A LIBRARY OF INDUCED PLURIPOTENT STEM CELLS FROM CLINICALLY WELL-CHARACTERIZED, DIVERSE HEALTHY HUMAN INDIVIDUALS

Christoph Schaniel[a,b,c,], Priyanka Dhanan[d], Bin Hu[d], Yuguang Xiong[d], Teeya Raghunandan[d], David M. Gonzalez[b,e], Sunita L. D'Souza[b,e], Arjun Yadaw[d], Jens Hansen[d], Gomathi Jayaraman[d], Bino Mathew[f], Moara Machado[f], Seth Berger[g], Joseph Tripodi[h], Vesna Najfeld[h], Jalaj Garg[i,j], Marc Miller[i,j], Colleen Lynch[k], Katherine Michelis[k], Neelima Tangirala[k], Himali Weerahandi[k], David C. Thomas[k], Robert Sebra[l], Milind Mahajan[f], Eric Schadt[f], Dusica Vidovic[m], Stephan C Schürer[m], Joseph Goldfarb[c,d], Evren U. Azeloglu[c,d,n], Marc R. Birtwistle[d,o], Eric A. Sobie[c,d], Jason C. Kovacic[i,j], Nicole C. Dubois[b,e,p], and Ravi Iyengar[c,d]

[a]Department of Medicine, Division of Hematology and Medical Oncology, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[b]Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[c]Mount Sinai Institute for Systems Biomedicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[d]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[e]Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[f]Sema4, Stamford, CT 06902, USA

[g]Center for Genetic Medicine Research & Rare Disease Institute, Children's National Hospital, Washington, DC, 20010, USA

[h]Tumor Cytogenomics Laboratory, Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[i]Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[j]Department of Medicine, Division of Cardiology, Icahn School of Medicine at Mount Sinai, and The Mount Sinai Hospital, New York, NY 10029, USA

[k]Department of Medicine, Division of General Internal Medicine, Icahn School of Medicine at Mount Sinai, and The Mount Sinai Hospital, New York, NY 10029, USA

[l]Department of Genetics and Genomic Siences, Icahn School of Medicine at Mount Sinai, and The Mount Sinai Hospital, New York, NY 10029, USA

[m]Institute for Data Science and Computing, University of Miami, Coral Gables, FL 33146, USA

[n]Department of Medicine, Division of Nephrology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[o]Chemical and Biomolecular Engineering, Clemson University, Clemson, SC, 29634, USA

[p]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Correspondence: Christoph Schaniel (christoph.schaniel@mssm.edu), Nicole C. Dubois (nicole.dubois@mssm.edu) & Ravi Iyengar (ravi.iyengar@mssm.edu)

Abstract

A library of well-characterized human induced pluripotent stem cell (hiPSC) lines from clinically healthy human subjects could serve as a powerful resource of normal controls for *in vitro* human development, disease modeling, genotype-phenotype association studies, and drug response evaluation. We report generation and extensive characterization of a gender-balanced, racially/ethnically diverse library of hiPSC lines from forty clinically healthy human individuals who range in age from 22-61. The hiPSCs match the karyotype and short tandem repeat identity of their parental fibroblasts, and have a transcription profile characteristic of pluripotent stem cells. We provide whole genome sequencing data for one hiPSC clone from each individual, ancestry determination, and analysis of Mendelian disease genes and risks. We document similar physiology of cardiomyocytes differentiated from multiple independent hiPSC clones derived from two individuals. This extensive characterization makes this hiPSC library a unique and valuable resource for many studies on human biology.

Introduction

Since their groundbreaking discovery (Park et al., 2008; Takahashi et al., 2007; Yu et al., 2007), hiPSCs and cells differentiated from hiPSCs have become a powerful system to model *in vitro* human phenotypes, disease etiology and mechanisms, genotype-phenotype correlations and drug responses. However, such studies are often hampered by the small number of hiPSC lines or respective controls used for comparative analysis, reported somatic variability of derived hiPSCs, and heterogeneity of differentiated cells (Dubois et al., 2011; Fusaki et al., 2009; International Stem Cell et al., 2011; Nazor et al., 2012; Witty et al., 2014). In recent years, several groups and consortia reported on the establishment of various hiPSC libraries with a wide range in the number of hiPSC lines (Carcamo-Orive et al., 2017; Kilpinen et al., 2017; Panopoulos et al., 2017; Park et al., 2017; Streeter et al., 2017). These libraries included mostly disease/disorder-specific hiPSCs and control hiPSCs. The control hiPSC lines in these studies were derived from subjects without the specific diseases/disorders studied, either relatives or non-related individuals (Carcamo-Orive et al., 2017; Panopoulos et al., 2017), who were self-declared healthy subjects, or individuals with no medical disease history (Kilpinen et al., 2017; Panopoulos et al., 2017; Rouhani et al., 2014). However, clinical documentation of health status was most often not explicitly established. Nevertheless, such libraries have been used for the study of how genetic variants associated with complex genomic traits and phenotypes drive molecular and physiological variation in hiPSCs and their differentiated cells (Carcamo-Orive et al., 2017; D'Antonio-Chronowska et al., 2019; DeBoever et al., 2017; Karch et al., 2019; Kaserman et al., 2020; Kilpinen et al., 2017; Panopoulos et al., 2017; Park et al., 2017; Rouhani et al., 2014). Larger scale comparative and effective disease modeling, drug discovery and evaluation, and genotype-phenotype association studies suffer from the limited availability and inclusion of hiPSC lines from clinically screened, healthy individuals of various racial and ethnic backgrounds and age. Here, as part of the NIH-Common Funds-funded Library of Integrated Network-based Cellular Signatures (LINCS) program (Keenan et al., 2018), we report the creation of a hiPSC

library from forty selected individuals of diverse racial/ethnic backgrounds and age who passed a rigorous clinical health evaluation. We provide the clinical characteristics of each of the participants from whom hiPSCs were derived, cytogenetics reports, short-tandem repeat (STR) authentication, and pluripotency analyses for all forty hiPSC lines, as well as whole genome sequencing data, ancestry determination, and Mendelian disease gene and risk assessment.

Several studies have suggested a potential impact of donor cell source, cellular heterogeneity of established hiPSCs, as well as sex on cellular differentiation and physiological behavior (D'Antonio-Chronowska et al., 2019; Hu et al., 2016; Pianezzi et al., 2020; Sanchez-Freire et al., 2014). Variability in measured physiological parameters might also be affected by the lack of cellular homogeneity in differentiated hiPSCs. It remains unclear however, whether cells differentiated using a well-established differentiation and selection protocol from multiple well-characterized hiPSC clones, established from the same healthy subject using the same non-integrating reprogramming method, physiologically behave the same. Therefore we studied the characteristics of ventricular and atrial cardiomyocytes differentiated from independent hiPSC clones from the same individual and from hiPSC clones from different individuals.

In summary, our diverse hiPSCs library from forty clinically well-characterized healthy human individuals contributes a valuable resource to the scientific community for a broad variety of biomedical and pharmacological research.

