**: 305.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Mackeh R, Boughorbel S, Chaussabel D, *et al.*: **-A curated transcriptomic dataset collection relevant to embryonic development associated with in vitro fertilization in healthy individuals and patients with polycystic ovary syndrome.** *F1000Res.* 2017; **6**: 181.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Blazkova J, Boughorbel S, Presnell S, *et al.*: **A curated transcriptome dataset collection to investigate the immunobiology of HIV infection.** *F1000Res.* 2016; **5**: 327.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Bougarn S, Boughorbel S, Chaussabel D, *et al.*: **A curated transcriptome dataset collection to investigate inborn errors of immunity.** *F1000Res.* 2019; **8**: 188.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. 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.** *F1000Res.* 2016; **5**: 414.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Rinchai D, Presnell S, Vidal M, *et al.*: **Blood Interferon Signatures Putatively Link Lack of Protection Conferred by the RTS,S Recombinant Malaria Vaccine to an Antigen-specific IgE Response.** *F1000Res.* 2015; **4**: 919.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Rawat A, Rinchai D, Toufiq M, *et al.*: **A Neutrophil-Driven Inflammatory Signature Characterizes the Blood Transcriptome Fingerprint of Psoriasis.** *Front Immunol.* 2020 [cited 2020 Dec 7]; **11**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Reference Source](#)
33. Rinchai D, Altman MC, Konza O, *et al.*: **Definition of erythroid cell-positive blood transcriptome phenotypes associated with severe respiratory syncytial virus infection.** *Clin Transl Med.* 2020; **10**(8): e244.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Chaussabel D: **Literature profiling workshop, introduction session slides (S File 1).** *figshare. Journal contribution.* 2021.
[Publisher Full Text](#)
35. Chaussabel D: **Literature profiling workshop: steps 1-3 (S File 2).** *figshare. Journal contribution.* 2021.
[Publisher Full Text](#)
36. Chaussabel D: **Literature Profiling Workshop: Step 5 (S File 3).** *figshare. Journal contribution.* 2021.
[Publisher Full Text](#)
37. Chaussabel D: **Literature profiling workshop: HBKU handout (S File 4).** *figshare. Journal contribution.* 2021.
[Publisher Full Text](#)
38. Chaussabel D: **Literature profiling workshop: HBKU intro presentation (S File 5).** *figshare. Journal contribution.* 2021.
[Publisher Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research