Research Article

Genetic Ancestry Analysis Reveals Misclassification of Commonly Used Cancer Cell Lines

Stanley E. Hooker, Jr.[1], Leanne Woods-Burnham[1], Madhavi Bathina[1], Stacy Lloyd[2], Priyatham Gorjala[3], Ranjana Mitra[3], Larisa Nonn[4], K. Sean Kimbro[5], and Rick A. Kittles[1]

Abstract

Background: Given the scarcity of cell lines from under-represented populations, it is imperative that genetic ancestry for these cell lines is characterized. Consequences of cell line mischaracterization include squandered resources and publication retractions.

Methods: We calculated genetic ancestry proportions for 15 cell lines to assess the accuracy of previous race/ethnicity classification and determine previously unknown estimates. DNA was extracted from cell lines and genotyped for ancestry informative markers representing West African (WA), Native American (NA), and European (EUR) ancestry.

Results: Of the cell lines tested, all previously classified as White/Caucasian were accurately described with mean EUR ancestry proportions of 97%. Cell lines previously classified as Black/African American were not always accurately described. For instance, the 22Rv1 prostate cancer cell line was recently found to carry mixed genetic ancestry using a much smaller panel of markers. However, our more comprehensive analysis determined the 22Rv1 cell line carries 99% EUR ancestry. Most notably, the E006AA-hT prostate cancer cell line, classified as African American, was found to carry 92% EUR ancestry. We also determined the MDA-MB-468 breast cancer cell line carries 23% NA ancestry, suggesting possible Afro-Hispanic/Latina ancestry.

Conclusions: Our results suggest predominantly EUR ancestry for the White/Caucasian-designated cell lines, yet high variance in ancestry for the Black/African American–designated cell lines. In addition, we revealed an extreme misclassification of the E006AA-hT cell line.

Impact: Genetic ancestry estimates offer more sophisticated characterization leading to better contextualization of findings. Ancestry estimates should be provided for all cell lines to avoid erroneous conclusions in disparities literature.

Introduction

In cancer research, biological model systems are used to understand the contribution of cellular mechanisms to tissue function, disease pathogenesis, and drug efficacy (1–4). Human cell lines representing various tissues in normal and disease states are vital tools necessary to establish preclinical research models. Effects and mechanisms of genetic, epigenetic, and chemical perturbations on cellular viability are explored *in vitro* and *in vivo* through gene knockdown, knockout, and transfection assays using human cell lines (4–7). Cellular assays employing human cell lines have proven essential to determine the impact of intra- and intergenetic polymorphisms on gene expression and cellular viability (8, 9). Although impactful mutations altering gene expression, cell function, or cell viability have been observed, there is significant variability in reproducibility among varying cell lines (10, 11). A portion of this variability is attributable to differences in genetic background (e.g., varying polymorphisms, epigenetic signatures, and gene expression patterns) among cell lines of the same ethnicity as well as between cell lines representing different ethnicities (12–16).

The examination of experimental outcomes using a racial/ethnic variety of cell lines exposed to the same conditions is a useful tool to elucidate genes or pathways driving the biological contribution to disease disparities (15, 17–19). Because genetic tapestries differ within and between populations, it is important to include a variety of cell lines derived from individuals within and between populations in an effort to provide an accurate representation of the genetic variability (20–22). Currently, there are many commercially available cell lines representing various tissues and disease stage. However, the majority of these cell lines are from patients of European descent, and there remains a critical need to diversify the selection (23, 24). As an example of the existing disparity in cellular model diversity, a recent search of the American Type Culture Collection (ATCC) website for cell lines derived from normal and malignant breast tissue revealed 59 specimens designated as Caucasian/White. Conversely, there were only 14 non-Caucasian/White cell lines consisting of 11

[1]Division of Health Equities, Department of Population Sciences, City of Hope Comprehensive Cancer Center, Duarte, California. [2]Department of Molecular and Cellular Biology and Alkek Center for Molecular Discovery, Baylor College of Medicine, Houston, Texas. [3]College of Medicine, Roseman University of Health Sciences, Las Vegas, Nevada. [4]The Department of Pathology, University of Illinois, Chicago, Illinois. [5]Biomedical/Biotechnology Research Institute (BBRI), North Carolina Central University, Durham, North Carolina.

S.E. Hooker Jr and L. Woods-Burnham contributed equally to this article and share first authorship.

K.S. Kimbro and R.A. Kittles contributed equally to this article and share last authorship.

Corresponding Author: Rick A. Kittles, City of Hope Comprehensive Cancer Center, 1500 E. Duarte Rd., Duarte, CA 91010. Phone: 626-218-9248; Fax: 626-301-8985; E-mail: rkittles@coh.org

designated as Black, 1 designated as Hispanic, and 1 designated as East Indian (24). Of note, 10 additional breast tissue cell lines lacked any ethnicity assignment (24).

