Lack of Population Diversity in Commonly Used Human Embryonic Stem-Cell Lines

TO THE EDITOR: Human embryonic stem-cell research may lead to new methods of drug discovery, insights into mechanisms of disease, and eventually, cellular therapies. The potential benefit to patient populations may depend partially on the diversity of the stem-cell lines that are available for research and clinical use. However, investigators have been unable to target their research to diverse subgroups of existing lines or to ensure the inclusion of lines from the human populations most relevant to their diseases of interest, because almost no information has been available on the human population origin of existing stem-cell lines.

Therefore, with the approval of the University of Michigan's Human Pluripotent Stem Cell Research Oversight Committee, we determined the genetic ancestry of a large collection of stem-cell lines, including the most commonly used lines that were approved for federally funded research under the Bush administration's policy, other lines derived in the United States that have been widely distributed,[1] and additional lines derived in other countries (for details, see the table in the Supplementary Appendix, available with the full text of this letter at NEJM.org).

Using the Illumina 660W genotyping platform, we genotyped genomewide single-nucleotide polymorphisms (SNPs) in each stem-cell line. Control experiments showed that the presence of mouse embryonic feeder cells did not affect the SNP genotypes (>99.99% identity of SNP genotypes between stem-cell lines that were grown with or without feeder cells) or the inferred ancestry (data not shown). Genotypes of the stem-cell lines were compared with previously obtained genotypes on a reference set of 2001 subjects from the HapMap Project and the Human Genome Diversity Project,[2,3] comprising 63 populations with worldwide representation. We analyzed 483,304 high-quality SNPs that had been genotyped in all sets of samples.

A cluster analysis[4] of combined stem-cell and worldwide reference genotypes showed that nearly all the stem-cell lines clustered exclusively with reference subjects of known European and Middle Eastern origin (Fig. 1). Two stem-cell lines clustered with East Asians. Using a European and Middle Eastern subgroup of the reference data, we found that most lines clustered primarily with subjects of northern and western European an-