Results

Recruitment, Health Evaluation, and Characterization of Individuals in the Mount Sinai Library of hiPSC Lines Derived from Diverse, Clinical Healthy Subjects

Potential healthy volunteers were approached across Mount Sinai through word-of-mouth and IRB-approved advertisements, and were pre-screened for potential inclusion in the study. Ninety six male and female individuals who satisfied the initial pre-screening consented (Document S1 - *Blank Study-Consent Form* and Document S2 -*Blank HIV consent Form*) to the study and their

sex, age, and race/ethnicity were recorded through an enrollment questionnaire (Figure 1A). Eighty-five underwent a formal and thorough evaluation by a screening study physician. Formal screening involved a full medical history, measurement of weight, height, waist and hip circumference, heart rate, blood pressure, respiratory rate, and oxygen saturation, a physical exam, and an electrocardiogram (EKG). Blood was drawn for analysis of clinically relevant parameters (Table S1 -*Excel file with all clinical parameters of the forty healthy subjects*). A pregnancy test was included for female participants. All blood draws and pregnancy tests were sent to a certified clinical laboratory for analysis. The more than twenty exclusion criteria included abnormal EKG, family history of any cardiovascular disorder excluding hypertension in any first or second degree relative at age <50, family history of non-ischemic cardiomyopathy in any first or second degree relative at any age, prior organ transplantation, HIV positive status, history of myopathy, obesity, renal impairment, autoimmune disease, abnormal blood test results including brain natriuretic peptide, body mass index of ≥30 kg/m^2, lifetime smoking of >2 pack-years or abnormality on physical exam. These data were entered onto a clinical report form (Document S3 -*Blank Clinical Report Form*) and a final assessment of study eligibility was then made by consensus of two senior study physicians (J.C.K., D.T.) and the screening study physician. If eligibility was approved by all three physicians, the subject then underwent skin biopsy and formally became one of the included study subjects. Of the eighty-five subjects who were screened, forty-two (48.3%) were deemed eligible for final inclusion. Reasons for a screened subject's exclusion from further participation included EKG abnormalities beyond very minor physiologic variations on normal such as bradycardia with heart rate 58/min in a healthy fit subject, anemia, body mass index (BMI) of ≥30 kg/m^2, arrhythmia, lifetime smoking of >2 pack-year and other reasons (Figure 1B). Of the forty-two eligible individuals, one declined to undergo skin biopsy and one didn’t show up for the scheduled biopsy, leaving forty subjects on whom a skin biopsy was performed. These forty healthy male (22) and female (18) individuals ranged in age from 22-61.

These observations demonstrate that an inclusive health evaluation and strict selection criteria for what constitutes a healthy individual, rather than simply relying on a participant's self-declaration or medical history is critical for selecting clinically healthy subjects.

Generation, Authentication and Characterization of hiPSC clones

We derived fibroblast lines from skin biopsy samples taken from the 40 eligible healthy subjects and banked them. We used integration-free reprogramming methods (mRNA with microRNA boost (R) (Warren et al., 2010)) and/or Sendai virus (S) (Fusaki et al., 2009)) to generate hiPSCs and establish multiple clones from all 40 fibroblast lines. Our present resource consists of one hiPSC clone per individual (Table S2). All data, as well as linked associated metadata, , can be found on the searchable LINCS data portal (Koleti et al., 2018; Stathias et al., 2020). All hiPSCs in our library match the karyotype of the parental fibroblast line. In all cases except one the karyotype was normal (Figure 2A, Table S3). The abnormal karyotype, which was a t(1;17)(p34;q23) translocation, was observed for MSN24. In total, we analyzed the karyotypes of 5 independent hiPSC clones derived from individual MSN24. They all carried the t(1;17)(p34;q23) translocation. It is not uncommon that a karyotype is abnormal (International Stem Cell et al., 2011; Mayshar et al., 2010; Peterson and Loring, 2014; Taapken et al., 2011); however, a frequency of 100% abnormal hiPSC clones is very unusual. Thus, we karyotyped the parental fibroblasts. Interestingly, the parental fibroblasts already carried this chromosomal abnormality. Therefore, this may represent a case of a healthy individual, who is the carrier of a balanced translocation. No report exists about this specific translocation. Of 84 hiPSC clones we karyotyped, 26 (in addition to the 5 analyzed MSN24 hiPSC clones) showed an abnormal karyotype (31%). This is consistent with published reports of abnormal karyotypes in up to one third of derived human embryonic stem cells (ESCs) and hiPSCs (International Stem Cell et al., 2011; Mayshar et al., 2010; Peterson and Loring, 2014; Taapken et al., 2011). All hiPSC clones in our library have been authenticated by STR analysis to match the profile of their respective

parental fibroblast line (Figure 2B, Table S4). This is an important and necessary quality control and assurance measure to ensure the origin and authenticity of the derived hiPSC clones.

One karyotypically normal (with the exception of MSN24) and authenticated hiPSC clone generated from each of the forty clinically healthy individuals was further characterized by immunocytochemistry for expression of the pluripotency-associated markers NANOG, OCT4, SOX2, TRA-1-60 and SSEA4 (Figure 2C, Figure S1). We also conducted the mRNA-seq-based PluriTest assay (Panopoulos et al., 2017) as an additional measure of the pluripotent status of all hiPSC clones (Table S5, Figure 2D, Figure S2). We included the pluripotency and novelty scores retrieved from fibroblasts of 66 individuals as comparative reference (Hagai et al., 2018).

We further performed ancestry determination using the whole-genome sequencing information from all forty hiPSC lines (Table S6). This analysis provided additional confirmation of the authenticity of the hiPSC lines. It additionally reflected the diverse nature of the subjects' racial and ethnic backgrounds. The selected individuals have origins in East Africa, West Africa, East Asia, the Central India subcontinent, the Southern India subcontinent, Eastern Mediterranean, Northeast Europe, Northern and Central Europe, Southwest Europe, and the Anatolia/Caucasus,/Iranian Plateau region.

Our data strongly support the fact that we established a gender-balanced, racially/ethnically diverse library of well-characterized hiPSCs from forty clinically healthy human individuals who range in age from 22-61.

Clinically Healthy Individuals Can Be Carriers Of Genetic Variants With Disease Risks

Although all forty individuals from whom we established the hiPSC library were determined to be clinically healthy, we next assessed whether they carry potential genetic disease risks. To assess this, we interrogated the whole-genome sequencing information by performing an expanded carrier screening analysis and annotation of gene variants associated with both dominant and recessive disorders (vcf files containing variant ranking (VaRank) analysis for each individual are

provided along with the whole genome sequencing data in dbGAP). We further analyzed these variants using ClinVar (Landrum et al., 2014) and extracted those that are associated with a potential dominant presentation of a disease as asserted by the American Board of Medical Genetics and Genomics. Six individuals carried a variant with a dominant disease association (Figure 3A). These data demonstrate that even clinically well-characterized healthy human individuals can still be carriers of genetic variants with disease risks.

hiPSC Gene Signature is Independent of Age or Sex of Individuals from whom hiPSCs are Derived

In order to assess whether age or sex of the included individuals could segregate the derived hiPSC lines based on global gene expression, we determined pairwise correlation between mRNA-seq raw read counts of duplicate samples of all forty hiPSC lines. This was followed by hierarchical clustering. The resulting dendrogram documents a high similarity between duplicate samples and across the forty hiPSC lines (Figure 3B). Similar to pluripotency marker expression (Figure 2C, Figure S1) and PluriTest scores (Figure 2D, Figure S2), we observed no significant influence of age or sex on global gene expression. These results indicate that global gene expression, within the wide age range we studied, is independent of age and sex of clinically well-characterized healthy human subjects from whom hiPSC lines are derived.

Cardiomyocytes Differentiated from hiPSC Lines From Different Clinically Well-Characterized Healthy Individuals As Well As Independent hiPSC Clones From The Same Individual Exhibit Physiologically Similar Behavior.

