Stem Cell Reports
Resource
OPEN ACCESS

Integrated Genomic Analysis of Diverse Induced Pluripotent Stem Cells from the Progenitor Cell Biology Consortium

Nathan Salomonis,[1,16] Phillip J. Dexheimer,[1,16] Larsson Omberg,[2] Robin Schroll,[3] Stacy Bush,[3] Jeffrey Huo,[4,5] Lynn Schriml,[6] Shannan Ho Sui,[7,8] Mehdi Keddache,[9] Christopher Mayhew,[10] Shiva Kumar Shanmukhappa,[11] James Wells,[10] Kenneth Daily,[2] Shane Hubler,[2] Yuliang Wang,[12] Elias Zambidis,[4,5] Adam Margolin,[2,12] Winston Hide,[7,8,13] Antonis K. Hatzopoulos,[14] Punam Malik,[3] Jose A. Cancelas,[3,15] Bruce J. Aronow,[1,17] and Carolyn Lutzko[3,15,17,*]

[1]Department of Biomedical Informatics, Cincinnati Children's Hospital, Cincinnati, OH 45229, USA
[2]Sage Bionetworks, Seattle, WA 98109, USA
[3]Division of Experimental Hematology and Cancer Biology, Cincinnati Children's Hospital, Cincinnati, OH 45229, USA
[4]Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[5]Division of Pediatric Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD 21205, USA
[6]Department of Epidemiology and Public Health, Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA
[7]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[8]Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[9]Division of Human Genetics
[10]Division of Developmental Biology
[11]Division of Pathology
Cincinnati Children's Hospital, Cincinnati, OH 45229, USA
[12]Computational Biology Program, Oregon Health & Science University, Portland, OR 97239, USA
[13]Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield S10 2HQ, UK
[14]Division of Cardiovascular Medicine, Departments of Medicine and Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37232, USA
[15]Hoxworth Blood Center, University of Cincinnati, Cincinnati, OH 45229, USA
[16]Co-first author
[17]Co-senior author
*Correspondence: carolyn.lutzko@cchmc.org
http://dx.doi.org/10.1016/j.stemcr.2016.05.006

SUMMARY

The rigorous characterization of distinct induced pluripotent stem cells (iPSC) derived from multiple reprogramming technologies, somatic sources, and donors is required to understand potential sources of variability and downstream potential. To achieve this goal, the Progenitor Cell Biology Consortium performed comprehensive experimental and genomic analyses of 58 iPSC from ten laboratories generated using a variety of reprogramming genes, vectors, and cells. Associated global molecular characterization studies identified functionally informative correlations in gene expression, DNA methylation, and/or copy-number variation among key developmental and oncogenic regulators as a result of donor, sex, line stability, reprogramming technology, and cell of origin. Furthermore, X-chromosome inactivation in PSC produced highly correlated differences in teratoma-lineage staining and regulator expression upon differentiation. All experimental results, and raw, processed, and metadata from these analyses, including powerful tools, are interactively accessible from a new online portal at https://www.synapse.org to serve as a reusable resource for the stem cell community.

INTRODUCTION

Pluripotent stem cells (PSC) have been used to study human development, model disease, and generate cellular tools for regenerative medicine. Human embryonic stem cells (hESC) have been considered the functional, genetic, and epigenetic gold standard in the field (Thomson et al., 1998). Methods of somatic cell reprogramming to generate induced PSC (iPSC) (Takahashi and Yamanaka, 2006) are continually being improved and have enabled the generation of iPSC using a variety of somatic cell sources, gene combinations, and methodologies. However, due to the intensive resources required for iPSC generation and characterization, direct comparisons of iPSC generated using a wide range of technologies and cell sources from multiple independent laboratories have rarely been performed, making it unclear whether all methodologies produce iPSC with a similar quality and stability.

A variety of studies have compared the expression profiles, pluripotentiality, and genetic and epigenetic stability of hESC and iPSC including lines generated using different strategies, distinct parental somatic cell types, or reprogramming methods (Bock et al., 2011; International Stem Cell Initiative et al., 2007; Müller et al., 2011; Rouhani et al., 2014; Schlaeger et al., 2015). However, these have been limited to a few variables, have multiple methods or laboratories collecting and processing samples, and typically employ a single genomics platform. "Multi-omics" analyses have proved to be essential in deciphering complex gene regulatory programs, as demonstrated by analyses of iPSC reprogramming transitional states (Clancy et al., 2014; Lee et al., 2014; Tonge et al., 2014).

