MEMORANDUM

Date: January 26, 2011

From: Alan Trounson, PhD
CIRM President

To: Independent Citizen's Oversight Committee

Subject: Extraordinary Petition for Application RT2-02054 **(LATE SUBMISSION)**

Enclosed is a petition letter from Dr. Alexander Urban of Stanford University, an applicant for funding under RFA 10-02, CIRM Tools and Technology II Awards. This letter was received at CIRM on January 26, 2011 and we are forwarding it pursuant to the ICOC Policy Governing Extraordinary Petitions for ICOC Consideration of Applications for Funding.

To: the Chairman of the ICOC and the President of CIRM
Application RT2-0 2054 for CIRM Tools and Technology Awards, RFA 10-02

Extraordinary Petition for ICOC Consideration of Applications for Funding

This RFA calls for novel tools and technology for the use in stem cell research and stem cell applications (i.e. regenerative medicine, RM). We propose to introduce to RM a revolutionary new technology, Next-Generation DNA Sequencing (NG-Seq), to carry out whole-genome comprehensive and high-resolution analysis of DNA sequence variation (SeqVar analysis). By bringing this new technology into the field of RM we will introduce a new tool that will allow the community of SC and RM researchers to address a critical and currently unmet need of SC research, that of comprehensive monitoring of genomic sequence variation and stability.

It is known that SC lines which have been kept in culture for an extended period of time will acquire large chromosomal aberrations. It is also known that the human genome is much more variable than was thought until recently and that large chromosomal aberrations present only a small fraction of the possible cumulative DNA sequence variation in a genome. Stem cell lines with a genome that has acquired sequence variation and that has generally become unstable are bound to yield grossly misleading results when used in research into the biology of stem cells; also they could cause grave harm when used in tissue grafting, since genomes with a large degree of acquired sequence variation and instability are known to make cell lines prone to develop cancer-like characteristics. However, the technologies that are currently the standard in stem cell research and RM for assessing the stability and integrity of stem cell culture genomes (karyotyping, FISH and recently some microarray based methods) possess only low or very low resolving power and are bound to fail to detect all except a few percent of genomic DNA sequence variation. The approach proposed by us would be a large step towards remedying this situation. As this is a new technology and approach for the field of RM and stem cell research we were aware that the page limitations of the main application did not allow for describing all aspects with sufficient detail. The comments of the reviewers have now highlighted the areas in our proposal where more detail of description would be most beneficial.

The reviewer's comments are inserted below, underlined and in italics.

"The reviewers acknowledged that this application addresses the need in the regenerative medicine field to assess the genomic integrity of human cells used for therapies. They, however, did not find the proposed approach particularly innovative, since others are already applying high throughput sequencing technology to stem cell biology." We are pleased that the reviewers acknowledge the need for advanced genomic analysis in stem cells and RM. High-throughput/next-generation DNA sequencing (NG-Seq) is a powerful new technology that is only now becoming more widely available. Its standards are constantly evolving and it will take the work of many researchers over many years to fully establish this technology in a manner that will unlock its potential also for researchers who are not experts in genome sequence analysis. The methods proposed by us are at the very cutting edge of human genome analysis. The market leading Illumina HiSeq 2000 instruments that are at our disposal represent the latest in genome analysis technology, having just become available during 2010, with their use still being limited to a few leading genome centers. Furthermore we have at our disposal the most recent computational algorithms for the analysis of HiSeq data. Most of these algorithms and software tools have become available as a result of the NIH sponsored *1000 Genomes Project*, the preliminary results of which were just published in late 2010 or are currently in press. We were and continue to be actively and significantly involved in the *1000 Genomes Project*, a result of our several years of expertise in the field of high-resolution genome analysis. While we are aware that others are also using NG-Seq, applying stem cell DNA samples to a NG-Seq instrument does not yet immediately reach the comprehensive nature and high standards of the various and integrated types of genome analyses which we are proposing. We will apply multiple different software tools to each DNA sequencing dataset. There is not yet a single computational approach or integrated suite of analytical algorithms that can uncover all the

different types and sizes of possible DNA sequence variation. It is our particular expertise to combine SeqVar analysis algorithms to create the ideal analytical "toolkit" for a given task.

"Based on the preliminary data, reviewers were confident that the applicants would be able to carry out the proposed experiments, which mostly involve next-generation sequencing. While reviewers supported the rationale for determining whether hPSC have acquired a detrimental genomic sequence variation load, they were strongly concerned that this proposal lacked a clear vision as to how the collected information will be utilized to come to a clear conclusion. The applicants will arrive at a stability score, but did not explain how they will determine what amount or type of sequence variation will be considered detrimental. Experiments that address how a certain stability score correlates with the cells' phenotype or safety profile are critical to validate the meaning of the score, but were not proposed. Reviewers expressed the opinion that measuring genetic drift over a large number of population doublings is not of great significance by itself,…" We are thankful for the confidence expressed by reviewers in our ability to carry out the proposed experiments and the recognition of the need for determining whether a stem cell line has acquired detrimental sequence variation. This detrimental sequence variation could come in the form of individual mutations in critical parts of the genome or as an overall load of sequence variation and a general instability of the genome that leads to an accelerating load of additional mutations. Defining the Quality-Control Standard Operating Procedure (QC-SOP) for the use of NG-Seq technology in stem cell research and RM will require to assess the effect of stem cell culture specific aspects on DNA sequencing, such as the potential for genomic heterogeneity within a given cell culture, in particular at high passage numbers, or the potential for activation of mobile, transposable genomic elements. This can successfully be carried out to conclusion within the duration of the proposed project.