A recent report determined that that sex affects cardiac ventricular and atrial differentiation outcomes and inherited genetic variation does not (D'Antonio-Chronowska et al., 2019). It remains unknown, however, whether hiPSCs derived from different clinically well-characterized healthy individuals when differentiated to either ventricular or atrial cardiomyocytes, respectively,

using well-established differentiation protocols (Devalla et al., 2015; Dubois et al., 2011; Lee et al., 2017) and an efficient selection regimen (Tohyama et al., 2013) display comparable atrial or ventricular physiological characteristics or not. Similarly, it remains unclear, albeit very important to know, whether the physiological measurements of independent hiPSC clones from the same clinically healthy subject show similar physiological characteristics. To test this, we differentiated three independent, karyotypically normal and STR-authenticated hiPSC clones derived from two racially/ethnically diverse, age-matched males into atrial and ventricular cardiomyocytes (Figure 4). We measured calcium transients and analyzed various physiological parameters (Figure 5). As expected, we observed some variability in differentiation efficiency pre-lactate selection between the two hiPSC lines as well as between the three independent clones of subject MSN14 (Figure 4B). Similar variability was present in both the atrial and ventricular cardiomyocytes' differentiation efficiency (Figure 5). Importantly however, we detected non-significant differences in the parameters among atrial or among ventricular cardiomyocytes derived from either hiPSC line/subject or the three independent hiPSC clones (Figure 5B). These data suggest that well-characterized hiPSC clones established from the same clinically healthy individual behave physiologically in a similar manner, at least, when differentiated under well controlled conditions to atrial and ventricular cardiomyocytes, respectively, Moreover, the data indicate that the physiological behavior of the atrial and ventricular cardiomyocytes, respectively, are similar irrespective of the hiPSC line, i.e., the clinically healthy individual, they were differentiated from.

Discussion

Previous efforts to create libraries of hiPSCs have focused on diseases and associated controls that included disease-unaffected relatives or non-related individuals, self-declared healthy subjects, or persons with no medical disease record (Carcamo-Orive et al., 2017; Kilpinen et al., 2017; Panopoulos et al., 2017; Park et al., 2017; Streeter et al., 2017). To our knowledge, no gender-balanced, racially/ethnically diverse hiPSC library of well-characterized clinically screened

healthy individuals exists. We have created such an hiPSC library consisting of well-characterized hiPSC lines that were established using integration-free reprogramming methods from forty healthy male and female individuals of diverse racial/ethnic backgrounds, who passed a rigorous health evaluation. As was done for other studies that generated “control” hiPSCs, our subjects had to complete a health questionnaire and had their medical histories evaluated. However, and in contrast to these other studies, the individuals who were selected for a skin biopsy underwent a detailed health screen conducted by an internal medicine physician, and their results were evaluated by a three member clinical panel. This screen included the measurement of general parameters, analysis of clinically relevant parameters present in the blood, a physical examination of the respiratory and gastrointestinal/abdominal systems, a neurological exam and a cardiac exam, including an EKG.

It is well established that race and ethnicity, which are used as a way of categorizing people of shared ancestry and physical traits (Sankar et al., 2004), as well as sex (Soldin and Mattison, 2009) are contributing factors to interindividual differences in drug exposure and/or response (Ramamoorthy et al., 2015; Wilson et al., 2001). These are important factors to consider during drug development and evaluation of drug responses/toxicity as they imply diverse risk-benefit ratios in specific populations. Therefore, any established hiPSC library, if it is not a library of hiPSC lines derived from individuals with a disease/disorder that is preferentially associated with a particular racial/ethnic group, should be gender-balanced and racially/ethnically diverse. Our hiPSC library with 18 females and 22 males that include Whites, White-Hispanics, Blacks and Asians from East and West Africa, East Asia, the Central and Southern India subcontinents, Eastern Mediterranean, Central, North, Northeast and Southwest Europe, and the Anatolia/Caucasus,/Iranian Plateau region fulfills this criteria.

While our hiPSC library was derived from well-characterized clinically healthy individuals, whole-genome sequencing analysis revealed that several individuals were carriers of genetic variants associated with recessive and dominant pathogenic disease risk. This is important and useful

information, particularly when using the hiPSC lines from this library, or for that matter any other hiPSCs, as controls for modeling of a particular disease/disorder, or drug discovery and response evaluation. Several expression quantitative trait loci (eQTL) studies in hiPSCs have shown that genetic variations and also non-genetic determinants account for the majority of gene expression differences between hiPSC lines (Carcamo-Orive et al., 2017; DeBoever et al., 2017; Kilpinen et al., 2017; Panopoulos et al., 2017; Rouhani et al., 2014). While we did not investigate eQTL on global gene expression in hiPSCs, we found that age and sex had no impact (Figure 3B). The latter is in agreement with Kilpinen and colleagues' study, which showed that sex as well as cell passage number did not substantially influence gene and protein expression in hiPSCs (Kilpinen et al., 2017).

One would agree that the physiological function of cells differentiated from disease-specific hiPSC lines should differ from those derived from control hiPSCs. On the other hand, one may expect that the "normal", non-perturbed physiological behaviors of cells of high quality and purity differentiated from various hiPSC clones derived from the same clinically healthy individual would be similar. We found this to be indeed the case in our case study of atrial and ventricular cardiomyocytes, respectively, differentiated from three hiPSC clones each of two age-matched, racially/ethnically diverse male hiPSC lines (Figures 4 and 5). This has scientific, practical as well as financial implications, especially for large-scale studies, where use of a single hiPSC clone per subject rather than multiple clones would curtail costs, reduce potential technical variability, and, increase confidence in the data obtained with a single hiPSC clone. Interestingly, and maybe surprisingly, the analyzed physiological parameters from atrial and ventricular cardiomyocytes, respectively, between these male hiPSC lines also showed great similarity (Figure 5B). This becomes relevant if one were to test drug responses across cell types differentiated from hiPSC derived from a range of racially and ethnically diverse, clinical well-characterized healthy male and female individuals. One could potentially expect different responses to the drugs investigated based on racial and ethnic background as well as genetic variants (pharmacogenomics). In this

context, it is noteworthy that a recent report studying ten control human pluripotent stem cell lines for drug screening in an engineered heart tissue format found that while there was great variability in baseline contractile parameters observed across the different lines, the variability appeared less relevant for drug screening (Mannhardt et al., 2020).

In summary, we provide a gender-balanced, racially/ethnically diverse library of hiPSCs from forty clinically well-characterized healthy human individuals ranging in age from 22-61 years. This library is accompanied with comprehensive quality control characterization of karyotype, STR-matching to each individual (authentication), pluripotency analysis, whole-genome sequencing and thereof ancestry determination and disease gene and risk analysis. This hiPSC library can be useful to investigators looking for control hiPSCs to study normal human development, investigate drug responses/evaluation or as controls for their specific disease model.

Experimental procedures

Recruitment and health screening of subjects

Female and male subjects of diverse racial and ethnic background were recruited and consented under Mount Sinai Institutional Review Board approved protocol (HS# 14-00530) (Documents S1 and S2). Consented subjects were evaluated for their health status at the Mount Sinai Clinical Research Unit by internal medicine physicians. The evaluation involved completion of a health questionnaire, clinical history, and measurement of weight, height, waist and hip circumference, heart rate, blood pressure, respiratory rate, and oxygen saturation. Blood was drawn for analysis of clinically relevant parameters, and the subjects underwent respiratory, gastrointestinal/abdominal, neurological and cardiovascular exams including an EKG (Document S3). A pregnancy test was included for female participants. Subjects with no clinical history of disease, normal physical exam, normal EKG and evaluated parameters within the normal range (Document S3) were classified as healthy by a medical panel consisting of two internal medicine physicians and an interventional cardiologist. Clinical health information remained within Mount

Sinai's protected electronic medical records system.