The lack of adequate genetic population representation among human cell lines leads to a cascade of downstream consequences. It is reasonable to assume that numerous variants, genes, and pathways involved in disease pathogenesis or drug efficacy remain undiscovered when the majority of cell lines used for research are derived from patients of European descent. Because there are lower levels of variation within the European population, we miss potential genetic indicators that steer molecular mechanisms, disease outcomes, and drug response (22, 25–27). Also, SNPs identified in European populations do not completely capture the genetic variation contributing to a phenotype in other racial groups (28). More disturbingly, entire demographics are ignored in key mechanistic studies hindered by the restraints monoethnic cellular models afford (29, 30). This is problematic as underrepresented racial/ethnic groups, who are more likely to suffer disproportionate disease incidence and mortality, remain also underrepresented in biospecimens available for research (31–34).

Complementary to the aspiration of health disparity researchers to diversify commercially available biospecimens, the NIH has currently established guidelines concerning the authentication of key biological resources to ensure identity and validity (35). In response to this recommendation, a recent publication has incorporated ancestry analysis using SNP genotyping to validate previously reported ethnicities of cell lines as well as elucidate the ancestral proportions of racially unidentified cell lines (23). These efforts are important and provide the scientific community with the (1) confirmation or rejection of the ethnicity designations assigned to these cell lines; (2) prevention of wasted time and money; (3) prevention of misleading publications leading to retractions; (3) confirmed ethnicity of previously unidentified cell lines; and (4) increased variation among cell lines.

In this study, we expand upon this recent focus and include 15 human cell lines representative of various cancer types for ancestry analysis. Using a set of 105 established Ancestry Informative Markers (AIMs) validated as SNP genotypes for population structure analyses, we assign respective West African (WA), Native American (NA), and European (EUR) ancestral proportions (36, 37). Key findings of our study include the high NA proportion found within the MDA-MB-468 breast cancer cell line suggesting this patient may have been an Afro-Hispanic/Latina female, although currently simply designated as Black. We also identified the 22Rv1 cell line, currently racially unidentified yet recently categorized as carrying mixed genetic ancestry in a study using 29 AIMs, as a predominantly EUR prostate cancer cell line (23). Furthermore, we conclusively establish that the E006AA-hT prostate cancer cell line, designated as African American, in fact carries majority EUR ancestry.

Materials and Methods

Cell culture

DU145 (ATCC, Cat. # HTB-81), 22Rv1 (ATCC, Cat. # CRL-2505), and HeLa (ATCC, Cat. # CRM-CCL-2) cell lines were purchased from ATCC and grown in a humidified incubator with 5% CO_2 at 37°C in R.A. Kittles' laboratory. Cells were routinely tested for mycoplasma contamination using the PCR Mycoplasma Detection Kit (ABM, Cat. # G238). Cell lines were cultured in RPMI 1640 medium (ATCC, Cat. # 30-2001) supplemented with

10% FBS (ATCC, Cat. # 30-2020), penicillin–streptomycin (Gibco, Cat. # 15140122), and gentamicin (Fisher Scientific, Cat. # 15710064) as recommended by the supplier. Note that 0.2% Normocin (Invivogen, Cat. # ANT-NR-1) was added to the medium to prevent contamination by mycoplasma, bacteria, or fungi. In accordance with NIH guidelines concerning the authentication of key biological resources and in order to ensure the identity and validity of the resource, cell lines cultured in this study were purchased from ATCC that performs cell line characterizations. DNA was extracted from these cell lines within 2 weeks from receipt and not exceeding 3 passages. DNA extracted from the RWPE1 cell line was provided by L. Nonn. DNA extracted from the E006AA-hT cell line was tested from three separate sources, including an original purchase from ATCC, and were kindly provided by S. Ambs, A. Sreekumar, and S. Patierno. DNA extracted from the HCC1500, HCC1806, MCF-10A, MDA-MB-453, MDA-MB-468, and T-47D cell lines was provided by S. Kimbro. DNA extracted from the MDA-PCa-2b cell line was provided by S. Lloyd. DNA extracted from the RC-77T/E, LNCaP, and PC3 cell lines was provided by R. Mitra.

SNP genotyping

SNPs that were previously identified and validated for estimating continental ancestry information in admixed populations were selected to identify the AIMs (36–38). The AIMs panel consisted of 105 SNPs and were genotyped using the Sequenom MassARRAY genotyping platform with iPLEX chemistry according to the manufacturer's recommendations. iPLEX assays were designed utilizing the Sequenom Assay Design software, allowing for single-base extension designs used for multiplexing. DNA was isolated from the cell lines, and multiplex assays were performed to amplify 10 ng of genomic DNA by PCR. PCR reactions were treated with shrimp alkaline phosphatase enzyme to neutralize the unincorporated deoxyribonucleotide triphosphate. A post-PCR single-base extension reaction was performed for each multiplex reaction using concentrations of 0.625 µmol/L for low mass primers and 1.25 µmol/L for high mass primers. Reactions were diluted with 16 µL of H2O, and fragments were purified with resin, spotted onto Sequenom SpectroCHIP microarrays (Agena Bioscience, Product 10500), and scanned by MALDI-TOF mass spectrometry. Individual SNP genotype calls were generated using Sequenom TYPER software, which automatically calls allele-specific peaks according to their expected masses. A genotype concordance rate of 99% was observed for all markers. Genotyping call rates that exceeded 98.5% were included in the analyses.

