UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

Los Angeles, CA
January 27, 2025

To the CIRM Board, Independent Citizens Oversight Committee, and the Application Review Subcommittee:

We submit this letter in complement to our application DISC4-16400 titled "High throughput, multi-modal analyses of neuropsychiatric disorder risk genes in a diverse cohort" which **received a Tier 1 score by the expert scientific review panel and the Grants Working Group**. Despite the recommendation that this application has exceptional merit and warrants funding from the reviewers, our application is not being recommended by the CIRM team for funding. We would like to highlight several reasons why this decision should be reconsidered and this important project funded.

(1) A preeminent goal of the 2020 California Proposition 14 is to dedicate "*$1.5 billion for the support of research and the development of treatments for diseases and conditions of the brain and central nervous system, such as…, schizophrenia, autism, and other diseases and conditions of the brain*." Although psychiatric disorders are highlighted as a focus in proposition 14, to date there has been little funding of research on schizophrenia and other disorders affecting mental health. Our proposal–which aims to identify the disease mechanisms underlying schizophrenia and autism by using stem cell models to functionally characterize high confidence genetic risk factors–is directly in alignment with this critically urgent Prop 14 priority and directly addresses this important gap.

(2) Increasing the representation of groups that are under-represented in biomedical research is recognized as a high priority, especially in neuropsychiatric disorders, where disparities are widespread. Our application focuses on one of the most under-represented populations in psychiatric disease research–African Americans and this was recognized as a strength of our application for many reasons. One example is in understanding of biological differences in response to drugs, which is critical for modern drug development, especially since toxicity is a major element in ending a drug development program. There is an enormous risk in ignoring this population in basic and translational research, as there are many examples where certain adverse drug reactions are more common in African Americans compared to other groups. The absence of cellular materials from African Americans further magnifies these disparities, since they preclude preclinical studies of efficacy or toxicity in this population and leave such studies performed only on white, European populations. As such, this lack of representation exacerbates existing health disparities. The CIRM/FujiFilm Cellular Dynamics biobank currently only offers 6 iPSC lines from African American donors with autism, and zero from African Americans diagnosed with schizophrenia. Our work would significantly address this disparity by

creating 45 iPSC patient lines and an appropriate number of controls that would serve as a unique resource for the entire stem cell and biomedical research community.

(3) Proposition 14 prioritizes "*direct patient engagement and outreach activities that engage California's diverse communities to ensure that all communities are aware of, and have access to, institute-funded treatments and cures shall be a priority outcome of this program*." A core aspect of our proposal uniquely leverages existing and emerging relationships with community groups to create new avenues for community outreach. We will create new and important relationships with community organizations that have a strong interest in being more included in the research process to benefit their constituents. Our efforts will combat the historically exclusion of African Americans from research and our educational outreach will remediate the lack of adequate information they have received as to the role of biomedical research in alleviating the suffering of their loved ones living with these conditions. Our proposal is the only one that brings together these community resources with our world-class expertise in neuropsychiatric disorder genetics, stem cell modeling, and AI technologies to execute one of the largest investigations to date of autism and schizophrenia disease mechanisms, and undoubtedly the first and largest of its kind in African American populations.

(4) A key, widely recognized hurdle for work with IPSC-based models of human disease, especially organoids, is increasing the throughput of these studies. We perform our analyses at large scale (>100 lines), leveraging unique assets, including advanced platforms for high throughput biological analysis, such as the STIMscope, which permits high throughput electrophysiological analysis and cell villages–to perform high throughput and highly paralleled analyses of neural models of these conditions derived from diverse individuals.

(5) Finally, the past couple of weeks have made it clear that CIRM's role in the U.S. and global biomedical research enterprise is more important now than ever. The recent disruptions in federal review processes and funding have highlighted the need for CIRM financial resources to be dispersed urgently to fill this gap and ensure that the goals of Proposition 14 are brought to fruition. While our application has been recommended for funding by the reviewers, we recognize that some easily addressable criticisms were raised about specific aspects of our proposal. Most of these concerns are unfounded and we have provided detailed responses to the reviewers' comments in Appendix 1 to address this.

We thank you for your time and attention. Since our proposal received a recommendation to be funded, we respectfully request that you seriously consider approving the funding for our ReMIND-L application. Given the clear allocation of funds and focus on psychiatric conditions in Proposition 14, we believe that it does not serve the purpose of advancing research in psychiatric

disorders, addressing health disparities, and engaging our communities in our mission, to withhold funding at this stage from an application for which funding was approved.

