Published in final edited form as:
Cancer Epidemiol Biomarkers Prev. 2014 December ; 23(12): 2681–2687. doi: 10.1158/1055-9965.EPI-14-0490.

SEER Cancer Registry Biospecimen Research: Yesterday and Tomorrow

Sean F. Altekruse[1], **Gabriel E. Rosenfeld**[1], **Danielle M. Carrick**[1], **Emilee J. Pressman**[1], **Sheri D. Schully**[1], **Leah E. Mechanic**[1], **Kathleen A. Cronin**[1], **Brenda Y. Hernandez**[2], **Charles F. Lynch**[3], **Wendy Cozen**[4], **Muin J. Khoury**[1,5], and **Lynne T. Penberthy**[1]

[1]National Institutes of Health, National Cancer Institute, Division of Cancer Control and Population Sciences, Rockville, Maryland [2]University of Hawaii Cancer Center, Honolulu, Hawaii [3]Department of Epidemiology, College of Public Health, The University of Iowa, Iowa City, Iowa [4]Departments of Preventive Medicine and Pathology, Keck School of Medicine of the University of Southern California (USC), USC Norris Comprehensive Cancer Center, Los Angeles, California [5]Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, Georgia

Abstract

The National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) registries have been a source of biospecimens for cancer research for decades. Recently, registry-based biospecimen studies have become more practical, with the expansion of electronic networks for pathology and medical record reporting. Formalin-fixed paraffin-embedded specimens are now used for next-generation sequencing and other molecular techniques. These developments create new opportunities for SEER biospecimen research.

We evaluated 31 research articles published during 2005–2013 based on author confirmation that these studies involved linkage of SEER data to biospecimens. Rather than providing an exhaustive review of all possible articles, our intent was to indicate the breadth of research made possible by such a resource. We also summarize responses to a 2012 questionnaire that was broadly distributed to the NCI intra- and extramural biospecimen research community. This included responses from 30 investigators who had used SEER biospecimens in their research. The survey was not intended to be a systematic sample, but instead to provide anecdotal insight on strengths, limitations, and the future of SEER biospecimen research. Identified strengths of this research resource include biospecimen availability, cost, and annotation of data, including demographic information, stage, and survival. Shortcomings include limited annotation of clinical attributes such as detailed chemotherapy history and recurrence, and timeliness of turnaround following biospecimen requests. A review of selected SEER biospecimen articles, investigator feedback, and

Corresponding Author: Sean Altekruse DVM, MPH, PhD, National Cancer Institute, Division of Cancer Control and Population Sciences, 9609 Medical Center Drive, Room 4E536, Rockville, MD 20850, Office: (240) 276-6933, altekrusesf@mail.nih.gov.

Conflicts of Interest and Financial Support: The authors report no financial interests.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention, the National Cancer Institute, or the US Department of Health and Human Services.

technological advances reinforced our view that SEER biospecimen resources should be developed. This would advance cancer biology, etiology, and personalized therapy research.

Keywords

SEER; Tissue Banks; Public Health Surveillance; Cancer

Introduction

The Surveillance, Epidemiology, and End Results (SEER) Program, funded by the National Cancer Institute (NCI) since 1973, collects data from 20 cancer registries that cover 28% of the United States population to monitor cancer incidence and survival in the population and advance cancer surveillance research. As an expansion of biospecimen research, in 2001 SEER launched the Residual Tissue Repository (RTR) resource in Hawaii, Iowa, and Los Angeles to obtain formalin-fixed paraffin-embedded (FFPE) tissues before laboratories discarded them (1). The RTR has supported research on prognostic and predictive biomarkers of cancer development and progression with de-identified cancer tissue samples linked to SEER demographic data, including race and age, and clinical data such as stage, treatment, and survival. With approximately 450,000 cancer diagnoses reported annually, SEER biospecimens have potential to serve as a resource for unbiased population-based studies, even of rare cancer histologies.

Advances in biomedical science and medical reporting during the past decade offer new opportunities for SEER-linked biobanking. Advances in electronic networks enable linkage of medical records to surgical or biopsy biospecimens (2). De-identified data linkages to sources of clinical data such as Medicare claims (3) permit rich annotation of biospecimens (4). Next-generation sequencing and other advances in molecular biology now allow FFPE tissues to be used for molecular analyses (5), including studies of DNA methylation (6) and microRNA expression (7) in cancer. These developments create an opportunity for SEER registries to be a resource for acquisition of annotated biospecimens. A need exists for custom annotation of data, including chemotherapy history and recurrence, to support current research hypotheses.

