CNS Consortium 2022 Workshop

California Institute for Regenerative Medicine

February 24 & 25, 2022

Virtual Workshop

DAY 1 SUMMARY – Shared Resources Laboratories for Stem Cell-Based Modeling

(Day 2 summary is presented in a separate document)

Table of Contents

Executive Summary

Day 1 & Day 2

As articulated in its 2022-2027 strategic plan, CIRM's mission is to accelerate world class science to deliver transformative regenerative medicine treatments in an equitable manner to a diverse California and world. One of the three major themes of the Strategic Plan calls for the advancement of world class science by leveraging collective scientific knowledge to inspire collaborative research that addresses Californian's unmet medical needs. To achieve this vision, the goals are (1) to develop next-generation technology **competency hubs** and (2) to build **knowledge networks,** fostering a culture of open science.

The goal of the CNS Consortium Workshop was to solicit feedback regarding the feasibility, opportunities, and best approaches to realize these goals of CIRM's Strategic Plan. **Day 1** discussions focused on **Shared Resources Labs** and **Day 2** discussions focused on **Data Infrastructure**.

Discussions at this workshop were focused on applications to the central nervous system (CNS) as a use case, but the resulting initiative(s) would be implemented broadly across cell types, organs, and diseases.

Day 1 – Shared Resources Labs for Stem Cell-Based Modeling

There is abundant interest and expertise in the California research community to capitalize on the promise of stem cell-based modeling. To assess needs in the field and utility of a possible shared resources funding program, the goal of the first day of the workshop was to identify challenges related to stem cell-based modeling and how a network of Shared Resources Labs may help to address them. To consider outcomes and lessons learned from previous efforts, **Session I** featured speakers who described their involvement and use of **shared resources that CIRM created under Proposition 71**, i.e., Shared Laboratories, a human induced pluripotent stem cell (hiPSC) Repository, and a Stem Cell Genomics Initiative. These shared resources provided the stem cell research community with access to infrastructure, tools, datasets and training, and fostered effective collaborations between laboratories with different areas of expertise. This made stem cell research more accessible to researchers at a time when federal funding for human embryonic stem cell (hESC) research was restricted, and made entry into the emerging field of human pluripotent stem cell (hPSC, umbrella term for hESC and hiPSC) research accessible to a broader research community. Even after CIRM funding for these programs ended, many remained sustainable and valuable resources for California's stem cell research community.

Session II on the first day featured a moderated discussion with 22 subject matter experts and stakeholders who were asked to identify hurdles to effective stem cell-based disease modeling and explore potential strategies for CIRM to help researchers overcome these challenges through the implementation of Shared Resources Labs. Two major hurdles to meaningful stem cell-based modeling were considered: **(1) limited reproducibility of findings** and **(2) uncertainty about the predictive value for human biology and disease.**

Discussants emphasized that **limited reproducibility** across projects employing similar stem cell-based models poses a major hurdle to scientific advancement. Suggested approaches to improving reproducibility ranged from: (a) technical solutions, such as automation and standardization of materials and protocols, (b) networking among researchers to share best practices and protocols, to train those new to the field and to replicate studies across labs, (c) scaling research for increased statistical power, and (d) data-related considerations, such as sharing outcomes data using FAIR (Findable, Accessible, Interoperable, Reusable) principles, providing detailed and consistent metadata, and deploying machine learning for analyses.

Discussants argued that there remains a need to continue to innovate and **improve stem cell-based models to increase their predictive value**. To better understand how a stem cell-based model relates to human biology and disease, discussants pointed to a need for deep clinical phenotyping of cell donors and for obtaining molecular and cellular information from relevant post-mortem human tissues as ground truth for analysis of various omics datasets generated from hPSC-based models. Another approach involves validating a candidate hPSC-based model by testing whether a drug elicits cellular and molecular phenotypes in vitro consistent with the drug's known effects in vivo. Developing more predictive hPSC-based disease models may be achieved by e.g., integrating multiple relevant cell types to better mimic complex biology, but this may also add variability to the experiment. Discussants commented that standardization of simpler models and innovation toward more complex models may both be needed to advance the field.

Discussants considered two distinct goals for a possible Shared Resources Labs network: **(i) to drive innovation toward optimizing and standardizing cutting-edge stem cell-based models** and **(ii) to lower barriers of entry into the stem cell-based modeling field**. An argument was made that both goals could be pursued. In addition to effectively sharing stem cell-based modeling expertise, recommendations for approaches to **building a network of Shared Resources Labs** included providing access to well characterized unmodified and modified hPSC collections, providing access to new technologies and equipment that may be too expensive or specialized for a single laboratory to acquire, and providing help with navigating the relevant stem cell and gene editing intellectual property (IP) landscapes. Discussants emphasized the importance of creating a network of Shared Resources Labs that will be sustainable in the long-term.