Respectfully Yours,

Daniel Geschwind, MD, PhD
Gordon and Virginia MacDonald Distinguished Professor, Neurology Psychiatry and Human Genetics, Senior Associate Dean and Associate Vice Chancellor, Precision Health
Director, Institute of Precision Health
UCLA

Scientific Director, UC Santa Cruz Genomics Institute
Distinguished Professor, Biomolecular Engineering, UC Santa Cruz
Adjunct Professor, Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco
Consulting Professor, Medical Informatics, Stanford University School of Medicine

Mohammed A. Mostajo-Radji, PhD
Assistant Research Scientist
UCSC Genomics Institute

Michael F. Wells, PhD
Assistant Professor
UCLA Department of Human Genetics

Appendix 1. Response to Reviewer Critiques

We emphasize that there was substantial enthusiasm for our proposal, and we put some of the summary comments below:

- ***"This proposal addresses the biggest bottleneck in the field right now: moving from gene discovery to etiological understanding."***
- ***"Interrogating mechanisms of disease deeply will yield important insights into understanding neuropsychiatric disease. This is a good team for this effort."***
- ***"The project will give important mechanistic insights to gene-to-phenotype relationships. Will establish needed diversity in iPSC-lines"***
- ***"The study follows a strong scientific rationale using novel technologies to advance mechanistic insights. "***
- ***"The author's rebuttal about the criticism around using cell lines from African American people is very strong. "***
- ***"The applicant makes a clear point that this choice addresses a major knowledge gap, is scientifically highly relevant and that nothing is known about genetic background***
- ***"Highly relevant - in fact it's such an important topic and approach that it's worth doing in a way positioned to succeed."***
- ***"Project involves global leaders across different areas and has highly synergistic parts. "Even if not all aspects will be well powered, this is well positioned to provide mechanistic insights."***
- ***"The analyses are highly advanced and innovative."***
- ***"Outstanding DEI proposal."***
- ***"Focus on samples from African Americans addresses a major gap.***
- ***"The project has a very relevant DEI part." "This is at the foundation of the proposal."***

We further emphasize the urgent need to reduce mental health disparities and to communicate science and biomedical research, including stem cell research, to our communities. Remarkably, for reasons that are not yet known, Black individuals have a substantially higher risk (> 2 fold) for being diagnosed with schizophrenia than those who are white. Yet, despite these numbers, African Americans with neuropsychiatric disease are grossly underrepresented in human cellular biobanks and psychiatric disease research more broadly. In addition to the important scientific advances and resources that we will create through this work, we also emphasize substantial community outreach and education, a feature that is often an afterthought, or entirely neglected in most biomedical research. We have dedicated resources to this effort and will develop educational materials and community forums, perform outreach and education, and in collaboration with several organizations, we will assemble a Community Advisory Board (CAB) composed of patient families to obtain feedback on recruitment practices, education materials, and dissemination of results and resources.

RESPONSE TO AIMS 1-4 CRITIQUES

The reviewers had three areas of criticism for these aims. We find that they either miss key aspects of our proposal or of the published literature cited. We address each of these in turn.

Insufficient power for idiopathic conditions

- ***Currently, the power analysis for the idiopathic cell lines does not adequately reflect complex disease finding with such sample sizes, and given the lack of effects typically seen in the transcriptome with n=45 case/control size, this really does need to be demonstrated with real analogous data***
- ***Still potentially under-powered.***
- ***Stating that the applicant's sample size is automatically powered because it is larger than some underpowered studies does not mean that it is actually adequately powered***

Our proposal includes power calculations for every major experiment and phenotype proposed, including transcriptomics and e-phys and we have substantial statistical support for these analyses (co-PI, Pimental). In each case, we show that our sample size for idiopathic lines (n=45) is sufficient to detect biologically meaningful signals and point to several published examples of conclusions being reached from much smaller *in vitro* cohorts of idiopathic cases. In contrast to the statement that 45 cases and controls is not sufficient to see transcriptomic effects, there are dozens of high-impact studies showing significant effects with far smaller samples (e.g. PMID: 31548722, PMID: 35491564, PMID: 38704507, PMID: 34493831, PMID: 30356048, PMID: 29263384, PMID: 21490598; PMID: 35017298; PMID: 36104286; PMID: 29849033; PMID: 27378147; PMID: 26186191 ...) – this critique is simply not true. We strongly contend that–given the level of detail we have provided on this issue using well-established methods for power calculation–that our proposed sample size is sufficient for the idiopathic studies.

