The Use of Biospecimens in Population-Based Research: A Review of the National Cancer Institute's Division of Cancer Control and Population Sciences Grant Portfolio

Danielle M. Carrick, Eliza Mette, Brittany Hoyle, Scott D. Rogers, Elizabeth M. Gillanders, Sheri D. Schully, and Leah E. Mechanic

Over the past two decades, researchers have increasingly used human biospecimens to evaluate hypotheses related to disease risk, outcomes and treatment. We conducted an analysis of population-science cancer research grants funded by the National Cancer Institute (NCI) to gain a more comprehensive understanding of biospecimens and common derivatives involved in those studies and identify opportunities for advancing the field. Data available for 1,018 extramural, peer-reviewed grants (active as of July 2012) supported by the Division of Cancer Control and Population Sciences (DCCPS), the NCI Division that supports cancer control and population-science extramural research grants, were analyzed. 455 of the grants were determined to involve biospecimens or derivatives. The most common specimen types included were whole blood (51% of grants), serum or plasma (40%), tissue (39%), and the biospecimen derivative, DNA (66%). While use of biospecimens in molecular epidemiology has become common, biospecimens for behavioral and social research is emerging, as observed in our analysis. Additionally, we found the majority of grants were using already existing biospecimens (63%). Grants that involved use of existing biospecimens resulted in lower costs (studies that used existing serum/plasma biospecimens were 4.2 times less expensive) and more publications per year (1.4 times) than grants collecting new biospecimens. This analysis serves as a first step at understanding the types of biospecimen collections supported by NCI DCCPS. There is room to encourage increased use of archived biospecimens and new collections of rarer specimen and cancer types, as well as for behavioral and social research. To facilitate these efforts, we are working to better catalogue our funded resources and make that data available to the extramural community.

Introduction

THE FIELD OF MOLECULAR EPIDEMIOLOGY has evolved considerably following significant advances in molecular techniques and analytical methods, such as high-throughput genomic and computational technologies and an increased willingness to share data.¹ As a result, researchers are increasingly using biospecimens to evaluate hypotheses related to disease risk, outcomes, and treatment. For the purposes of this article, we included biospecimens (e.g., tissue, blood, urine) and their common derivatives (e.g., DNA),² and refer to all as "biospecimens."

In addition to advances in molecular epidemiology, the number of cancer population-research publications involving biospecimens has more than doubled in the past 10 years.³ Advances in biospecimen collection and molecular techniques have enabled researchers to more readily involve

biospecimens in their research. Moreover, there are a number of large, population-science biobanks that link biospecimens with epidemiologic, clinical, genomic, and other data that can be utilized for many investigations.^{4–7} Many of these biobanks were established for particular diseases, but others have biospecimens to support general biomedical research.^{4,7}

The Division of Cancer Control & Population Sciences (DCCPS) at the National Cancer Institute (NCI) funds population-science research studies focused on reducing the risk, incidence, and deaths from cancer as well as enhancing the quality of life for cancer survivors. The grants cross many disciplines, such as genetic, epidemiological, behavioral, social, applied, and surveillance cancer research. Many of these NCI supported population-science studies involve biospecimens. Such studies involve a variety of molecular epidemiology approaches, including genome-wide association studies (GWAS) for cancer risk or outcomes, analysis of cotinine

Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Rockville, Maryland.

BIOSPECIMENS IN POPULATION-BASED RESEARCH

levels in saliva samples to verify smoking status, and tissue based analyses to examine molecular changes that occur in cancer subtypes.

Appreciating that biospecimens are becoming an integral part of population-science studies, we conducted an analysis of extramural grants funded by DCCPS; DCCPS holds the majority of population-science cancer grants funded by the NCI. The objectives of this analysis were to: 1) characterize types of biospecimen-related population-science cancer grants and 2) identify potential research gaps and opportunities to facilitate advancements in population-based cancer research. In addition, recognizing that there are many preexisting collections of biospecimens, we investigated efficiencies observed with studies involving use of existing biospecimens versus collecting new samples.