DNA ancestry analysis

Individual admixture estimates for each cell line were calculated using a model-based clustering method as implemented in the program STRUCTURE v2.3 (39). STRUCTURE 2.3 was run using parental population genotypes from WAs, EURs, and NAs (36) under the Admixture model using the Bayesian Markov chain Monte Carlo method and a burn-in length of 30,000 for 70,000 repetitions. Because we are unsure about the ancestries of our cell line samples, we used the admixture model to determine which estimation of K (number of sub populations) is the best fit for the data. We set K from 2 to 5 and ran 100 iterations. We determined that K = 3 had the best fit. We used the K = 3 estimates for our analyses.

Multidimensional scaling analysis

Multidimensional scaling (MDS) analysis was performed to visualize the genetic similarity of the cell lines to worldwide populations. Note that 1000 genomes variant data were downloaded, and 824 individuals representing 8 groups were selected to represent worldwide populations—Mende in Sierra Leone (MSL); Luhya in Webuye, Kenya (LWK); Toscani in Italia (TSI); British in England and Scotland (GBR); Indian Telugu in the UK (ITU); Han Chinese in Beijing, China (CHB); Chinese Dai in Xishuangbanna, China (CDX); and Gujarati Indian in Houston, TX (GIH; ref. 40). Markers matching the 105 AIMs were then selected for the MDS analysis. Individuals were removed with missingness > 0.05. Markers were removed with missing genotypes > 0.05 or minor allele frequency < 0.05. Though sparsely located across the genome, PLINK software was still used to remove markers in linkage disequilibrium using an r^2 greater than 0.5 in a 50 SNP window with a 5 SNP sliding window in the combined cell line and 1 KG variant data (41). PLINK software was also used to perform the MDS analysis on the remaining 87 markers.

Results

One hundred five AIMs were selected from a larger previously validated set of markers to define critical genome candidate regions and characterize samples from diverse population groups (36, 37). This subset of AIMs contains specific SNPs capable of distinguishing WA, NA, and EUR genetic ancestry. Our ancestry estimates suggest cell lines previously classified as White/Caucasian by ATCC were accurately described, with mean EUR ancestry proportions of 97% (range, 92%–99%; Table 1). MCF-10A and MDA-MB-453 breast cancer cell lines were found to carry 97% and 99% EUR ancestry, respectively. PC3, DU145, LNCaP, and RWPE1 prostate cancer cell lines were found to carry 98%, 99%, 92%, and 96% EUR ancestry, respectively.

Similarly, results of some ancestry estimates for cell lines previously classified as Black/African American were accurately described. For example, the MDA-PCa-2b prostate cancer cell line was found to carry 86% WA ancestry (Table 1). Likewise, the HCC1500, HCC1806, and MDA-MB-468 breast cancer cell lines

were found to carry 79%, 80%, and 77% WA ancestry, respectively. Interestingly, the MDA-MB-468 breast cancer cell line was also found to carry an appreciable amount of NA ancestry (23%) suggesting possible Afro-Hispanic/Latina ancestry. Although the HeLa cervical cancer cell line was found to carry majority WA ancestry (66%), the WA proportion falls below the mean of approximately 80% WA ancestry typically observed in U.S.-born African Americans.

There are several cell lines that were included in our study that do not have racial identifiers specified by ATCC. For example, T-47D is a breast cancer cell line previously racially unidentified. Our ancestry analysis revealed that T-47D carries 100% EUR ancestry (Table 1). Similarly, the racial identification of the 22Rv1 prostate cancer cell line was never included within the biological characteristics specified when originally derived (42, 43). For this reason, no racial identifier has been available through ATCC (23, 24). Although a recent publication found this cell line to carry mixed genetic ancestry, our ancestry analysis revealed a majority EUR ancestry (99%; ref. 23).

Our genetic ancestry analysis of the E006AA-hT prostate cancer cell line revealed an extreme misclassification of racial identity compared with what has been reported in the literature as well as what is described by ATCC (24, 44). Although this cell line is sold commercially as an African American prostate cancer cell line, ATCC provides a disclaimer within the "Characteristics" section stating "during the accessioning of this line ATCC ran a short tandem repeat (STR) profile for the original starting material. The results match the characterization data in the cited references, however the STR profile was also found to match the STR profile (an 86% match) of another ATCC cell line, 786-O, a cell line derived from a renal cell carcinoma. The originating laboratory did not use the ATCC cell line, 786-O" (45). Due to the perplexity surrounding the true classification of this cell line, we included three separate DNA extraction samples provided by three separate laboratory groups of the E006AA-hT cell line. The SNP genotyping was found to be identical in all three E006AA-hT samples, revealing this cell line carries 91% EUR ancestry (Table 1).