Inadequate model for schizophrenia

- ***Reviewer is skeptical of the idea that schizophrenia-related biology/phenotype will be uncovered in iPSC cultures as they represent too early a time point***
- ***Model for schizophrenia is more challenging.***

There is over a decade of research–including in hiPSC models made by top researchers Fred Gage, Kristen Brennand, Steven Goldman, Sangmi Chung, Ralda Nehme, and many others–that strongly indicates a large degree of prenatal origins for schizophrenia risk, thus providing confidence that the established iPSC models we leverage are suitable for disease mechanism discovery (e.g. PMID: 34876311; PMID: 39245692, PMID: 22009633, PMID: 35017298). We discussed these findings at length in our proposal and focused on 3 key findings with substantial published support: (1) schizophrenia risk genes show enriched expression during prenatal brain development; and (2) there are significant overlap between the risk genes for autism, a known

prenatal disorder, and schizophrenia; (3) there is convergence at the level of the transcriptome and epigenome between autism and schizophrenia in postmortem human tissue. *Furthermore, transcriptional signatures in IPSC from SCZ patients are concordant with post-mortem adult brains, further emphasizing the relevance utility of these models (PMID: 29263384).*

In addition to these key points that we mentioned in the proposal, there is a large body of literature involving many different approaches and methods that indicate schizophrenia has substantial prenatal risk: (1) Genetic risk for schizophrenia is enriched in genes, or their regulatory elements, with maximal expression in during fetal brain development, during the peak of cortical neurogenesis, as demonstrated by multiple analyses performed in *in vivo* and *in vitro* data sets, which are highly relevant to the models being used in this application (PMID: 34172755, PMID: 34017130, PMID: 32359439, PMID: 31835028, PMID: 30545851, PMID: 29307494, PMID: 27760116). (2) epidemiological, population-level and twin studies consistently demonstrate strong maternal *in utero* (shared environmental effects) risk for schizophrenia. For example, the concordance in siblings that share the *in utero* environment (DZ twins) to siblings that are born separately, but have the same level of genetic risk sharing as DZ twins, is nearly 2-fold. (3) there are many known prenatal risk factors for schizophrenia, including maternal infections and severe maternal stress (PMID: 20955757); (4) Post-mortem studies of schizophrenia patients show disorganization in cortical neurons, abnormal layering, and altered migration patterns, suggesting prenatal disruptions (PMID: 14702264; PMID: 36223459); (5) Retrospective studies have found subtle neurodevelopmental delays in childhood among individuals who later develop schizophrenia, such as delayed motor and cognitive milestones, and atypical social behaviors (PMCID: PMC3676634). Collectively, this body of work clearly justifies our exploration of schizophrenia phenotypes using 2D and 3D *in vitro* models of the prenatal human brain. *Moreover, given the genetic overlap with ASD, the direct comparison of these two disorders is essential to understanding their shared and distinct mechanisms. Such a comparison at reasonable scale in IPSC-based models has never been done.*

Insufficient budget

- ***There is not adequate funding/effort to complete two key and arguably most important components of the project that are going to be of great interest: the cell line transcriptomics and deep learning.***

Our budget details exactly how much we intend to spend on transcriptomics, a method and we account for every sample (plus some overage) we intend to characterize. The PI responsible for this (Geschwind), has literally performed hundreds of such experiments and oversees a core facility that provides this service for dozens of laboratories around the world. Not only are we well aware of the costs, but we leverage our scale to perform these studies at costs that are apparently lower than this reviewer is used to. Furthermore, it should be noted that the cost of library preparation and sequencing continues to decrease, increasing the likelihood that we can

execute on these experiments within our proposed budget. In addition, we are, and have been for a long time, at the forefront of single cell transcriptomics in the brain and as such collaborate closely with multiple companies (10X Genomics, Illumina, BD Biosciences to name a few), which give us early access to cheaper and more throughput technologies. Our scale also results in substantial discounts as mentioned above. We address the deep learning concern in a later section of this rebuttal.

