BRIEF COMMUNICATION

Tumor Specimen Biobanks: Data Gaps for Analyzing Health Inequities—the Case of Breast Cancer

Nancy Krieger, Jaquelyn L. Jahn

See the Notes section for the full list of authors’ affiliations.

Correspondence to: Nancy Krieger, PhD, Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Kresge 717, 677 Huntington Avenue, Boston, MA 02115 (e-mail: nkrieger@hsph.harvard.edu).

Abstract

Biobanks are increasingly recognized to be vital for analyzing tumor properties, treatment options, and clinical prognosis, yet few data exist on whether they are equipped to enable research on cancer inequities, that is, unfair and unnecessary social group differences in health. We conducted a systematic search of global biobanks, identified 46 that have breast tumor tissue and share data externally with academic researchers, and e-mailed and called to obtain data on the sociodemographic, socioeconomic, and geospatial data included, plus time span encompassed. Among the 32 biobank respondents, 91% housed specimens solely from the Global North, only 31% obtained socioeconomic data, 63% included racial/ethnic data (of which 55% lacked socioeconomic data), 44% included limited geographic data, and 55% had specimens dating back at most to 2000. To enable research to address cancer inequities, including trends over time, biobanks will need to address the data gaps documented by our study.

In an age of ever-expanding biomarker, genomic, and other –omics data, biobanks are a vital resource for research on tumor properties, clinical treatment, and prognosis (1–4). Indicative of the keen interest in such data, the US National Cancer Institute is funding research to improve biobanks so as “to accelerate and/or enhance research in cancer biology, early detection and screening, clinical diagnosis, treatment, epidemiology, or address issues associated with cancer health disparities” (5).

Analysis of cancer health inequities, however, requires more than the biological specimens. Also needed are the social data used to characterize and quantify the inequities (6–8), that is, group differences in health that are unfair, unnecessary, and in principle preventable (6–8). For example, recent research in the United States has shown that understanding the etiology of, and temporal changes in, differences in breast cancer estrogen receptor status among US black and white women requires data on not only race/ethnicity and socioeconomic position but also time and place of birth in relation to the pre-1965 existence of legal racial discrimination (“Jim Crow”) in the District of Columbia and 21 of the 50 US states (11–15). The well-documented need for socioeconomic, sociodemographic, and geographic data (eg, neighborhood characteristics, urban vs rural location) to analyze cancer inequities (6,7,16,17), however, remains largely unaddressed in the literature on minimum data sets for biospecimen repositories (1–3,18–21).

Motivated by our prior US-based research on assessing the feasibility of analyzing long-term trends in disparities in breast cancer biomarkers using archival tissue specimens (22,23), we sought to assess the feasibility of using tumor biobank data for research on cancer inequities and their trends over time. Our a priori hypotheses were that most biobanks 1) would not routinely collect socioeconomic, sociodemographic, or geographic data needed to quantify cancer inequities and 2) would primarily include specimens only from the Global North, with few specimens available before the 1990s.

We used the following international and US biobank directories to identify biobanks: specimencentral.com (24), IARC biobank membership (25), National Institute of Health (NIH) Cancer Specimens Search (26), and the NIH Cooperative Human Tissue Network (27). Inclusion criteria were that the biobanks had available breast cancer tumor tissue (not collected as part of randomized controlled trials) whose data they would share externally with academic institutions (ie, excluding tissue banking services that do not share data for research). We sent out initial
Table 1. Cancer biobank data, on biobanks with breast cancer specimens that share data with academic collaborators