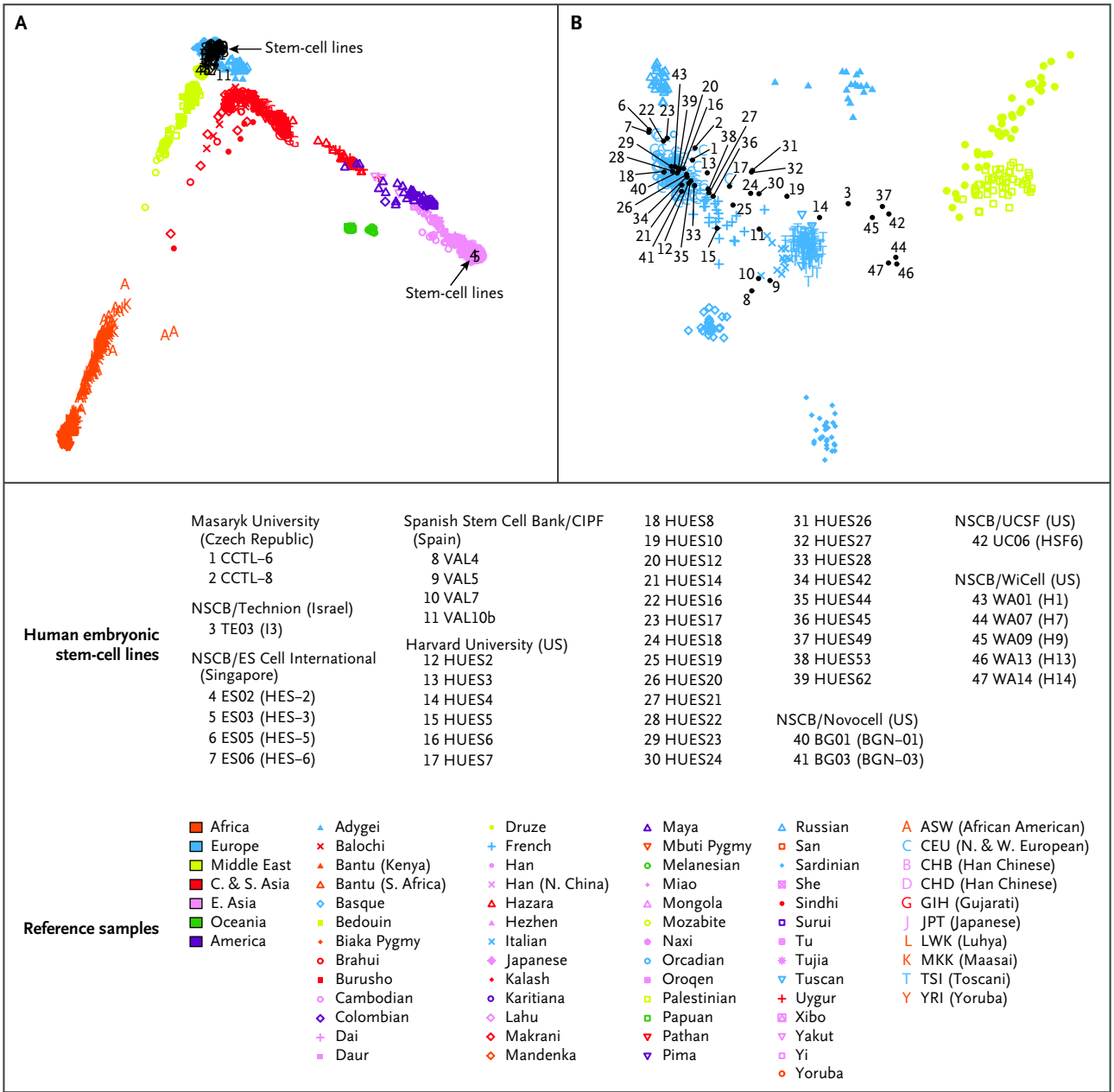

Figure 1. Cluster Analysis of Combined Stem-Cell and Worldwide Reference Genotypes.

Shown are the clustering of human embryonic stem-cell lines with subjects of known origin, including 2001 worldwide subjects (Panel A) and 458 European and Middle Eastern subjects (Panel B). In both plots, classic metric multidimensional scaling analysis was performed on pairwise individual genetic distance matrixes that were computed with the use of identity-by-state allele sharing.[4] Each sample (stem-cell lines and reference subjects) is depicted as a point so that proximate placement reflects genetic similarity. Each stem-cell line has been given a distinct numeric label. Reference subjects appear as either colored symbols (for subjects from the Human Genome Diversity Project) or letters (for subjects from the HapMap Project). Inferred sets of stem-cell lines that derive from the same gamete donors include ES05 and ES06; HUES16 and HUES17; HUES22 and HUES23; HUES26 and HUES27; and WA07, WA13, and WA14. The Mozabite population from North Africa is included as part of the "Middle East" reference sample, but close clustering of stem-cell lines with this population was not observed. CIPF denotes Prince Felipe Research Center, C. & S. Asia Central and South Asia, NSCB National Stem Cell Bank, and UCSF University of California, San Francisco.

cestry. The remaining lines clustered with Middle Eastern and southern European populations, a finding that was compatible with the derivation of some of these lines from embryos with likely origins in Israel and Spain. Interestingly, an analysis of genotype sharing identified several sets of

lines for which all lines in a given set had the same gamete donors (Fig. 1).

We have found that widely distributed stem-cell lines lack population diversity and that none of these lines derive from populations with recent African ancestry. Other existing lines that we did not analyze probably derive from populations that were not represented in our study, but most published stem-cell studies have used the lines that we investigated.[5]

Efforts to derive and disseminate new stem-cell lines should now emphasize underrepresented populations, to allow researchers to assess the extent to which the ancestry of stem-cell lines influences disease models, cellular therapies, and drug screening with the use of stem cells. Availability of more diverse lines will reduce the risk that the potential benefits of stem-cell research will be limited to patients with certain ancestries. Another promising approach to increasing the diversity of pluripotent human cell lines is to derive induced pluripotent stem-cell lines from diverse donors. It is not yet clear, however, whether certain types of studies and therapies will be more readily performed with human embryonic stem cells.

Jack T. Mosher, Ph.D.
Trevor J. Pemberton, D.Phil.
Kristina Harter
Chaolong Wang, B.S.
Erkan O. Buzbas, Ph.D.
University of Michigan
Ann Arbor, MI

Petr Dvorak, Ph.D.
Masaryk University
Brno, Czech Republic

Carlos Simón, M.D., Ph.D.
Valencia University
Valencia, Spain

Sean J. Morrison, Ph.D.
Noah A. Rosenberg, Ph.D.
University of Michigan
Ann Arbor, MI
seanjm@umich.edu

Supported by the Howard Hughes Medical Institute, the Alfred P. Sloan Foundation, the Generalitat Valenciana, the Instituto de Salud Carlos III, Fondo de Investigaciones Sanitarias grant P1081134, and donors to the University of Michigan's Center for Stem Cell Biology, especially the Jeffrey and Susan Liss Fund for the Life Sciences.

Financial and other disclosures provided by the authors are available with the full text of this letter at NEJM.org.

This letter (10.1056/NEJMc0910371) was published on December 16, 2009, at NEJM.org.

1. McCormick JB, Owen-Smith J, Scott CT. Distribution of human embryonic stem cell lines: who, when, and where. Cell Stem Cell 2009;4:107-10.

2. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851-61.

3. Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science 2008;319:1100-4.

4. Jakobsson M, Scholz SW, Scheet P, et al. Genotype, haplotype, and copy-number variation in worldwide human populations. Nature 2008;451:998-1003.

5. Scott CT, McCormick JB, Owen-Smith J. And then there were two: use of hESC lines. Nat Biotechnol 2009;27:696-7.