Furthermore we propose to develop, as part of the QC-SOP, a method to determine a Stem Cell Genome Stability Score which can then be used by stem cell researchers to determine whether a given stem cell line is still suitable for further use. Our project will lay the groundwork for the metric of determining the Stem Cell Genome Stability Score but to come to a final conclusion as to which SeqVar will have which effect on stem cell biology, for all possible SeqVar and in all possible applications of RM, is far beyond the scope of this or any single research proposal. We will introduce the technology and define the basic variables for the analysis of NG-Seq based SeqVar in stem cell biology and RM upon which the stem cell research community can then build in the years to come.

The basic variables for Stem Cell Genome Stability Score will include: the length of a sequence variant, whether the SeqVar changes the copy number of the given allele and whether the SeqVar affects genes, conserved or regulatory regions of the genome. Genes that are judged to be of critical importance for stem cell function and differentiation, based on the information already produced by others and by our own gene expression experiments, will be given an increased weight in this calculation. We agree with the reviewers that eventually the association of a genomic SeqVar to a functional phenotype will be the defining criterion for determining whether a SeqVar is detrimental. We would like to point out that we did propose (and budget for) experiments to lay the foundation of this association between genotype and phenotype: we will carry our gene expression analysis by RNA-Seq and will correlate the changes in gene expression with the detected SeqVar, along the principles described in our recent publication in Science [Kasowski, Grubert et al., Science, 2010]. Also, the large dataset collected by us during this project will allow us to determine at which point in time the acquisition of SeqVar in a stem cell genome will accelerate, which has not been done before and which should be in itself, regardless of the phenotypic effect of individual SeqVar, an important parameter for assessing the status of a stem cell line.

"…and felt that the proposal would have benefited from an elaboration on the final paragraph in which the applicants briefly mention additional benefits that might arise from this research." The additional benefits from introducing NG-Seq to stem cell research and RM

which we briefly mentioned were: determining the best match for transplantations, detect possible infections of the tissue culture, extend the QC-SOP to other types of stem cells beyond embryonal and induced pluripotent SC, generate a basis for epigenomic analysis and monitor the effects of potential stem cell genetic engineering projects. Just as in the main proposal we do here not have space to elaborate on these points but would like to point out that all these benefits would result from introducing the technology to stem cell research without adding cost or effort to our proposal.

"With regard to the details of the proposed experimental approach, reviewers cautioned that the applicants did not consider how long the hPSC lines to be analyzed have been in culture at the start of the experiments, which could influence the practical significance of the stability score." We thank reviewers for pointing out to us the need to clarify this aspect. We will indeed include the already pre-existing number of passages for a given stem cell line in our total count of passage numbers. For the embryonic stem cell lines we will start the project with lines with the lowest possible passage numbers while for induced pluripotent stem cell lines this issue will not be of significant concern since the lines will have been freshly generated.

"The PI and research team were judged to be well qualified to execute this proposal. The applicant is a young PI and has strong expertise in the sequence analysis aspect of the proposal. A Partner-PI, with whom the PI has successfully worked together in the past, is included in this proposal, but reviewers felt the proposed collaboration required more justification, since both labs will perform similar experiments." We thank the reviewers for their positive assessment of our abilities and skills with regard to the requirements of our proposal. We are glad to provide further justification for the proposed collaboration between the laboratories in California (Urban, Stanford) and Germany (Korbel, Heidelberg). We consider this aspect of our proposal both essential for the success of the proposed project and a particular strength of our application. We have a history of working together productively in the field of high-resolution analysis of the human genome going back several years, beginning when we were both doing research at Yale. We have established a very productive division of labor where one side (California) carries out more of the experimental work and the other (Heidelberg) is in charge of more of the data processing work while the final interpretation of results is conducted jointly. As also depicted by the flow-chart on page 14 of our proposal (and reflected in our proposed budget), the group in California would carry out all the experimental stem cell work and 2/3 of the DNA sequencing work; Heidelberg would carry out the remaining 1/3 of DNA sequencing (this distribution of DNA sequencing also adds a quality control feature for the raw DNA sequencing data). Raw DNA data is pre-processed at both sites while the bulk of computational sequence variation detection is carried out at Heidelberg (and not at Stanford, the statement to this end on page 14 is a typographical error). Detected DNA sequence changes are experimentally confirmed at Stanford, and the final conclusions are reached jointly.

"Although two stem cell experts have been recruited to the team, reviewers were not convinced their level of commitment was sufficient to support that aspect of this application."

We would like to point out that we have recruited not two but three collaborating laboratories with extensive expertise in stem cell work (please see the three letters of collaboration). With one of those groups (the laboratory at Yale) we (Urban) already have an ongoing collaboration in the field of human induced pluripotent stem cells (iPSC). With the second of the groups (the neurobiology laboratory at Stanford) we are currently setting up several projects. And we have a particularly strong connection to the third group that has pledged to support the stem cell aspect of our proposal, the Snyder laboratory at Stanford, which is directed by the former doctoral and postdoctoral mentor and advisor of one of us (Urban) and we both (Urban and Korbel) have numerous completed and ongoing collaborative projects and shared publications with this research group. We therefore are confident that we will have access to the stem cell lines and training in stem cell work that will be necessary for this project.