| Name                                                                 | Country       | Collection years | No. of cases | Preservation type | Organization type | Racial/ethnic data | Socioeconomic data | Geospatial information | Vital status | Treatment received |
|----------------------------------------------------------------------|---------------|------------------|--------------|-------------------|-------------------|--------------------|--------------------|------------------------|--------------|-------------------|
| 1) Australian Breast Cancer Tissue Bank                              | Australia     | 2004+            | 8149         | FFPE              | Academic          | No                 | No                 | No                     | Yes          | Yes               |
| 2) Biobank Graz                                                      | Austria       | 1985+            | 462 881 (includes noncase biopsy samples) | FFPE              | Academic          | No (reported "99% Caucasian") | Yes (education, occupation, birthplace, "confession," marital status, income) | Yes (birthplace) | Yes               |
| 3) Biobank of Hospital Clinic—IDIBAPS                                | Spain         | 2008+            | Data not shared† | FFPE              | Academic          | No                 | No                 | No                     | Yes          | Yes               |
| 4) Breast Cancer Family Registry                                      | Australia, Canada, USA | 1995+            | Current data not shared† (4293 as reported in 2004‡) | FFPE              | Academic          | Yes (black, white, Filipino, Japanese, Chinese, Vietnamese, other East Asian, South Asian, Middle Eastern, Hispanic) | Yes (education, marital status, language, and English proficiency) | Yes (country of birth, years in current country) | Yes               |
| 5) CHTN/NCI Breast (Prospective Procurement)                         | USA           | 1987+            | Temporally dynamic§ | Frozen tumor tissue | Academic          | Yes (categories not shared)† | No                 | No                     | Yes          | Yes               |
| 6) CHTN/NCI Specialized Tissue Microarray Resource Stage I, II, III Breast Prognostic TMAs | USA           | 1985+            | Stage I: 590 Stage II: 398 Stage III: 181 | TMA              | Academic/governmental | Yes (categories not shared)† | No                 | No                     | Yes          | Yes               |
| 7) Duke Cancer Institute and Dept. of Pathology Biobank              | USA           | 1967+            | Data not shared† | FFPE              | Academic          | Yes (categories not shared)† | Yes (occupation)     | No                     | Yes          | Yes               |

(continued)
| Name                                                                 | Country              | Collection years | No. of cases | Preservation type* | Organization type | Racial/ethnic data                             | Socioeconomic data | Geospatial information | Vital status | Treatment received |
|----------------------------------------------------------------------|----------------------|------------------|--------------|-------------------|-------------------|-----------------------------------------------|--------------------|-----------------------|--------------|--------------------|
| 8) East-West Biopharma                                                | Switzerland          | 2005+            | 17           | FFPE              | Private           | Yes (categories not shared†)                  | Yes (occupation before surgery) | No                     | Yes          | Yes                |
| 9) Hawaii Tumor Registry                                             | USA                  | 1992+            | >6000        | FFPE              | Academic          | Yes (white, black, American Indian/Alaska Native, Asian, Chinese, Filipino, Japanese, Native Hawaiian/Pacific Islander, other, multiracial, Hispanic) | No                     | Yes                    | Yes          | Yes                |
| 10) Interdisciplinary Center for Biobanking-Lübeck                   | Germany              | Data not shared† | Data not shared† | FFPE              | Academic          | No                                            | No                  | No                     | Yes          | Yes                |
| 11) Iowa Residual Tissue Repository                                  | USA                  | 1973+            | 11 064       | FFPE              | Academic          | Yes (white, black, American Indian/Alaska Native, Asian, Chinese, Filipino, Japanese, Native Hawaiian/Pacific Islander, other, multiracial, Hispanic) | No                     | Yes                    | Yes          | Yes                |
| 12) Karolinska Mammography Project for Risk Prediction of Breast Cancer | Sweden               | 2011+            | 70 877       | FFPE              | Academic          | Yes (European/not European, nativity)         | Yes (education)      | No                     | Yes          | Yes                |
| 13) Los Angeles Residual Tissue Repository                            | USA                  | 1970+            | 9972         | FFPE              | Academic/governmental | Yes (white, black, American Indian/Alaska Native, Asian, Chinese, Filipino) | No                     | Yes                    | Yes          | Yes                |
| Name | Country | Collection years | No. of cases | Preservation type* | Organization type | Racial/ethnic data | Socioeconomic data | Geospatial information | Vital status | Treatment received |
|------|---------|------------------|--------------|-------------------|-------------------|-------------------|-------------------|----------------------|--------------|---------------------|
| 14) LBIH Biobank | UK | 1993+ | Data not shared† | FFPE | Academic/ governmental | Yes (categories not shared)† | Yes (occupation) | No | Yes | Yes |
| 15) Lifelines Cohort Study | Netherlands | 2006+ | 167 000 (cases and noncases) | FFPE | Academic | Yes (white [East and West European, Mediterranean, or Arabic], Black, Asian) | Yes (employment, work, harassment, structure, social support) | Yes (longitudinal neighborhood data + baseline GIS data on air pollution, noise exposure) | Yes | Yes |
| 16) MRC Brain Banks | UK | 2013+ | 25 | FFPE | Academic | Yes (Arab, Bangladeshi, black other, black African, black Caribbean, Chinese, Indian, other, Other Asian, multiracial, Pakistani, white European, white other, Asian) | No | No | Yes | Yes |
| 17) National University of Singapore Tissue Repository | Singapore | 2002+ | 613 | FFPE | Academic | Yes (racial categories not shared but collects nationality) | No | Yes (place of death) | Yes | Yes |
| 18) NHS Grampian Biorepository | UK | 2016–2020 | < 100 | FFPE | Academic/ governmental | No | No | Yes (postal code at diagnosis) | Yes | Yes |