Materials and Methods

This commentary includes a review of 31 selected original research articles published during 2005–2013 that linked SEER data to biospecimens. Articles were selected after obtaining author confirmation that biospecimens were linked to SEER data in these articles. The articles illustrate the breadth of research that SEER biospecimens already support. A summary of responses to an October 2012 questionnaire from cancer biospecimen researchers also is presented. Researchers were asked about their awareness of the SEER RTR and its strengths, limitations, and future directions. The results of the review of research articles and the survey suggest that further refinement and development of SEER biospecimen resources is warranted.

Results

Published Research

Selection and characteristics—We evaluated 31 original research articles (8–38) published during 2005–2013. Articles were selected based on the authors' confirmation that cancer biospecimens were linked to SEER data in the studies. Research topics addressed in these SEER biospecimen research articles included cancer classification, epidemiology, and therapeutic targets. Table 1 summarizes aspects of these articles, including cancer and biospecimen type, research focus, and patient demographics.

Cancer type—Many studies examined SEER-linked biospecimens for the leading malignant cancer diagnoses: female breast (9, 23, 24, 27, 32, 37), colon and rectum (18, 29, 30, 38), lung (10), and prostate(16) cancers. Other studies focused on cancer types responsible for an increasing proportion of cancer deaths, including pancreas (8, 21, 35, 36) and liver cancer (19, 22, 33). Other cancer types of interest were lymphomas (11–13, 26, 28); cancers of the ovaries (25), oral cavity, and pharynx (14, 31); and cervical cancer (20). Although rare, vulvar (15, 17) and anal (34) cancer biospecimen collections could be assembled from across multiple registries.

Biospecimens—The principal biospecimen type at the registries was FFPE tissues (10–17, 19, 20, 23, 25, 27–31, 34, 37, 38). FFPE cores also were used to construct tissue microarrays (TMA) (8, 9, 18, 21, 22, 24, 32, 35, 36) and multiple tumor block arrays (26, 28). In one study, FFPE tissues were used for pathology review, and frozen normal tissues were used for DNA methylation studies (33). The most common source of biospecimens was a physical repository co-located with the registry (8–15, 17–19, 21–30, 32, 34–38); however, biospecimens maintained in pathology laboratories distributed across registry catchment areas also were used in many studies (16, 17, 20, 29–31, 33, 34).

Biomarkers—Biospecimen research topics (Table 1) included assessment of immunohistochemical markers (8, 9, 18, 21–23, 30, 32, 35, 36) and genetic sequences (10–17, 19, 20, 26, 28, 29, 31, 34). Several studies examined histopathology markers (24, 25, 27, 37) and DNA methylation in cancer tissue (33, 38).

Demographics—Several studies reported cancer subtype distributions across racial and ethnic populations (9, 12–14, 17, 18, 20–22, 24, 26, 35–38). Biospecimens also were used to study molecular subtype distributions in the populations for cancer of the breast (9) and colon and rectum (30), and for non-Hodgkin lymphoma (11–13), including subtypes of diffuse large B-cell lymphoma (28) and Burkitt lymphoma (26). One publication leveraged the population-based characteristics of SEER RTR biospecimens to project the increase in future oropharyngeal cancer incidence due to human papillomavirus (HPV) infection (14). Another study reported HPV genotype distribution among vulvar cancer cases in 39 countries (15). Both of these HPV-related articles addressed the implications of their findings for ongoing HPV vaccination efforts.

Table 2 summarizes additional aspects of selected SEER biospecimen articles, including study purpose, risk factors of interest, participating SEER registries, and sources of funding.

Study purpose—One common research topic addressed in the SEER biospecimen research articles was cancer etiology (10–15, 17, 19–20, 24, 26–29, 31, 33, 34, 37, 38). Several studies used SEER survival data to assess the prognostic value of biomarkers (8, 11, 14, 21–23, 35, 36). In other instances, biospecimens were used to evaluate biomarkers that potentially were associated with the detection, development, and progression of malignancy (16, 18, 24, 33, 37, 38).

Risk factors—SEER-linked biospecimens were used in studies of risk factors associated with cancer. Although the majority of studies used de-identified linked data, several studies obtained patient consent to link their biospecimens to questionnaire responses (10–13, 24, 27, 29, 33, 37, 38). A range of exposures were examined. These included studies of HPV in head and neck (14, 31) and anogenital cancer cases (15, 17, 20, 34). Detroit registry researchers studied relationships between HPV genotypes in oropharyngeal cancers and area-level smoking data (31). Two other studies, of lung (10) and colorectal cancer (29), respectively, included individual data on tobacco use. Two studies tested hepatocellular carcinoma tumor blocks for the presence of Hepatitis B and C viruses (19, 33). The Iowa registry participated in studies of agricultural exposures associated with non-Hodgkin lymphoma (12, 13). Other exposures of interest were radon (10), soy intake (24), Epstein-Barr virus infection (26), parity (27), HIV infection (28), alcohol use (33), mammographic tissue density (37), and menopausal hormone therapy (38).

