

# Demographic Bias in Human Cell Studies

Jessica Snyder<sup>1</sup>, Iva Bojic<sup>1,2</sup>, Aarom Gerow<sup>3</sup>, Carlo Ratti<sup>1,2</sup>

<sup>1</sup>*Massachusetts Institute of Technology, SENSEable City Lab*

<sup>2</sup>*Singapore-MIT Alliance for Research and Technology*

<sup>3</sup>*The University of Chicago, Knowledge Lab*

October 26, 2016

## Abstract

### Rationale

The long term consequences of medical research is those demographics most heavily represented as donors to biopsies used in lab work have the best prognosis for treatment. Targeted breast cancer therapies and preventative screening sensitive to various tumor are required to absolve demographic inequalities in breast cancer treatment [1]. When breast cancer patients are grouped by ethnicity, black patients are less likely to survive than other ethnicities. Leading research hypothesizes some part of black patient's decreased breast cancer survival rate is due to due to diagnosis at a later stage, as well as the cancer growing at a faster rate and were more likely to spread [1]. Intrinsic factors, biological properties of the disease itself, and not just extrinsic factors, access to care, effect the efficacy of breast cancer treatment. The importance of intrinsic factors justifies a proportional and parallel research effort using cell lines from black patients for effective treatment.

[2]

Generating credible cell-based data has historically relied on consensus-building using a catalog of synonymous cell lines [3].

#### Biomedical vocabularies

Environmental exposure and genetic background determine the breast cancer's subtypes, aggressiveness and prognosis. Alternative treatment options for patients with a genetic predisposition for hormone receptor-negative breast cancers, which are also likely black, is limited [4]. Highly aggressive tumor subtypes are coorelated to risk factors such as obesity, which disproportionately effects black women.

Access to high quality hospital care is known to effect breast cancer survival rate, coupling economic status with prognosis, adversely effecting black patients [5]. Self-reported race coorelated to breast cancer properties, such as estrogen receptor status and stage of diagnosis, more than percent African ancestry, suggesting social determinants shape the disease physiology more than genetic predispositions [6].

A data-driven retrospective on disease burden and treatment effectiveness to guide medical research funding and health policy decisions towards the most productive population-level outcomes [7]. Racial difference in the efficacy of breast cancer treatment cause black woman to have a higher risk of dying than other ethnicites, the higher risk for death is consistent in all-cause and cardiovascular disease death [8].

Methylation along the DNA strand stabilizes gene expression, without methylation irregular gene expression increases the risk of cancer [9]. Demographic and environmental factors dictate the gradation of methylation in peripheral blood cells, and therefore cancer risk and tumor properties.

1. CDC statistic on on incidence and morbidity rates expose a disparity in the efficacy of treatment among demographic groups.
2. Center for Disease Control (CDC) statistics show, since 1999, Caucasian patients are more likely to survive a breast cancer diagnosis than Black patients.
3. The demographic most cited in human cell studies of breast cancer is the demographic with the highest survival rate.
4. The same positive correlation, between the demographic-specificity of the source tissue and the efficacy of the eventual medical treatment, is observed in prostate cancer

## Building the database

Cell lines are well-recognized models to study disease, especially cancer.

Cross-contamination of a cell line by another cell line cancels its utility as a standard, representative of the original donor, tissue and cell type. Therefore, accepted practice requires such cell lines be omitted from research. [10]. In this retrospective, we disregard contaminated cell lines, regardless of when the contamination occurred. Of the 336 most popular cell lines surveyed, 23 were disregarded, due to inclusion on International Cell Line Authentication Committee's Database of Cross-contaminated or Misidentified Cell Lines.

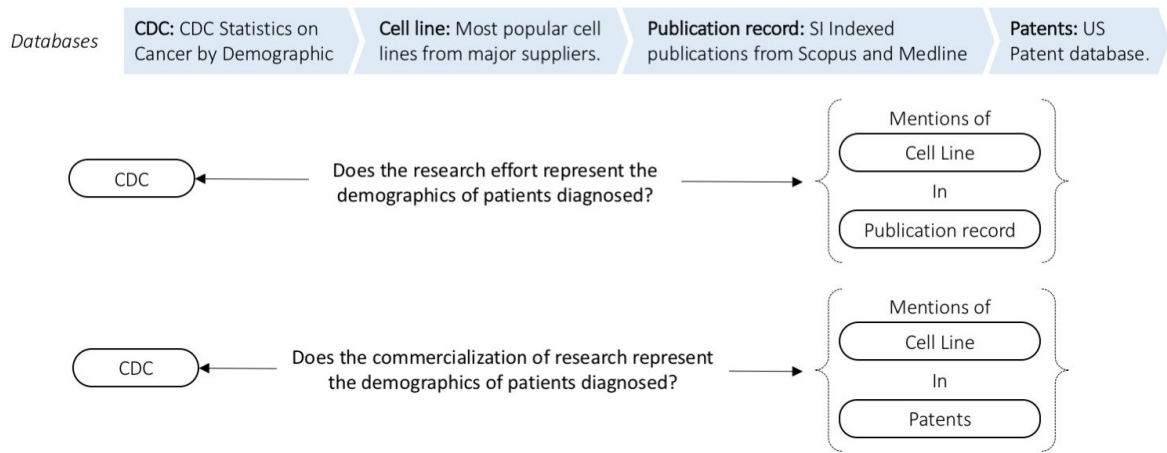


Figure 1: Diagram of databases to compare the cancer patient population to research efforts and commercialization.

