





DATA NOTE

CFM: a database of experimentally validated protocols for chemical compound-based direct reprogramming and transdifferentiation [version 1; peer review: 2 approved with reservations]

Alexey Sizykh ¹, Khalimat Murtazalieva^{1,2}, Yulia Vyshkvorkina¹, Alexey Stupnikov ¹, Yulia A. Medvedeva^{1,3}

¹Moscow Institute of Physics and Technology, Dolgoprudny, Moscow region, 141701, Russian Federation

²Vavilov Institute of General Genetics, Moscow, 117971, Russian Federation

³Center for Biotechnology of Russian Academy of Sciences, Moscow, 119071, Russian Federation

V1 First published: 16 Apr 2021, 10:295
<https://doi.org/10.12688/f1000research.28439.1>
 Latest published: 16 Apr 2021, 10:295
<https://doi.org/10.12688/f1000research.28439.1>

Abstract



Cell fate engineering technologies are critically important for basic and applied science, yet many protocols for direct cell conversions are still unstable, have a low yield and require improvement. There is an increasing need for a data aggregator containing a structured collection of protocols - preprocessed, verified, and represented in a standardized manner to facilitate their comparison, and providing a platform for the researchers to evaluate and improve the protocols. We developed CFM (cell fate mastering), a database of experimentally validated protocols for chemical compound-based direct reprogramming and direct cell conversion. The current version of CFM contains 169 distinct protocols, 113 types of cell conversions, and 158 small molecules capable of inducing cell conversion. CFM allows stem cell biologists to compare and choose the best protocol with high efficiency and reliability for their needs. The protocol representation contains PubChem CIDs and Mechanisms Of Action (MOA) for chemicals, protocol duration, media, and yield with a comment on a measurement strategy. Ratings of the protocols and feedback from the community will help to promote high-quality and reproducible protocols. We are committed to a long-term database maintenance strategy. The database is currently available at <https://cfm.mipt.ru>


Keywords

Cell Fate Engineering, Chemical reprogramming, Direct cell conversion, Transdifferentiation

Open Peer Review

Approval Status ? ?

	1	2
version 1		
16 Apr 2021	view	view

1. **Erdem B. Dashinimaev** , Pirogov Russian National Research Medical University, Moscow, Russian Federation
2. **Stephanie M. Willerth** , University of Victoria, Victoria, Canada

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



This article is included in the **Cheminformatics** gateway.



This article is included in the **Bioinformatics** gateway.

Corresponding author: Yulia A. Medvedeva (medvedeva.ua@mipt.edu)

Author roles: **Sizykh A:** Conceptualization, Data Curation, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Murtazalieva K:** Conceptualization, Data Curation, Investigation, Methodology, Resources, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Vyshkvorkina Y:** Resources; **Stupnikov A:** Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration; **Medvedeva YA:** Funding Acquisition, Project Administration, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The work was supported by the Ministry of Science and Higher Education of the Russian Federation. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2021 Sizykh A *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: Sizykh A, Murtazalieva K, Vyshkvorkina Y *et al.* **CFM: a database of experimentally validated protocols for chemical compound-based direct reprogramming and transdifferentiation [version 1; peer review: 2 approved with reservations]** F1000Research 2021, 10:295 <https://doi.org/10.12688/f1000research.28439.1>

First published: 16 Apr 2021, 10:295 <https://doi.org/10.12688/f1000research.28439.1>

Introduction

Cell engineering technologies possess a tremendous potential for basic and applied science. Recent discoveries in this field pushed forward regenerative and personalized medicine, drug discovery, and toxicology. In the future, cell engineering technologies can enable regeneration of altered tissues and extend the human lifespan. Now they serve as a platform for drug testing and toxicology experiments. Furthermore, these approaches enable molecular mechanisms to be revealed that drive normal development and differentiation. Transcription factors (TFs) have been used to manipulate cell fate in the majority of published protocols.¹ Low efficiency and safety concerns restrict the application of cell fate engineering approaches in clinical practice.² Small molecules that target signaling pathways or regulate the epigenetic state offer powerful tools for refined manipulation of cell fate to the desired outcome.³ There is an increasing number of protocols that utilize small molecules in order to induce lineage differentiation or to facilitate transdifferentiation by increasing efficiency or by replacing genetic reprogramming factors.

