California Institute for Regenerative Medicine
Overview of Current ESC Databanks

Introduction

The following overview provides a high-level summary of current activities related to human embryonic stem cell (hESC) databanks and their characterization efforts. It is designed to provide some background for our discussion and is not intended to be comprehensive or exhaustive.

I. Current Status

A. Overview of hESC databanks

1. National Institutes of Health - Stem Cell Registry

- The National Institutes of Health (NIH) Stem Cell Registry is a PDF file providing information such as passage number, karyotype, and status for stem cell immuno markers, on lines approved for federal funding (1).

- The NIH Stem Cell Unit consists of four scientists dedicated to characterizing and comparing the available federally approved hESC lines. The primary aim is to establish standards for all aspects of the culture process as well as quality control and monitoring. In the future, three specific lines will be selected for use in larger-scale gene expression studies (2).

2. WiCell Research Institute - National Stem Cell Bank (3)

- In October 2005 the NIH awarded $16.1 million over four years to the WiCell Research Institute in Wisconsin to fund a National Stem Cell Bank (NSCB) to perform a more detailed characterization on hESC lines approved for federal funding (3).

- The goal of establishing a National Stem Cell Bank is to consolidate many of the federally funded eligible hESC lines in one location, reduce the costs that researchers have to pay for the cells, and maintain quality control over the cells. This resource is intended to enable researchers (at the NSCB) to fully analyze, characterize, and control the quality of approved cell lines that are shipped out to investigators. The hope is this will optimize and standardize the techniques used for comparing the properties of these hESC lines (4).

- The National Stem Cell Bank's goal is to provide scientists affordable and timely access to federally approved hESCs and other technical support that will make it easier for scientists to obtain the cell lines currently listed on the NIH Human

Embryonic Stem Cell Registry (1). The Stem Cell Bank will also make efforts to ensure consistent quality of the lines by analyzing and comparing existing cell lines; documenting the growth characteristics of cell lines; assessing the cells' genetic stability; and determining the molecular background and basic characteristics of the different cell lines (3).

3. **International Stem Cell Forum (5)**
   - **International Stem Cell Forum (ISCF) Characterization Project -** In 2003, the ISCF started work to develop a set of global criteria for the derivation, characterization, and maintenance of human stem cell lines. ISCF has invited research groups worldwide to submit stem cell lines from their laboratories for inclusion in a large-scale characterization project using ISCF criteria.
     - The ISCF is setting up a registry to record details of these stem cell lines, to be made available online to the international scientific community. The registry will contain the data generated by the characterization project and is expected to be launched later in 2006 when the project nears completion.
     - The characterization project will be organized on a 'hub-and-spoke' principle. The hub, the UK Stem Cell Bank (discussed below), will collect and prepare antibodies in research-grade facilities for distribution to participating laboratories throughout the world. These laboratories will use the antibodies in specified assays of their cell lines under defined conditions.
     - For each stem cell line, characterization will involve:
       - Establishing the expression patterns of selected surface antigens.
       - Establishing the expression patterns of genes marking undifferentiated stem cells and specific lineages of differentiation.
       - Comparing the changes in marker expression in response to a simple differentiation protocol.
       - Establishing the degree of correlation between expression of different potential markers of undifferentiated ES cells.
       - Assessing "epigenetic" status with respect to genes subject to imprinting.
       - Providing a DNA fingerprint of each line.
       - Assessing microbiological status, particularly with respect to endogenous retroviruses.
       - Assessing the histology of the cell line's xenograft tumors.
       - Collating information regarding origins, karyotype, culture conditions, etc.

   - **UK Stem Cell Bank (6) -** The Medical Research Council (MRC) has awarded a £2.6 million contract to the National Institute for Biological Standards and Control (NIBSC) to set up a UK Stem Cell Bank. The Bank will provide a vital resource centre to support progress towards realizing the potential of stem cell research to deliver advances in treating a wide range of diseases.
     - The Bank is part of a national stem cell initiative being developed by the MRC, along with the Department of Health, Human Fertilization and Embryology Authority, Medicines Control Agency, other research councils

and relevant charities. It will hold new and existing adult, fetal, and embryonic stem cell lines, making them available to researchers and private companies from the UK and abroad. The MRC is meeting 75% of the costs of the initial three year contract to establish the bank, and the Biotechnology and Biological Sciences Research Council is contributing 25%.

- The guiding principles of the bank include:
  - Progress with responsibility.
  - Establish standards and controls.
  - Improve research efficiency.

4. National Institute of Diabetes and Digestive and Kidney Disease - Stem Cell Database (7)

- The Stem Cell Database (SCDb) is one component of the Stem Cell Genome Anatomy Projects (SCGAP) supported by the US National Institute of Diabetes and Digestive and Kidney Disease (NIDDK). The database is made available on the Web by the Department of Molecular Biology, Princeton University (8).

- The SCDb contains a data set representing the molecular phenotype of the hematopoietic stem cell. In addition to an overview of the project and an introduction to hematopoietic stem cells, information is provided on materials and methods used (including high-density arrays, semi-quantitative PCR, and low-density hybridization screens), along with a hierarchy overview, transplantation data, an explanation of the database, library statistics, and sequence homologies.

- From a characterization standpoint, a genome-wide gene expression analysis was performed in order to define biological regulatory pathways in stem cells as well as their global genetic program. The complete data, available in a biological process-oriented database, represent the molecular phenotype of the hematopoietic stem cell. (7)

5. National Institute of Diabetes and Digestive and Kidney Disease - Stromal Cell Database (9)

- The Stromal Cell Database (StroCDB) is another component of the Stem Cell Genome Anatomy Projects supported by the NIDDK. Like SCDb, this database is made available on the Web by the Department of Molecular Biology, Princeton University (8).

