

6:43 PM DEC 01, 2022

THE MOMENT YOUR GUT INSTINCT
WAS BACKED BY SCIENTIFIC
RESULTS_



THE DIFFERENCE OF BREAKTHROUGH MOMENTS

WITH COMPLETE SOLUTIONS FOR GROUNDBREAKING DISCOVERIES FROM A TRUSTED PARTNER.

Your next breakthrough could be closer than you imagine, especially with the right resources to help you advance your research. At BD, we are dedicated to helping you get the data you need, when, where and how you need it. Our integrated solutions in instrumentation, software and reagents are optimized to work together to bring you closer to your next breakthrough moment. And you can depend on us for world-class training, service and support to help you get the most from the results your research depends on. Discover a range of optimized solutions that open endless possibilities for your future research. **Discover the new BD.**

Learn how you can advance your research >



FlowRepository: A Resource of Annotated Flow Cytometry Datasets Associated with Peer-Reviewed Publications

DATA associated with peer-reviewed manuscripts should be easily available and accessible; however, the rapid expansion of flow cytometry (FCM) applications has outpaced the development of tools for storage, analysis, and data representation (1–3). In addition, data associated with peer-reviewed manuscripts are rarely available publicly and even then are usually not stored with explicit links to experimental metadata, such as data analyses procedures, experimental conditions used, or information about the samples processed. This link between data and metadata is crucial as it facilitates the understanding of analysis approaches and reproducible research. Having datasets linked to figures and summaries through a detailed explanation of the sample processing and analysis pipeline allows other scientists to ask additional questions and build upon the published findings.

MIFLOWCYT

To efficiently exchange information, the community must first agree on the kind of information they want to store, although the precise content of documentation will vary by scientific area, study design, the type of data collected, and characteristics of the dataset. In flow cytometry, this need is addressed by the Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) (4), which is accepted as a standard by the International Society for Advancement of Cytometry and several journals (e.g., *Cytometry A* and the *Nature Publishing Group* journals). However, providing the information specified by MIFlowCyt, including the raw data (FCS files), is often impractical within the context of a peer-reviewed manuscript, and searching and retrieving data files of interest across various journals' supplementary information can be cumbersome.

Scientific communities in the fields of microarray, proteomics, and sequencing provide public repositories for pub-

lished data and are benefiting from reuse and re-exploration of existing data to test new or alternative hypotheses and methods of analysis. Until now, no such repository existed for flow cytometry, constraining open scientific inquiry and progress in the field. To address these needs, we developed a public resource (FlowRepository.org) for authors to deposit their FCM data, provide the MIFlowCyt information, and share annotated datasets upon publication.

FLOW REPOSITORY

We developed FlowRepository by extending and adapting CytoBank (5), an online tool for storage and collaborative analysis of cytometric data. FlowRepository uses the object-oriented Ruby programming language with the Ruby on Rails application framework. In addition, the Java implementation of the Ruby language (i.e., JRuby) allows for the seamless integration with Java code used for some of FlowRepository components. The user interface was encoded by the HTML Abstraction Markup Language (HAML) and the resulting HTML is supported by series of Javascript functions and Java applets. Its functionality is powered by JQuery libraries, which allow for the integration of components and features otherwise very difficult to provide within a web browser environment. In addition, the Asynchronous JavaScript and XML (AJAX) is used to exchange data with the server without the need of reloading full web pages. This technology allows FlowRepository to implement interactivity and various auto-complete functionality.

The amount of annotation required to annotate datasets so that they can be understood by third parties represents a challenge for data submitters. MIFlowCyt formalizes these requirements, for example, each sample needs to be described for each submitted FCS data file. This includes descriptions of the sample source, characteristics, treatment, and fluorescence

Additional Supporting Information may be found in the online version of this article.

Received 15 May 2012; Revision Received 5 June 2012; Accepted 27 June 2012

Grant sponsor: Michael Smith Foundation (Health Research Scholar Award [R.B]).