## The available human cell lines

The first human cells to maintain phenotypic stability after a year in vitro were from a cervical cancer biopsy, achieving the first human "cancer in a test tube", used since as a standard for medical research [11].

1. A timeline of when cell lines were introduced, and the demographics of the donor.
2. Conclusion: Beginning in the 1950s, cell line were established from black females, followed by Caucasian men and women in the 60's, Asian in 80's, and Hispanic in the 90's.

55 A significant majority of publications use cell lines [12].

## 56 **Demographic inclusiveness of the publication record**

57 "Human cancer in a test tube" provide a standard for genotypic and phenotypic stability. Cross-contamination  
58 of cell lines causes falsified data, this is a failing of the peer review process to require authentication of the  
59 cell line as part of the experimental method [13].

60 [?]

61 1. A timeline of the number of publications in the Medline and/or Scopus database, indexed with the  
62 MESH term "Breast Neoplasms" shows >120k publications citing human cell lines derived from  
63 breast cancer since the late 70's. The demographic representation in the publication record shows  
64 >95% of publications cite a Caucasian cell line, despite the availability of alternatives.

65 2. The same trend, a majority of publications citing a human cell line were sourced from Caucasian  
66 donors, is true for prostate and lung cancer. Calculate the demographic bias in the publication records.  
67 Subtract the proportion of people diagnosed with cancer for each demographic from the proportion of  
68 papers citing a human cell line derived from a donor of the same demographic.

## 69 **Defining bias.**

$$\alpha = \sqrt{\beta} \quad (1)$$

70 1. According to the afore defined demographic bias, Caucasian are over-represented compared to other  
71 races. The bias in favor of Caucasian in breast cancer research using human cell lines has risen from  
72 15-20% since 1999.

73 2. A similar bias exists in prostate cancer research, the bias in favor of Caucasian has risen from 10-15%  
74 since 1999.

## Commercialization of the research effort

US Patents cite fewer unique cell lines than the publication database, meaning patents rely on a more uniform set of cell line models than publications, homogenizing the research standards used to substantiate claims, increasing the gap between population-level genetic diversity and medical research based on cell models.

Figure 2 presents the cumulative distribution function

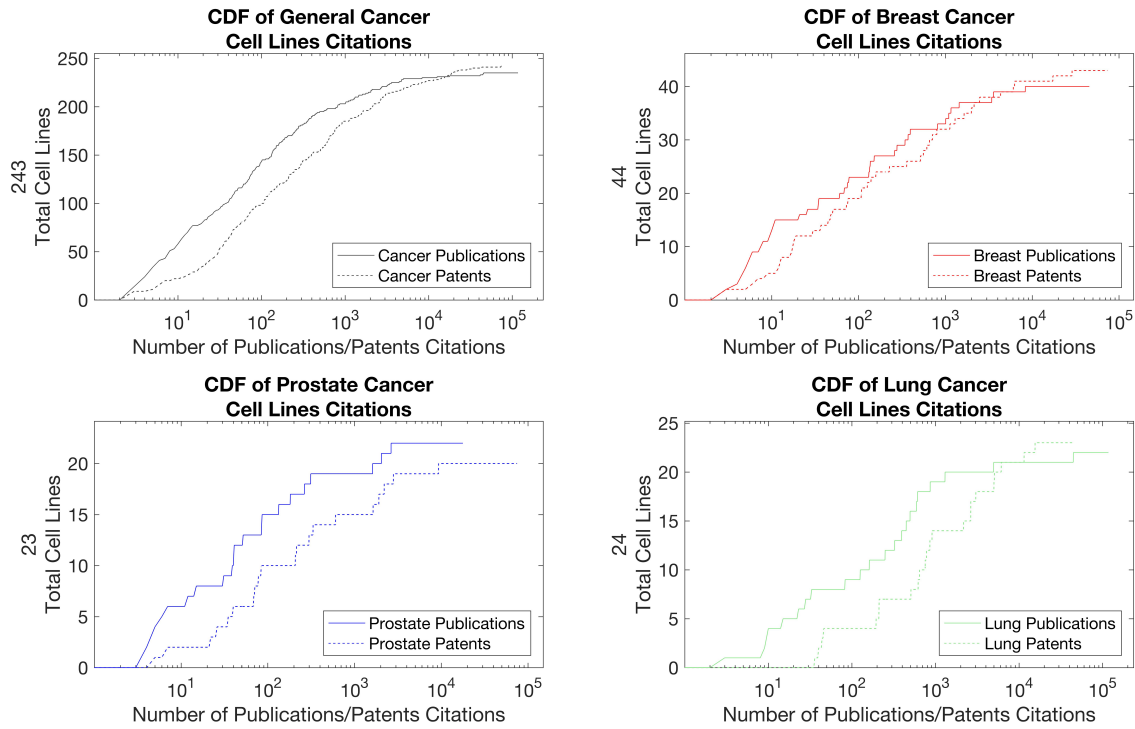


Figure 2: Cumulative distribution function (CDF) of the number of publications and patents citing human cell lines for all cancer types, breast, prostate and lung.