Materials and Methods

Identification of biospecimen research grants and associated data

In July 2012, the NIH IMPAC II grants database (contains information about all NIH extramural research projects) was queried using the NCI's Portfolio Management Application (PMA) to identify grants for inclusion. Search criteria included all active NCI DCCPS grants, excluding supplements⁸. It revealed a total of 1,018 active grants—awarded, currently receiving funding, or completing work-as of July 2012. The grants were categorized in a two-stage process, with quality control checks implemented at each stage, as either involving biospecimens or not (Fig. 1; Supplemental Document 1). (Supplementary material is available in the online article at www.liebertpub.com/bio.) In the first stage, titles, abstracts, and NCI funding program were reviewed for all grants. 242 grants were determined to not involve biospecimens. This left 776 grants to be reviewed in the second stage, where 320 of the 776 grants were determined to not involve biospecimens based on a full review of the grant application text.

The remaining 455 grants were classified as using existing (i.e., biospecimens were already obtained from participants), collecting new, or using existing and collecting new biospecimens. Specific biospecimen types, primary as well as derived, were recorded in the following categories (Supplemental Document 1): blood (whole blood), serum and/or plasma, buccal and/or saliva for DNA, cell lines, DNA, buffy coat or leukocytes, peripheral blood lymphocytes, solid tissue, tissue culture, urine, cervical biospecimen, RNA, saliva for biomarkers, blood spot, stool, blood clot, red blood cells, or other. (Supplementary material is available in the online article at www.liebertpub.com/bio.) The biospecimen categories were not mutually exclusive. For example, if a grant proposed collecting blood and processing the blood into plasma and DNA, the grant was categorized as involving blood, serum/plasma, and DNA. This was done to keep categories consistent for both studies collecting new and those using existing biospecimens, which may utilize already collected and processed blood products (e.g., plasma, blood clots), and to further elucidate the types of biospecimens being stored/analyzed.

Additional data about cancer type, direct costs, last competing year of funding (most recent year when the application underwent peer-review), and project start and end dates were obtained from IMPAC II. Target enrollment, or anticipated number of participants, was obtained from the NIH eRA Population Tracking module in the grants management system database. This additional data was supplemented by manual data abstraction when needed to ensure completeness.

Publications associated with the biospecimen research grants

The 455 biospecimen grants were linked to the internal NIH Scientific Publication Information Retrieval & Evaluation System (SPIRES) to evaluate the number of publications cited by authors as receiving support from the grants. This analysis only included grants from the year in

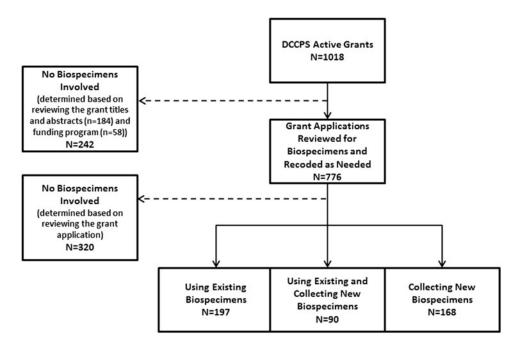


FIG. 1. Biospecimen grant review process. 1018 grants were included. Of these, 242 were determined to have no biospecimens involved based on scanning of the grant titles and funding program. The remaining 776 grants were reviewed for biospecimens. which the current funding cycle began. Manuscripts with a publication date at least one year after the last competing year for these associated grants were included. A total of 4,349 references were linked to 285 of the grants. Two variables were then derived: a) total number of PubMed IDs publications, per grant and b) length of time of the grant from the last competing year through to 2013 or the project end year, whichever came first. The average number of publications per year was calculated for each grant as follows: total number of publications/length of time.

Direct costs per participant

Budgets for grant applications include direct and indirect (or F&A) costs. Direct costs are costs specifically identified with a particular project, in contrast to indirect costs.⁹ Since the costs of a study can rise dramatically when the number of participants increases, direct costs of the studies were normalized by dividing by the targeted enrollment from the grant application.

Data Analyses

SAS Enterprise Guide (version 4.3) was used for all analyses. Fisher's Exact Test was used to compare differences in proportions of biospecimen types by collection status; Bonferroni corrected alpha level (0.0026) was used to account for the 19 independent comparisons. Wilcoxon nonparametric tests were used to compare average direct costs per participant (cost efficiencies) between grants by collection method, average publications per year (productivity/ time efficiencies) between grants by collection method, average direct costs per participant by activity code (P01 vs. R01, etc.; activity codes are used by NIH to differentiate research-related grant programs NIH supports,⁸ and average direct costs per participant for R01 grants by collection method (collecting new, using existing, or both). All reported *p*-values were two sided.