(continued)
| Name | Country | Collection years | No. of cases | Preservation type | Organization type | Racial/ethnic data | Socioeconomic data | Geospatial information | Vital status | Treatment received |
|------|---------|------------------|--------------|------------------|-------------------|-------------------|-------------------|----------------------|---------------|-------------------|
| 19) NHS Greater Glasgow & Clyde | UK | 2005+ | Data not shared† | FFPE | Academic/geographical | | Yes (categories not shared)† | Yes | No | Yes |
| 20) NHS Tayside | UK | 2006+ | <1000 | Data not shared† | FFPE | Academic/geographical | | Yes (categories not shared)† | Yes | Yes |
| 21) Northern Ireland Biobank | UK | 2011+ | Data not shared† | FFPE | Academic/geographical | | Yes (categories not shared)† | Yes | No | Yes |
| 22) OHSU Knight BioLibrary | USA | 2011+ | Data not shared† | FFPE | Academic/geographical | | Yes (categories not shared)† | Yes | No | Yes |
| 23) Ontario Tumour Bank | Canada | 2014+ | 666 tumor, 201 normal | FFPE | Academic/geographical | | Yes (birthplace, current address) | Yes | No | Yes |
| 24) Sapien Biosciences | India | 1997 | 17/09 | FFPE | Academic | | Yes (categories not shared)† | Yes | No | Yes |
| 25) Southwest France Tumour Bank | France | 1987+ | 3419 | FFPE | Academic/geographical | | Yes | Yes | No | Yes |
| 26) Taiwan Biobank | Taiwan | 2016+ | Data not shared† | FFPE | Hospital consortia | | Yes (education, occupation) | Yes | No | Yes |
| 27) Tumor Bank of Provence | France | 2015–2016 | 3419 | Data not shared† | Academic/geographical | | Yes (categories not shared)† | Yes | No | Yes |
| 28) University of Alabama Breast Oncology Tissue Bank | USA | 1991+ | <1000 | Data not shared† | FFPE | Academic | | Yes (categories not shared)† | Yes | Yes |
| 29) University of Leuven Tissue Bank | Belgium | 1987-2017 | 4381 | FFPE | Academic | | Yes (categories not shared)† | Yes | No | Yes |

* Note: FFPE = formalin-fixed paraffin-embedded.
| Name | Country | Collection years | No. of cases | Preservation type* | Organization type | Racial/ethnic data | Socioeconomic data | Geospatial information | Vital status | Treatment received |
|------|---------|------------------|--------------|-------------------|-------------------|--------------------|--------------------|----------------------|--------------|-------------------|
| 29) University of Arizona Cancer Prevention & Control (CPC) Biorepository http://uacc.arizona.edu/research/programs/cpc | USA | 2006+ | Data not shared† | FFPE | Academic | Yes (categories not shared)† | No | No | Yes | Yes |
| 30) BioServe https://www.bioserve.com/human-samples/ | International | 1999+ | 1271 | FFPE and frozen tumor tissue | Private | Yes (Arab, Asian, Hispanic/Latino, Black/African American, American Indian/Alaskan Native, Czech, Eastern European, Filipino, Greek, Jewish, Native Hawaiian/other Pacific Islander, Russian, Vietnamese, Kau, white/Caucasian) | Yes (occupation) | Yes (clinic location at treatment or biopsy, birth country, parents' birth country) | Yes | Yes |
| 31) Victorian Cancer Biobank https://viccancerbiobank.org.au/ | Australia | 1999+ | Data not shared† | FFPE | Government | Yes (categories not shared)† | No | Yes (state, county) | Yes | Yes |
| 32) Wales Cancer Bank http://walescancerbank.com/ | UK | 2005+ | 1400 | FFPE | Academic/governmental | No | No | No | Yes | Yes |