Registries—Residual biospecimens were acquired from the SEER RTR in Hawaii (8, 9, 14, 15, 17–22, 24–27, 32–37), Iowa (8, 10–17, 20, 21, 23, 25, 26, 29, 30, 34–36), and Los Angeles (8, 11, 13, 14, 17, 20, 21, 25, 26, 28, 34–36, 38). In some instances, biospecimen collection was supplemented to include tissues retrieved from pathology laboratories. SEER registries in Seattle, Detroit, Kentucky, and Louisiana also contributed biospecimens (11, 13, 17, 20, 31, 34). Other studies were collaborations between SEER and other registries (10, 12, 17, 20, 34).

Funding—SEER contracts were a major source of funding for RTR-based studies (8–30, 32–38). Specific hypothesis-driven research often was performed with targeted support. Sources of funding for this purpose were provided via NCI's SEER Rapid Response Surveillance Study mechanism (31), Intramural Program (9, 10, 12–14, 19, 25, 26, 28, 36), Office of HIV/AIDS Associated Malignancies (28), R01 research grants (16, 29, 30); an international research consortium (15); and the Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries (NPCR) (17, 20, 34).

Although not presented in table form, biospecimens linked to SEER data supported research by investigators affiliated with many institutions, including the University of Hawaii (18, 19, 22, 24, 25, 27, 32, 33, 37), University of Iowa (10, 16, 23), Mayo Clinic (11, 29, 30), University of Kentucky (20), University of Southern California (38), University of Utah (30), Wayne State University (31), Case Western Reserve University (35), University of Arkansas (21), Hospital Clinic de Barcelona (8), Institut Catala d'Oncologia (15), and University of Toronto (21).

Biospecimen Researcher Questionnaire

Questionnaire—After Office of Management and Budget clearance was obtained, a Web-based questionnaire was distributed by NCI to investigators with a known interest in biospecimen research. Responses were provided during October 2012. The goal was to assess awareness of the RTR, views on its strengths and limitations, and recommendations on the future direction of SEER biospecimen efforts. NCI's Surveillance Research Program sent an email invitation to 70 co-authors of articles that used SEER-linked biospecimens and investigators affiliated with the 20 SEER registries. NCI's Epidemiology and Genomics Research Program (EGRP) sent the invitation to investigator LISTSERV groups, including the American Association for Cancer Research Molecular Epidemiology Working Group, NCI Biospecimens, and the Division of Cancer Control and Population Sciences' Friends of EGRP. Recipients were asked to forward the email invitation to peers in the cancer research community. The online questionnaire was not intended to provide systematic information. Instead, anecdotal input from the research community was meant to provide insight on strengths, limitations, and the future of SEER biospecimen research.

Questionnaire responses—The questionnaire included 10 questions (Table 3). The number of responses varied for each question.

Respondents' backgrounds (Table 3A): The majority of the 174 overall respondents (67%) were affiliated with academic institutions (Q1). A total of 30 respondents indicated that they had accessed SEER biospecimens (Q2). Of 90 respondents who had not used the resource, 40% were waiting to obtain preliminary results or funding before applying, and the resource did not meet the needs of 31% (Q3). Some but not all of the 90 respondents continued with question 8, after skipping questions 4–7, which were directed at investigators who had used the SEER RTR.

Responses of SEER RTR researchers (Table 3B): Only investigators who indicated that they had used the RTR were asked questions 4–7. Among 26 RTR users who responded to the question about their research (Q4), interests included biomarker identification/validation (31%), whole-genome analysis (27%), multivariate molecular profiling (23%), and other uses (19%). Among 22 RTR users who answered whether or not the RTR met their needs, 19 (86%) indicated that the RTR meet their research needs (Q5). A total of 24 previous RTR researchers provided comments on the benefits of SEER RTR biospecimens (Q6a), which included population coverage (42%), the number of biospecimens available (29%), and cost (13%). Limitations listed by 22 previous users (Q6b) included sample size (36%), quality control documentation (36%), and incomplete clinical annotation (27%). In response to question 7, a total of 24 previous RTR investigators provided recommendations for improving access to SEER-linked biospecimens. Recommendations included increasing the number of biospecimens available (25%) and developing a more streamlined application process (21%).