We also included samples in our study of cell lines that are commonly used for research but are not commercially available.

Table 1. Ancestral proportions of commonly used human cancer cell lines

Cell line	ATCC catalog number	Gender	Age	Tissue	Designated race/ethnicity	WA	NA	EUR	Putative ancestry
PC3	CRL-1435	M	62	Prostate cancer	White/EA	0.3%	0.4%	99.3%	EA
DU145	HTB-81	M	69	Prostate cancer	White/EA	0.4%	0.5%	99.1%	EA
LNCaP	CRL-1740	M	50	Prostate cancer	White/EA	6.5%	1.9%	91.7%	EA
RWPE1	CRL-11609	M	54	Prostate cancer	White/EA	2.2%	1.7%	96.1%	EA
22Rv1	CRL-2505	M	N/A	Prostate cancer	N/A	0.3%	0.4%	99.3%	EA
E006AA-hT	CRL-3277	M	50	Prostate cancer[a]	Black/AA	1.3%	7.4%	91.3%	EA
MDA-PCa-2b	CRL-2422	M	63	Prostate cancer	Black/AA	86.5%	3.0%	10.6%	AA
RC-77T/E	N/A	M	63	Prostate cancer	Black/AA	88.9%	0.4%	10.7%	AA
HeLa	CCL-2	F	31	Cervical cancer	Black/AA	66.5%	0.4%	33.1%	AA
MCF-10A	CRL-10317	F	36	Breast disease	White/EA	0.9%	2.4%	96.7%	EA
MDA-MB-453	HBT-131	F	48	Breast cancer	White/EA	0.2%	0.4%	99.4%	EA
HCC1500	CRL-2329	F	32	Breast cancer	Black/AA	78.8%	0.5%	20.7%	AA
HCC1806	CRL-2335	F	60	Breast cancer	Black/AA	80.2%	4.4%	15.4%	AA
MDA-MB-468	HTB-132	F	51	Breast cancer	Black/AA	77.1%	22.7%	0.2%	AA/HA
T-47D	HTB-133	F	54	Breast cancer	N/A	0.1%	0.3%	99.7%	EA

Abbreviations: AA, African American; EA, European American; EUR, European; F, female; HA, Hispanic American; M, male; NA, Native American; N/A, not available; WA, West African.

[a]E006AA-hT is sold by the ATCC as a prostate cancer cell line; however, ATCC acknowledges that the STR profile of this cell line is an 86% match to the 786-O renal cell carcinoma cell line (44).

Hooker et al.

For example, the RC-77T/E prostate cancer cell line was developed from an African American prostate cancer patient, and we confirmed the WA ancestral proportions of this cell lines at 89% (46).

As an additional method to verify the accuracy of cell line genetic ancestry characterization, MDS analysis was performed to visualize the genetic similarity of the cell lines against worldwide populations from the 1000 genomes project (Fig. 1). As expected, when plotting the cell lines and 1KG groups using the first two MDS dimensions, cell lines with predominantly EUR ancestry clustered with individuals from the GBR and TSI groups and admixed cell lines with predominantly WA ancestry clustered near the MSL and LWK groups. The first two MDS dimensions provide evidence that E006AA-hT is a cell line of predominantly European descent, due to its clustering with the European 1KG individuals, and the MDA-MB-468 cell line as possibly of Afro-Hispanic/Latina descent, due to its position on the axis between the WA and East Asian groups (a proxy for NA ancestry). As the MDA-MB-468 and HeLa cell lines are more heavily admixed, their genetic similarity lies between the East Asian to WA axis and between the EUR to WA axis, respectively.

Discussion

In order to promote meaningful and quality health disparities research, there has been a recent interest in incorporating increased racial diversity among human cell lines (17, 19, 23). Racial classification remains extremely useful for describing general patterns of health as most data are reported by self-identified race (47). However, we recognize that race also embodies social and cultural constructs, and most commonly used human cell lines were developed at a time when self-reported race was considered a sufficient demographic detail (47). As the role of

Figure 1.
Genetic similarity of cell lines with worldwide populations. MDS analysis was used to determine the genetic similarity of the cell lines to groups in the 1000 genomes (1KG) project. Eighty-seven markers were used to generate MDS dimensions for 15 cell lines and 824 1KG individuals. Cell lines and 1KG groups are color-coded. Here, the East Asian group serves as a proxy for NA groups. CELL, a human cancer cell line.

biological determinants in disease acquisition and progression becomes better defined, we cannot undercut the importance of individual genetic background (48–50). Because the use of cell lines is necessary to elucidate genes and pathways driving disease disparities, it is imperative to ascertain the accurate genetic ancestry of cell lines used in preclinical research in order to adequately explore the impact of genetic contributions on incidence and progression. Recognizing the importance of precise genetic assignment for research biospecimens, we sought to confirm or negate the current racial identification of commonly used cell lines, provide accurate and robust global ancestry estimates for cell lines from admixed individuals, while also revealing the genetic ancestry of previously racially unidentified cell lines.