*Preservation type: FFPE, TMA. CHTN = Cooperative Human Tissue Network; FFPE = formalin-fixed paraffin-embedded; IDIBAPS = Institut d’Investigacions Biomèdiques August Pi i Sunyer; LBIH = Liverpool Bio-Innovation Hub; MRC = Medical Research Council; NCI = National Cancer Institute; NHS = National Health Service; OHSU = Oregon Health & Science University; SPORE = Specialized Program of Research Excellence; TMA = tumor microarray.
†Data not shared indicates that the biobank did not provide the requested information.
‡See reference 36 (John et al. 2004).
§The CHTN banks specimens prospectively, and they are provided to investigators on a first come, first served basis, resulting in a temporally dynamic number of specimens.
inquiries by e-mail in July 2017, e-mailed nonresponders twice (September and October 2017), and attempted to call once (October 2017). We asked about data collected for year of diagnosis and death, specimen preservation, race/ethnicity, socioeconomic position, and residential address/geographic information. No institutional review board (IRB) review was needed, as no human subjects were involved.

Of the 46 initial eligible biobanks we identified, 32 had specimen data available online or responded to e-mail or phone calls. We excluded 12 biobanks (10 from the United States and Europe, one from Iran, one from China) because they either did not respond (n = 10) or would not provide information without a formal data request and IRB approval (n = 2). The Global North included Europe, North America, Australia, New Zealand, and Japan; other countries comprised the Global South (28).

Table 1 presents results for the 32 included biobanks, listed in alphabetical order. Virtually all (94%) housed specimens solely from the Global North (North America [United States: 9; Canada: 2]; Western Europe: 16; or Australia: 3); two encompassed multiple countries, and three respectively included specimens from India, Singapore, and Taiwan. All 32 biobanks included data on age at diagnosis, treatment, and mortality. Among the 21 biobanks that shared data on their number of specimens (primarily cases but also some noncase biopsies), the range was from 25 to approximately 463,000. Among the 30 biobanks reporting preservation type, 29 (97%) stored formalin-fixed paraffin-embedded (FFPE) specimens, three had frozen tissue, and one had tumor microarray (TMA) specimens.

Only 10 (31%) of the 32 biobanks included any socioeconomic information, mainly education and occupation. Among the 20 biobanks (63%) with racial/ethnic data, 12 (60%) did not share the categories employed, the rest employed diverse categories (with two [10%] using the category “Caucasian”), and 11 (55%) lacked socioeconomic data. Only 14 of the 32 biobanks (44%) included any geographic information, primarily pertaining to birth place, with only eight (25%) having residential location at time of diagnosis.

Among the 31 biobanks providing the dates of their specimens, the number (and percentage) whose earliest dates of diagnosis by decade were: 1960s = 1 (3%); 1970s = 2 (6%); 1980s = 4 (13%); 1990s = 7 (23%); 2000s = 10 (32%); and 2010s = 7 (23%). Thus, the majority (55%) had specimens dating back at most to 2000, that is, less than two decades.

Supporting our a priori hypotheses, few biobanks included the social and geographic data required to analyze cancer inequities and their trends (6–8,11–17,29–31). Raising concerns about continuing the racialization of cancer data (6–8,32), among the 32 biobanks we surveyed with breast cancer tissue specimens available for use in research studies, 63% obtained racial/ethnic data (with two using the scientifically invalid and discredited spurious category of “Caucasian” [32–35]), only 31% had socioeconomic data, and only 28% had both racial/ethnic and socioeconomic data. The majority (>55%) also lacked geographic data. Additionally, these biobanks overwhelmingly (94%) contained specimens solely from countries of the Global North, primarily from the past two decades.