*Correspondence to: Ryan R. Brinkman, Terry Fox Laboratory, BC Cancer Research Centre 675 West 10th Avenue, Vancouver, BC V5Z 1L3, Canada.

E-mail: rbrinkman@bccrc.ca

Published online 4 August 2012 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.22106

© 2012 International Society for Advancement of Cytometry

1 - Experiment Overview - 40.00% provided i

Items considered relatively based on importance, 30% contribution to total score.

Item	Compliance [+ -]	Improve
Purpose	✖ Purpose description does not seem sufficient	Edit description
Keywords	✔ Not enough keywords provided	Edit description
Experiment variables	✖ Missing experiment variables	Edit experiment variables
Organization	✔ Provided organization name Provided organization address	Fully provided
Contact	✔ First name of primary researcher provided Last name of primary researcher provided Email of primary researcher provided	Fully provided
Date	✔ Provided experiment starting date Missing experiment end date	Edit description
Conclusions	✖ Missing conclusions	Edit description
Quality control measures	✖ Missing quality control description Missing quality control experiment	Edit description

Figure 1. Example of the first section of a MIFlowCyt score report showing details about provided versus missing information and providing direct links for annotation improvements. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

reagents. Each sample source also needs to be described by providing (as applicable) a description, taxonomy, age, gender, phenotype, genotype, treatment, and other relevant organism information. Details such as characteristics being measured, analyte, analyte detector, analyte reporter (fluorochrome), clone name or number, manufacturer name, catalog number, and other relevant information shall be provided for each fluorescence reagent for each sample. While experts agree that all this information is essential to interpret FCM datasets and experimental findings, providing this information requires a significant amount of time and effort, which may be discouraging for some data submitters, especially those with very large datasets or complicated experimental designs. Moreover, providing this information in a structured way requires typically more effort than a free text description as commonly present in the methods section of a manuscript. Consequently, the design of a user interface allowing for the input of all required annotations with minimal effort proved to be the biggest challenge when implementing FlowRepository. We tackled this issue by separating the annotation process into two steps: creation of annotations (e.g., in the form of templates) and annotation application with sample-specific adjustments. This approach simplifies dataset annotating as several aspects are

commonly repeated for multiple samples (e.g., same set of reagents, same treatment). Additionally, FlowRepository can also extract some annotation from spreadsheets (e.g., MS Excel) created by users locally on their computers.

Detailed specification of acquisition instrumentation configuration may also be cumbersome, especially if customized instrumentation is used. We have attempted to minimize the amount of effort required by implementing a rich user interface where users can apply *Drag & Drop* techniques to build and configure optical paths and reuse existing components (e.g., light sources, optical filters, or detectors). Currently, we are assessing options to further simplify the instrumentation description based on common instrumentation settings of widely used instrument models.

The quality of the provided annotation is automatically assessed by the server and provided in the form of a *MIFlowCyt Score* that approximates the level of compliance with the MIFlowCyt standard as a percentage. The release of a dataset is not restricted based on any specific cutoff; however, users can easily identify datasets that match their desired thresholds, and the system also helps reviewers decide whether the annotation is sufficient for publication. As shown in the example in Figure 1, overview tables are displaying required