Day 2 – Data Infrastructure

The second day of the workshop focused on evaluating the best approach to promoting data sharing in California and determining which role, if any, CIRM should have. The day was divided into 2 sessions. During the first session (**Session III**), presenters outlined the **principles of the data biosphere**, which provides a framework for creating an open, compatible, and secure approach to data storage and collaboration designed for biomedical research, and they described **examples** of scientific initiatives that have **successfully implemented Knowledge Platforms**, deploying a cloud-based data and software ecosystem based on data biosphere principles. The session wrapped up with an overview of the "Data Use Oversight System" (DUOS) to semi-automate and efficiently manage compliant sharing of human subjects data. The goal of Session III was to illustrate how the concept of a Knowledge Platform could be applied to address

technical and collaborative needs of California researchers, many of which were identified during Day 1 of the workshop.

The second session (**Session IV**) was a moderated discussion dedicated to understanding the best approach to promoting data sharing in California. The discussion was framed around the current needs and obstacles for existing collaborative Knowledge Platforms. A pre-workshop survey provided insight into the limited knowledge with regard to Knowledge Platforms among respondents, who were mainly researchers in the regenerative medicine field but who also showed a general interest in and openness to sharing data and collaborating across laboratories.

The moderated discussion with 22 subject matter experts and stakeholders focused on considerations for the most optimal design and implementation of a potential collaborative Knowledge Platform and for designing a Data Coordination and Management Center (DCMC).

The Knowledge Platform would provide a cloud-based data and software ecosystem, including tools, applications, and data processing workflows, to allow collaborative analysis in a shared computing environment with defined data access and security protocols. For a Knowledge Platform to be successful, discussants commented that it is important to understand and consider the needs and goals of those who will contribute data and those who will use the data. To allow efficient data sharing, it is also important to address administrative and technical challenges related to accessing and retrieving data, and discussants stated that data processing standards, metadata specifications, and naming conventions for each of the anticipated data types should be addressed early in a program's execution to enable interoperability of data from different sources. Discussants argued for tiered metadata to accommodate core information needed to replicate analyses and ensure interoperability of datasets while also allowing optional metadata as needed for specialized experiments. Importantly, discussants emphasized that cloud-based collaboration is new to many researchers, and that those who generate and contribute data need to be supported during data submission, and researchers should be incentivized to collaborate in the cloud.

The final part of the discussion was focused on considering several options for structuring data coordination and management responsibilities among researchers who generate data and a DCMC that would be responsible defining the conventions used, and for storing the data and making it available to researchers. Discussants considered several DCMC models and favored a model that would entail the inclusion of data type-specific expertise within the DCMC and also within the sites that produce raw data.

Conclusion

Overall, the presentations and discussions during this 2-day workshop reaffirmed the needs of the research community for shared competency hubs and a collaborative data infrastructure. By supporting these resources, CIRM can help democratize data analysis, improve access to hPSC models for human biology and disease, and increase collaboration across laboratories with diverse areas of expertise.

Workshop Summary

The mission of CIRM is to accelerate world class science to deliver transformative regenerative medicine treatments in an equitable manner to a diverse California and world. One of the three major themes of the Strategic Plan calls for the advancement of world class science by leveraging collective scientific knowledge to inspire collaborative research that addresses Californian's unmet medical needs. To achieve this vision, the goals are (1) to develop next-generation technology **competency hubs** and (2) to build **knowledge networks,** fostering a culture of open science.

Box 1 - What is a competency hub? An entity that shares a specialized skill or resource (competency) at any stage of the drug development pipeline with other investigators in a collaborative manner.

The goal is to empower and connect California's research ecosystem and facilitate validation and standardization of research platforms.

Shared Resources Labs are one form of competency hub.

Box 2 - What is a knowledge network? Shared scientific knowledge across discovery, translational, and clinical research.

The goal is to maximize the impact of research by facilitating and incentivizing data sharing.

The Data Infrastructure (see Day 2, Figure 4) would enable cloud-based, collaborative analyses across data shared from CIRM-funded and other research through the creation of a **Data Coordination and Management Center** (DCMC) and a **Knowledge Platform**.

The goal of the CNS Consortium Workshop was to solicit feedback regarding the feasibility, opportunities, and best approaches to realize these goals of CIRM's Strategic Plan. **Day 1** discussions focused on **Shared Resources Labs** and **Day 2** discussions focused on **Data Infrastructure**.

Day 2 summary presented in a separate document

Day 1 – Shared Resources Laboratories

The first day of the workshop focused on obtaining feedback to inform CIRM about opportunities to share resources and promote collaborative research in California and help assess feasibility and best approaches toward building a California network of Shared Resources Labs. Discussions centered on stem cell-based models as shared resources, based on results from outreach to directors of past CIRM Shared Labs (Proposition 71) and a CIRM townhall meeting, that highlighted a critical need for access to validated stem cell-based models (Figure 1).

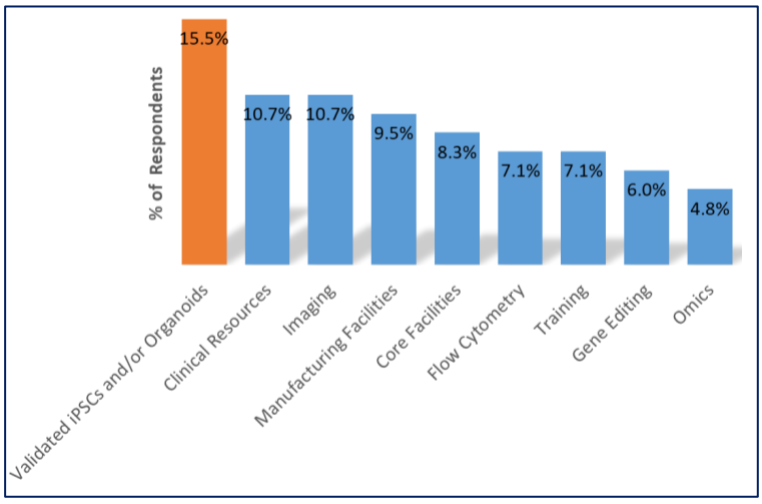


Figure 1: Preference for sharable resources. Directors of past CIRM Shared Labs and users and providers of shared resources indicated which shared resources they would like to access, n = 84.

