Title: A new NHGRI Sample Repository for Human Genetic Research collection of induced pluripotent stem cell lines.

Authors: Tatyana Pozner[1], Christine Grandizio[1], Matthew W Mitchell[1], Nahid Turan[1], Laura Scheinfeldt[1,*]

Affiliations: [1]Coriell Institute for Medical Research, Camden, NJ 08003, USA.

* Corresponding author

Abstract:

We describe here a new NHGRI Sample Repository for Human Genetic Research collection of induced pluripotent stem cell (iPSC) lines reprogrammed from whole blood derived peripheral blood mononuclear cells (PBMCs). The PBMCs were reprogrammed using Sendai viral vectors carrying transcription factors OCT4, SOX2, KLF4, and c-MYC. The iPSC lines exhibit a normal karyotype, express common stemness and pluripotency markers, and demonstrate the ability to differentiate into cell types representing all three germ layers. These iPSCs (n=5) will have accompanying public, near telomere to telomere genomic data through the Human Pangenome Reference Consortium, and will provide an invaluable in vitro resource for studying common genetic and genomic variation and its functional implications.

Introduction

The NHGRI Sample Repository for Human Genetic Research (NHGRI Repository) housed at the Coriell Institute for Medical Research facilitates studies of human genetic and genomic variation by establishing, characterizing and distributing a large (n>3,700) publicly available collection of renewable biospecimens donated by communities living around the world, including the communities that have participated in the 1000 Genomes Project and the Human Pangenome Reference Consortium [1,2]. NHGRI Repository participants have consented to their biospecimens being used for a wide range of general research and to public data sharing of large-scale genomic data collected from their samples.

This new collection of five induced pluripotent stem cell (iPSC) lines was prepared from peripheral blood samples collected as part of a larger effort to improve the human reference genome to better represent human genetic variation [2-4]. This collection is an important new resource to improve our understanding of tissue-specific gene regulation and cellular function.

Results

We reprogrammed five iPSC lines (NHGRI Repository identifiers: HG06800, HG06801, HG06802, HG06803, and HG06804) from peripheral blood mononuclear cells (PBMCs) using the CytoTune 2.0 Sendai Reprogramming Kit (Thermo Fisher Scientific), which includes the

pluripotency factors OCT4, SOX2, KLF4, and c-MYC. We assessed successful reprogramming and viability with cell examination (Table 1).

The cells have passed all internal quality control assessments [5]: they display typical iPSC morphology under phase contrast microscopy, exhibit alkaline phosphatase activity (Fig. S1), and express pluripotency markers (Fig. 1A). We assessed post-thaw cell viability by culturing a frozen vial, with iPSC colony area increasing 9-28-fold over a 3-5 day period (Table S1). We used quantitative flow cytometry to assess SSEA-4 expression (94.78% - 99.22%; Fig. 1B). We performed cytogenomic G-banding analysis to confirm normal karyotypes (Fig. 1C). We did not detect any Sendai virus (SeV) genome or transgenes by qRT-PCR using SeV-specific primers (Table S2) after passage 10. We confirmed pluripotency by an embryoid body (EB) formation assay (Fig. 1D). The lines are free of mycoplasma contamination (Table 1, Table S3), and microsatellite profiling has confirmed iPSC authenticity (Table 1, Table S3).

Methods

Samples

The samples included in the new NHGRI Repository iPSC resource were collected as part of a larger effort to improve the human reference genome with more comprehensive coverage of the genome, as well as to increase representation of human genetic variation [2-4]. The participants that generously agreed to donate to the repository consented to the creation of immortalized cell lines, including iPSCs, to their cell lines and DNA being made available for general research usage, and to large-scale genomic data collected from their samples being made public. In addition, these samples are in the process of being characterized by the HPRC with near telomere to telomere genomic assemblies, and these public data will add important genetic and genomic characterizations to this iPSC collection. The recommended language for referring to these samples is: African Americans living in St. Louis, Missouri, which may be abbreviated to ASL after the full description is provided (https://catalog.coriell.org/1/NHGRI/About/Guidelines-for-Referring-to-Populations).