Clinical data of all consented subjects were extracted from the individual's electronic medical record through the Electronic Privacy Information Center server of the Mount Sinai Data Warehouse. Clinical data for each participant were downloaded manually. Subjects' age, sex, race/ethnicity and selected healthy/passed or unselected/unpassed status were transcribed to a database and healthy subjects were de-identified with a unique sample ID (MSNxx, where xx is a numerical two digit number ranging from 01 to 40) (Tables S1 and S2). The primary reason for exclusion of forty-two individuals from the final pool of eligible healthy subjects' was plotted as a pie-chart(Figure 1B-E). To visualize the distribution of eligible healthy subjects vs. excluded subjects, we plotted violin swarm plots of (i) body mass index (BMI) with median value for eligible healthy subjects of 24.15 vs. excluded of 24.4 kg/m^2 (ii) Hemoglobin with median value for healthy of 14.5 vs excluded of 13.5 g/dL, (iii) Red Blood Cells with median value for healthy of 4.7 vs excluded of 4.6 x 10^6/uL by using Seaborn python data visualization library. All analyses were done using Python programing.

Biopsy punch, derivation and culture of fibroblasts

Our detailed SOP can be found in Supplementary Experimental Procedures. Briefly, a skin sample was taken from each of 40 clinically healthy subjects using a 3 mm sterile disposable biopsy punch (Integra Miltex, Cat# 98PUN3-11) during a second visit at the Clinical Research Unit. Each sample was cut into smaller pieces and placed into gelatin-coated tissue culture dishes with DMEM supplemented with 20% FBS, Penicillin/Streptomycin, non-essential amino, acids, 2mM L-glutamine, 2 mM sodium pyruvate (all from Thermo Fisher Scientific) and 100 µM 2-mercaptoethanol (MP Biomedicals, Cat# 194705) to establish fibroblast lines. Fibroblasts were harvested using TrypLE Express (Thermo Fisher Scientific, Cat# 12605010) and passaged at a 1 to 4 split ratio. Fibroblasts were cryopreserved in 40% DMEM, 50% FBS and 10% DMSO (Millipore Sigma, Cat# D2438).

Reprogramming fibroblasts to hiPSCs

Mycoplasma-free fibroblasts at passage number 3-5 were reprogrammed using the mRNA reprogramming Kit (Stemgent, Cat # 00-0071) in combination with the microRNA booster kit (Stemgent, Cat# 00-0073) and/or the CytoTune™-iPS 2.0 Sendai Reprogramming Kit (Thermo Fisher Scientific, Cat# A16517) according to our SOPs that can be found in Supplementary Experimental Procedures. hiPSC clones will be deposited with, and made available through, WiCell.

hiPSC culture

Established hiPSCs were cultured on plates coated with Matrigel (Corning, Cat# 254248) in mTeSR medium (STEMCELL Technologies, Cat# 05850) in a humidified incubator at 5% CO2, 37˚C. Cells were passaged using ReLeSR (STEMCELL Technologies, Cat# 05872) according to the manufacturer's instructions and grown for 16-24 hrs in mTeSR medium supplemented with 2 μM Thiazovivin (Millipore Sigma, Cat# 420220).

Author Contributions

C.S. recruited, enrolled, scheduled health evaluation and biopsies of study participants and de-identified and coded samples; C.L., K. M., N.T. H. W. conducting the clinical health evaluations and biopsy punches of study participants; J.C.K. conceived the clinical health evaluation; D.T. and J.C.K. supervised the clinical health evaluation, evaluated all clinical data and made final determination on health-pass fail decision supervised and oversaw all clinical aspects of the study; A.Y. performed clinical health data extraction, processing and analysis; C.S., P.D., and T.R. established fibroblast lines; C.S., P.D., T.R., B.H., and S.L.D. generated hiPSCs; P.D., D.M.G. and N.C.D. differentiated hiPSCs to cardiomyocytes, and performed and analyzed the physiology experiments; C.S., P.D., T.R. and G.J. performed immunostaining and imaging; B.H.

and G.J. isolated mRNA and genomic DNA; J.T. and V.N. supervised and interpreted Mount Sinai karyotype analysis; K.M., C.L., H.W and M.T. conducted the clinical health evaluations; J.G. and M.A.M. read and interpreted ECG data; J.C.K. and D.T supervised clinical health evaluations and made final determination on subject eligibility; Y.X. and J.H. performed mRNA-seq data processing and computational analysis; B.M, M.M., M.M. E.S. performed whole genome-sequencing data processing, ancestry, variant and recessive disorder carrier mutation analysis, and interpretation; S.B. performed direct dominant genetic clinical risk analysis; J.G. provided quality assurance and control; D.V. and S.C.S. coordinated and integrated data into the online LINSC data portal; C.S. and A.E. managed the project; C.S., M.B., E.A.S., N.C.D. and R.I. conceived, designed and supervised the study; C.S. and R.I. wrote the manuscript with input from all authors.

Declaration of Interests

The authors declare no competing financial interests.

Acknowledgements

This work was supported by NIH Common Funds LINCS program awards U54HG008098 (R.I.) and U54HG127624 (S.C.S). We thank the Mount Sinai Clinical Research Unit for providing exam rooms and nurses, and the Mount Sinai Genomics Core for mRNA sequencing.

References

Carcamo-Orive, I., Hoffman, G.E., Cundiff, P., Beckmann, N.D., D'Souza, S.L., Knowles, J.W., Patel, A., Papatsenko, D., Abbasi, F., Reaven, G.M.*, et al.* (2017). Analysis of Transcriptional Variability in a Large Human iPSC Library Reveals Genetic and Non-genetic Determinants of Heterogeneity. Cell Stem Cell *20*, 518-532 e519.

D'Antonio-Chronowska, A., Donovan, M.K.R., Young Greenwald, W.W., Nguyen, J.P., Fujita, K., Hashem, S., Matsui, H., Soncin, F., Parast, M., Ward, M.C.*, et al.* (2019). Association of Human iPSC Gene Signatures and X Chromosome Dosage with Two Distinct Cardiac Differentiation Trajectories. Stem Cell Reports *13*, 924-938.

DeBoever, C., Li, H., Jakubosky, D., Benaglio, P., Reyna, J., Olson, K.M., Huang, H., Biggs, W., Sandoval, E., D'Antonio, M.*, et al.* (2017). Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene Expression in Human Induced Pluripotent Stem Cells. Cell Stem Cell *20*, 533-546 e537.

Devalla, H.D., Schwach, V., Ford, J.W., Milnes, J.T., El-Haou, S., Jackson, C., Gkatzis, K., Elliott, D.A., Chuva de Sousa Lopes, S.M., Mummery, C.L.*, et al.* (2015). Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology. EMBO Mol Med *7*, 394-410.

Dubois, N.C., Craft, A.M., Sharma, P., Elliott, D.A., Stanley, E.G., Elefanty, A.G., Gramolini, A., and Keller, G. (2011). SIRPA is a specific cell-surface marker for isolating cardiomyocytes derived from human pluripotent stem cells. Nat Biotechnol *29*, 1011-1018.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad Ser B Phys Biol Sci *85*, 348-362.

Hagai, T., Chen, X., Miragaia, R.J., Rostom, R., Gomes, T., Kunowska, N., Henriksson, J., Park, J.E., Proserpio, V., Donati, G.*, et al.* (2018). Gene expression variability across cells and species shapes innate immunity. Nature *563*, 197-202.

Hu, S., Zhao, M.T., Jahanbani, F., Shao, N.Y., Lee, W.H., Chen, H., Snyder, M.P., and Wu, J.C. (2016). Effects of cellular origin on differentiation of human induced pluripotent stem cell-derived endothelial cells. JCI Insight *1*.

International Stem Cell, I., Amps, K., Andrews, P.W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M.B.*, et al.* (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nat Biotechnol *29*, 1132-1144.

Karch, C.M., Kao, A.W., Karydas, A., Onanuga, K., Martinez, R., Argouarch, A., Wang, C., Huang, C., Sohn, P.D., Bowles, K.R.*, et al.* (2019). A Comprehensive Resource for Induced Pluripotent Stem Cells from Patients with Primary Tauopathies. Stem Cell Reports *13*, 939-955.