The workshop was framed around applications to the central nervous system (CNS), given Proposition 14's allocation of more than a quarter of all funds to diseases and conditions of the CNS, and as a use case to generate ideas that can then be implemented broadly across cell types, organs, and diseases.

The day was divided into 2 sessions. **Session I** provided an overview of previous research resources, funded by CIRM under Proposition 71, and **Session II** was a moderated discussion among researchers and other stakeholders to define hurdles facing the stem cell-based modeling field and potential solutions, and to delineate approaches for sharing resources and promoting collaborative research going forward under Proposition 14.

Session I: Previous CIRM-funded Research Resources

Under Proposition 71, CIRM funded a variety of research resources, including (A) Shared Laboratories, (B) a human induced pluripotent stem cell (hiPSC) Repository, and (C) the Center of Excellence in Stem Cell Genomics (CESCG). Lessons learned from these resources will contribute to the development of the next generation of Shared Resources Laboratories.

I.A. Shared Laboratories

The objectives of CIRM's Shared Laboratories (Shared Labs) were to provide California researchers with (1) dedicated research space, specialized instrumentation, cell lines, and cell culture materials free from federal limits placed on human embryonic stem cell (hESC) research at the time, and (2) training in stem cell culture and related technologies. CIRM funded 17 Shared Labs from 2007-2016, providing a total of $68.5 million to build and operate the facilities, making stem cell research more accessible to California researchers. Six of those labs received additional CIRM funding to develop an advanced stem cell techniques course for researchers and trainees new to the field of stem cell research.

Presentation: Prop 71 Shared Labs
David Schaffer, UC Berkeley

David Schaffer presented an overview of the UC Berkeley hESC Shared Lab he directed. CIRM funding supported regional access to stem cell culture equipment and expertise and enabled shared access to research equipment that may have been cost prohibitive for a single laboratory to purchase, including a cell sorter and specialized microscopes. Access extended beyond UC Berkeley and included Children's Hospital Oakland Research Institute, now University of California, San Francisco (UCSF) Benioff Children's Hospital Oakland.

The UC Berkeley Shared Lab is an example of a sustainable shared resource with lasting impact that was established with CIRM funding; in 2021 alone, 43 principal investigators and 100 researchers used the facility. After CIRM funding ended, UC Berkeley and California Institute for Quantitative Biosciences (QB3) began subsidizing the facility (the facility name was changed to QB3 Cell and Tissue Analysis Facility), and the facility now supports work with cells other than stem cells to expand its user base. Recharge income combined with subsidies provide funds for facility management, which is carried out by a part-time staff member and facility "super users."

I.B. hiPSC Repository

The CIRM hiPSC Repository established a comprehensive collection of research-grade hiPSC lines from 2,184 unique cell donors representing mostly prevalent, genetically complex diseases. The CIRM hiPSC collection also serves as a valuable resource for obtaining close to 300 healthy control lines with diverse ancestral backgrounds. Each hiPSC line is linked to demographic, medical, and/or diagnostic information about the cell donor. Cellular Dynamics Inc., now part of FUJIFILM (FCDI), produced these cell lines using a standardized protocol and distributes them worldwide. Through a collaboration with the Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard University, single nucleotide polymorphism (SNP) data for all CIRM hiPSC lines, and whole genome sequence data for 299 CIRM hiPSC lines has been made available to the research community.

Three presentations served as examples for the use of the CIRM hiPSC lines in large scale disease modeling efforts.

Presentation: hiPSC-based Population Genetics
Ralda Nehme and Sulagna Ghosh, Broad Institute

Sulagna Ghosh and Ralda Nehme spoke about their work using CIRM hiPSC lines to create population scale stem cell models to study the effect of genetic variation on cellular phenotypes. A prepublication manuscript describes their "village in a dish" approach, which entails culturing dozens (up to ~100) of different cell lines in the same dish and conducting single cell sequencing to deconvolute analysis of phenotypic outcomes.

Presentation: NAFLD Lines for Disease Modeling
Jacquelyn Maher, UC San Francisco

Jacquelyn Maher contributed cell samples for hiPSC line generation to the CIRM hiPSC Repository, representing patients with nonalcoholic fatty liver disease (NAFLD, a genetically complex disease) and healthy controls. Dr. Maher described her initial analysis of hiPSC-derived hepatocytes (44 NAFLD, 19 controls), showing that cells from NAFLD patients, but not those from healthy controls, robustly developed spontaneous steatosis (fat buildup). This published finding provides the rationale for using this hiPSC collection to gain new insights into disease mechanisms such as elucidating distinct contributions of different cell types in the liver and the role of individual genetic risk factors in disease development. However, given the large number of cell lines that need to be analyzed to interrogate a genetically complex disease like NAFLD, obtaining adequate funding remains a challenge.