Overall, our ancestry analysis mostly confirmed the racial classifications previously assigned to cell lines used in this study. In other words, most cell lines classified as "Caucasian/White" carried majority EUR genetic ancestry, whereas most cell lines classified as "African American/Black" carried majority WA genetic ancestry. However, a few key findings from our study refuted what has been reported in the literature and/or by ATCC. Most erroneously, we found E006AA-hT to carry 91% EUR genetic ancestry. This finding is problematic as E006AA-hT is currently marketed and commercially available as an "African American" prostate cancer cell line (24). Although ATCC provides a disclaimer that E006AA-hT matches the STR profiling of a renal cell carcinoma cell line, 786-O, there is no indication provided that the racial identifier of "African American" is not accurate (45). Because of the increasing mistrust of the E006AA-hT cell line we have observed in our interactions with others within the prostate cancer research community, we sought to include multiple samples of this cell line from three different laboratories located in distant geographical locations to firmly establish the true genetic ancestry. The implications of our finding are considerable as many laboratories have published or are in the process of publishing manuscripts exploring prostate cancer health disparities incorporating E006AA-hT as an African American prostate cancer cell line when, in fact, the cell line is neither African American nor likely derived from the prostate. The revelation of E006AA-hT as carrying predominant EUR ancestry is further disappointing as the field of prostate cancer health disparities research is already hindered by a lack of commercially available African American/Black cell lines, and this new EUR ancestral assignment leaves MDA-PCa-2b as the sole commercially available Black prostate cell line (23, 24).

Our study also highlights the advantage of utilizing an ample number of AIMs when conducting ancestry analysis. Recently, the 22Rv1 prostate cancer cell was found to carry mixed genetic ancestry using a smaller subset of 29 AIMs (23, 36). Although subsets of AIMs as small as 24 have been shown to be useful tools for ascertaining the origin of subjects from particular continents and correct for population stratification in admixed population sample sets, we suggest a more comprehensive subset of at least 100 AIMs as a large decrease in EUR performance has been observed in marker sets smaller than 64 (36). Using a validated set of 105 AIMs, we found 22Rv1 to carry 99% EUR ancestry. Because no racial identifier has ever been assigned to the 22Rv1 cell line, our genetic ancestry results clarify the ambiguity.

Given the high heterogeneity of African Americans, it is imperative to tease out the WA ancestral proportions of individual cell lines (51). We are the first to report that the commonly used HeLa cell line derived from African American cervical cancer patient Henrietta Lacks carries 66% WA ancestry (52). The average African
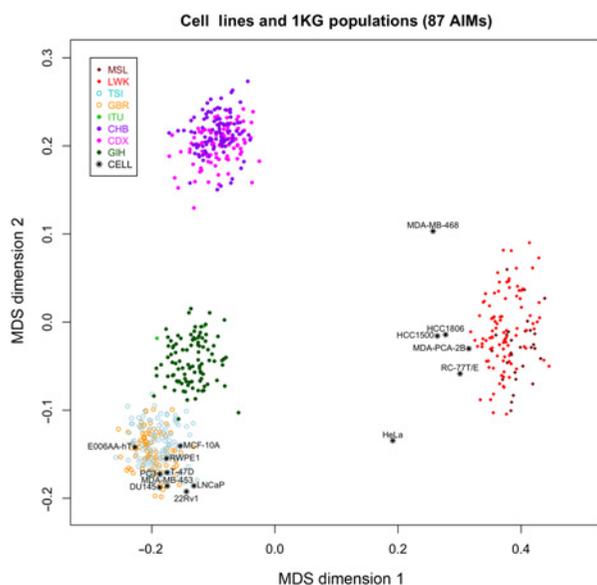
American carries approximately 80% WA genetic ancestry, yet historically, African Americans have been likely to self-identify as Black regardless of how much or little European background they may possess (53). This self-identification as Black has followed the rule of hypodescent, under which any amount of Black ancestry warrants an association as African American (54, 55). Although the social and cultural influences of race on disease are undeniable, the role of genetics cannot be ignored. For example, an admixed individual with 20% EUR genetic ancestry may be at greater or lesser risk of certain diseases than an individual with 40% EUR genetic ancestry (56–60). In addition, the implications for pharmacogenomics exist as the nuances of drug effects or mechanisms may not be generalizable in the African American population due to high WA genetic variance (28, 61). For these reasons, there remains a burden on the scientific community to expand the collection of currently available human cell lines by incorporating more non-European options to capture the complete picture of genetic contribution to disease incidence, aggressiveness, progression, and response to treatment.