Our study has several strengths and limitations. First, we employed a systematic and replicable approach to identifying, contacting, and requesting information from eligible biobanks. Nevertheless, among the 46 eligible biobanks, we were unable to obtain information from 12 (of which only two were countries in the Global South), and among the 32 that provided information, not all provided information on all variables. As there is no a priori reason to believe that nonrespondents would obtain more or better-quality data than the respondents, our findings thus likely provide conservative estimates of data deficiencies.

In summary, for biobanks to enable research to address cancer inequities (6–8,11–17,29–32), the social data gaps we document require remedy. Greater expansion of the time frame also warrants support, given evidence of long-term trends and variations in the magnitude of cancer inequities (12,14,28–31).

**Funding**

This work was supported by the American Cancer Society Clinical Research Professor Award (to NK).

**Note**

Affiliation of authors: Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA.

Contributions: NK conceived the study, designed and supervised the analyses, and drafted the manuscript; JJ obtained the data and contributed to the manuscript; both authors reviewed and approved the final version prior to submission.

**References**

1. Vaught J. Biobanking comes of age: The transition to biospecimen science. *Annu Rev Pharmacol Toxicol* 2016;56:211–228.

2. Carrick DM, Mette E, Hoyle B, et al. The use of biospecimens in population-based research: A review of the National Cancer Institute’s Division of Cancer Control and Population Sciences grants portfolio. *Biogreserv Biobank* 2014;12(4):240–245.

3. Imboden M, Probst-Hensch NM. Biobanking across the phenotype—at the center of chronic disease research. *BMC Public Health* 2013;13(1):1094.

4. Kokkat TJ, Patel MS, McGarvey D, LiVolsi VA, Baloch ZW. Archived formalin-fixed paraffin-embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA, and protein. *Biogreserv Biobank* 2015;13(2):101–106.

5. National Cancer Institute. RFA-CA-17-012: innovative technologies for cancer-relevant biospecimen science. Posted November 7, 2016. https://grants.nih.gov/grants/guide/rfa-files/RFA-CA-17-012.html. Accessed March 5, 2018.

6. Krieger N, Emmons K, Williams D. Defining, investigating, and addressing cancer inequities: Critical issues. In: Koh H, ed. *Toward the Elimination of Cancer Disparities: A Clinical and Public Health Perspective*. New York: Springer; 2009:3–38.

7. Polite BN, Adams-Campbell LL, Brawley OW, et al. Charting the future of cancer health disparities research: A position statement from the American Association for Cancer Research, the American Cancer Society, the American Society of Clinical Oncology, and the National Cancer Institute. *CA Cancer J Clin* 2017;67(5):353–361.

8. Krieger N. Follow the North Star: Why space, place, and power matter for geographic approaches to cancer control and health equity. *Cancer Epidemiol Biomarkers Prev*. 2017;26(4):476–479.

9. Whitehead M. *The concepts and principles of equity and health*. Health Promot Int. 1991;6(3):217–228.

10. Braveman P, Gruskin S. Defining equity in health. *J Epidemiol Community Health*. 2003;57(4):539–545.

11. Krieger N, Jahn JL, Waterman PD. Jim Crow and estrogen-receptor negative breast cancer: US-born black & white non-Hispanic women, 1992-2012. *Cancer Causes Control*. 2017;28(3):49–59.

12. Krieger N, Jahn JL, Waterman PD, Chen JT. Breast cancer estrogen receptor by biological generation: US black & white women, born 1915-1979. [published online ahead of print September 20, 2017]. *Am J Epidemiol*. 2017. doi 10.1093/aje/kwx312.

13. Krieger N, Chen JT, Ware JH, Raddour A. Race/ethnicity and breast cancer estrogen receptor status: Impact of class, missing data, & modeling assumptions. *Cancer Causes Control*. 2008;19(10):1305–1318.

14. Krieger N, Chen JT, Waterman PD. Temporal trends in the black/white breast cancer case ratio for estrogen receptor status: Disparities are historically contingent, not innate. *Cancer Epidemiol Biomarkers Prev*. 2013;10(3):251–254.

15. Krieger N. History, biology, and health inequities: Emergent embodied phenotypes & the illustrative case of the breast cancer estrogen receptor. *Am J Public Health*. 2013;103(2):22–27.