Presentation: Machine Learning & Engineered iPSCs for Unraveling the Complex Biology of CNS Disease
Ajamete Kaykas, insitro

Ajamete Kaykas described insitro's use of machine learning (ML) and engineered hiPSC, differentiated into relevant cell types, for unraveling the complex biology of CNS and liver disease. By integrating analysis of varied data types, such as cellular morphology, single cell

transcriptomics, and longitudinal live cell imaging, insitro builds a disease state manifold, i.e., a multi-faceted representation of disease combining data from molecular, cellular, tissue, and organ levels. To build predictive models, data from patient samples are used to validate in vitro findings. Automated hiPSC model culture at scale and unbiased ML-enabled phenotyping approaches are used to interrogate the effects of chemical and genetic perturbations to identify targets for drug development. Insitro uses hiPSC lines from the CIRM repository and other sources for these large scale efforts.

I.C. Center of Excellence in Stem Cell Genomics

CIRM funded the Center of Excellence for Stem Cell Genomics (CESCG) from 2014-2019 to enable the application of state-of-the-art genomic approaches to substantive problems of human stem cell biology. CESCG involved seven institutions that, among other efforts, provided support for emerging single cell-sequencing approaches to stem cell researchers in California. Data sharing was mandated and supported by this CIRM Initiative, and the CESCG included a Data Coordination and Management Center at the University of California (UC), Santa Cruz (UCSC) that was responsible for housing and making available 84 terabytes of data. CESCG institutions conducted three center-initiated research projects and provided sequencing and omics analysis support to 14 collaborative stem cell projects at eight institutions.

Presentation: Experimental-Computational Collaboration to Characterize Cortical Organoids
Aparna Bhaduri, UC Los Angeles; Max Haeussler, UC Santa Cruz

The Genomics Initiative was collaborative by design, and Aparna Bhaduri (former post-doc in Kriegstein lab) and Max Haeussler (staff in Kent group) described an experimental-computational collaboration that emerged between the Kriegstein lab (UCSF, one of the CESCG collaborative projects) and the Kent group (UCSC, Data Coordination and Management Center) to investigate the validity of cerebral organoids as a model system for studying cortical development. As part of a single cell RNA sequencing characterization, the Kriegstein lab found that stress pathways were upregulated in organoids, and subsequent analysis of transplanted organoids in mice suggested that the cell culture environment triggers these stress responses.

The Kent group used its expertise in the creation of visualization tools to create the Cell Browser, in part to help analyze the Kriegstein laboratory's data. This tool enables single-cell transcriptomic data visualization without requiring users to have formal computational expertise or access to large servers. This resource can be used to visualize cell type and gene expression annotations for any single-cell dataset. The data browser currently hosts more than 100 datasets, supports over 3,000 monthly users, and has been cited in 180 publications. This collaboration was successful in part due to the unique expertise of both groups, as well as CIRM's dedicated funding for development of new data analysis tools that helped support long-term computational staff.

Session II: Moderated Discussion

The stem cell research resources funded by CIRM under Proposition 71 facilitated effective collaborations between laboratories with different areas of expertise, provided access to a rich collection of tools and datasets, and made the field of stem cell-based disease modeling more accessible to researchers new to the field.

To inform CIRM about opportunities to support California researchers, under Proposition 14, in the use of the next generation of stem cell-based models as shared resources, session II provided (A) an overview of the results from a pre-workshop survey, and featured a moderated discussion to address (B) Overcoming hurdles to effective stem cell-based modeling, and (C) Building Shared Resources Labs for stem cell-based modeling.

II.A. Pre-Workshop Survey Results

Prior to the workshop, CIRM conducted a survey of invitees (CIRM listserv) to assess the needs of stem cell researchers and potential opportunities to share resources for stem cell-based modeling through the implementation of Shared Resources Labs. To help frame the discussion, Session II began with a presentation of survey results. Of the respondents who currently use stem cell-based models in their labs (56 percent), 61 percent receive requests to share their models (including 2D and organoid models), which is typically achieved by providing protocols, offering advice, shipping cells, initiating formal collaborations, or training lab members (

Figure **2**).