The need for increased diversity in research biospecimens is glaringly obvious when noting the near-complete lack of commercially available Hispanic/Latino cell lines (24). In our own ancestry analysis of the cell lines included in this study, we observed very low proportions of NA ancestry in all cell lines except the MDA-MB-468 breast cancer cell line. Although the MDA-MB-468 cell line is reported as Black, we found that it carries 77% WA ancestry and 23% NA ancestry (24). Based on these genetic ancestry proportions, it is reasonable to presume that the breast cancer patient from whom this cell line was derived may have been Afro-Hispanic/Latina as recent studies have highlighted this admixture proportion within a Hispanic-Caribbean population (62, 63). Thus, our genetic ancestry analysis may have uncovered an additional cell line with which to measure the impact of NA ancestry on breast cancer disease incidence, progression, and treatment.

As we progress further into an era of personalized medicine, the importance of racially diverse cell lines will grow clearer. Although it is imperative that research biospecimens are designated accurately in terms of race/ethnicity, it is crucial that they be characterized globally and locally for their genetic ancestry so that findings can be properly contextualized for the representative populations. In the future, it would be ideal for commercial companies to report these global and local findings. In addition, genetic distance mapping of cell lines with the 1000 genomes or Human Genome Diversity Panel populations, using dimensional reduction techniques such as MDS or principal component analysis, should be performed to further determine the race/ethnicity of cell lines through their clustering with worldwide populations (40, 64). These techniques should further avoid misclassification that could occur with relying solely on AIMs designed to discern genetic ancestry proportions for a few discrete ancestral populations. We intend for the results of this study to encourage the scientific community to pursue ancestry analysis of additional cell lines as well as develop a wider range of diverse biospecimens.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: S. Lloyd, K.S. Kimbro, R.A. Kittles
Development of methodology: S.E. Hooker Jr, K.S. Kimbro, R.A. Kittles
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Woods-Burnham, M. Bathina, S. Lloyd, L. Nonn, K.S. Kimbro, R.A. Kittles
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.E. Hooker Jr, R.A. Kittles
Writing, review, and/or revision of the manuscript: S.E. Hooker Jr, L. Woods-Burnham, S. Lloyd, R. Mitra, L. Nonn, K.S. Kimbro, R.A. Kittles
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.E. Hooker Jr, P. Gorjala, R. Mitra, R.A. Kittles
Study supervision: K.S. Kimbro, R.A. Kittles

Acknowledgments

References

1. Grandori C, Kemp CJ. Personalized cancer models for target discovery and precision medicine. Trends Cancer 2018;4:634–42.
2. Galuschka C, Proynova R, Roth B, Augustin HG, Muller-Decker K. Models in translational oncology: a public resource database for preclinical cancer research. Cancer Res 2017;77:2557–63.
3. Kim HS, Sung YJ, Paik S. Cancer cell line panels empower genomics-based discovery of precision cancer medicine. Yonsei Med J 2015;56:1186–98.
4. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.
5. Shearer RF, Saunders DN. Experimental design for stable genetic manipulation in mammalian cell lines: lentivirus and alternatives. Genes Cells 2015;20:1–10.
6. Miest T, Saenz D, Meehan A, Llano M, Poeschla EM. Intensive RNAi with lentiviral vectors in mammalian cells. Methods 2009;47:298–303.
7. Kim TK, Eberwine JH. Mammalian cell transfection: the present and the future. Anal Bioanal Chem 2010;397:3173–8.
8. Fang X, Li X, Yin Z, Xia L, Quan X, Zhao Y, et al. Genetic variation at the microRNA binding site of CAV1 gene is associated with lung cancer susceptibility. Oncotarget 2017;8:92943–54.
9. Dansonka-Mieszkowska A, Szafron LM, Moes-Sosnowska J, Kulinczak M, Balcerak A, Konopka B, et al. Clinical importance of the EMSY gene expression and polymorphisms in ovarian cancer. Oncotarget 2018;9: 17735–55.
10. Freedman LP, Gibson MC, Ethier SP, Soule HR, Neve RM, Reid YA. Reproducibility: changing the policies and culture of cell line authentication. Nat Methods 2015;12:493–7.
11. Eisner DA. Reproducibility of science: fraud, impact factors and carelessness. J Mol Cell Cardiol 2018;114:364–8.
12. Cajigas-Du Ross CK, Martinez SR, Woods-Burnham L, Duran AM, Roy S, Basu A, et al. RNA sequencing reveals upregulation of a transcriptomic program associated with stemness in metastatic prostate cancer cells selected for taxane resistance. Oncotarget 2018;9: 30363–84.
13. Barrett CS, Millena AC, Khan SA. TGF-beta effects on prostate cancer cell migration and invasion require FosB. Prostate 2017;77:72–81.
14. Liu X, Chen X, Rycaj K, Chao HP, Deng Q, Jeter C, et al. Systematic dissection of phenotypic, functional, and tumorigenic heterogeneity of human prostate cancer cells. Oncotarget 2015;6:23959–86.

