A comprehensive platform for induced pluripotent stem cell research data

Kirill Borziak, Irena Parvanova, Joseph Finkelstein

Icahn School of Medicine at Mount Sinai, New York, New York, USA

Background

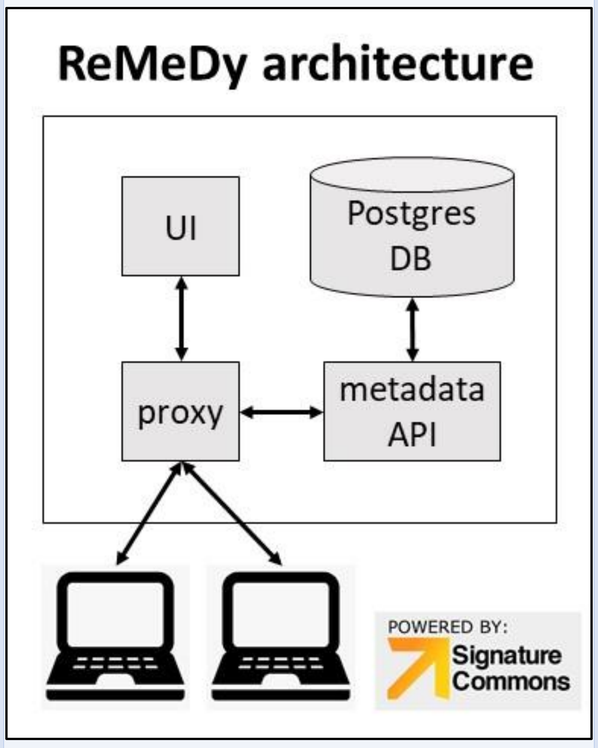
Stem cell therapies, particularly those involving induced pluripotent stem cells (iPSCs), are a growing new therapeutic field for treating, repairing, or replacing patient cells, with the aim of addressing physical damage or disease. Our previous work examined how the voluminous amount of data generated by stem cell research is maintained across the many publically-available stem cell databases[1]. However, the diverse forms of data resulting from stem cell research are not consolidated, stored, and available for access by researchers in a centralized and harmonized manner. To address this we developed the Regenerative Medicine Data Repository (ReMeDy). The potential users of ReMeDy include clinical and pre-clinical researchers interested in understanding the similarities and trends in the derivation of iPSC products. ReMeDy is publically accessible at https://remedy.mssm.edu/.

Objectives

The objectives of this projects were to perform a feasibility study of our new platform, ReMeDy, for storing diverse metadata associated with iPSC research. Additional objectives included developing the multi-modular CDE framework tasked with storing the diverse data types in a harmonized and flexible manner.

Research Design

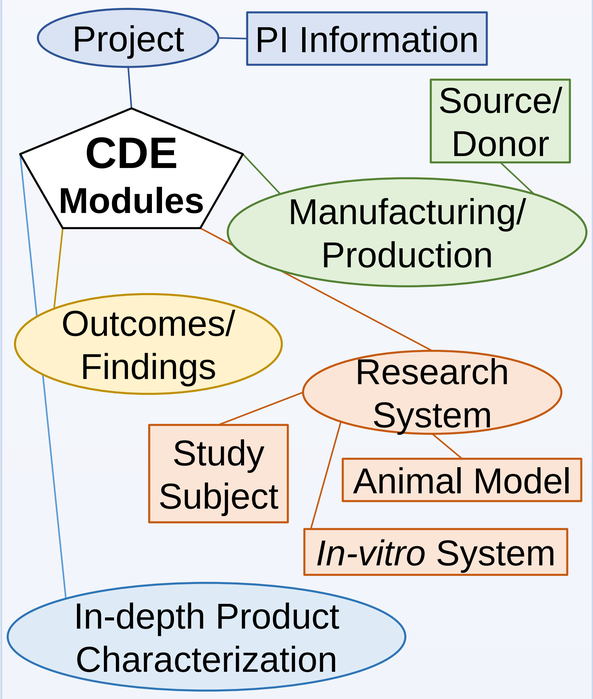
Our platform, Regenerative Medicine Data Repository (ReMeDy), is an implementation of the Signature Commons (https://github.com/MaayanLab/signature-commons), which is a BD2K-LINCS DCIC platform[2], installed through Docker. To improve the utility of the API, we developed an upload interface, which automated the ingestion process.