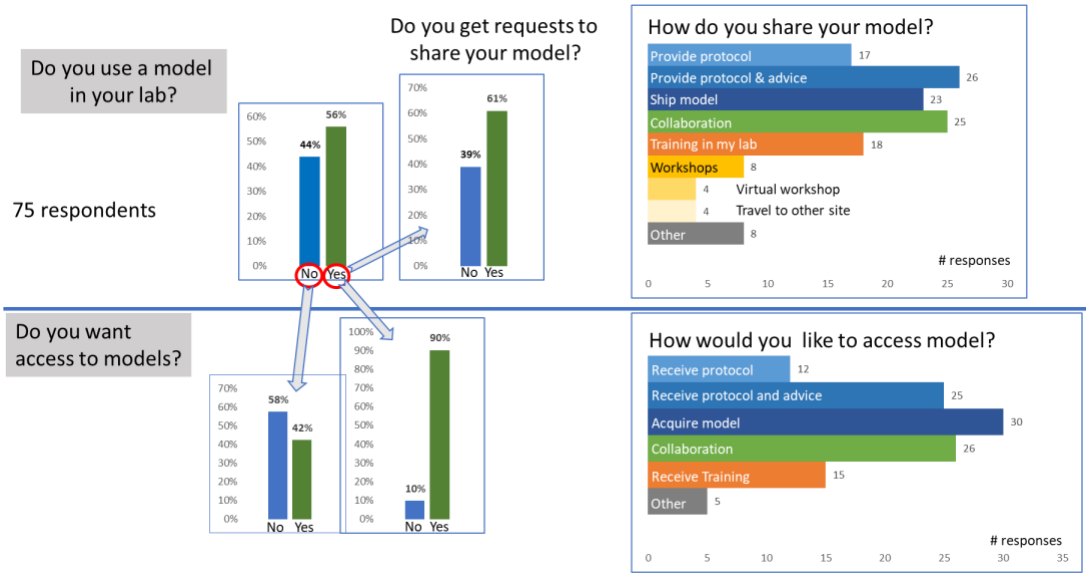


Figure 2: Stem cell model sharing and accessibility.

Many respondents, including 90 percent of those who already use stem cell-based models and 42 percent of those who do not, expressed interest in gaining access to 2D and organoid models of a variety of different cell types.

II.B. Overcoming Hurdles to Effective Stem Cell-Based Modeling

While there is abundant interest and expertise in the California research community to capitalize on the promise of hPSC-based modeling, two major themes emerged during the moderated discussion that represent significant hurdles to meaningful stem cell-based modeling, i.e.

1. Limited Reproducibility of Findings and
2. Uncertainty about the Predictive Value for Human Biology and Disease.

Discussants considered various efforts that could contribute to overcoming these hurdles and will be described in the following section.

1. Limited Reproducibility of Findings from Stem Cell-Based Models

hPSC-based models require the differentiation of hPSC into desired cell types. There are various types of differentiated products that can be obtained for a given cell type (2D single cell type, 2D multiple cell types, organoid, assembloid, 3D engineered constructs and others) and none of the differentiation protocols have been standardized. Even if two different laboratories, or two researchers in the same laboratory, are using a similar differentiation approach, the resulting models likely differ in cellular identity, composition and function.

Approaches toward improving reproducibility can be divided into several categories, highlighted in Box 3.

Box 3 – Overcoming hurdles to effective stem cell-based modeling

1. Approaches to improving reproducibility of findings from hPSC-based models

1a. Technical
- Automation
- Standardization – **see Box 4**

1b. Networking
- Share best practices
- Training
- Replication of protocols and modeling studies across multiple labs

1c. Scaling
- Larger scale research (many cell lines)
- Diversity of hPSC lines

1d. Data-related
- Share outcomes data – FAIR
- Detailed and consistent metadata
- Machine learning for analyses

1a. Improving reproducibility: technical approaches

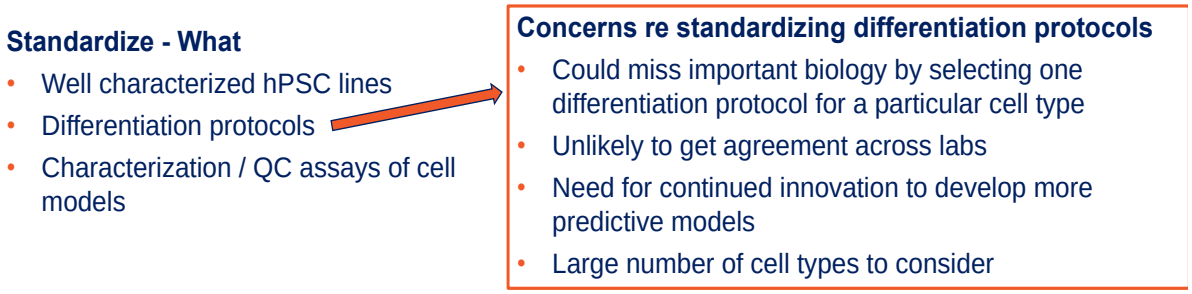
Technical efforts toward improving reproducibility include automation of experimentation to reduce human-introduced variability, and standardization of materials and protocols. The need for standardization was discussed at some length and the main discussion points are described below and summarized in Box 4.

Standardization of differentiation protocols can improve reproducibility of results across laboratories. Certain 2D differentiation protocols are relatively simple; however, optimization and standardization of these protocols is not incentivized, and specific funding opportunities to do so may be needed. More complex differentiation protocols still require optimization and researchers are unlikely to agree on a single protocol for a given cell type or organoid. Discussants

acknowledged that overemphasis on standardization could discourage researchers from developing improved differentiation protocols that better reflect human biology. For instance, a lack of typical mechanical forces and integration with other cell types limit the capacity of many current hPSC-based models to properly reflect human biology and disease, and development of next generation stem cell models needs to be encouraged. To strike a balance between protocol standardization and innovation, CIRM could incentivize standardization of simpler differentiation protocols and encourage individual laboratories to continue to innovate more complex protocols. One approach CIRM could take to provide an opportunity for some standardization is to create a centralized resource that distributes hPSC partially differentiated to progenitor states using standardized protocols, which researchers could then further differentiate according to their own specific protocols.