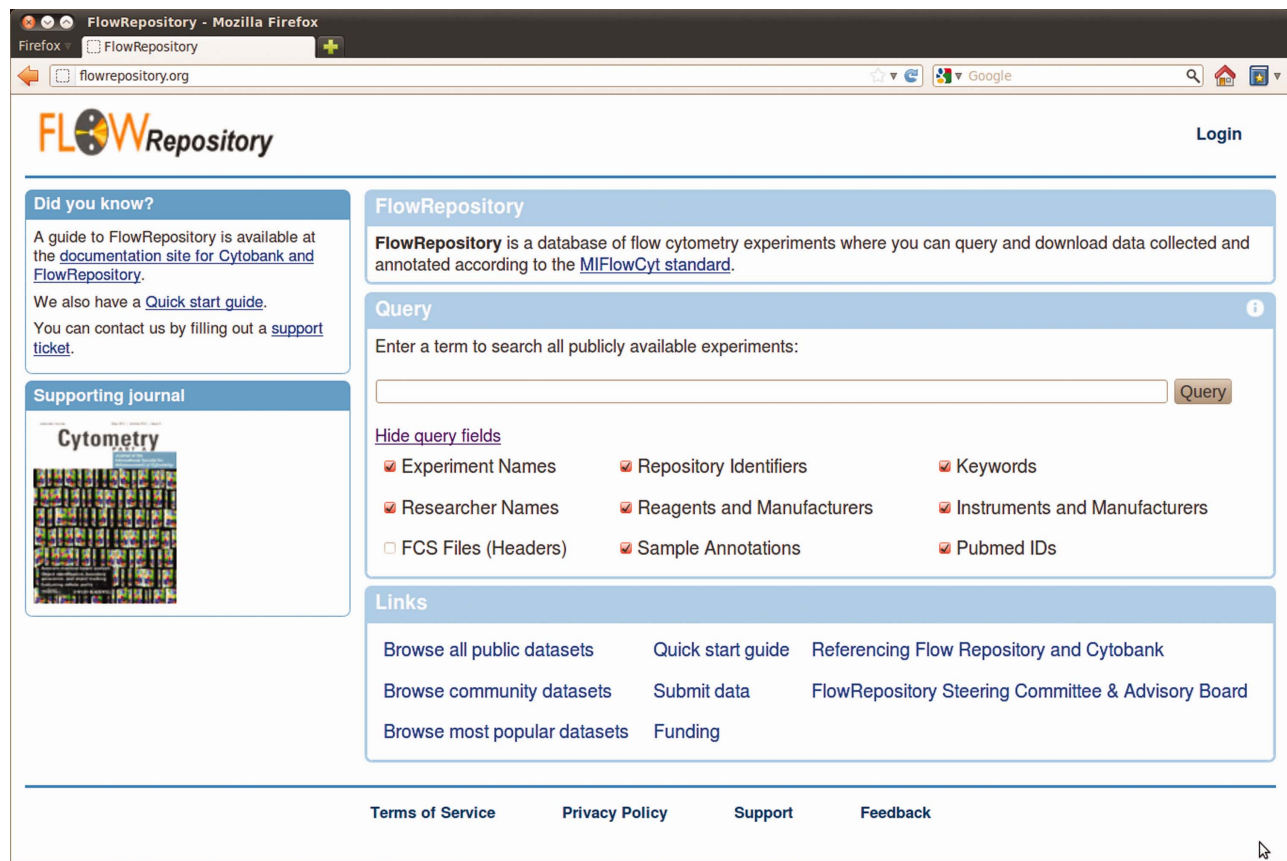


Figure 2. Landing web page of FlowRepository that offers mainly querying or browsing of available datasets. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

items, annotation compliance levels, and direct links to pages that may be used to improve the score by providing additional details. Supporting Information File 1 shows a screen shot of the form that is displayed when users follow the *Edit description* link.

FlowRepository source code is available under the Affero General Public License (6) using the Git version control system accessible through GitHub (<https://github.com/>). The Affero license is an Open Source Initiative compatible software license based on the GNU General Public License copyright from the Free Software Foundation, Inc.

Getting Started

To access the application, users need a computer with Windows (XP, Vista, or Windows 7), Mac OS X (10.3 or above), or Linux (Ubuntu, OpenSUSE, Fedora, or Red Hat), an Internet connection, a web browser, and Java version 1.5 or higher installed. Most data processing is performed on the server with only scaled-down images and results transported to the client. Therefore, even slower Internet connection speed is typically not a problem for most use cases while working with FlowRepository. Large data upload or the download of very large raw data files may be a bottleneck if the Internet connection is extremely slow. The latest version of the Firefox web browser is recommended, but other major web browsers,

such as Apple Safari or Google Chrome, or even using other types of devices (e.g., iPhones, iPads, Android phones and tablets) will work; however, the functionality of the website may be limited and/or visual appearance impaired. In this manuscript, we provide an overview of the main FlowRepository features. For further details, we encourage readers to consult the guide detailing the preparation of a MIFlowCyt compliant dataset annotation using FlowRepository (7), or the online documentation and the online *Quick Start Guide* that are available as part of FlowRepository. In addition, since FlowRepository is built on top of Cytobank, most of Cytobank User's Guide (5) and documentation is also applicable to FlowRepository.