Small-molecule-mediated approaches have more potential for clinical applications since:³

1. The biological effects of small molecules are typically rapid, reversible, dose-dependent and predictable, allowing precise control over specific outcomes and precise resembling of target cell type;
2. Compared with genetic interventions, the relative ease of handling and administration of small molecules make them more practical for further therapeutic development;
3. Small molecules do not evoke safety concerns about possible genome integration in contradistinction to genetic interventions.

Chemical cell conversion protocols vary not only in chemicals used for reprogramming but also in media, making it difficult to choose the optimal protocol in practice. A database integrating cellular reprogramming protocol has been already published.⁴ Yet, the protocol representation is not standardized and the database has not been updated recently. Here emerges the need for aggregation of protocols in a standardized form that would contain all relevant and specific information from them. Also, aggregators should provide user-friendly access to enable expeditious retrieval of information.

In this work, we present a protocol aggregator, CFM (cell fate mastering), with a user-friendly interface that allows stem cell biologists to compare protocols, select the best one for their needs with high efficiency and reliability as well as leave feedback for their experience regarding the protocol for the research community. Additionally, we provide information about a potential mechanism of drug action and pathways involved in cell conversion. Data on molecular mechanisms affected by specific compounds could help to improve the protocol. We use expert opinions to develop an original standard for protocol representation, provide public ratings for protocols calculated based on users' evaluation, and maintain a discussion forum. Protocol ratings as well as detailed feedback from the community will help to promote high-quality and reproducible protocols. Standardized and detailed protocol representation will facilitate development and validation of systematic computational approaches for cell fate engineering (e.g. DECODE⁵). The current version of CFM contains 169 distinct protocols, 113 types of cell conversions, and 158 small molecules. Summary statistics are available in [Figure 1](#).

We are committed to a strong database maintenance strategy, which means:

1. We manually check data in order to provide comprehensive information. Public rating is conducive to promoting reproducible protocols;
2. We communicate with experimental biologists and permanently improve and extend the functionality of the database to meet their needs. Hence, we created a special feedback form where users can inform us what features they would like to add;
3. We support protocol updating by users (after manual curation of the suggested protocol);
4. We also plan to update the content of the database on a regular basis (at least bi-annually).

Methods

Implementation

We collected protocols from PubMed and Google Scholar using the keywords 'direct reprogramming' AND 'chemicals OR small molecules', 'transdifferentiation' AND 'chemicals OR small molecules', 'direct cell conversion' AND 'chemicals OR small molecules'. In this way, more than 1000 papers were obtained. Experts prefiltered papers based

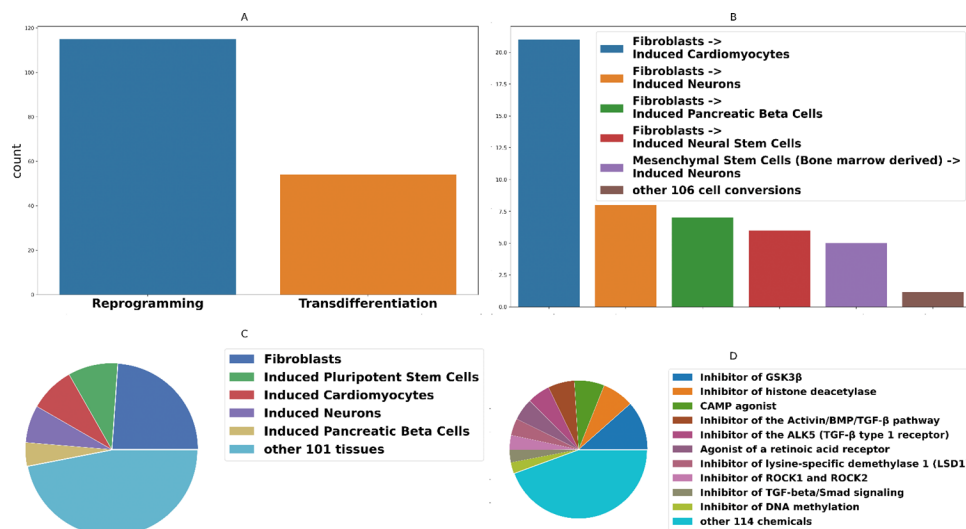


Figure 1. Overview of the database contents. (A) Number of types of cell conversions. (B) Number of cell type conversion pairs in CFM (the bar for “other 106 cell conversions” represents an averaged number of a protocols for each other cell conversions type). (C) Frequencies of chemical compounds used in the protocols. (D) Frequencies of initial cell types used in protocols.