Kaserman, J.E., Hurley, K., Dodge, M., Villacorta-Martin, C., Vedaie, M., Jean, J.C., Liberti, D.C., James, M.F., Higgins, M.I., Lee, N.J.*, et al.* (2020). A Highly Phenotyped Open Access Repository of Alpha-1 Antitrypsin Deficiency Pluripotent Stem Cells. Stem Cell Reports *15*, 242-255.

Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D., Wang, Z., Dohlman, A.B., Silverstein, M.C., Lachmann, A.*, et al.* (2018). The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. Cell Syst *6*, 13-24.

Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F.P., Culley, O.J.*, et al.* (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. Nature *546*, 370-375.

Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D.J., Turner, J.P., Vidovic, D., Forlin, M., Kelley, T.T., D'Urso, A.*, et al.* (2018). Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. Nucleic Acids Res *46*, D558-D566.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res *42*, D980-985.

Lee, J.H., Protze, S.I., Laksman, Z., Backx, P.H., and Keller, G.M. (2017). Human Pluripotent Stem Cell-Derived Atrial and Ventricular Cardiomyocytes Develop from Distinct Mesoderm Populations. Cell Stem Cell *21*, 179-194 e174.

Mannhardt, I., Saleem, U., Mosqueira, D., Loos, M.F., Ulmer, B.M., Lemoine, M.D., Larsson, C., Ameen, C., de Korte, T., Vlaming, M.L.H.*, et al.* (2020). Comparison of 10 Control hPSC Lines for Drug Screening in an Engineered Heart Tissue Format. Stem Cell Reports *15*, 983-998.

Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.C., Yakir, B., Clark, A.T., Plath, K., Lowry, W.E., and Benvenisty, N. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell *7*, 521-531.

Nazor, K.L., Altun, G., Lynch, C., Tran, H., Harness, J.V., Slavin, I., Garitaonandia, I., Muller, F.J., Wang, Y.C., Boscolo, F.S.*, et al.* (2012). Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell *10*, 620-634.

Panopoulos, A.D., D'Antonio, M., Benaglio, P., Williams, R., Hashem, S.I., Schuldt, B.M., DeBoever, C., Arias, A.D., Garcia, M., Nelson, B.C.*, et al.* (2017). iPSCORE: A Resource of 222 iPSC Lines Enabling Functional Characterization of Genetic Variation across a Variety of Cell Types. Stem Cell Reports *8*, 1086-1100.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141-146.

Park, S., Gianotti-Sommer, A., Molina-Estevez, F.J., Vanuytsel, K., Skvir, N., Leung, A., Rozelle, S.S., Shaikho, E.M., Weir, I., Jiang, Z.*, et al.* (2017). A Comprehensive, Ethnically Diverse Library of Sickle Cell Disease-Specific Induced Pluripotent Stem Cells. Stem Cell Reports *8*, 1076-1085.

Peterson, S.E., and Loring, J.F. (2014). Genomic instability in pluripotent stem cells: implications for clinical applications. J Biol Chem *289*, 4578-4584.

Pianezzi, E., Altomare, C., Bolis, S., Balbi, C., Torre, T., Rinaldi, A., Camici, G.G., Barile, L., and Vassalli, G. (2020). Role of somatic cell sources in the maturation degree of human induced pluripotent stem cell-derived cardiomyocytes. Biochim Biophys Acta Mol Cell Res *1867*, 118538.

Ramamoorthy, A., Pacanowski, M.A., Bull, J., and Zhang, L. (2015). Racial/ethnic differences in drug disposition and response: review of recently approved drugs. Clin Pharmacol Ther *97*, 263-273.

Rouhani, F., Kumasaka, N., de Brito, M.C., Bradley, A., Vallier, L., and Gaffney, D. (2014). Genetic background drives transcriptional variation in human induced pluripotent stem cells. PLoS Genet *10*, e1004432.

Sanchez-Freire, V., Lee, A.S., Hu, S., Abilez, O.J., Liang, P., Lan, F., Huber, B.C., Ong, S.G., Hong, W.X., Huang, M.*, et al.* (2014). Effect of human donor cell source on differentiation and function of cardiac induced pluripotent stem cells. J Am Coll Cardiol *64*, 436-448.

Sankar, P., Cho, M.K., Condit, C.M., Hunt, L.M., Koenig, B., Marshall, P., Lee, S.S., and Spicer, P. (2004). Genetic research and health disparities. JAMA *291*, 2985-2989.

Soldin, O.P., and Mattison, D.R. (2009). Sex differences in pharmacokinetics and pharmacodynamics. Clin Pharmacokinet *48*, 143-157.

Stathias, V., Turner, J., Koleti, A., Vidovic, D., Cooper, D., Fazel-Najafabadi, M., Pilarczyk, M., Terryn, R., Chung, C., Umeano, A.*, et al.* (2020). LINCS Data Portal 2.0: next generation access point for perturbation-response signatures. Nucleic Acids Res *48*, D431-D439.

Streeter, I., Harrison, P.W., Faulconbridge, A., The HipSci, C., Flicek, P., Parkinson, H., and Clarke, L. (2017). The human-induced pluripotent stem cell initiative-data resources for cellular genetics. Nucleic Acids Res *45*, D691-D697.

Taapken, S.M., Nisler, B.S., Newton, M.A., Sampsell-Barron, T.L., Leonhard, K.A., McIntire, E.M., and Montgomery, K.D. (2011). Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. Nat Biotechnol *29*, 313-314.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Tohyama, S., Hattori, F., Sano, M., Hishiki, T., Nagahata, Y., Matsuura, T., Hashimoto, H., Suzuki, T., Yamashita, H., Satoh, Y.*, et al.* (2013). Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. Cell Stem Cell *12*, 127-137.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A.*, et al.* (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell *7*, 618-630.

Wilson, J.F., Weale, M.E., Smith, A.C., Gratrix, F., Fletcher, B., Thomas, M.G., Bradman, N., and Goldstein, D.B. (2001). Population genetic structure of variable drug response. Nat Genet *29*, 265-269.

Witty, A.D., Mihic, A., Tam, R.Y., Fisher, S.A., Mikryukov, A., Shoichet, M.S., Li, R.K., Kattman, S.J., and Keller, G. (2014). Generation of the epicardial lineage from human pluripotent stem cells. Nat Biotechnol *32*, 1026-1035.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Figure Legends

Figure 1. Description of subject selection for the Mount Sinai clinically healthy hiPSC library.

(A) Flow chart for subject selection and hiPSC generation and characterization. See Table S1 for the summary characterizations performed for all 40 selected, clinically healthy subjects and derived hiPSCs.

(B) Pie diagram of reasons for excluding screened participants from the final group of eligible subjects. The percentage as well as the number of participants (in parenthesis) excluded for each specific main reason is presented.

(C-E) Violin plots of clinical selection criteria of body mass index (BMI, kg/m^2), red blood cell counts (RBC; x 10^6/uL) and hemoglobin level (g/dL) of healthy/selected and excluded participants. See Document S3 for the clinical report form and Table S1 for the clinical characteristics of all clinically healthy/selected individuals.

Figure 2. Characterization of two representative hiPSC lines/clones.

(A) G-banded karyotypes for female hiPSC clone MSN07-07S and male hiPSC clone MS34-05S. See Table S3 for the cytogenetics of all 40 fibroblast lines and one associated derived hiPSC clone.

(B) STR analysis on 16 genomic loci of MSN07 and MS34 fibroblasts (FB) and derived hiPSC clones, MSN07-07S and MS34-05S, respectively. See Table S4 for the STR analysis of all 40 fibroblast lines and one associated derived hiPSC clone.