RESPONSE TO AIM 5 CRITICISMS

The reviewers have three areas of criticism regarding data analysis. We address each of these in turn.

Qualifications of the Investigators

- Novel approach, highly relevant and addressing unmet medical need. However, the applicant's expertise and ability to synthesize and analyze obtained data is of concern.
- Overall the applicants have relevant background except for the AI part.
- Applicant has limited AI-related publications and lacks demonstrated technical ability on the specific issue of setting up deep neural network topologies.
- While Machine Learning (ML) is commonly confused and conflated with AI, they are not the same thing. The repeated assertions that an AI project is feasible based on [redacted investigator's name] ML experience does not follow. While helpful, it is not sufficient.

The PI leading this aim, Prof. David Haussler, has over 40 years of experience in machine learning and AI. His published work has been cited more than 300,000 times, including five machine learning papers with over 2,500 citations each and an additional five with over 1,000 citations each. **He is also a recipient of the ACM-AAAI Allen Newell Award, granted by the American Association for Artificial Intelligence (AAAI), in recognition of his contributions to the field**.

Prof. Razvan Marinescu, in turn, has focused extensively on generative AI applications in biomedical research over the past five years, starting with his postdoctoral work at MIT's CSAIL. In addition, he has experience building multimodal AI models, with some of his published works integrating up to seven different data modalities.

We acknowledge and agree with the reviewers' assertion that machine learning and generative AI are distinct fields. *This distinction is precisely why the complementary expertise of Haussler and Marinescu makes them an ideal team*. Haussler's decades of leadership in foundational machine learning, combined with Marinescu's cutting-edge work in generative AI for biomedical applications, ensures a well-rounded and highly capable leadership team.

The conclusion that the team lacks experience is factually incorrect. Together, Haussler and Marinescu represent a uniquely qualified team with the expertise to successfully execute this project.

Lack of Preliminary Data and Feasibility

- ***The novel Artificial Intelligence (AI) analysis claims to be 10x+ faster than experts in the field have ever experienced. The approach would violate findings of the largest studies on complex disease with regard to effect sizes. While it might find convergent mechanisms of disease, the capstone analysis does not show any preliminary results, intermediate constructs, or relevant success to support this.***

- A lot of that cross-cutting depends on a highly speculative AI model that lacks any preliminary data.

- Although a schematic AI model can be written down, that is miles away from demonstrating that it is possible. Direct evidence from the exact data is what that's needed. Failing that, the next step is to move to analogous data sources.

- Weakness is aim 5 which is still under-developed and with little preliminary data.

- However, the most nominally synergistic aim - the deep learning model that will synthesize everything together is very vaguely described. It lacks the preliminary data found in essentially every other grant, even those much smaller, and thus is both relatively and absolutely not demonstrated to be synergistic.

Generating data for training the foundation model is precisely the goal of aims 1-4, so it is impossible for us to provide preliminary results in aim 5 without having acquired the data first. However, it is worth noting that Prof. Marinescu has developed algorithms capable of integrating up to seven data modalities in neuroscience, well beyond the four modalities proposed here.

Regarding the feasibility of building the foundation model itself, it is not as risky as some reviewers suggest. As noted in our proposal, we will start from an ***already-trained*** LLAMA model and only fine-tune it with a low-rank adaptation protocol on our data. This strategy has already been successfully done not only by large industry players, but also academic groups, as exemplified in the recently published single-cell foundation models. Source code for doing fine-tuning with LoRA is already available online, and most of our efforts will actually be on pre-processing and cleaning the data, and generating the standardized embeddings that can be fed into the LLAMA model for fine tuning. Our team has 20+ years of experience working with and cleaning such data. Given this experience and preparation, we are confident that the proposed project is not only feasible but also positioned for success.

Lack of Resources

- ***The applicant will quite likely have an unfinished deep learning model due to insufficient resources, and hence the main piece of this proposal that is supposed to unify the findings***

is not functional, preventing the extraction of coherent disease mechanisms across diverse lines.
- There is not adequate funding/effort to complete two key and arguably most important components of the project that are going to be of great interest: the cell line transcriptomics and deep learning. Other aspects of the project are very strong, but the issue is these weak areas are rate-limiting for the overall impact of the project.
- Looking at what is proposed for Aim 5, which is both novel and speculative, 8% of a professor (+85% of a junior engineer) is insufficient for the inevitable process of refining the base model. Based on related projects, a minimum of three (world-class and experienced) Full Time Equivalents (FTEs) who are not torn in multiple directions are needed to make headway on such a major project.