15. Shiina M, Hashimoto Y, Kato T, Yamamura S, Tanaka Y, Majid S, et al. Differential expression of miR-34b and androgen receptor pathway regulate prostate cancer aggressiveness between African-Americans and Caucasians. Oncotarget 2017;8:8356–68.

16. Teslow EA, Bao B, Dyson G, Legendre C, Mitrea C, Sakr W, et al. Exogenous IL-6 induces mRNA splice variant MBD2_v2 to promote stemness in TP53 wild-type, African American PCa cells. Mol Oncol 2018;12:1138–52.

17. Panigrahi GK, Praharaj PP, Peak TC, Long J, Singh R, Rhim JS, et al. Hypoxia-induced exosome secretion promotes survival of African-American and Caucasian prostate cancer cells. Sci Rep 2018;8:3853.

18. Sanchez TW, Zhang G, Li J, Dai L, Mirshahidi S, Wall NR, et al. Immunoseroproteomic profiling in African American men with prostate cancer: evidence for an autoantibody response to glycolysis and plasminogen-associated proteins. Mol Cell Proteomics 2016;15:3564–80.

19. Woods-Burnham L, Cajigas-Du Ross CK, Love A, Basu A, Sanchez-Hernandez ES, Martinez SR, et al. Glucocorticoids induce stress oncoproteins associated with therapy-resistance in African American and European American prostate cancer cells. Sci Rep 2018;8:15063.

20. Shriver MD, Mei R, Parra EJ, Sonpar V, Halder I, Tishkoff SA, et al. Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. Hum Genomics 2005;2:81–9.

21. Manojlovic Z, Christofferson A, Liang WS, Aldrich J, Washington M, Wong S, et al. Comprehensive molecular profiling of 718 multiple myelomas reveals significant differences in mutation frequencies between African and European descent cases. PLos Genet 2017;13:e1007087.

22. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet 2008;9:403–33.

23. Woods-Burnham L, Basu A, Cajigas-Du Ross CK, Love A, Yates C, De Leon M, et al. The 22Rv1 prostate cancer cell line carries mixed genetic ancestry: implications for prostate cancer health disparities research using pre-clinical models. Prostate 2017;77:1601–8.

24. American Type Culture Collection. 2018. [cited 2018 September 25]. Available from: https://atcc.org/products/all/CRL-3277.aspx#generalinformation.

25. Brockmoller J, Tzvetkov MV. Pharmacogenetics: data, concepts and tools to improve drug discovery and drug treatment. Eur J Clin Pharmacol 2008;64:133–57.

26. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 2000;66:979–88.

27. Lee SS. Racializing drug design: implications of pharmacogenomics for health disparities. Am J Public Health 2005;95:2133–8.

28. Hernandez W, Gamazon ER, Aquino-Michaels K, Patel S, O'Brien TJ, Harralson AF, et al. Ethnicity-specific pharmacogenetics: the case of warfarin in African Americans. Pharmacogenomics J 2014;14:223–8.

29. Mosher JT, Pemberton TJ, Harter K, Wang C, Buzbas EO, Dvorak P, et al. Lack of population diversity in commonly used human embryonic stem-cell lines. N Engl J Med 2010;362:183–5.

30. Tofoli FA, Dasso M, Morato-Marques M, Nunes K, Pereira LA, da Silva GS, et al. Increasing the genetic admixture of available lines of human pluripotent stem cells. Sci Rep 2016;6:34699.

31. Hagiwara N, Berry-Bobovski L, Francis C, Ramsey L, Chapman RA, Albrecht TL. Unexpected findings in the exploration of African American underrepresentation in biospecimen collection and biobanks. J Cancer Educ 2014;29:580–7.

32. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7–30.

33. Ragin C, Park JY. Biospecimens, biobanking and global cancer research collaborations. Ecancermedicalscience 2014;8:454.

34. Singh GK, Jemal A. Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950–2014: over six decades of changing patterns and widening inequalities. J Environ Public Health 2017;2017:2819372.

35. National Institutes of Health. Implementing rigor and transparency in NIH and AHRQ research grant applications. October 9, 2015;NOT-OD-16-011.

36. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat 2009;30:69–78.

37. Nassir R, Kosoy R, Tian C, White PA, Butler LM, Silva G, et al. An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. BMC Genet 2009;10:39.

38. Torres JB, Stone AC, Kittles R. An anthropological genetic perspective on Creolization in the Anglophone Caribbean. Am J Phys Anthropol 2013;151:135–43.

39. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003;164:1567–87.

40. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.

41. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7.

42. Sramkoski RM, Pretlow TG 2nd, Giaconia JM, Pretlow TP, Schwartz S, Sy MS, et al. A new human prostate carcinoma cell line, 22Rv1. In Vitro Cell Dev Biol Anim 1999;35:403–9.