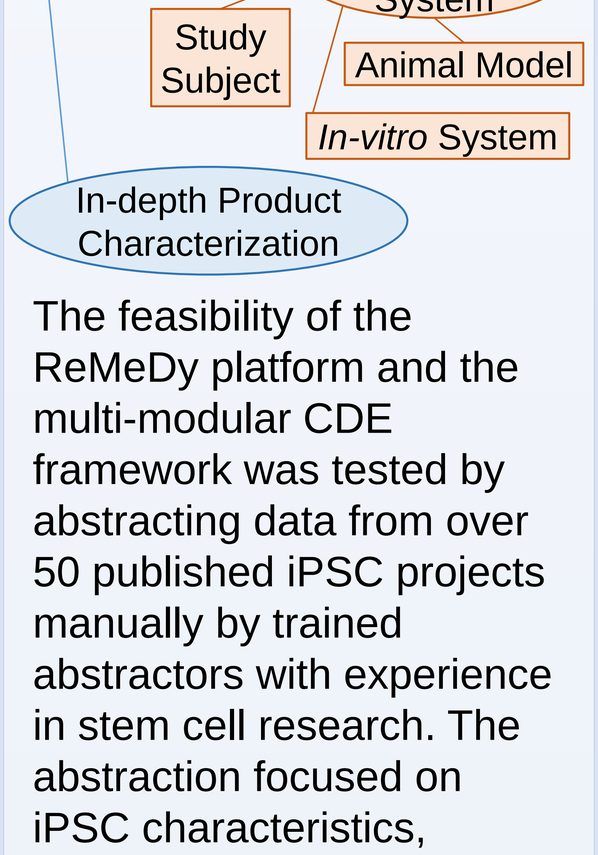
To test the ability of ReMeDy to handle heterogeneous stem cell data, we selected a set of 51 iPSC original research publications. The randomized selection process was designed to ensure the inclusion of the full range of stem cell research. Following the selection of our publication set, the data from the publications was abstracted into the multi-modular Common Data Elements (CDE) framework. The abstraction process was conducted manually by trained abstractors with experience in regenerative medicine, and stem cell research.

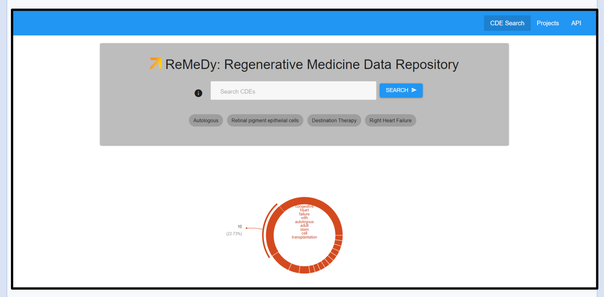
Results

The multi-modular CDE framework is designed to store all data related to iPSC research in a harmonized and flexible manner. The CDEs are organized into 5 modules: Project, Manufacturing /Production, In-depth Product Characterization, Research System and Outcomes/Findings.



The feasibility of the ReMeDy platform and the multi-modular CDE framework was tested by abstracting data from over 50 published iPSC projects manually by trained abstractors with experience in stem cell research. The abstraction focused on iPSC characteristics, patients or animal model and the research finding and outcomes.





The ReMeDy landing page provides access to the comprehensive key and value search functionality, the projects page, and the fully functioning RESTful API.

Conclusions

ReMeDy provides a comprehensive, centralized and unified platform for data generated by iPSC research, as seen in our feasibility study. The improved access to aggregated CDE values and statistics has the potential to drive knowledge discovery.

Future Plans

Future aims for ReMeDy include increasing the database size to include all published iPSC research. This will be accomplished by implementing natural language processing and crowdsourcing functionalities. To automate data abstraction through NLP, we aim to use MeSH terminology and ontology-driven functionalities. Further, we aim to establish an automated pipeline for the uploading of iPSC data through CDE templates.

References

1. Finkelstein J, Parvanova I, Zhang F. Informatics Approaches for Harmonized Intelligent Integration of Stem Cell Research. Stem Cells Cloning. 2020;13: 1-20.

2. Keenan AB, Jenkins SL, Jagodnik KM, et al. The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. Cell Syst. 2018;6(1):13-24.