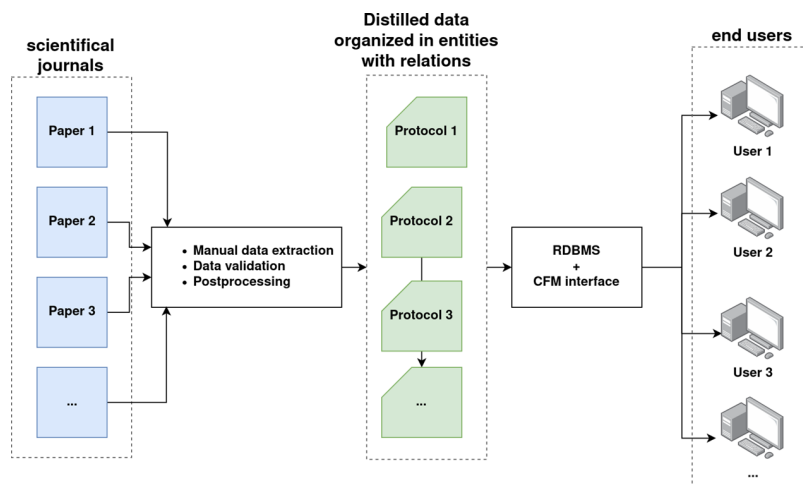


Figure 2. Data processing workflow diagram.

on the content of key sections (abstract, introduction, method) discarding irrelevant papers. We manually extracted information from each relevant article and added data about small molecules implicated in protocol from PubChem as a part of postprocessing. The overview of the whole workflow can be seen in Figure 2.

The database can be queried at cfm.mipt.ru/query page (Figure 3). By default, all protocols are listed. If no protocols match a query, the whole database is displayed (default view). Search is case-independent. The first column contains a link to protocol description in the cfm.mipt.ru/viewProtocol/<protocolId> format, where protocol ID is an inner identifier. The protocol description page contains article information (DOI, article title, and authors), source and target cell lines, initial cell culture description, the species from which the source cell line was obtained, the total protocol duration, the media in which the protocol was implemented, protocol yield (with comments on how it was calculated), chemicals (with their methods of effect is known), transcription factors, stress factors and growth factors used during the procedure and some comments on the protocol in general. Also, we provide a simple interface for rating protocols for registered participants. After curation, the rating will be available on the protocol page. All rated protocols are shown in a personal user account (cfm.mipt.ru/login). A researcher can add a protocol (and a published related paper) to the CFM database by submitting it through cfm.mipt.ru/add after the email is verified by a CFM administrator. Detailed information about verification can be found on the Add Protocol page.

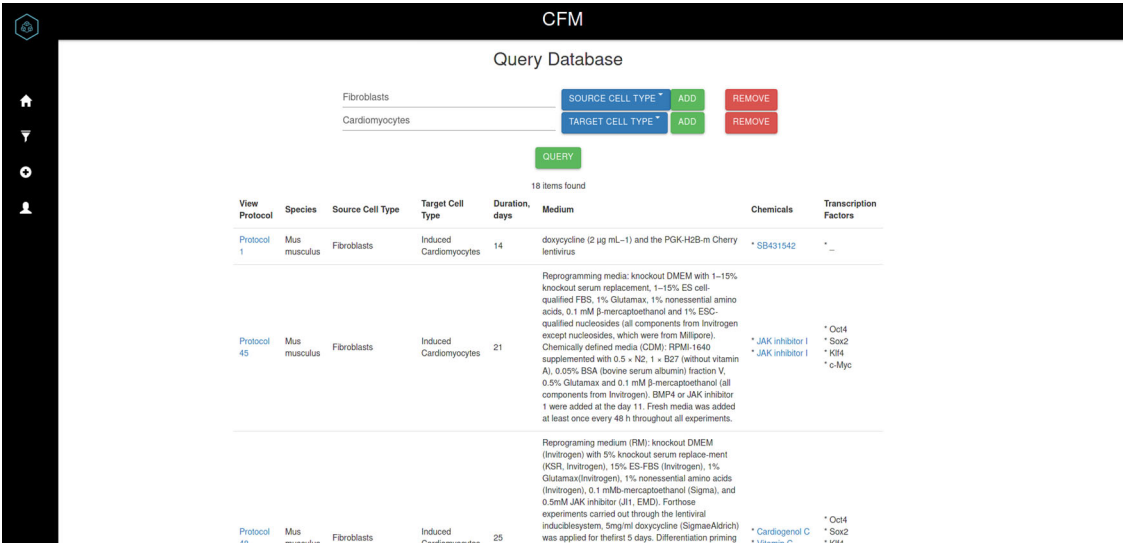


Figure 3. Database interface screenshot.