A *Public View* is automatically accessed when connecting to FlowRepository's main web page. As shown in Figure 2, this page offers mostly browsing and searching for datasets (experiments) of interest. These may be located based on keywords, researcher names, sample descriptions, and other experiment annotations. In addition, each experiment can be identified and referenced by a unique repository identifier. These identifiers are typically in the form of FR-FCM-xxxx, where xxxx is a sequence of alphanumeric characters. A Uniform Resource Locator (URL) accessing directly a specified dataset may be created by attaching /id/FR-FCM-xxxx to FlowRepository's main URL. For example, <https://flowreposit>

tory.org/id/FR-FCM-ZZZ3 will navigate the browser directly to a dataset related to identification of B cells through negative gating (8). At the dataset description page, users can review all the annotations including experiment overview, flow sample/specimen details (specifically for each uploaded data file), data acquisition (instrumentation) details, as well as analysis details and illustrations attached to the experiment. Files associated with the experiment may be downloaded to users local computer. Besides FCS data files, download options typically include illustrations provided by the authors and attachments (e.g., MS Word and Excel documents, FlowJo workspaces). Various types of PDF reports can also be generated. In addition, gates created online in FlowRepository may be exported in the Gating-ML (9) file format.

While anonymous read-only access is supported, we encourage users to register to gain full access to FlowRepository. Only registered users can upload their datasets and make these accessible to general public or reviewers prior to publication. In order to register, users need an *OpenID* account, freely provided by services such as Google, Yahoo, and others.

To bank a new experiment, users need to provide basic details such as name, primary researcher, and a brief description of the purpose of the experiment. While this information is sufficient to upload a dataset, it is recommended (and required by MIFlowCyt) that users provide additional details consisting of experiment dates, conclusion, quality control measures, specific keywords, and organization contact details. Users can also reference related publications if these exist already. Related publications may also be linked to the experiment later on, which is commonly the case since data is typically shared (at least with the reviewers) prior to the publication of the related manuscript.

To annotate a dataset, users should provide as much of the MIFlowCyt required details as possible. Typically, users may want to navigate to the *annotation data* section and create templates that will be further applied on their samples. For large experiments, it has been shown useful to apply the common part of the annotation to all samples (or samples from a single panel) and then adjust the actual sample-specific details by extracting these automatically from an attached spreadsheet (7).

Data analysis details are part of the required dataset annotations as specified by MIFlowCyt. There are two options of providing these to FlowRepository. Researchers may either use the online analytical capabilities as described in Cytobank User's Guide (5) and share dynamically created tables, illustrations, and statistics or they may upload images and relevant project files (workspaces, experiment files, etc.) from third party software tools if these were used to analyze their data. Currently, FlowRepository recognizes files from FlowJo[®], FCS Express[®] and BD FACSDiva[®]; however, images and third party project files are only stored as attachments and considered as additional supporting information that cannot be used to reproduce the analysis online. In these cases, users may download the project files along with the dataset and review the analysis locally on their computers provided that they have the required analysis tool. Consequently, if users are perform-

ing data analysis locally, they should upload and share supporting files (e.g., spreadsheets with statistics) that will allow others to review the original analysis without the need of obtaining additional software.

The MIFlowCyt score of an experiment is displayed in the top right corner of the *Experiment Overview* page. The bar changes its background color from red to orange and finally to green depending on the level of compliance with MIFlowCyt. Users can review the details of their compliance by clicking on the *MIFlowCyt score* bar. Where applicable, direct links are provided to assist users in increasing their score by providing the missing information.