Reprogramming and Cell Culture

We reprogrammed PBMCs using CytoTune Sendai vectors encoding OCT4, SOX2, KLF4, C-MYC, and EmGFP. We cultured cells with medium changes every 48 hours and replated on day 3. After 2-3 weeks, we selected 24 colonies and expanded them. At passage 10, we cryopreserved the lines.

We thawed cells by warming at 37°C, tested them for sterility, and cultured them in mTeSR1 with ROCK inhibitor for 24 hours. We maintained the cells on mouse embryonic fibroblasts (MEFs) with daily media changes. We passaged the cells at 75-85% confluency using Versene (2 min, RT) or ReLeSR (5-7 min, 37°C). We expanded the cell lines for master (5 vials) and distribution (26 vials) banks and transitioned the cells to Matrigel before distribution availability through the NHGRI Repository.

Alkaline Phosphatase Staining

We stained iPSCs using the StemTAG™ Alkaline Phosphatase Staining Kit (Cell Biolabs, Inc.).

Immunocytochemistry

We characterized the cells using a PSC Immunocytochemistry Kit with SOX2/TRA-1-60 and SSEA4/OCT4 antibody pairs. After fixation, permeabilization, and blocking, we applied primary

antibodies (3h, 4°C), followed by secondary antibodies (1h, RT). We added DAPI during the final PBS wash (5 min).

Flow Cytometry
We quantified surface antigen expression of iPSC markers by flow cytometry. We dissociated the cells with Trypsin, washed the cells with PBS, and incubated the cells with fluorophore-conjugated antibodies (Table S2) for 15 min at room temperature. We used the MACSQuant Flow Cytometer and MACSQuantify software for all related analyses.

G-Banded Karyotyping
Twenty metaphase cells were counted and examined, and five selected metaphase cells were karyotyped.

Cell Line Authentication
We assessed cell line identity and authenticity by extracting DNA from each line and comparing the microsatellite (MSAT) marker profile to the DNA extracted from the whole blood sample submission as previously described [6].

Mycoplasma Detection and Sterility Testing
We tested for mycoplasma contamination using the MycoSEQ™ Mycoplasma Detection Kit, a real-time PCR assay detecting over 90 species with <10 copies sensitivity. We assessed cell culture sterility via growth assays on trypticase soy agar and Sabouraud dextrose.

Sendai Virus Detection
We extracted total RNA using the RNeasy Plus Mini Kit. We synthesized cDNA from 1 µg RNA using the High Capacity cDNA Reverse Transcription Kit. We quantified gene expression by qRT-PCR using TaqMan® Gene Expression Master Mix on a QuantStudio 6 Flex, following a standard PCR program (Table S2).

Differentiation Potential
We assessed pluripotency via embryoid body (EB) formation. We cultured EBs in DMEM with 10% FBS, 1% L-glutamine, NEAA, and sodium pyruvate for 10 days. We extracted RNA using the RNeasy Mini Kit and quantified RNA by NanoDrop One. We synthesized cDNA, analyzed gene expression by qRT-PCR, and normalized Ct values to GAPDH. We calculated relative expression as fold change compared to undifferentiated cells using the Livak-Schmittgen method [7].

Discussion

The present study introduces a new collection of five iPSCs incorporated into the NHGRI Repository. These cell lines were created from whole blood samples donated by African Americans living in St. Louis, MO. The iPSC lines are currently available to the research community through the NHGRI Repository, housed at the Coriell Institute for Medical Research (https://www.coriell.org/1/NHGRI).

A distinctive feature of this collection is that each iPSC line has a corresponding lymphoblastoid cell line (LCL), and DNAs extracted from these LCLs are in the process of being extensively genome sequenced through the Human Pangenome Reference Consortium. This genomic data will be publicly accessible and will include high-quality, robust, long-read, near telomere-to-

telomere (T2T) assemblies [2]. The availability of matched iPSC and LCL lines enables comparative studies of cellular phenotypes and lineage-specific gene regulation. Moreover, the accompanying genomic data will facilitate a more comprehensive investigation of genetic and genomic variation and its functional consequences.