(C) Immunocytochemistry of pluripotency markers, NANOG, OCT4, SOX2, TRA-1-60 and SSEA4 in representative hiPSC clones, MSN07-07S and MS34-05S. DAPI is used to stain nuclei. Bar, 400 μm. See Figure S1 for the immunocytochemistry of one representative hiPSC clone derived from each of the 40 fibroblast lines.

(D) PluriTest summary plot of mRNA-seq based transcriptomic analyses performed in duplicate for one hiPSC clone per clinically healthy individual (colored circles). As a reference, the PluriTest results of transcriptomic data of fibroblasts from 66 individuals (Hagai et al., 2018) are plotted in black/grey. See Figure S2 for the PluriTest plots of one representative hiPSC clone derived from each of the 40 fibroblast lines.

Figure 3. Genetic variants with disease risks and global gene expression independence of age and sex.

(A) Subjects with genetic disease variants associated with a dominant presentation.

(B) Pairwise correlation was determined between mRNA-seq raw read counts of duplicate samples of all forty hiPSC lines, followed by hierarchical clustering. The resulting dendrogram documents high similarity between duplicate samples and hiPSC lines. No influence of age or sex was found on gene expression.

Figure 4. Atrial and ventricular differentiation of hiPSCs.

(A) Representative flow cytometry plots of KDR and PDGFR expression on cells at day 5 of atrial (ACM) and ventricular (VCM) cardiomyocyte differentiation from two racially/ethnically distinct, age-matched male clinically healthy subjects (MSN14 and MSN25, respectively) (top).

(B) Percentage of cells expressing both KDR and PDGFR across three independent hiPSC clones each established from subjects MSN14 (dark grey) and MSN25 (light grey). Symbols (circles, squares and triangels) represent independent differentiations. Data is represented as the mean±SD.

(C) Representative flow cytometry plots of SIRPA and CD90 expression at day 20 of atrial (ACM, right) and ventricular (VCM ,left) differentiation from hiPSC clone MSN14-01 pre-lactate and post-lactate selection.

(D) Percentage of cells expressing SIRPA but not CD90 across the three independent hiPSC clones established from the same two racially/ethnically distinct, age-matched male clinically healthy subjects as in (A) pre-lactate (light grey) and post-lactate (dark grey) selection. Individual dots represent independent differentiations. Data is represented as the mean±SD. The red dotted line indicates 95% purity, which was the cut off set for a differentiated line to be used for the subsequent physiology assays.

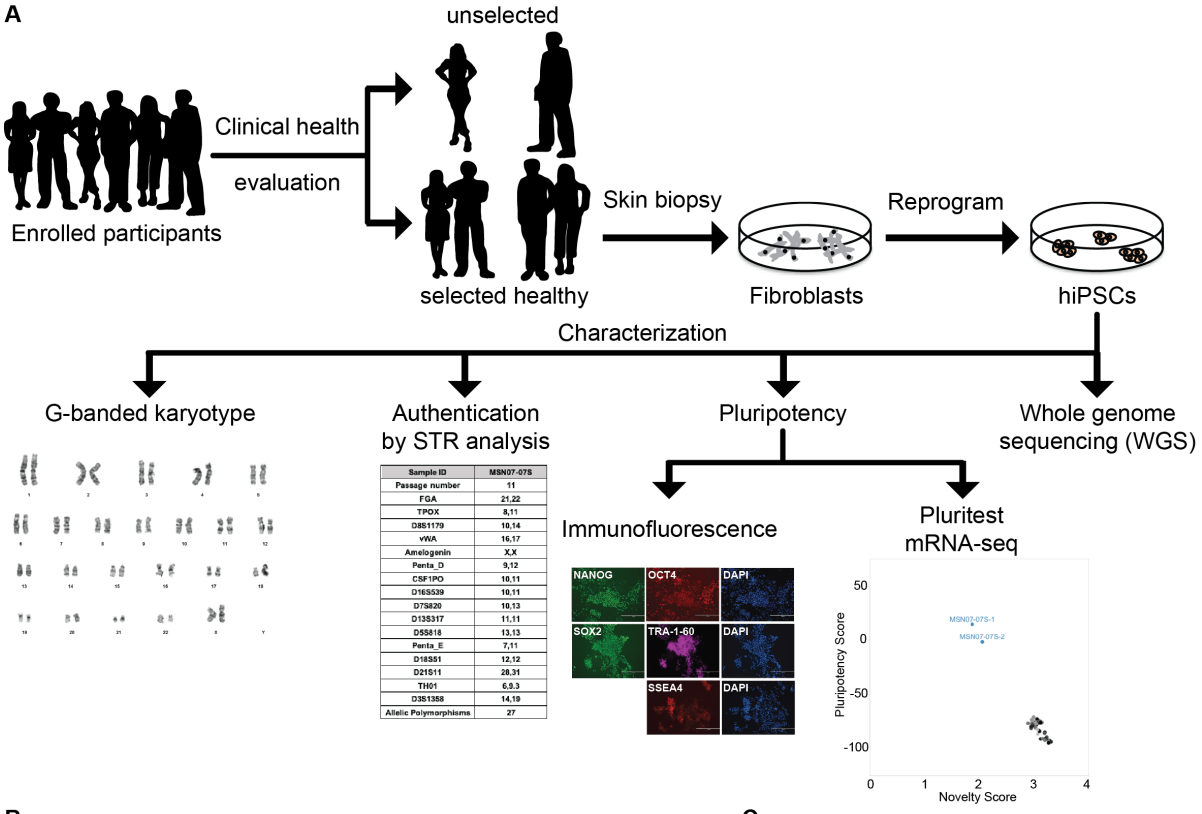
Figure 5. Calcium transient analysis of hiPSC-derived cardiomyocytes from two racially/ethnically distinct, age-matched males.

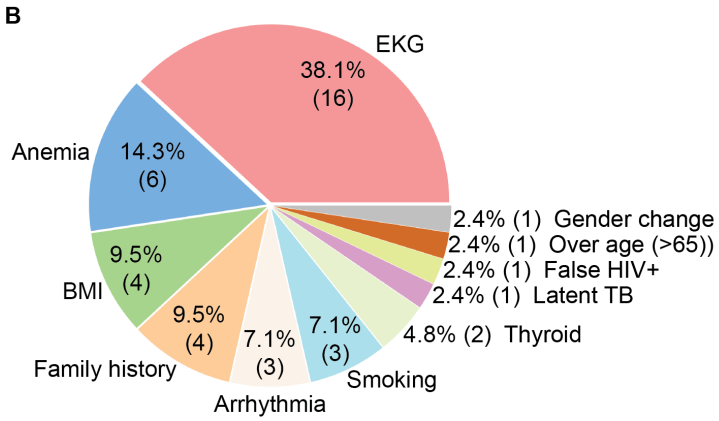
(A) Representative calcium transient analysis traces of ventricular (VCM, left) and atrial (ACM, right) hiPSC-derived cardiomyocytes. hiPSC-CMs were lactate-purified at day 20 of differentiation and analyzed at day 30 of differentiation.

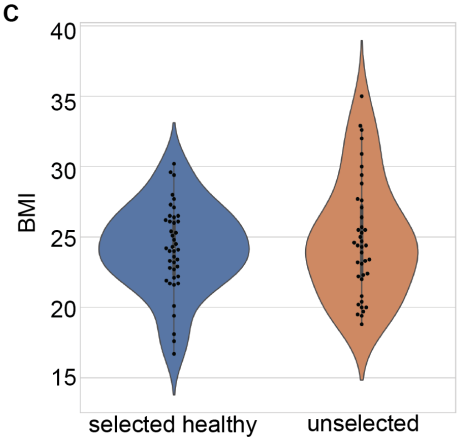
(B) Calcium transient analysis of VCM (light gray) and ACM (dark gray) differentiated from three independent hiPSC clones established from each of the two (MSN14, MSN25) racially/ethnically distinct, age-matched male clinically healthy subjects (same as in Figure 4) at day 30 of differentiation after lactate selection at day 20. Individual dots represent technical replicates (individual cells); and colors (red and blue) indicate biological replicates.