Sharing Datasets

Experiments in FlowRepository can be either public or private. However, they can only be kept private for up to 1 year after the first creation of the experiment or until an associated manuscript referencing the experiment data is published. This requirement has been introduced to avoid usage of FlowRepository as a private data store. A list of "old" unpublished datasets is available to FlowRepository administrators, who contact the data owners with a request to resolve the situation. Dataset owners may choose to delete their unpublished datasets at any time. Data from public experiments are available to all users without the need of authentication. Private experiments are accessible to their authors, members of the authors' team as explicitly specified by the authors, and FlowRepository web site administrators. By default, new experiments are considered as private. Authors may choose to make their experiments public by clicking on the *Share with Everyone* button in the *Sharing Permissions* panel on the *Experiment Overview* page. Alternatively, using the *Share with a User* field, authors may specify a list of FlowRepository users who shall gain full access to the experiment. These users may change annotations, perform their own analysis, and create their own figures. In addition, authors may choose to make private experiments available to journal reviewers by clicking on the *Share with Reviewers* button. This generates a secret access code and a secret URL that can be used to anonymously access the experiment. Typically, authors would share this information with the editor (e.g., in their cover letter) when submitting a related manuscript for peer review. The editor can pass this information on to designated reviewers, who can use it to access the particular dataset. Reviewers can review the dataset with associated annotations but they do not have write access to the experiment.

In Conclusion

Transparency and public availability of protocols, data, analyses, and results are crucial to make sense of the complex biology of human diseases, and research efforts should be integrated across teams in an open environment (10). Private funding agencies, regulatory agencies, publishers, and the scientific community have all recognized the importance of protecting cumulative data outputs to accelerate subsequent exploitation through the community-based development of public data repositories (11). In a recent re-evaluation of 18

peer-reviewed *Nature Genetics* microarray articles, the inability of researchers to reproduce analyses was directly linked to data unavailability, incomplete data annotation, or specification of data processing and analysis (12). Data sharing allows scientists to expedite the translation of research results into knowledge, products, and procedures to improve human health. It reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis. It also supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new datasets when data from multiple sources are combined. However, in flow cytometry, researchers faced technical difficulties when attempting to share their datasets, as required by many granting agencies and journals.

FlowRepository is based on open source source code donated under the Affero General Public License by Cytobank Inc. The original source code has been extended and adapted extensively, mainly to incorporate MIFlowCyt. FlowRepository is independently managed by the FlowRepository Steering Committee with representation from the International Society for Advancement of Cytometry (ISAC), International Clinical Cytometry Society and the flow cytometry community. This Committee makes strategic and technical decisions regarding all aspects of the policies and operations of FlowRepository that do not require expenditures and makes recommendations regarding financial issues to ISAC and other groups that may be involved in financially supporting the repository. A larger Advisory Committee is called upon as needed to provide advice to the Steering Committee on technical issues, regional considerations, and publisher interfaces.

The endorsement of MIFlowCyt as the community standard for the annotation of FCM datasets provided the first and necessary step toward addressing FCM data sharing issues. It filled the gap in biomedical data reporting standards by outlining the information required to interpret FCM experiments, understand the conclusions reached, and make comparisons to experiments performed by different laboratories. However, providing raw data linked to MIFlowCyt annotations was still cumbersome and impractical within the context of peer-reviewed manuscripts. This issue is now addressed by FlowRepository—a public online resource for researchers to deposit FCM data so that these are easily available to others. In addition, FlowRepository's annotation capabilities and MIFlowCyt support finally allow users to properly annotate their data making these not just available but also understandable to others. As it is the case with microarray, proteomics, and other types of datasets, we believe that underlying FCM data will now become routinely available along with manuscripts publishing FCM-based findings, to the benefit of the entire flow cytometry community.

ACKNOWLEDGMENTS

Development of FlowRepository was supported by the International Society for Advancement of Cytometry, the Wallace H. Coulter Foundation, the Terry Fox Foundation, and the Terry Fox Research Institute. Cytobank Incorporated donated source code for the underlying technological platform and assisted in the initial implementation. N.K. is one of the founders of Cytobank, Inc. C.R. is an employee of Cytobank, Inc.