This resource will be particularly valuable for investigating tissue-specific gene regulation, cellular differentiation pathways, and the functional consequences of genetic and genomic variation on cellular function. Furthermore, these well-characterized lines will serve as important reference materials for future studies involving iPSCs.

Table 1 – Characterization and validation.

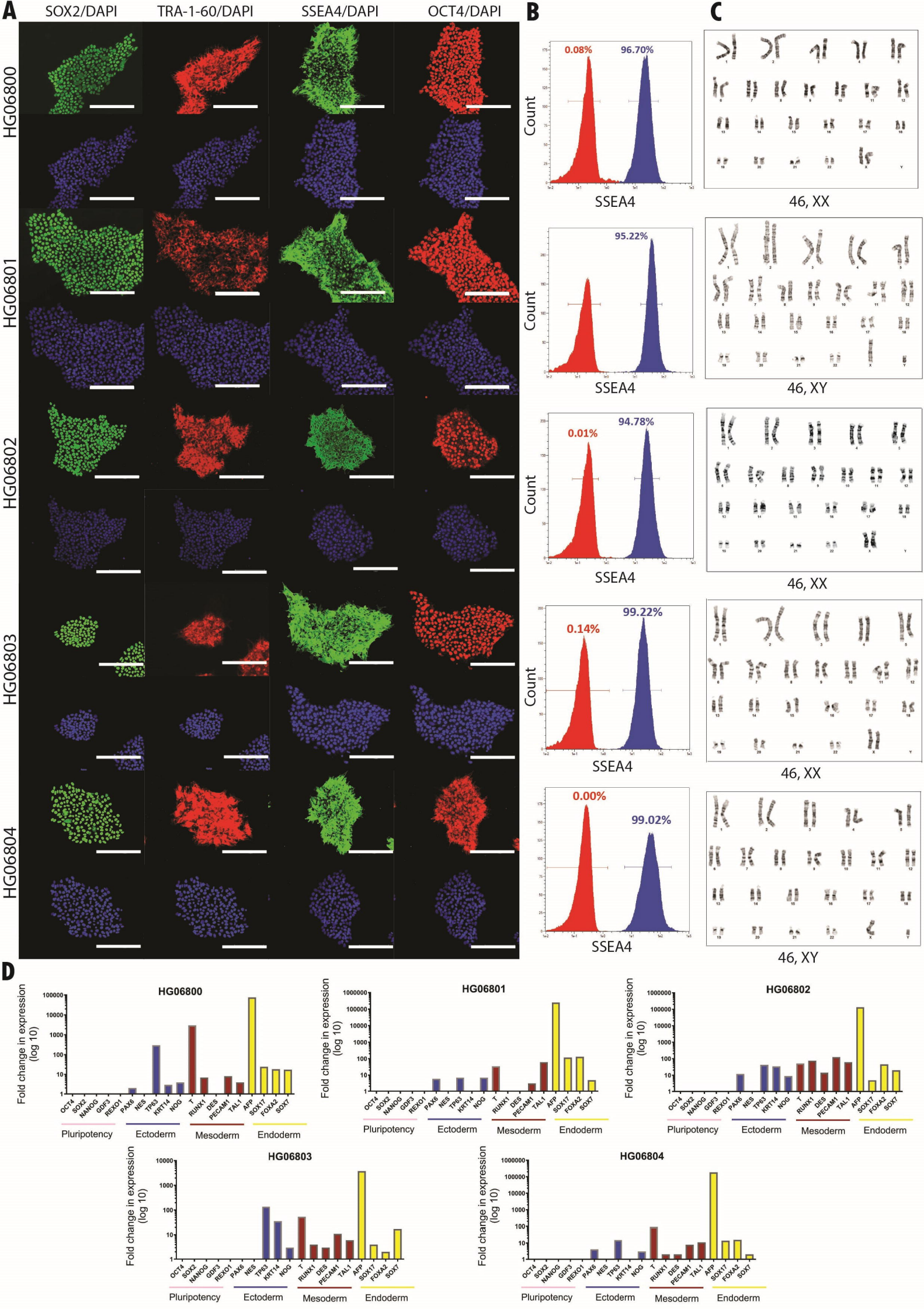
Classification	**Test**	**Result**	**Data**
Morphology	Photography	Normal	*Fig. S1*
Phenotype	Qualitative analysis (Immunocytochemistry)	Assess staining/expression of pluripotency markers: SOX2, OCT4, SSEA4, & TRA-1-60	*Fig. 1 panel A*
	Quantitative analysis (i.e. Flow cytometry)	SSEA-4 (94.78% - 99.22%)	*Fig. 1 panel B*
Genotype	Karyotype (G-banding and microarray) and resolution	HG06800 – 46, XX HG06801 – 46, XY HG06802 – 46, XX HG06803 – 46, XX HG06804 – 46, XY	*Fig. 1 panel C*
Identity	MSAT analysis	Six sites tested, all sites matched	*N/A*
Microbiology and virology	Mycoplasma	Mycoplasma testing by qRT-PCR. Negative	*N/A*
Differentiation potential	Embryoid body formation	Embryoid body (EB) with three germ layers formation (ectoderm, mesoderm, and endoderm)	*Fig. 1 panel D*