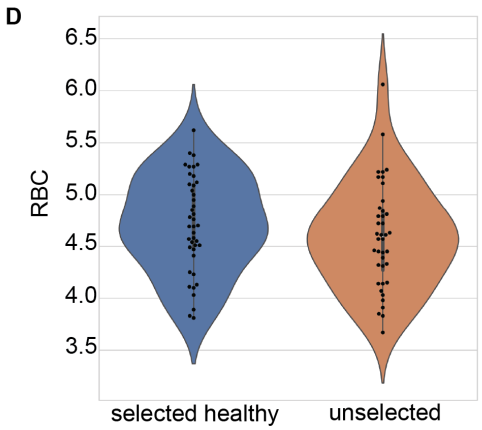
(C) Multielectrode array analysis (Maestro, Axion Biosystems) on day 30 of differentiation from three independent hiPSC clones established from each of the two racially/ethnically distinct, age-matched male clinically healthy subjects. Individual dots represent technical replicates (individual cells); and colors (red and blue) indicate biological replicates.

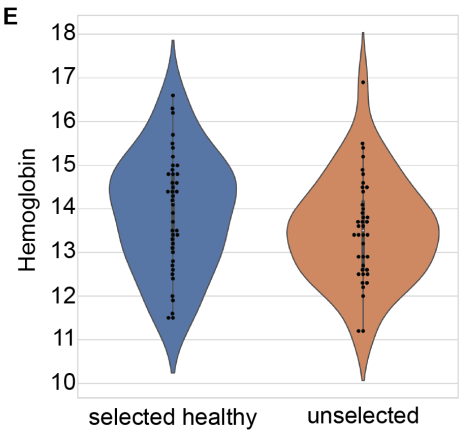
Schaniel et al. Figure 1



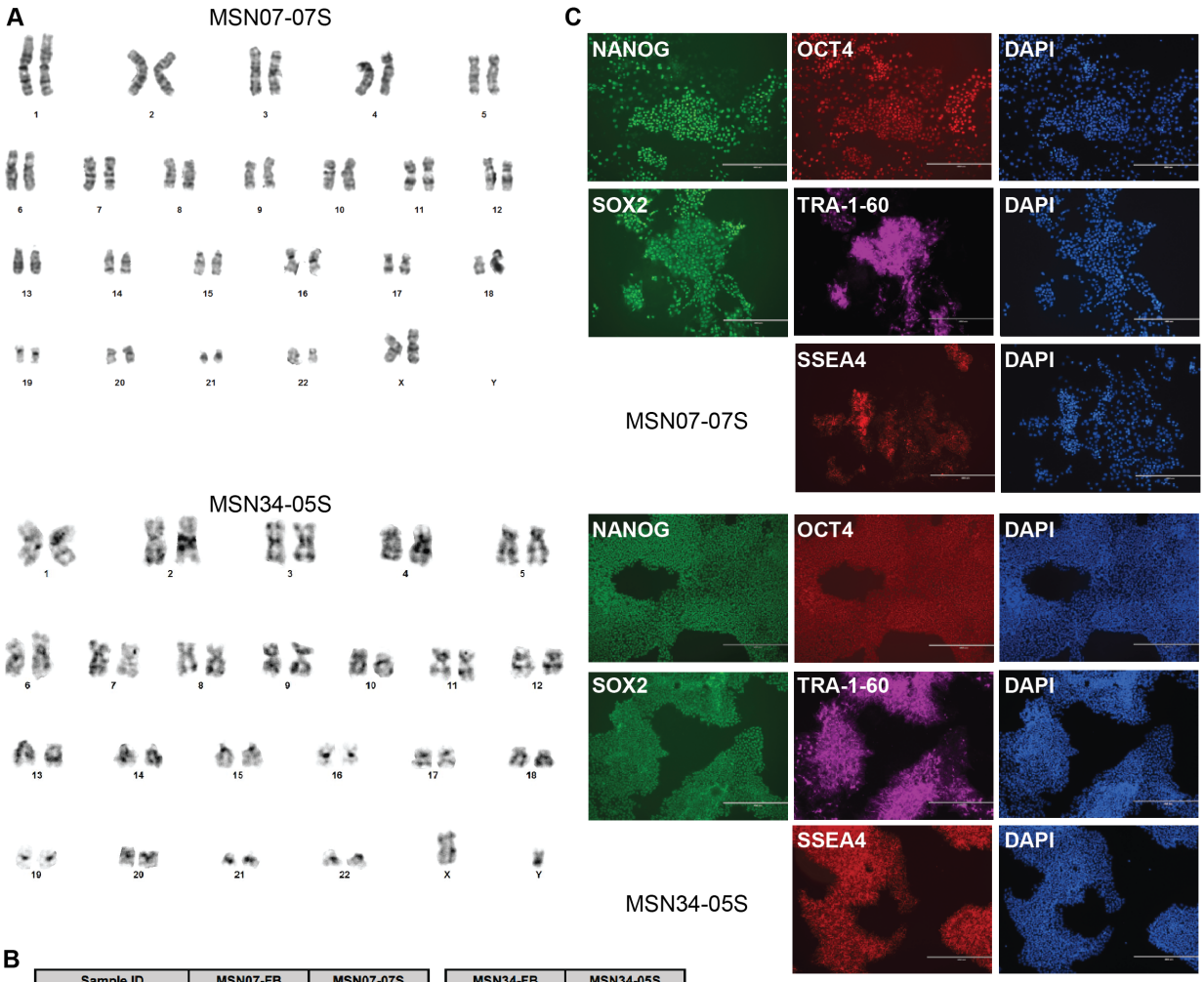








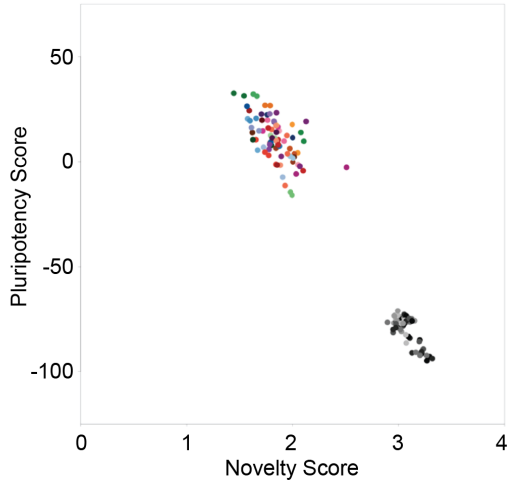
Schaniel et al. Figure 2



B

Sample ID	MSN07-FB	MSN07-07S	MSN34-FB	MSN34-05S
Passage number	4	11	4	12
Assay Date	5/18/18	5/18/18	6/20/18	8/14/18
Report Date	5/22/18	5/22/18	6/26/10	8/22/18
STR performed by	WiCell Research Institute, Inc		WiCell Research Institute, Inc	
WiCell ID	71650	71649	72133	72794
WiCell File name	STR 180521 wmr	STR 180521 wmr	STR 180622 wmr	STR 180815 wmr
FGA	21,22	21,22	21,22	21,22
TPOX	8,11	8,11	8,8	8,8
D8S1179	10,14	10,14	11,12	11,12
vWA	16,17	16,17	16,17	16,17
Amelogenin	X,X	X,X	X,Y	X,Y
Penta_D	9,12	9,12	12,12	12,12
CSF1PO	10,11	10,11	11,11	11,11
D16S539	10,11	10,11	11,13	11,13
D7S820	10,13	10,13	10,12	10,12
D13S317	11,11	11,11	9,14	9,14
D5S818	13,13	13,13	11,11	11,11
Penta_E	7,11	7,11	7,13	7,13
D18S51	12,12	12,12	13,16	13,16
D21S11	28,31	28,31	28,29	28,29
TH01	6,9.3	6,9.3	6,9	6,9
D3S1358	14,19	14,19	16,16	16,16
Allelic Polymorphisms	27	27	25	25