Box 4 – Overcoming hurdles to effective stem cell-based modeling

1. Approaches to improving reproducibility of findings from hPSC-based models



While the field appears unready for standardized differentiation protocols, discussants stated that reproducibility could be improved if laboratories would all use the same set of well characterized hPSC lines, at least as controls, and if all investigators used rigorous, standardized methods to characterize the cellular composition and identity of their models, using quantitative assays when possible. Use of a centralized characterization facility would help overcome experimental variability. Discussants also suggested the inclusion of functional assays in standard characterization schemes but acknowledged that a single laboratory may not have the expertise to conduct a full range of useful functional assays, suggesting a need for more collaborations.

1b. Improving reproducibility: networking among researchers

Networking across experienced laboratories to share and define best practices may help identify sources of variability among differentiation protocols. Importantly, there is an urgent need to support expert stem cell-based modeling labs to train researchers who are interested in entering the field. This could be through courses offered at future CIRM Shared Resources Labs, or through funding to promote new collaborations among researchers who are new to the field and those with hPSC modeling expertise. Video tutorials for learning standard operating procedures for work with various cell lines and differentiation protocols would broaden reach of uniform training.

Discussants also considered that replication of protocols and of modeling studies across multiple laboratories would boost confidence in scientific conclusions.

1c. Improving reproducibility: scaling

An approach to addressing high variability across cell lines and limited reproducibility of findings involves scaling efforts to study a large number of hPSC lines that are representative of a particular disease or condition, to improve the statistical power of experiments. However, many laboratories do not have the capacity or funding to scale their efforts in this way. A CIRM Shared Resources Lab could provide e.g., automation capabilities that would make scaling more feasible.

While keeping experimental variability to a minimum as best as possible, the analysis of collections of hPSC lines representing diverse genetic ancestries would either provide confidence in the robustness of findings or could help delineate important differences among populations.

1d. Improving reproducibility: data-related approaches

The benefits of sharing research data in a manner that supports collaborative cross-lab analyses (FAIR principles) are manifold and may include the opportunity to better understand variability between experiments if data are accompanied by detailed metadata. For instance, deep clinical data from cell donors might help explain variability in outcomes from stem cell-based models that reflect biological variation. Importantly though, detailed information about the experiments themselves is critical to track experimental differences between studies conducted in different laboratories. Therefore, shared data on stem cell-based modeling experiments needs to include metadata on the derivation, maintenance and quality control of the hPSC lines, the differentiation protocols, and the characterization of the differentiated cell models.

Analysis of extensive data and associated metadata shared across many laboratories studying similar questions may help shed light on the contribution of biological versus experimental variability to differences in study outcomes. Machine learning approaches may help extract knowledge related to reproducibility.

Detailed information about the clinical phenotypes of cell donors and about experimental approaches will only be useful in cross-study analyses, if types and structure of metadata associated with hPSC-based modeling are consistent across experiments. Ensuring quality metadata may require any new CIRM data repositories to establish required metadata fields prior to soliciting data submissions, similar to the strategy employed by the CIRM-funded Center of Excellence for Stem Cell Genomics, which created and used 'Minimal Information about a Stem Cell Experiment' (MISCE) metadata modules for data submissions. Discussants acknowledged that managing and sharing metadata is a challenge for researchers and associated costs should be included in research and data infrastructure budgets. Part of this budget should support staff who would be responsible for monitoring data submissions and ensuring metadata completeness. In addition, discussants emphasized the value of streamlining and incentivizing metadata submission to encourage the inclusion of quality metadata.

2. Uncertainty about the Predictive Value of Stem Cell-Based Models for Human Biology and Disease

Any model system used in biomedical research has its limitations, but many models are useful for understanding certain aspects of human biology. The predictive value of animal and cell-based models of disease that have been used for decades in basic and preclinical research are reasonably well understood, but hPSC-based models are new, and there is a need to better understand what aspects of human biology and disease they represent well.

To assess the predictive value of stem cell-based models, discussants emphasized the need for deep clinical phenotyping of cell donors, which may reveal clinical subtypes that may explain differences among cell-based models and help guide better informed in vitro experimentation. Since clinical profiles are not static, some research questions may require updates to clinical data over time as a disease progresses. The accuracy of clinical data for control cells from healthy donors must also be ensured; for example, deriving hiPSC from donors who reach old age cognitively intact would boost confidence that control lines were not derived from individuals who would eventually develop a neurodegenerative disease.

Another important goal for the field is to ascertain molecular and cellular information from relevant post-mortem human tissue, healthy and diseased, as ground truth for analysis of various omics datasets generated from hPSC-based models.

In addition to considering clinical data from cell donors and molecular and cellular data from authentic human tissues, proof-of-concept approaches can further help determine the validity of hPSC-based models; for example, a candidate hPSC-based model could be treated with a drug known to be effective for a disease of interest in order to confirm that cellular and molecular phenotypes are also effectively targeted in vitro.