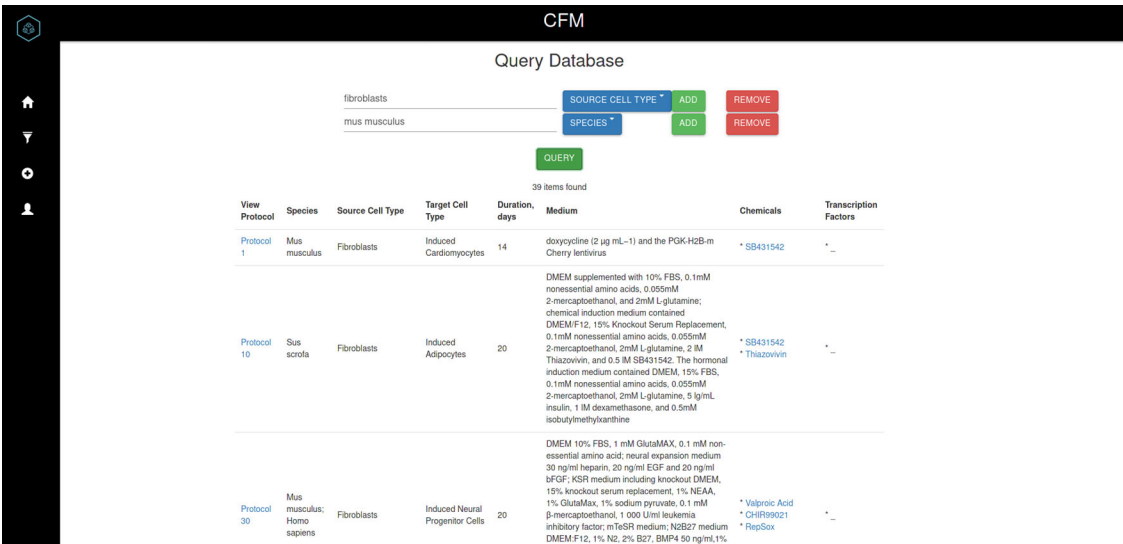


Figure 4. Query to the database.

We strongly believe that the development and maintenance of such a database will encourage researchers to keep improving protocols for cell conversion. Among all cell conversions, reprogramming protocol still dominates over trans-differentiation [Figure 1](#). As it is clear from the [Figure 1B and C](#), fibroblasts are widely used as an initial cell type in various protocols, while other cell types are underrepresented. Although fibroblasts might be an easy cell type to obtain in the clinic, other initial cell types such as blood cells could be also of use. As can be seen in [Figure 1D](#), there is a small group of chemicals effective in several protocols, suggesting that they target a general mechanism. Possibly other chemicals from the same functional class, such as epigenetic drugs, should also be tested for their cell conversion potential. Further research on the regarding relations between cell conversions and chemicals, media, transcription factors, etc. should lead to the development of a computational framework that will facilitate acceleration of the search for conditions conducive to new cell conversions which are highly applicable in medicine (for example, ⁶⁻⁸).

Operation

We perform all computations on the server side, so the user only needs to have a web browser to use our service. However, we recommend not to access the service from smartphones because our interface is not optimized for them.

Figure 5. SMILES similarity interface.

Use cases

Querying the database

In [Figure 4](#), as well as on [Figure 3](#), you can see examples of querying the database. New fields of search can be added using the green ADD button; unnecessary fields can be removed using the red DELETE button.

Molecular similarity calculation

We calculate the similarity of molecules based on their SMILES. Retrieving Morgan fingerprints from them, we calculate the Tanimoto similarity and provide the result to the user. The interface can be seen in [Figure 5](#).

Data availability

All data was gathered from open sources such as PubMed; links to the source articles can be found on protocol pages accessible from cfm.mipt.ru/query page.

We collected protocols from [PubMed](#) and [Google Scholar](#), and downloaded SMILES for chemical compounds from [PubChem](#). All data can be downloaded from our [GitHub Repo](#).