Future SEER biospecimen resources (Table 3C): Forty-three investigators provided recommendations on the future direction of SEER biospecimen research (Q8). Priorities included prognostic biomarker studies (33%), biomarker identification and validation (21%),

and molecular profiling for the purpose of tumor classification (19%). Among 25 investigators who commented on tissue quality issues (Q9), efforts to ensure biospecimen quality control were seen as useful (64%), and integration of pathology review and more detailed annotation were both recommended (12%). Availability of age-matched controls and adjacent normal tissue were additional recommended enhancements. More than one response could be selected for question 10, pertaining to annotation needs, and a total of 70 responses were provided. In descending frequency, researchers listed these items as very important: tissue collection, processing, and storage conditions (58% of researchers); type of treatment received (56% of researchers); age of biospecimen (52% of researchers); risk factors associated with cancer diagnoses (42% of researchers); and type of health insurance (4% of researchers).

Discussion

A review of a selection of SEER registry-based biospecimen articles demonstrates the breadth of research that this resource can support. Innovations in molecular biology are expanding the potential value of FFPE biospecimens as a resource for biomedical research. Advances in electronic medical record reporting also can assist registries in locating and annotating tissues that meet study criteria.

Although fresh frozen tissue collections are a gold standard for preserving nucleic acids and proteins, the expense of procurement and maintenance may not be feasible in many clinical or research settings. Fortunately, methods of nucleic acid and protein analysis using FFPE samples have advanced rapidly, expanding their potential for research on the molecular mechanisms of cancer (39), including microRNA profiles (40, 41), genome-wide analysis of copy number and mutations (42), whole-genome methylation (6), other epigenetic markers (43), and proteomic studies with FFPE samples (44, 45). Thus, FFPE biospecimens, drawn from unbiased SEER catchments, hold promise for cancer research. The potential to annotate these biospecimens with detailed demographic and clinical data from electronic records is another compelling aspect of performing biospecimen research using data from SEER registries.

Based on anecdotal information gained from the investigator questionnaire, several key goals were identified for future registry-based biospecimen research. These include implementation of an efficient, centralized process with consistent methods for tissue acquisition to support hypothesis-driven biospecimen research. Linkage to external data sources would enhance biospecimen annotation with detailed information on risk factors, co-morbidities, and treatment. The use of SEER-linked biospecimens could be an efficient mechanism to reduce research costs by assisting in case ascertainment, biospecimen acquisition, annotation, and follow-up of vital status. To realize this goal, Institutional Review Board (IRB) and material transfer agreement (MTA) processes should be simplified and expedited to the extent possible. In this way, SEER biospecimen processes could increase sample size, statistical power, and diligent completion of biospecimen acquisition for case-only, case-control, and cohort studies, as well as clinical trials.

A combination of centralized processes and dedicated registry staff is recommended to facilitate SEER multiregistry biospecimen activities. Central coordination processes can help to locate and coordinate acquisition of biospecimens that meet specific study criteria. Dedicated personnel at the registry level are essential to developing trusting relationships between collaborating pathology laboratories to retrieve, annotate, and transfer biospecimens to investigators. Ethical issues involving informed consent should be addressed to make these processes run smoothly. The engaged support of registries, medical facilities, providers, patients, and community advocates will be essential for this large-scale, population-based biospecimen resource to be successful (46). In summary, registry-linked biospecimens hold promise as a resource for cancer research. Carefully developing this resource is a priority of NCI's SEER cancer registry program.