Once tools and approaches are in place to better assess the predictive value of an hPSC-based model, improvements of the model can be reasonably undertaken. Discussants suggested that a multifaceted approach to developing predictive hPSC-based disease models may be needed to capture the complexity of human biology, such as providing integration of multiple cell types and mechanical forces to better mimic the in vivo context. However, while adding more dimensions to a predictive model may increase its physiological relevance, that approach may also add variability, undermining the sensitivity of the assays. It is also critical to select appropriate experimental readouts to obtain data relevant for a given disease phenotype.

II.C. Building Shared Resources Labs for Stem Cell-Based Modeling

California is home to a robust stem cell-based modeling research infrastructure, and as indicated by pre-workshop survey results and the discussions, there is a strong interest in access to stem cell-based models, both by researchers who do not have such models established in their labs, but also by experts who would like access to different models than their own. Experts already share their stem cell-based modeling expertise, by training others and through collaborations, but need support to cope with high demand, especially since training others is time- and resource-intensive.

The approach to building and supporting a network of Shared Resources Labs should be informed by the overall goals that CIRM plans to pursue. A distinction was made between driving optimization and standardization of cutting-edge stem cell-based models and democratization of the stem cell-based modeling field. For example, if CIRM intends for the Shared Resources Labs to support scalability of research efforts, a smaller network of large Labs that include automation may be needed. On the other hand, if the goal is to lower barriers for researchers to enter the stem cell-based modeling field, establishing multiple localized hubs that provide expertise and training would be preferable. CIRM could feasibly pursue multiple strategies for its Shared Resources Labs network simultaneously, and enabling academic – industry partnerships may provide additional resources and opportunities toward a multi-pronged approach. CIRM's past funding of shared resources was praised for having lowered the barriers for new and occasional users of stem cell technologies to enter the field and enabled experienced users to develop new resources and technologies.

CIRM could also consider pursuit of specific research goals with Shared Resources Labs, such as identification of new disease signatures, especially for psychiatric diseases, or drug screening. To broaden the potential for impact, one suggestion was to consider supporting the investigation of unifying disease mechanisms, such as cellular responses to stress.

In addition to sharing expertise and optimizing stem cell-based models, Shared Resources Labs could enable the field by providing access to well characterized control hPSC lines, high value edited hPSC collections and other relevant resources, such as CRISPR libraries. Another important function of shared resources, similar to previous CIRM Shared Labs, is to provide researchers at grantee and other local institutions with access to new technologies and equipment that may be too expensive for a single laboratory to acquire.

Using and sharing unmodified and modified hPSC lines is subject to navigating the relevant intellectual property (IP) landscape, and CIRM could have a role in streamlining resource sharing by e.g., negotiating and bundling IP rights. As an example, the Allen Institute has adopted a tiered approach that prioritizes making the technologies available to nonprofits and academia for research purposes, which is often easier to negotiate than uses for commercial purposes. The institute then works to negotiate access for companies to use these technologies for research purposes only.

Discussants emphasized the importance of creating Shared Resources Labs that will be sustainable in the long-term and suggested that CIRM should carefully consider lessons learned from the experiences of Shared Labs that were initially funded under Proposition 71. Although these labs are no longer funded by CIRM, many have adapted their operational models to continue successfully sharing resources with stem cell researchers.

Concluding Remarks

CIRM thanks the presenters and discussants for the time spent to prepare and participate in the workshop, and very much appreciates the dynamic and informative discussions and the invaluable insights provided. The outcomes of this workshop, summarized in this document, will inform CIRM as we develop and implement our strategic vision to advance world class science.

Appendix A - Acronym Definitions

ADDI	Alzheimer's Disease Data Initiative
AnVIL	Analysis Visualization and Informatics Lab-space
AMP PD	Accelerating Medicines Partnership Parkinson's Disease
CESCG	Center of Excellence in Stem Cell Genomics
CIRM	California Institute for Regenerative Medicine
CNS	central nervous system
FAIR	findable, accessible, interoperable, and reusable
GA4GH	Global Alliance for Genomics and Health
hESC	human embryonic stem cell
hiPSC	human induced pluripotent stem cell
hPSC	human pluripotent stem cell
iPSC	induced pluripotent stem cell
NAFLD	non-alcoholic fatty liver disease
NIA	National Institute on Aging
NHGRI	National Human Genome Research Institute
NIH	National Institutes of Health
SNP	single nucleotide polymorphism
UC	University of California
UCSF	University of California, San Francisco
UCSC	University of California, Santa Cruz

Appendix B: Agenda

February 24, 2022

11:00 – 11:25 AM Introduction, Background, Purpose, and Goals for the Workshop
Rosa Canet-Avilés, CIRM

Session I: Overview of CIRM-funded Research Resources

11:25 – 11:40 AM Overview of CIRM-funded Research Resources
Uta Grieshammer, CIRM

Case Studies

11:40 – 11:55 AM RFA 07-01: CIRM Shared Research Laboratory Grants and Stem Cell Techniques Course
David Schaffer, UC Berkeley

11:55 – 12:10 PM Leveraging Large iPSC Cohorts and Population Scale Stem Cell Models to Study the Effect of Genetic Variation on Cellular Phenotypes
Sulagna Ghosh and Ralda Nehme, Broad Institute

12:10 – 12:25 PM CIRM hiPSC Repository: NAFLD Lines for Disease Modeling
Jacquelyn Maher, UC San Francisco