The reviewers' concerns about resource constraints seem to overlook our ability to optimize and scale resources effectively by balancing cost with computational complexity. We are fully aware of the high computational demands associated with generative AI, particularly GPU-intensive tasks. To address these challenges, we leverage extensive collaborations across multiple projects. By integrating data from other initiatives, we not only distribute costs through joint analyses and shared compute runs but also enhance the robustness of our results. Access to diverse datasets further strengthens our ability to uncover meaningful patterns specific to this CIRM project, leading to deeper insights and better outcomes. In addition, for fine-tuning large language models, we will also use low-rank adaptations such as LoRA, which have proven extremely successful in reducing the GPU requirements needed for fine-tuning LLMs.

Our team leads and participates in several major initiatives that directly support this work, including two additional ReMIND applications, the SSPsyGene consortium, PsychENCODE, and others. Additionally, we benefit from access to scalable computational infrastructure through partnerships like the NSF's National Research Platform (NRP). This allows us to train large-scale generative AI models on more than 1,000 GPUs without maintaining our own GPU cluster. For example, graduate students and staff in our Braingeneers group regularly design and train advanced models using NRP's resources, showcasing our ability to efficiently tackle complex computational challenges.

We also emphasize that this project is not an isolated effort; it leverages and is strengthened by the collective expertise, infrastructure, and history of success at the UCSC Genomics Institute. We have a proven track record of overcoming computationally intensive challenges with limited resources. For instance, the first draft of the human genome was assembled by a single graduate student, Jim Kent, while the telomere-to-telomere chromosome assembly was led by then-research scientist Karen Miga. Dr Marinescu's student Najmeh Mashhadi has been able to train 15 different neural networks comprising state-of-the-art diffusion models, GANs, U-Nets and ResNets across 10 different datasets and integrate them into a directed graph of AI models. These

achievements highlight our team's ability to deliver transformative results under resource constraints.

Our efforts are further bolstered by UCSC's leadership in cutting-edge initiatives like the Generative AI Center, one of the first of its kind globally. This center brings together experts across academic disciplines, fostering interdisciplinary collaboration and innovation in generative AI. As active participants in this center, we regularly engage with its network to incorporate the latest advancements into our work.

With a team of two engineers, one graduate student, and four faculty and research staff members, we are confident in our ability to meet and exceed the demands of this project. By strategically leveraging our team's expertise, institutional resources, and external partnerships, we ensure that resource constraints will not impede the success of this project. We are fully equipped to deliver innovative, impactful, and cost-effective results.

Overall summary of response to aim 5 criticisms

This project will generate incredibly valuable data, and we are committed to ensuring that its analysis is both rigorous and impactful. We do not believe we will fail because we relied solely on experimental methods that previously fell short. Our approach is never to "shoot for the stars or fail." With decades of experience in machine learning, we always ensure that new methods are supported by fallback options: simpler, well-established machine learning and statistical approaches. This dual-layered strategy is a fundamental part of our process.

At a minimum, we will perform standard machine learning and statistical analyses to establish a baseline and evaluate whether the newer methods we propose offer any improvements. We have extensive experience with this approach, and we deeply value the experimental data itself, especially in a project like this, where the data's value is exceptionally high. Once baseline results are established, we will iteratively refine our analyses using innovative and customized methods.

There is an urgent need for the development of therapeutics for neuropsychiatric disorders, and success on this front will require bold ideas and actions, such as the cutting-edge AI methods we propose. We categorically reject the "don't try anything new unless you've already proven it works" mentality that has stifled innovation, discouraged exploration, and resulted in our lack of therapeutic options for these conditions.

CIRM has an opportunity to support our forward-thinking perspective that balances tried-and-true methods with the potential breakthroughs that only cutting-edge approaches can deliver. By including Aim 5 in the current grant proposal, we have taken a bold step towards building something that can provide a truly ground-breaking contribution to medical researchers and to society as a whole. While not easy, in the updated proposal we have provided extreme clarity on how the foundation model will be built, including precisely defining the latent embeddings for each

modality, the training and inference regime of the foundation model, and the mathematical formulas for the drug optimizations, and we have no doubt that our team can successfully deliver the proposed project.