References

1. Goodman MT, Hernandez BY, Hewitt S, Lynch CF, Cote TR, Frierson HF Jr, et al. Tissues from population-based cancer registries: a novel approach to increasing research potential. Hum Pathol. 2005; 36:812–20. [PubMed: 16084952]
2. Olson JE, Bielinski SJ, Ryu E, Winkler EM, Takahashi PY, Pathak J, et al. Biobanks and personalized medicine. Clin Genet. 2014; 86:50–55. [PubMed: 24588254]
3. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care. 2002; 40(8 Suppl):IV-3–18.
4. Weiner MG, Lyman JA, Murphy S, Weiner M. Electronic health records: high-quality electronic data for higher-quality clinical research. Inform Prim Care. 2007; 15:121–7. [PubMed: 17877874]
5. Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M, et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med. 2012; 18:382–4. [PubMed: 22327622]
6. Gu H, Bock C, Mikkelsen TS, Jager N, Smith ZD, Tomazou E, et al. Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat Methods. 2010; 7:133–6. [PubMed: 20062050]
7. Weng L, Wu X, Gao H, Mu B, Li X, Wang JH, et al. MicroRNA profiling of clear cell renal cell carcinoma by whole-genome small RNA deep sequencing of paired frozen and formalin-fixed, paraffin-embedded tissue specimens. J Pathol. 2010; 222:41–51. [PubMed: 20593407]
8. Altirriba J, Garcia A, Sanchez B, Haba L, Altekruse S, Stratmann T, et al. The sole presence of CDK4 is not a solid criterion for discriminating between tumor and healthy pancreatic tissues. Int J Cancer. 2012; 130:2743–5. [PubMed: 21792898]
9. Anderson WF, Luo S, Chatterjee N, Rosenberg PS, Matsuno RK, Goodman MT, et al. Human epidermal growth factor receptor-2 and estrogen receptor expression, a demonstration project using the residual tissue repository of the Surveillance, Epidemiology, and End Results (SEER) Program. Breast Cancer Res Treat. 2009; 113:189–96. [PubMed: 18256926]
10. Bonner MR, Bennett WP, Xiong W, Lan Q, Brownson RC, Harris CC, et al. Radon, secondhand smoke, glutathione-S-transferase M1 and lung cancer among women. Int J Cancer. 2006; 119:1462–7. [PubMed: 16642467]
11. Cerhan JR, Natkunam Y, Morton LM, Maurer MJ, Asmann Y, Habermann TM, et al. LIM domain only 2 protein expression, LMO2 germline genetic variation, and overall survival in diffuse large B-cell lymphoma in the pre-rituximab era. Leuk Lymphoma. 2012; 53:1105–12. [PubMed: 22066713]
12. Chang CM, Schroeder JC, Huang WY, Dunphy CH, Baric RS, Olshan AF, et al. Non-Hodgkin lymphoma (NHL) subtypes defined by common translocations: utility of fluorescence in situ hybridization (FISH) in a case-control study. Leuk Res. 2010; 34:190–5. [PubMed: 19505720]

13. Chang CM, Wang SS, Dave BJ, Jain S, Vasef MA, Weisenburger DD, et al. Risk factors for non-Hodgkin lymphoma subtypes defined by histology and t(14;18) in a population-based case-control study. Int J Cancer. 2011; 129:938–47. [PubMed: 20949561]

14. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. J Clin Oncol. 2011; 29:4294–301. [PubMed: 21969503]

15. de Sanjose S, Alemany L, Ordi J, Tous S, Alejo M, Bigby SM, et al. Worldwide human papillomavirus genotype attribution in over 2000 cases of intraepithelial and invasive lesions of the vulva. Eur J Cancer. 2013; 49:3450–61. [PubMed: 23886586]

16. Esser AK, Miller MR, Huang Q, Meier MM, Beltran-Valero de Bernabe D, Stipp CS, et al. Loss of LARGE2 disrupts functional glycosylation of alpha-dystroglycan in prostate cancer. J Biol Chem. 2013; 288:2132–42. [PubMed: 23223448]

17. Gargano JW, Wilkinson EJ, Unger ER, Steinau M, Watson M, Huang Y, et al. Prevalence of human papillomavirus types in invasive vulvar cancers and vulvar intraepithelial neoplasia 3 in the United States before vaccine introduction. J Low Genit Tract Dis. 2012; 16:471–9. [PubMed: 22652576]

18. Hernandez BY, Frierson HF, Moskaluk CA, Li YJ, Clegg L, Cote TR, et al. CK20 and CK7 protein expression in colorectal cancer: demonstration of the utility of a population-based tissue microarray. Hum Pathol. 2005; 36:275–81. [PubMed: 15791572]

19. Hernandez BY, Zhu X, Kwee S, Chan O, Tsai N, Okimoto G, et al. Viral hepatitis markers in liver tissue in relation to serostatus in hepatocellular carcinoma. Cancer Epidemiol Biomarkers Prev. 2013; 22:2016–23. [PubMed: 23983238]

20. Hopenhayn CA, Christian WJ, Watson M, Unger E, Lynch CF, Peters ES, et al. Prevalence of human papillomavirus types in invasive cervical cancers in seven US states prior to vaccine introduction. J Low Genit Tract Dis. 2014; 18:182–9. [PubMed: 24477171]

21. Iakovlev V, Siegel ER, Tsao MS, Haun RS. Expression of kallikrein-related peptidase 7 predicts poor prognosis in patients with unresectable pancreatic ductal adenocarcinoma. Cancer Epidemiol Biomarkers Prev. 2012; 21:1135–42. [PubMed: 22573795]