12:25 – 12:40 PM CIRM hiPSC Repository: Machine Learning & Engineered iPSCs for Unraveling the Complex Biology of CNS Disease
Ajamete Kaykas, insitro

12:40 – 1:00 PM CIRM Genomics Stem Cell Hub: Experimental-Computational Collaboration to Characterize Cortical Organoids
Aparna Bhaduri, UC Los Angeles; and Max Haeussler, UC Santa Cruz

1:00 – 1:30 PM BREAK

Session II: Moderated Discussion – Building Shared Resources for Stem Cell-Based Modeling

1:30 – 1:45 PM Summary of Pre-Workshop Survey Results
Uta Grieshammer, CIRM

1:45 – 3:45 PM Discussion
Moderated by Uta Grieshammer, CIRM

3:45 – 4:00 PM Summary and Closing Remarks for Day 1
Rosa Canet-Avilés, CIRM

4:00 PM ADJOURN FOR DAY

February 25, 2022

9:00 – 9:20 AM Introduction to Data Infrastructure: Outcomes from September 2021 Expert Meeting
Rosa Canet-Avilés, CIRM

Session III: Data Infrastructure Overview and Examples

9:20 – 9:50 AM Data Biosphere: An Introduction
Benedict Paten, UC Santa Cruz; Brian O'Connor, Broad Institute/SageBionetworks; and Timothy Tickle, Broad Institute

9:50 – 10:00 AM Data Biosphere Q&A

User Experiences: Examples of Cloud Collaboration

10:00 – 10:30 AM Collaborating in the Cloud – AMP PD/Terra
Matt Bookman, Verily; David Craig, University of Southern California; and Barry Landin, Technome

10:30 – 10:45 AM Cloud-based Collaborative Research in Neurodegenerative Diseases
Patrick Brannelly, ADDI

10:45 – 11:15 AM NHGRI Analysis Visualization and Informatics Lab-space (AnVIL)
Ken Wiley, NHGRI/NIH; and Cornelis Blauwendraat, CARD, LNG, NIA/NIH

11:15 – 11:30 AM User Experiences Q&A

11:30 – 11:40 AM BREAK

Data Access

11:40 – 12:00 PM DUOS & GA4GH Standards
Jonathan Lawson, Broad Institute

12:00 – 12:10 PM Data Access Q&A

12:10 – 12:40 PM LUNCH BREAK

Session IV: Moderated Discussion – CIRM CNS Data Infrastructure

12:40 – 2:40 PM Discussion
Moderated by Rosa Canet-Avilés, CIRM

2:40 – 3:00 PM Summary and Closing Remarks
Rosa Canet-Avilés, CIRM

3:00 PM ADJOURN

Appendix C: Meeting Participants

Day 1 Presenters and Discussants

David Amaral, PhD, Professor of Psychiatry and Behavioral Sciences, UC Davis
Aileen Anderson, PhD, Director, Sue and Bill Gross Stem Cell Research Center, UC Irvine
Aparna Bhaduri, PhD, Assistant Professor, UC Los Angeles
Matthew Blurton-Jones, PhD, Associate Professor, UC Irvine
Bruce Conklin, MD, Senior Investigator, Gladstone Institutes
Steven Finkbeiner, MD, PhD, Senior Investigator, Gladstone Institutes
Sulagna Ghosh, PhD, Computational Biologist, Broad Institute
Ru Gunawardane, PhD, Executive Director, Allen Institute for Cell Science
David Haussler, PhD, Scientific Director, UC Santa Cruz Genomics Institute
Ajamete Kaykas, PhD, Chief Exploration Officer and Head of Neuroscience, insitro
Arnold Kriegstein, MD, PhD, Professor, UC San Francisco
Stuart Lipton, MD, PhD, Professor, Scripps Research
Jacquelyn Maher, MD, Professor, UC San Francisco
Alysson Muotri, PhD, Professor, UC San Diego
Ralda Nehme, PhD, Group Leader, Stanley Center, Broad Institute
Ruth O'Hara, PhD, Senior Associate Dean for Research, Stanford University School of Medicine
David Panchision, PhD, Chief, Developmental & Genomic Neuroscience Research Branch, National Institute of Mental Health
Sergiu Pasca, MD, Associate Professor, Stanford University
Viji Santhakumar, PhD, Associate Professor, UC Riverside
David Schaffer, PhD, Director, UC Berkeley Stem Cell Center and QB3-Berkeley
Ilyas Singeç, MD, PhD, Director, Stem Cell Translation Laboratory, National Center for Advancing Translational Sciences
George Slavich, PhD, Professor, UC Los Angeles
Michael Snyder, PhD, Director, Center for Genomics and Personalized Medicine, Stanford University
Clive Svendsen, PhD, Executive Director, Board of Governors Regenerative Medicine Institute, Cedars-Sinai

Meeting Organizers

Rosa Canet-Avilés, PhD, Vice President of Scientific Programs, CIRM
Uta Grieshammer, PhD, Senior Science Officer, Discovery Program, CIRM
Mitra Hooshmand, PhD, Senior Science Officer, Special Projects and Initiatives, CIRM
Shyam Patel, PhD, Director of Business Development, CIRM