## Forging ahead

Summary and call to action Transformative approaches to health care, such as personalized medicine, authenticate the importance of genetic variation on the effectiveness of treatment, physicians use genomic analysis to determine the course of treatment, circumventing the limitations of generalized treatments with tailored treatments. Here, we survey the genetic variation in breast cancer research with human cell models

by characterizing the demographics of cell line donors represented in breast cancer research.

## References

- [1] Nataliya G Batina, Amy Trentham-Dietz, Ronald E Gangnon, Brian L Sprague, Marjorie A Rosenberg, Natasha K Stout, Dennis G Fryback, and Oguzhan Alagoz. Variation in tumor natural history contributes to racial disparities in breast cancer stage at diagnosis. *Breast cancer research and treatment*, 138(2):519–528, 2013.
- [2] Kevin Brennan and James M Flanagan. Is there a link between genome-wide hypomethylation in blood and cancer risk? *Cancer prevention research*, 5(12):1345–1357, 2012.
- [3] Mamie Yu, Suresh K Selvaraj, May MY Liang-Chu, Sahar Aghajani, Matthew Busse, Jean Yuan, Genee Lee, Franklin Peale, Christiaan Klijn, Richard Bourgon, et al. A resource for cell line authentication, annotation and quality control. *Nature*, 520(7547):307–311, 2015.
- [4] Dezheng Huo, Francis Ikpatt, Andrey Khramtsov, Jean-Marie Dangou, Rita Nanda, James Dignam, Bifeng Zhang, Tatyana Grushko, Chunling Zhang, Olayiwola Oluwasola, et al. Population differences in breast cancer: survey in indigenous african women reveals over-representation of triple-negative breast cancer. *Journal of Clinical Oncology*, 27(27):4515–4521, 2009.
- [5] Vickie L Shavers, Linda C Harlan, and Jennifer L Stevens. Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35. *Cancer*, 97(1):134–147, 2003.
- [6] Kerryn W Reding, Christopher S Carlson, Orsalem Kahsai, Christina C Chen, Andrew McDavid, David R Doody, Chu Chen, Kimberly Lowe, Leslie Bernstein, Linda Weiss, et al. Examination of ancestral informative markers and self-reported race with tumor characteristics of breast cancer among black and white women. *Breast cancer research and treatment*, 134(2):801–809, 2012.

- [7] Jane J Kim, Anna NA Tosteson, Ann G Zauber, Brian L Sprague, Natasha K Stout, Oguzhan Alagoz, Amy Trentham-Dietz, Katrina Armstrong, Sandi L Pruitt, Carolyn M Rutter, et al. Cancer models and real-world data: Better together. *Journal of the National Cancer Institute*, 108(2):djv316, 2016.
- [8] Amy Berkman, Bernard F Cole, Philips A Ades, Samantha Dickey, Stephen T Higgins, Amy Trentham-Dietz, Brian L Sprague, and Susan G Lakoski. Racial differences in breast cancer, cardiovascular disease, and all-cause mortality among women with ductal carcinoma in situ of the breast. *Breast cancer research and treatment*, 148(2):407–413, 2014.
- [9] Fang Fang Zhang, Roberto Cardarelli, Joan Carroll, Kimberly G Fulda, Manleen Kaur, Karina Gonzalez, Jamboor K Vishwanatha, Regina M Santella, and Alfredo Morabia. Significant differences in global genomic dna methylation by gender and race/ethnicity in peripheral blood. *Epigenetics*, 6(5):623–629, 2011.
- [10] Amanda Capes-Davis, George Theodosopoulos, Isobel Atkin, Hans G Drexler, Arihiro Kohara, Roderrick AF MacLeod, John R Masters, Yukio Nakamura, Yvonne A Reid, Roger R Reddel, et al. Check your cultures! a list of cross-contaminated or misidentified cell lines. *International journal of cancer*, 127(1):1–8, 2010.
- [11] GOea Gey, Ward D Coffman, and Mart T Kubicek. Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. In *Cancer research*, volume 12, pages 264–265. AMER ASSOC CANCER RESEARCH PO BOX 11806, BIRMINGHAM, AL 35202, 1952.
- [12] Colin F Arlett. The use of dubious cell lines in research: is trust enough? *The Lancet Oncology*, 2(8):467, 2001.
- [13] John R Masters. Hela cells 50 years on: the good, the bad and the ugly. *Nature Reviews Cancer*, 2(4):315–319, 2002.