22. Kwee SA, Hernandez B, Chan O, Wong L. Choline kinase alpha and hexokinase-2 protein expression in hepatocellular carcinoma: association with survival. PLoS One. 2012; 7:e46591. [PubMed: 23071593]

23. Lal G, Hashimi S, Smith BJ, Lynch CF, Zhang L, Robinson RA, et al. Extracellular matrix 1 (ECM1) expression is a novel prognostic marker for poor long-term survival in breast cancer: a Hospital-based Cohort Study in Iowa. Ann Surg Oncol. 2009; 16:2280–7. [PubMed: 19521735]

24. Maskarinec G, Erber E, Verheus M, Hernandez BY, Killeen J, Cashin S, et al. Soy consumption and histopathologic markers in breast tissue using tissue microarrays. Nutr Cancer. 2009; 61:708–16. [PubMed: 19838945]

25. Matsuno RK, Sherman ME, Visvanathan K, Goodman MT, Hernandez BY, Lynch CF, et al. Agreement for tumor grade of ovarian carcinoma: analysis of archival tissues from the Surveillance, Epidemiology, and End Results Residual Tissue Repository. Cancer Causes Control. 2013; 24:749–57. [PubMed: 23378140]

26. Mbulaiteye PS, Nathwani BN, Weiss LM, Rao N, Emmanuel B, Lynch CF, et al. Epstein-Barr virus patterns in U.S. Burkitt lymphoma tumors from the SEER Residual Tissue Repository during 1979-2009. APMIS. 2014; 122:5–15. [PubMed: 23607450]

27. Morimoto Y, Killeen J, Hernandez BY, Mark Cline J, Maskarinec G. Parity and expression of epithelial histopathologic markers in breast tissue. Eur J Cancer Prev. 2013; 22:404–8. [PubMed: 23222412]

28. Morton LM, Kim CJ, Weiss LM, Bhatia K, Cockburn M, Hawes D, et al. Molecular characteristics of diffuse large B-cell lymphoma in HIV-infected and HIV-uninfected patients in the pre-HAART and pre-rituximab era. Leuk Lymphoma. 2014; 55:551–7. [PubMed: 23772639]

29. Samadder NJ, Vierkant RA, Tillmans LS, Wang AH, Lynch CF, Anderson KE, et al. Cigarette smoking and colorectal cancer risk by Kras mutation status among older women. Am J Gastroenterology. 2012; 107:782–9.