Software availability

Our service is available at: <http://cfm.mipt.ru>

Source code is available from: <https://github.com/lubitelpospat/CFM-source>

Archived source code as at time of publication: <http://doi.org/10.5281/zenodo.4500125>.⁹

License: MIT

Acknowledgements

The authors are grateful to Valeria Tsvelaya for her help with the standardized description of the protocols. We are also grateful to the creators and maintainers of <https://pubchem.ncbi.nlm.nih.gov>.

References

1. Fu Y, Xie X, Liu J: **Chemical reprogramming and transdifferentiation**. *Curr Opin Genet Dev.* 2017; **46**: 104–113.

0959-437X.
[PubMed Abstract](#) | [Publisher Full Text](#)

2. Xu J, Li X, Deng H: **Small molecule-induced cellular fate reprogramming: promising road leading to rome.** *Curr Opin Genet Dev.* 2018; **52**: 29–35. 0959-437X.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Zhang Y, *et al.*: **Small molecules, big roles – the chemical manipulation of stem cell fate and somatic cell reprogramming.** *J Cell Sci.* 2012; **125**: 5609–5620.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Shen Y, *et al.*: **Rpdb: a database of experimentally verified cellular reprogramming records.** *Bioinformatics* 2015; **31**: 3237–3239.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Napolitano F, *et al.*: **Automatic identification of small molecules that promote cell conversion and reprogramming.** *bioRxiv* 2020.04.01.021089. 2020.
[Publisher Full Text](#)
6. Zhu S, Wei W, Ding S: **Chemical strategies for stem cell biology and regenerative medicine.** *Annu Rev Biomed Eng.* 2011; **13**: 73–90.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Xu A, Cheng L: **Chemical transdifferentiation: closer to regenerative medicine.** *Front Med.* 2016; **10**(2): 152–165.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Takeda Y, Harada Y, Yoshikawa T, *et al.*: **Chemical compound-based direct reprogramming for future clinical applications.** *Biosci Rep.* 2018; **38**(3).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Dmitrievich SA: *Cfm-source.* Feb 2021.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 28 June 2021

<https://doi.org/10.5256/f1000research.31475.r88121>

© 2021 Willerth S. This is an open access peer review report 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.



Stephanie M. Willerth 

Department of Mechanical Engineering, University of Victoria, Victoria, BC, Canada

The paper details the CFM database that enables users to explore, compare and evaluate different methods for chemical reprogramming cells from one phenotype into another. Cellular reprogramming is becoming an increasingly popular technique used for both understanding biology and for potential applications in regenerative medicine. The purpose of this database is clear as it provides a standard template for methodologies and evaluation of the effectiveness of these protocols for easy comparison for researchers who wish to use these methods. It also allows the community to rate the protocols. The database focuses on the use of small molecules to reprogram as opposed to transcription factor mediated reprogramming.

The database contains links to the papers containing the original protocols and the amount of data presented varies depending on the protocol. The methods I looked at had not yet been rated by users. The datasets have a standard format, but the listed information is not present for all protocols.

I would suggest writing out Cell Fate Mastering (CFM) in the title of the article so it is easier for readers to understand. The references could be expanded to add more on the use of small molecules for reprogramming. I would suggest using consistent capitalization in Figure 2 and improving the resolution of Figure 1, the font size for the numbers on the y-axis in Figure 1 could be increased to improve readability.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cell reprogramming, tissue engineering

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

Reviewer Report 11 May 2021

<https://doi.org/10.5256/f1000research.31475.r83456>

© 2021 Dashinimaev E. This is an open access peer review report 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.



Erdem B. Dashinimaev 

Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Pirogov Russian National Research Medical University, Moscow, Russian Federation

This work is devoted to creating a database of direct reprogramming protocols for human and mammalian cells. As a researcher working in this field, I welcome such initiatives to streamline and make sense of the large amounts of data emerging in this area. The database created by the authors is quite handy, illustrative, and can be used by other researchers as a convenient resource. However, I have a comment/suggestion to improve the content. It would be nice if two more parameters were offered for each protocol - the number of citations of the original article and the number of citations of the original article in reviews. Why I think this would be useful. It is no secret that the reproducibility of many published methods (even in top scientific journals) is very low. In my opinion, the citation rate of any article (not in reviews) can serve as a good marker that the method is reproducible with a high probability.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: cell biology

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

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