43. Pretlow TG, Wolman SR, Micale MA, Pelley RJ, Kursh ED, Resnick MI, et al. Xenografts of primary human prostatic carcinoma. J Natl Cancer Inst 1993;85:394–8.

44. Koochekpour S, Willard SS, Shourideh M, Ali S, Liu C, Azabdaftari G, et al. Establishment and characterization of a highly tumorigenic African American prostate cancer cell line, E006AA-hT. Int J Biol Sci 2014;10:834–45.

45. American Type Culture Collection. E006AA-hT. 2018 [cited 2018 September 25]. Available from: https://atcc.org/products/all/CRL-3277.aspx#characteristics.

46. Theodore S, Sharp S, Zhou J, Turner T, Li H, Miki J, et al. Establishment and characterization of a pair of non-malignant and malignant tumor derived cell lines from an African American prostate cancer patient. Int J Oncol 2010;37:1477–82.

47. DeSantis CE, Siegel RL, Sauer AG, Miller KD, Fedewa SA, Alcaraz KI, et al. Cancer statistics for African Americans, 2016: progress and opportunities in reducing racial disparities. CA Cancer J Clin 2016;66:290–308.

48. Bird A. Genetic determinants of the epigenome in development and cancer. Swiss Med Wkly 2017;147:w14523.

49. Ozdemir BC, Dotto GP. Racial differences in cancer susceptibility and survival: more than the color of the skin? Trends Cancer 2017;3:181–97.

50. Jing L, Su L, Ring BZ. Ethnic background and genetic variation in the evaluation of cancer risk: a systematic review. PLoS One 2014;9:e97522.

51. Halder I, Yang BZ, Kranzler HR, Stein MB, Shriver MD, Gelernter J. Measurement of admixture proportions and description of admixture structure in different U.S. populations. Hum Mutat 2009;30:1299–309.

52. C BJ. HeLa (for Henrietta Lacks). Science 1974;184:1268.

53. Zakharia F, Basu A, Absher D, Assimes TL, Go AS, Hlatky MA, et al. Characterizing the admixed African ancestry of African Americans. Genome Biol 2009;10:R141.

54. Hollinger DA. The one drop rule and the one hate rule. American Academy of Arts and Sciences 2005;134:18–28.

55. National Research Council (US) Panel on Race E, and Health in Later Life. Critical perspectives on racial and ethnic differences in health in late life. Washington, DC: National Academies Press; 2004.

56. Fejerman L, John EM, Huntsman S, Beckman K, Choudhry S, Perez-Stable E, et al. Genetic ancestry and risk of breast cancer among U.S. Latinas. Cancer Res 2008;68:9723–8.

57. Al-Alem U, Rauscher G, Shah E, Batai K, Mahmoud A, Beisner E, et al. Association of genetic ancestry with breast cancer in ethnically diverse women from Chicago. PLoS One 2014;9:e112916.

58. Klimentidis YC, Arora A, Zhou J, Kittles R, Allison DB. The genetic contribution of West-African ancestry to protection against central obesity in African-American men but not women: results from the ARIC and MESA Studies. Front Genet 2016;7:89.

59. Cappetta M, Berdasco M, Hochmann J, Bonilla C, Sans M, Hidalgo PC, et al. Effect of genetic ancestry on leukocyte global DNA methylation in cancer patients. BMC Cancer 2015;15:434.

60. Bress A, Kittles R, Wing C, Hooker SE Jr, King A. Genetic ancestry as an effect modifier of naltrexone in smoking cessation among African Americans: an analysis of a randomized controlled trial. Pharmacogenet Genomics 2015; 25:305–12.

61. Kaye JB, Schultz LE, Steiner HE, Kittles RA, Cavallari LH, Karnes JH. Warfarin pharmacogenomics in diverse populations. Pharmacotherapy 2017;37:1150–63.

62. Irizarry-Ramirez M, Kittles RA, Wang X, Salgado-Montilla J, Nogueras-Gonzalez GM, Sanchez-Ortiz R, et al. Genetic ancestry and prostate cancer susceptibility SNPs in Puerto Rican and African American men. Prostate 2017;77:1118–27.

63. Via M, Gignoux CR, Roth LA, Fejerman L, Galanter J, Choudhry S, et al. History shaped the geographic distribution of genomic admixture on the island of Puerto Rico. PLoS One 2011;6:e16513.

64. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, et al. A human genome diversity cell line panel. Science 2002;296: 261–2.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

Genetic Ancestry Analysis Reveals Misclassification of Commonly Used Cancer Cell Lines

Stanley E. Hooker, Jr., Leanne Woods-Burnham, Madhavi Bathina, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/1055-9965.EPI-18-1132

Cited articles	This article cites 60 articles, 5 of which you can access for free at: http://cebp.aacrjournals.org/content/28/6/1003.full#ref-list-1
Citing articles	This article has been cited by 1 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/28/6/1003.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/28/6/1003. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.