30. Samadder NJ, Vierkant RA, Tillmans LS, Wang AH, Weisenberger DJ, Laird PW, et al. Associations between colorectal cancer molecular markers and pathways with clinicopathologic features in older women. Gastroenterology. 2013; 145:348–56 e2. [PubMed: 23665275]
31. Sethi S, Ali-Fehmi R, Franceschi S, Struijk L, van Doorn LJ, Quint W, et al. Characteristics and survival of head and neck cancer by HPV status: a cancer registry-based study. Int J Cancer. 2012; 131:1179–86. [PubMed: 22020866]
32. Shimizu Y, Luk H, Horio D, Miron P, Griswold M, Iglehart D, et al. BRCA1-IRIS overexpression promotes formation of aggressive breast cancers. PLoS One. 2012; 7:e34102. [PubMed: 22511931]
33. Song MA, Tiirikainen M, Kwee S, Okimoto G, Yu H, Wong LL. Elucidating the landscape of aberrant DNA methylation in hepatocellular carcinoma. PLoS One. 2013; 8
34. Steinau M, Unger ER, Hernandez BY, Goodman MT, Copeland G, Hopenhayn C, et al. Human papillomavirus prevalence in invasive anal cancers in the United States before vaccine introduction. J Low Genit Tract Dis. 2013; 17:397–403. [PubMed: 23609590]
35. Sy MS, Altekruse SF, Li C, Lynch CF, Goodman MT, Hernandez BY, et al. Association of prion protein expression with pancreatic adenocarcinoma survival in the SEER Residual Tissue Repository. Cancer Biomark. 2011; 10:251–8. [PubMed: 22820080]
36. Takikita M, Altekruse S, Lynch CF, Goodman MT, Hernandez BY, Green M, et al. Associations between selected biomarkers and prognosis in a population-based pancreatic cancer tissue microarray. Cancer Res. 2009; 69:2950–5. [PubMed: 19276352]
37. Verheus M, Maskarinec G, Erber E, Steude JS, Killeen J, Hernandez BY, et al. Mammographic density and epithelial histopathologic markers. BMC Cancer. 2009; 9:182. [PubMed: 19523235]
38. Wu AH, Siegmund KD, Long TI, Cozen W, Wan P, Tseng CC, et al. Hormone therapy, DNA methylation and colon cancer. Carcinogenesis. 2010; 31:1060–7. [PubMed: 20064828]
39. Klopfleisch R, Weiss AT, Gruber AD. Excavation of a buried treasure—DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. Histology Histopathol. 2011; 26:797–810.
40. Liu A, Xu X. MicroRNA isolation from formalin-fixed, paraffin-embedded tissues. Methods Mol Biol. 2011; 724:259–67. [PubMed: 21370018]
41. Szafranska AE, Davison TS, Shingara J, Doleshal M, Riggenbach JA, Morrison CD, et al. Accurate molecular characterization of formalin-fixed, paraffin-embedded tissues by microRNA expression profiling. J Mol Diagn. 2008; 10:415–23. [PubMed: 18687792]
42. Schweiger MR, Kerick M, Timmermann B, Albrecht MW, Borodina T, Parkhomchuk D, et al. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. PLoS One. 2009; 4:e5548. [PubMed: 19440246]
43. Fanelli M, Amatori S, Barozzi I, Minucci S. Chromatin immunoprecipitation and high-throughput sequencing from paraffin-embedded pathology tissue. Nature Protoc. 2011; 6:1905–19. [PubMed: 22082985]
44. Lemaire R, Desmons A, Tabet JC, Day R, Salzet M, Fournier I. Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections. J Proteome Res. 2007; 6:1295–305. [PubMed: 17291023]
45. Negishi A, Masuda M, Ono M, Honda K, Shitashige M, Satow R, et al. Quantitative proteomics using formalin-fixed paraffin-embedded tissues of oral squamous cell carcinoma. Cancer Sci. 2009; 100:1605–11. [PubMed: 19522851]
46. Olson JE, Ryu E, Johnson KJ, Koenig BA, Maschke KJ, Morrisette JA, et al. The Mayo Clinic Biobank: a building block for individualized medicine. Mayo Clin Proc. 2013; 88:952–962. [PubMed: 24001487]

Table 1

Cancer, biospecimen, research focus, and demography, 31 SEER biospecimen articles, 2005–2013

Attribute	Category	N	References
Cancer Type			
	Breast	6	(9, 23, 24, 27, 32, 37)
	Lymphoma	5	(11–13, 26, 28)
	Anogenital	5	(15, 17, 20, 25, 34)
	Colon and Rectum	4	(18, 29, 30, 38)
	Pancreas	4	(8, 21, 35, 36)
	Liver	3	(19, 22, 33)
	Oropharynx	2	(14, 31)
	Lung, Prostate	2	(10, 16 [respectively])
Biospecimen			
	FFPE Tissues Other Than Arrays	19	(10–17, 19, 20, 23, 25, 27, 29–31, 34, 37, 38)
	Tissue Microarray/Multiple Tumor Arrays	11	(8, 9, 18, 21, 22, 24, 26, 28, 32, 35, 36)
Biomarker			
	Immunohistochemical Staining	10	(8, 9, 18, 21–23, 30, 32, 35, 36)
	Genetic Sequences	15	(10–17, 19, 20, 26, 28, 29, 31, 34)
	Histopathology Markers	4	(24, 25, 27, 37)
	DNA Methylation	2	(33, 38)
Demographics			
	Racial and Ethnic Distribution	15	(9, 12–14, 17, 18, 20–22, 24, 26, 35–38)
	Molecular Subtype Distribution	7	(9, 11–13, 26, 28, 30)
	Trend Projection, Subtype Distribution	2	(14, 15)

Table 2

Study purpose, risk factors, registries, and funding, 31 SEER biospecimen studies, 2005–2013

Attribute	Category	N	References
Study Purpose			
	Etiology	19	(10–15, 17, 19, 20, 24, 26–29, 31, 33, 34, 37, 38)
	Prognosis	8	(8, 11, 14, 21–23, 35, 36)
	Detection, Development, Progression	6	(16, 18, 24, 33, 37, 38)
Risk Factors			
	Human Papillomavirus	6	(14, 15, 17, 20, 31, 34)
	Tobacco	3	(10, 29, 31)
	Viral Hepatitis	2	(19, 33)
	Agricultural Chemical Exposure	2	(12, 13)
	Other (one each)[a]	8	(10, 24, 26, 27, 28, 33, 37, 38)
Registries			
	Hawaii	20	(8, 9, 14, 15, 17, 22, 24–27, 32–37)
	Iowa	19	(8, 10–17, 20, 21, 23, 29, 30, 34–36)
	Los Angeles	15	(8, 11, 13, 14, 17, 20, 21, 25, 26, 34–36, 38)
	Detroit, Kentucky, Seattle, Louisiana	5	(11, 13, 17, 20, 31)
	SEER/NPCR Collaboration	3	(17, 20, 34)
Funding			
	SEER Contract	30	(8–30, 32–38)
	NCI Intramural Program	10	(9, 10, 12–14, 19, 25, 26, 28, 36)
	CDC NPCR	3	(17, 20, 34)
	Other[b]	3	(15, 16,28-31)