- The StroCDB contains a data set representing the first molecular profile of a stem cell supportive cell line, AFT024, derived from the murine fetal liver microenvironment. In addition to a perspective on stem cell biology in vivo and in vitro, information is provided on materials and methods, along with supplementary data.

- To determine the molecular signals elaborated, a functional genomics approach was used that coalesces extensive sequence mining of a subtracted cDNA library, high-density array hybridization and in-depth bioinformatic analyses. The data has been assembled into a biological process oriented database, and represent the first molecular profile of a candidate stem cell niche.

6. Burnham Institute - Stem Cell Community (10)

- This stem cell database was developed in 2005 by the Burnham Institute and currently describes more than 240 stem cell lines and categorizes them into groups that are and are not approved for study with U.S. government funds. To characterize the cells and fill out portraits on them, site curators gather microarray measurements of gene activity and data on genetic variability (11).

- Users of the database can find the following information:
  - Where the cells came from.
  - What protein markers they express.
  - Whether these cells have been frozen.
  - How the cells were grown and nurtured (e.g., with mouse feeder cells).

- The site also includes a Community Information section where researchers can track down courses on growing and deriving stem cells or view a news archive.

7. European Consortium for Stem Cell Research (12)

- This new project, headed by Simon Tomlinson at the University of Edinburgh, aims to establish a stem cell database (Stem DB) containing a wide range of information about stem cells - from basic biology to clinical applications. The data will be derived from new findings generated by the EuroStemCell consortium and existing published data (12).

- "Workpackages" are designed to study the fundamental biology of stem cells and are focused in the following five areas:
  - Identification and isolation of stem cells from neural lineages
  - Identification and isolation of stem cells from non-neural lineages
  - Lineage analysis and differentiation potential
  - Self renewal and up-scaling
  - Control of differentiation

- This project will facilitate comparative analysis of the stem cell molecular profiling data generated by the EuroStemCell workpackages, and foster bioinformatics collaborations among different participating groups.

8. **National Institute of Diabetes and Digestive and Kidney Disease - The Stem Cell Genome Anatomy Project (13)**
   - The Stem Cell Genome Anatomy Project (SCGAP) initiated by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NIH) has developed a web portal to provide access to results from studies on stem and progenitors cells for a variety of organ systems including blood, liver, bone, prostate, bladder, kidney, and gut.
     - The web site org provides links and information to the current seven groups funded to form a research consortium.
     - A "portal" tab leads to a Gene Search that allows one to view the expression profile for a gene of interest in three different species (human, mouse, and zebra fish) using different technologies (e.g., microarrays).
     - The site compiles responses gathered in real time from computer servers distributed across the United States and Australia and uses an architecture designed to add new sites easily.
     - A goal of the SCGAP web portal is to become an interactive resource for the international stem cell community.

B. Challenges with Establishing and Maintaining an ESC Databank

1. Limitations of Databanks

- Some databanks, such as the NIH's Stem Cell Registry and WiCell's National Stem Cell Bank, are limited to providing information only on federally funded hESC lines (1)(4). Other databanks, such as those to be hosted by the ISCF and European Consortium for Stem Cell Research are currently still under development (5) (12).

2. General Limitations of Databanks

- Data Collection - How do you get people to come to you and submit information?
  - Incentives to submit information, such as open and equitable access to the data, are designed to encourage cooperation and collaboration.

- Data Relevance - How do you ensure information remains current?

3. Characterization of ESC Lines

- Lack of standardized protocols.
  - Some organizations have very elaborate and detailed characterization procedures, such as International Stem Cell Forum (ISCRF), whereas others currently do not.
  - There is also a lack of consistency in characterizing ESC lines, as most databanks request that researchers perform their own characterization.
- Lack of a cost effective approach for characterization.

- Length of time required for characterization.
  - Characterization can be a lengthy process. For example, the International Stem Cell Forum has pushed back the launch of its databank due to delays in the characterization of its lines.(5)

II. Potential Roles for CIRM

1. Topics for Discussion

- Is there a need for CIRM to provide an information resource on human embryonic stem cell lines?
  - If yes, what type of information resource would it be and what kinds of information should be included?
  - Would we provide the resource to all interested parties or only for our grantees?

- Should CIRM create its own ESC characterization database using a best practices approach?
  - CIRM can establish its own characterization protocols?
  - CIRM can establish its own guidelines for access to the database?
  - Do we make participation mandatory for our grantees who develop cell lines?

- Should CIRM provide a web page offering links to sources of ESC line and other information?
  - CIRM could partner with an existing group or groups and support further development of an existing resource.
    - This would allow CIRM to take advantage of existing infrastructure and leverage existing expertise.
  - What should CIRM's role be, if any, in describing and validating link content?

References

(1) http://stemcells.nih.gov/staticresources/research/registry/PDFs/Registry.pdf
(2) http://stemcells.nih.gov/research/nihresearch/scunit/
(3) http://www.nih.gov/news/pr/oct2005/od-03.htm
(4) http://www.wicell.org/
(5) http://www.stemcellforum.org.uk/
(6) www.nibsc.ac.uk/divisions/cbi/stemcell.html
(7) http://stemcell.princeton.edu/v1/
(8) http://bioresearch.ac.uk/browse/mesh/D020869.html
(9) http://stromalcell.princeton.edu/
(10) www.stemcellcommunity.org
(11) Science, vol 311, Feb 3, 2006
(12) http://www.eurostemcell.org/index.htm
(13) http://www.scgap.org