D

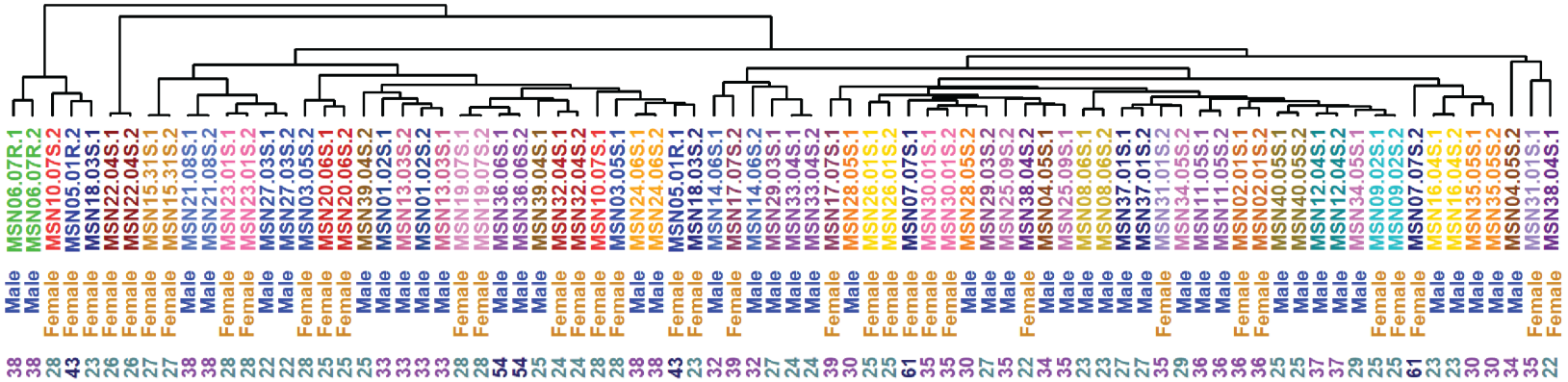


Schaniel et al. Figure 3

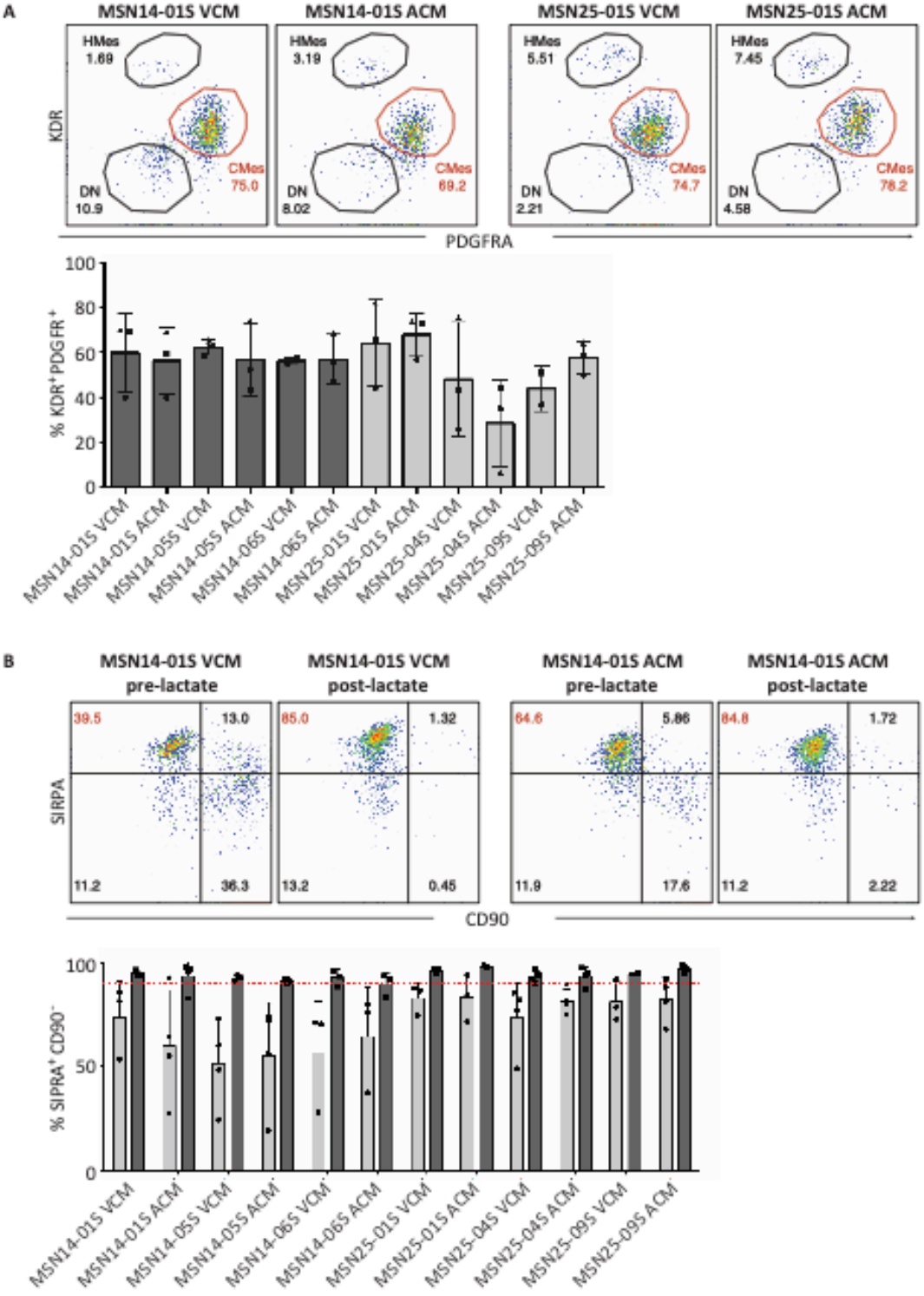
A

Subject ID	Chr	Position	Ref	Var	AD ref	AD var	GQ	Gene	Gene Function	Coding Change	Protein Change	Clinvar Calssification	Disease Association
MSN08	11	46761055	G	A	26	24	50	F2	UTR3	c.*97G>A		Pathogenic, risk factor	Venous Thromboembolysis - Stroke Risk
MSN10	17	7577539	G	A	6	27	33	TP53	Nonsynonymous SNV	c.C346T	p.R116W	Pathogenic	Li Fraumeni Syndrome - Cancer Risk
MSN18	1	152277415	G	C	18	21	39	FLG	Stopgain	c.C9947G	p.S3316X	Pathogenic	Ichthyosis Vulgaris - eczema
MSN19	7	143027960	C	T	29	14	43	CLCN1	Stopgain	c.C949T	p.R317X	Pathogenic	Myotonia Congenita
MSN29	18	29178618	G	A	37	28	65	TTR	Nonsynonymous SNV	c.G424A	p.V142I	Pathogenic	TTR-associated cardiac amyloidosis
MSN39	1	152286804	G	GT	19	18	37	FLG	Framshift insertion	c.557dupA	p.N186Kfs*4	Pathogenic	Ichthyosis Vulgaris - eczema

B



Schaniel et al. Figure 4



Schaniel et al. Figure 5