[a] Radon (10), soy intake (24), Epstein-Barr virus infection (26), parity (27), HIV infection (28), alcohol use (33), mammographic tissue density (37), and menopausal hormone therapy (38).

[b] International Public and Private Consortium (15), NCI Office of HIV & AIDS Associated Malignancies (28), RO1 grant from NCI (16, 29, 30), NCI Rapid Response Surveillance Studies (31)

Table 3

Responses of biospecimen researchers to questions on the SEER Residual Tissue Repository (RTR)

A. Background of biospecimen research questionnaire respondents	
Q1. If you conduct scientific research, please indicate your primary affiliation. *(174 responses)*	
Academic	117 (67%)
Government	24 (14%)
Other	33 (19%)
Q2. Have you worked with biospecimens from the SEER RTR in the past? *(159 responses)*	
Yes	30 (19%)
No	129 (81%)
Q3. If you are aware of the SEER RTR but have not worked with this resource in the past, please indicate why and continue with Question 8. *(90 responses)*	
Plan to apply once preliminary results are obtained or obtain funding	36 (40%)
Did not meet research needs	28 (31%)
Unaware of RTR resource	15 (17%)
Other	11 (12%)

B. Responses of SEER Biospecimen researchers on SEER RTR research use and potential *(n = 30)*	
Q4. If you answered YES to question 2, what were your research objectives in using the SEER RTR resource? *(26 responses)*	
Biomarker identification/validation	8 (31%)
Whole-genome analysis	7 (27%)
Multivariate molecular profiling	6 (23%)
Other	5 (19%)
Q5. Did the SEER RTR resource enable you to achieve your research goals? *(22 responses)*	
Yes	19 (86%)
No	3 (14%)
Q6a. Please comment on any advantages (strengths) of using the SEER RTR as a research resource. *(24 responses)*	
Population coverage	10 (42%)
Number of biospecimens	7 (29%)
SEER annotation (demographic, clinical, and survival data)	4 (17%)
Cost/speed of access (convenience)	3 (13%)

B. Responses of SEER Biospecimen researchers on SEER RTR research use and potential *(n = 30)*

Q6b. Please comment on any disadvantages (weaknesses) of using the SEER RTR as a research resource. *(22 responses)*	
Insufficient sample size	8 (36%)
Incomplete QC documentation	8 (36%)
Incomplete clinical annotation	6 (27%)
Q7. Please provide suggestions for improving your ability to access and utilize the SEER RTR biospecimens and associated data. *(24 responses)*	
Increase number of biospecimens	6 (25%)
Improve efficiency of access to biospecimens and associated data	6 (25%)
Streamline application process (IRB/MTA)	5 (21%)
Increase RTR funding/staff	4 (17%)
More targeted annotation of clinical data	3 (13%)

C. Future development of SEER biospecimen resources

Q8. Please elaborate on specific research objectives that you would like to see addressed in the future using the SEER RTR. *(43 responses)*	
Prognostic studies	14 (33%)
Other	12 (28%)
Biomarker identification/validation	9 (21%)
Molecular profiling for tumor classification	8 (19%)
Q9. Please comment on methods or techniques that could be used to assess the tissue quality of SEER RTR biospecimens to enhance their utility for advanced research applications, such as next-generation sequencing. *(25 responses)*	
Sample QC*	16 (64%)
Pathology review	3 (12%)
Upgraded annotation	3 (12%)
Age-matched control	2 (8%)
Adjacent tissue samples	1 (4%)
Q10. Please indicate the importance of the following standard SEER data items for research using SEER RTR biospecimens. *(70 responses, selection of multiple categories allowed)*	
Tissue collection, processing, and storage	41 (58%)
Type of treatment	39 (56%)
Age of specimens	37 (52%)
Risk factors	30 (42%)

C. Future development of SEER biospecimen resources	
Type of health insurance	3 (4%)

* Immunohistochemistry, *In situ* hybridization, Polymerase chain reaction