Josef Spidlen

Terry Fox Laboratory, BC Cancer Agency
Vancouver, British Columbia, Canada

Karin Breuer

Department of Molecular Biology and
Biochemistry, Simon Fraser University
Burnaby, British Columbia, Canada

Chad Rosenberg, Nikesh Kotecha

Cytobank Inc., Mountain View, California

Ryan R. Brinkman*

Terry Fox Laboratory, BC Cancer Agency
Vancouver, British Columbia, Canada
Department of Medical Genetics
University of British Columbia, Vancouver
British Columbia, Canada

LITERATURE CITED

1. Krutzik PO, Irish JM, Nolan GP, Perez OD. Analysis of protein phosphorylation and cellular signaling events by flow cytometry: Techniques and clinical applications. *Clin Immunol* 2004;110:206–221.
2. Krutzik PO, Crane JM, Clutter MR, Nolan GP. High-content single-cell drug screening with phosphospecific flow cytometry. *Nat Chem Biol* 2008;4:132–142.
3. Irish JM, Kotecha N, Nolan GP. Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nat Rev Cancer* 2006;6:146–155.
4. Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, Furlong J, Gasparetto M, Goldberg M, Goralczyk EM, Hyun B, Jansen K, Kollmann T, Kong M, Leif R, McWeeney S, Moloshok TD, Moore W, Nolan G, Nolan J, Nikolich-Zugich J, Parrish D, Purcell B, Qian Y, Selvaraj B, Smith C, Tchuvatkina O, Wertheimer A, Wilkinson P, Wilson C, Wood J, Zigon R, for Advancement of Cytometry Data Standards Task Force IS, Scheuermann RH, Brinkman RR. MIFlowCyt: the Minimum Information about a Flow Cytometry Experiment. *Cytometry Part A* 2008;73:926–930.
5. Kotecha N, Krutzik PO, Irish JM. Web-based analysis and publication of flow cytometry experiments. *Curr Protoc Cytom*, 2010; Chapter 10: Unit 10.17.
6. Free Software Foundation, Inc. GNU Affero General Public License, version 3.0 2007. Available at <http://www.gnu.org/licenses/agpl-3.0.html>, 2007. Accessed on July 2012.
7. Spidlen J, Breuer K, Brinkman R. Preparing a Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) compliant manuscript using the International Society for Advancement of Cytometry (ISAC) FCS File Repository (FlowRepository.org). *Curr Protoc in Cytom* 2012; Chapter 10:Unit 10.18.
8. Blimkie D, Fortunato ES 3rd, Thommai F, Xu L, Fernandes E, Crabtree J, Rein-Weston A, Jansen K, Brinkman RR, Kollmann TR. Identification of B cells through negative gating—An example of the MIFlowCyt standard applied. *Cytometry A* 2010;77:546–551.
9. Spidlen J, Leif RC, Moore W, Roederer M, Brinkman RR, the International Society for the Advancement of Cytometry Data Standards Task Force. Gating-ML: XML-based gating descriptions in flow cytometry. *Cytometry A* 2008;73:1151–1157.
10. Ioannidis JP. Molecular evidence-based medicine: evolution and integration of information in the genomic era. *Eur J Clin Invest* 2007;37:340–349.
11. Field D, Sansone SA, Collis A, Booth T, Dukes P, Gregurick SK, Kennedy K, Kolar P, Kolker E, Maxon M, Millard S, Mugabushaka AM, Perrin N, Remacle JE, Remington K, Rocca-Serra P, Taylor CF, Thorley M, Tiwari B, Wilbanks J. Omics data sharing. *Science* 2009;326:234–236.
12. Ioannidis JP, Allison DB, Ball CA, Coulbaly I, Cui X, Culhane AC, Falchi M, Furlanello C, Game L, Jurman G, Mangion J, Mehta T, Nitzberg M, Page GP, Petretto E, van Noort V. Repeatability of published microarray gene expression analyses. *Nat Genet* 2009;41:149–155.