

Executive Summary

Over the past two years, CIRM has convened a few workshops and panels of leading experts to help develop successful strategies for team-based scientific and collaborative research. Given Prop 14's prioritization around research and therapy development on diseases and conditions of the brain and CNS, the initial focus of these efforts has been neuroscience research. As such, in 2019, CIRM hosted the [Brainstorming Neurodegeneration Workshop](#) to discuss consortia-based research models. At the 2020 CIRM Grantee Meeting, an expert panel discussed [knowledge networks](#) and data sharing. At the recent Scientific Strategy Advisory Panel [meeting](#), panelists discussed how CIRM can best support California's stem cell and gene therapy research. The outcome of these workshops was for CIRM to gather value of a consortium approach and a CIRM knowledge platform for CNS diseases. In conjunction with the University of California, Santa Cruz (UCSC), CIRM therefore convened a CNS Consortium Advisory Committee to discuss and evaluate opportunities for a data infrastructure that could accelerate research by empowering broad and meaningful data sharing. Ultimately, the outcomes of this meeting will inform the development of a larger workshop (to be held in 2022) that will engage more stakeholders in discussion about a CIRM CNS Consortium and knowledge platform.

Participants discussed the unique potential value of a CIRM knowledge platform. While other large-scale data sharing initiatives often focus on the aggregation of data from normative tissues, CIRM is uniquely positioned to contribute a wealth of data from diseased tissues to the broader scientific community. A knowledge platform would provide an avenue for data sharing and collaboration with other groups that are dedicated to accelerating progress in therapeutics for CNS disorders. The aggregation of such data from all parts of the research pipeline would support efforts to: (1) explore commonalities across neurodegenerative, neuropsychiatric, and neurodevelopmental disorders; (2) identify and validate novel disease mechanisms and targets; and (3) identify and validate diagnostic, prognostic, progression, and predictive biomarkers of disease. Participants recognized the need for a data infrastructure that would, amongst other, support large-scale characterization of the variability that underlies current induced pluripotent stem cell models from diverse genetic backgrounds, which has the potential to improve the translation of new therapeutic candidates to broader clinical populations.

A CIRM knowledge platform would embrace the 4 principles of the [data biosphere](#): ingested data would not enter a static database but rather a dynamic cloud-based ecosystem with secure access that provides a collaborative, scalable, and interoperable computing environment for shared datasets and analysis workflows. The platform could leverage existing data biosphere modules with discrete capabilities that allow for meaningful data analysis while remaining fully customizable to meet the specific needs of CIRM researchers and collaborators. Critically, the platform would enable researchers to expand their analyses beyond the data collected in their own projects to include data from other projects within the CIRM-funded research landscape as well as datasets housed by other resources that have partnered with CIRM to ensure interoperability. The platform would also support efforts to improve scientific rigor and reproducibility by creating an environment to catalog, share, and collaborate on all aspects of the scientific process, including standardized workflows for common analyses. All of

these benefits would be accompanied by a reduction in the overall cost of large-scale data storage and computational analysis by reducing redundancy in data storage and leveraging cloud-based storage and compute resources.

Participants favored a hub-and-spoke architecture for the knowledge platform, in which a centralized data management team would work with dedicated program managers at different data generating locations (i.e., CIRM-funded labs and/or competency hubs) to ingest data into the biosphere. The data management team would perform initial processing of datasets, while program managers would be responsible for reporting data progress and maintaining the integrity of data during ingestion. To facilitate engagement with the platform, participants supported funding for “data wranglers”: members of this central data management team who would not only work with program managers to gather data for the platform but would also facilitate entry of those data into other non-CIRM databases. In addition to these data managers, CIRM would also need to invest in hiring data scientists and engineers to develop and optimize the pipelines that would be used for data processing and analysis. Participants agreed that the development of these analytical frameworks should be closely connected to the sources of data generation, which would include both CIRM competency hubs as well as CIRM-funded labs (through CIRM’s funding pillars).

While a CIRM knowledge platform would ultimately host numerous data types, some kinds of data may be more readily accessible or amenable to centralization on a CIRM knowledge platform. Participants generally agreed that omics data would be the easiest type of data to work with at the outset, because these data are plentiful and useful for CIRM-funded research. The knowledge platform would also create an opportunity to link data from CIRM’s robust basic and translational science portfolios with its growing number of clinical trials to support the characterization, validation, and predictive capacity of disease models.

In addition to expressing opinions about the feasibility and value of a CIRM knowledge portal, participants offered specific recommendations for a larger workshop on this subject in 2022. Individuals and groups recommended for inclusions at the workshop included: (1) large-scale projects that currently leverage data biosphere tools and could become potential data sharing partners, (2) initiatives that are collecting data on or banking normative tissues that could be used for benchmarking, (3) representatives from major medical centers who are leaders in stem cell research and neuroscience, (4) Dr. Kevin Eggan of the Harvard Stem Cell Institute, and (5) Dr Jonathan Lawson of the Broad Institute. Additionally, participants also recommended the following topics to be added for discussion: (1) historical use of CIRM’s iPSC repository, (2) specific guidelines for central data management and responsibilities in a CIRM knowledge platform, and (3) Data sharing and management policies.