16. Kopervinas M, Pearce N, Susser M, Boffetta P, eds. *Social Inequalities and Cancer*. IARC Scientific Publication No. 138. Lyon, France: International Agency for Research on Cancer; 1997.
17. National Cancer Institute. Cancer health disparities research. https://www.cancer.gov/research/areas/disparities. Accessed March 5, 2018.

18. Roberts JN, Karvonen C, Graham K, et al. Biobanking in the twenty-first century: Driving population metrics into biobanking quality. In: Karimi-Busheri F, eds. Biobanking in the 21st Century. Advances in Experimental Medicine and Biology. Vol. 864. Cham, Switzerland: Springer; 2015:95–114.

19. National Cancer Institute, Biorepositories Biospecimen Research Branch. NCI Best Practices for Biospecimen Resources. Bethesda, MD: National Cancer Institute; 2011. http://biospecimens.cancer.gov/bestpractices/2011-NCIBestPractices.pdf. Accessed March 5, 2018.

20. Norlin L, Fransson MN, Eriksson M, et al. A minimum data set for sharing biobank samples, information, and data: MIABIS. Biopreserv Biobank. 2012;10(4):343–348.

21. Merino-Martinez R, Norlin L, van Enckevort D, et al. Toward global biobank integration by implementation of the minimum information about biobank data sharing (MIABIS 2.0 Core). Biopreserv Biobank 2016;14(4):298–306.

22. Krieger N, Habel LA, Waterman PD, et al. Analyzing historical trends in breast cancer biomarker expression: A feasibility study (1947-2009). NPJ Breast Cancer. In press.

23. Krieger N, Nabavi S, Waterman PD, et al. Feasibility of analyzing DNA copy number variation in breast cancer tumor specimens from 1950-2010: How old is too old? Cancer Causes Control. 2018;29(3):305–314.

24. SpecimenCentral.com. http://specimencentral.com/. Accessed March 5, 2018.

25. International Agency for Research on Cancer (IARC). IARB Biobank. http://ibb.iarc.fr/collections/index.php. Accessed March 5, 2018.

26. National Institutes of Health (NIH). Specimen Resource Locator: A service of the National Cancer Institute (NCI). https://specimens.cancer.gov/search/. Accessed March 5, 2018.

27. National Institutes of Health (NIH)/National Cancer Institute (NCI). Cooperative Human Tissue Network (CHTN). https://www.chtn.org/. Accessed March 5, 2018.

28. Beckfield J, Krieger N. Epi + demos + cracy: Linking political systems and priorities to the magnitude of health inequities - evidence, gaps, and a research agenda. Epidemiol Rev. 2009;31:152–177.

29. Krieger N, Chen JT, Kosheleva A, Waterman PD. Shrinking, widening, reversing, and stagnating trends in US socioeconomic inequities in cancer mortality: 1960-2006. Cancer Causes Control. 2012;23(2):297–319.

30. Brown SB, Mallon EA, Edwards J, et al. Is the biology of breast cancer changing? A study of hormone receptor status 1984–1986 and 1996–1997. Br J Cancer. 2009;100(5):807–810.

31. Dal Bello R, Spinnillo A, Alberizzi P, Cesari S, Gardella R, Silini EM. Time trends of human papillomavirus type distribution in Italian women with cervical intraepithelial neoplasia (CIN). Gyn Oncol. 2009;115(2):262–266.

32. Krieger N. Stormy weather: “Race,” gene expression, and the science of health disparities. Am J Public Health. 2005;95(12):2155–2160.

33. Augustin HF. From the land of the Bible to the Caucasus and beyond: The shifting ideas of the geographical origin of humankind. In: Harris EB, ed. Race, Science, and Medicine, 1700-1960. London: Routlege; 1999:58–79.

34. Scheibinger L. Nature’s Body: Gender in the Making of Modern Science. Boston: Beacon Press; 1993:115–142.

35. Painter N. The History of White People. New York: W. W. Norton & Co; 2010.

36. John EM, Hopper JL, Beck JC, et al; Breast Cancer Family Registry. The Breast Cancer Family Registry: An infrastructure for cooperative multinational interdisciplinary and translational studies of the genetic epidemiology of breast cancer. Breast Cancer Res. 2004;6(4):R375–R389.