The Progenitor Cell Biology Consortium (PCBC) of the National Heart, Lung and Blood Institute was founded to study iPSC reprogramming and differentiation and develop strategies to address the challenges presented by the transplantation of these cells. These questions include, but are not limited to: (1) Do iPSC consistently generate all three germ layers? (2) How prevalent is copy-number variation (CNV) in iPSC generated using different reprogramming methodologies? (3) Do different reprogramming methods affect global methylation, gene, splicing and microRNA (miRNA) expression profiles? (4) Can aberrant PSC gene regulation be identified on a global basis? (5) How do variables such as X-chromosome inactivation (XCI) affect iPSC quality, stability, and differentiation potential? To advance these goals, the PCBC developed a Central Cell Characterization Core and Bioinformatics Core to perform standardized and comprehensive characterization of iPSC generated using different somatic cell sources, methodologies, and vectors. The characterized iPSC are being made available through WiCell Research Institute.

Using integrative analyses across genomic analysis platforms, we present comparative results on phenotype, genetics, epigenetics, and gene regulation for a diverse panel of iPSC and hESC. Standardized methods and strict control of reagents during cell culture, sample collection, and assay performance were used to evaluate the innate potential and limitations of these cells with fewer confounding factors. Our use of this uniform analytical methodology allowed us to discover candidate regulators of the fate of reprogrammed cells. To maximize the utility of this resource, we developed an interactive open data portal for access to the raw data, metadata, results, and protocols from these experiments for further analysis (https://www.synapse.org/PCBC).

RESULTS

Study Design and Synapse Analysis Portal

An overview of the study is presented in Figure 1. The evaluation of iPSC from multiple laboratories and methodologies required highly structured cell-line annotations and well-documented protocols to make comprehensive comparisons possible. Metadata standards were developed to capture the origin of each line, starting cell type, donor demographics, and reprogramming parameters (derivation method, vector type, reprogramming genes, culture conditions). These metadata were provided by the originating laboratory and confirmed and augmented with in vitro genetic and experimental characterization of the line. RNA sequencing (RNA-seq) was performed at an acceptable depth to facilitate accurate gene-expression quantification (Supplemental Experimental Procedures). To facilitate use of the protocols, genomic analyses, and metadata produced through this effort, we developed a sophisticated interactive data portal, the interface of which is exemplified in Figure 1. In addition to integrated provenance annotations for every raw data file, script, or processed result file, data can be queried through an interactive heatmap viewer that displays and inter-relates gene expression, DNA methylation, and miRNA expression for queried genes, pathways, and gene signatures produced in the analyses described here. These signatures have been further propagated into ToppGene (Chen et al., 2009) for interactive queries. Synapse IDs are included to access the resources, data, metadata, ontologies, and other information through the Synapse online repository.

Screening of Lines

The data from the first 64 lines (58 iPSC and 6 hESC) enrolled in the study are presented here with their characteristics outlined in Figure 2A (details in syn2767694). All lines completed a standardized screen to ensure they met a basic set of criteria. This included self-renewal in defined feeder-free conditions, expression of markers of pluripotency and a lack of expression of markers of differentiation, a normal karyotype, and the ability to grow sufficient quantities of cells for the analyses (Tables S2 and S3; Figure S1). Overall, 6 hESC and 35 iPSC (64%) met these criteria and 23 iPSC did not (36%) (Table S4). Abnormal karyotypes were observed in seven lines (Table S5), with karyotypes for all lines available (syn2679104). The most consistent flow cytometry anomalies were TRA-1-81 and TRA-1-60 below 90% or an increase in SSEA-1 above 5% (Figure 2B). Due to contamination, difficulty in expanding cells, and/or abnormal karyotype, not all lines were included in functional pluripotency assays.

Pluripotency Analysis

Pluripotency was evaluated in a teratoma assay on 49 lines. Forty-six of the lines met the screening criteria outlined in Table S3 and 45 of these lines generated teratomas. Three lines did not meet the PSC screening criteria with decreased expression of self-renewal markers and increased differentiation in culture (SC12-021, SC12-023, and SC14-082), and all three successfully generated teratomas. All teratomas were scored by a clinical pathologist, and representatives of all three embryonic germ layers were identified in all tumors (detailed information is available at Synapse syn2882785). We performed immunostaining analysis on teratomas from a subset of lines to confirm pluripotency (muscle-specific actin [MSA], neurofilament, and α-fetoprotein) and OCT4 to evaluate the presence of undifferentiated PSC (Figure S1). This included two lines that did not meet the screening criteria and independent iPSC from the same donor as controls (Table S7), and three teratomas that