Figure 1

Characterization and quality control assessment of induced pluripotent stem cells (iPSCs). A) Characterization of iPSC pluripotency markers by immunofluorescence. Representative immunofluorescence images showing expression of key pluripotency markers SOX2, OCT4, SSEA4, and TRA-1-60 in iPSCs. Scale bar: 200 µm. B) Surface antigen expression of stem cell markers. Undifferentiated cells are stained for stage specific embryonic antigen 4 (SSEA4) which is expressed on the surface of undifferentiated human pluripotent stem cells C) G-banding karyogram. D) Differentiation potential. Cells are differentiated by embryoid body (EB) formation to assess pluripotency. RNA is extracted and gene expression is measured by quantitative RT-PCR. Ct values are adjusted to the endogenous housekeeping gene GAPDH. Relative gene expression is shown as the fold difference in expression compared to undifferentiated cells. Expression of at least one gene per germ layer should increase by 2 fold or higher.

References

1. Byrska-Bishop M, Evani US, Zhao X, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell 2022;185(18):3426-3440 e19, doi:10.1016/j.cell.2022.08.004
2. Liao WW, Asri M, Ebler J, et al. A draft human pangenome reference. Nature 2023;617(7960):312-324, doi:10.1038/s41586-023-05896-x
3. Jarvis ED, Formenti G, Rhie A, et al. Semi-automated assembly of high-quality diploid human reference genomes. Nature 2022;611(7936):519-531, doi:10.1038/s41586-022-05325-5
4. Wang T, Antonacci-Fulton L, Howe K, et al. The Human Pangenome Project: a global resource to map genomic diversity. Nature 2022;604(7906):437-446, doi:10.1038/s41586-022-04601-8
5. Pozner T, Grandizio C, Mitchell MW, et al. Human iPSC Reprogramming Success: The Impact of Approaches and Source Materials. Stem Cells Int 2025;2025(2223645, doi:10.1155/sci/2223645
6. Smith G, Mathews D, Sander-Effron S, et al. Microsatellite Markers in Biobanking: A New Multiplexed Assay. Biopreserv Biobank 2021;19(5):438-443, doi:10.1089/bio.2021.0042
7. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25(4):402-8, doi:10.1006/meth.2001.1262


A
SOX2/DAPI
TRA-1-60/DAPI
SSEA4/DAPI
OCT4/DAPI
HG06800
HG06801
HG06802
HG06803
HG06804
B
Count
SSEA4
0.08%
96.70%
95.22%
0.01%
94.78%
0.14%
99.22%
0.00%
99.02%
C
46, XX
46, XY
46, XX
46, XX
46, XY
D
HG06800
HG06801
HG06802
HG06803
HG06804
Fold change in expression (log 10)
OCT4
SOX2
NANOG
GDF3
REXO1
PAX6
NES
TP63
KRT14
NOG
T
RUNX1
DES
PECAM1
TAL1
AFP
SOX17
FOXA2
SOX7
Pluripotency
Ectoderm
Mesoderm
Endoderm