Results

Overview of biospecimen related grant applications in DCCPS

Of 1,018 grants reviewed, 455 (45%), in the DCCPS grant portfolio, involved biospecimens. The grants were characterized as collecting new (n=168, 37%), using existing (n=197, 43%), or using existing and collecting new (n=90, 20%). The 455 biospecimen grants were spread across funding mechanisms and DCCPS program areas, with the majority being R01s and in the Epidemiology and Genomics Research Program (EGRP) and Behavioral Research Program (BRP): 72% EGRP, 20% BRP, 4% Office of Cancer Survivorship (OCS), 3% Applied Research Program (ARP), and 1% Office of the Director (OD) (Supplemental Tables 1A and 1B) (Supplementary material is available in the online article at www.liebertpub.com/bio.).

The four most commonly used biospecimens were DNA (66%), blood (51%), serum/plasma (40%), and tissue (39%) (Fig. 2A). Other frequently studied biospecimens included buccal cells/saliva for DNA (23%), buffy coat/leukocyte (16%), saliva for biomarkers (13%), peripheral blood lymphoyctes (12%), and urine (14%). Among grants involving use of existing biospecimens, DNA was the most common

(72%), followed by tissue (45%), serum/plasma (29%), and blood (23%). Grants that involved a combination of using existing and collecting new biospecimens mainly involved blood (90%), followed by DNA (89%), tissue (81%), and serum/plasma (56%). Among grants involving new collection of biospecimens (Figure 2B), blood was the most common (63%), followed by DNA (47%), serum/plasma (45%), and saliva for biomarkers (29%). Studies that involved new collections of DNA (n=79) tended to also involve blood (n=20), buccal cells/saliva for DNA (n=12), blood and buccal cells/saliva for DNA (n=20), blood and buffy coat/leukocyte (n=11), or blood and tissue (n=6).

The most common cancer types in the 455 biospecimen studies were breast (29%), colorectal (13%), blood cancers (13%), prostate (10%), and lung (7%) (see Supplemental Table 2) (Supplementary material is available in the online article at www.liebertpub.com/bio.). DNA, blood, serum/ plasma, and tissue were the most common biospecimens in studies of those cancers. Additionally, smoking or substance abuse-related research with no specified cancer type (12%) was also heavily studied and predominately involved saliva.

Efficiencies of using existing versus collecting new biospecimens

The average direct cost per targeted enrollment count was \$1,102 per participant (95% CI: \$268 to 1936) for the 455 grants. Average direct costs per target enrollment count were not statistically significantly different by extramural research activity code – e.g. direct costs per target enrollment R01 vs. R21 vs. R03 grants (p=0.03).

Costs of studies can vary greatly depending on the type of biospecimen involved. For example, the average direct costs per target enrollment count for studies involving DNA were \$499/subject, compared with \$922/subject for serum/plasma studies. Therefore, we restricted the cost efficiency analyses to a single biospecimen type. As an example, grants using existing serum/plasma were 4.2 times less expensive (average direct costs per participant based on target enrollment = \$362;95% CI: \$0 to 1212) than studies collecting new serum/plasma (\$1508; 95% CI: \$926 to 2196). Grants using a combination of existing serum/plasma and new were 2.3 times less expensive (\$669; 95% CI: \$470 to 585) than studies collecting new serum/plasma (p < 0.0001). This cost savings for using existing biospecimens versus collecting new was also observed when we accounted for activity code. Analysis of the costs per participant in R01 grants, which accounted for the majority of biospecimen grants (68%), revealed a statistically significant difference (p < 0.0001) between R01 grants collecting new (mean = \$926/participant), using a combination (mean = \$647/participant)participant), and using existing (mean = $\frac{262}{participant}$).

In addition to realizing cost savings, it is foreseeable that using existing biospecimens permits investigators to more quickly obtain research results. Overall, the average number of publications per year was 2.0 (95% CI: 1.7–2.4) for grants involving biospecimens. This number reflects the relatively large number of publications resulting from grants involving both the use of existing and collection of new biospecimens, which tended to be larger, comprehensive grants (such as specialized center grants); average number of publications per year were 2.9 (95% CI: 2.0–3.9). The average number of publications for each grant per year were 1.4 times higher for studies only using existing (2.1 publications per year;

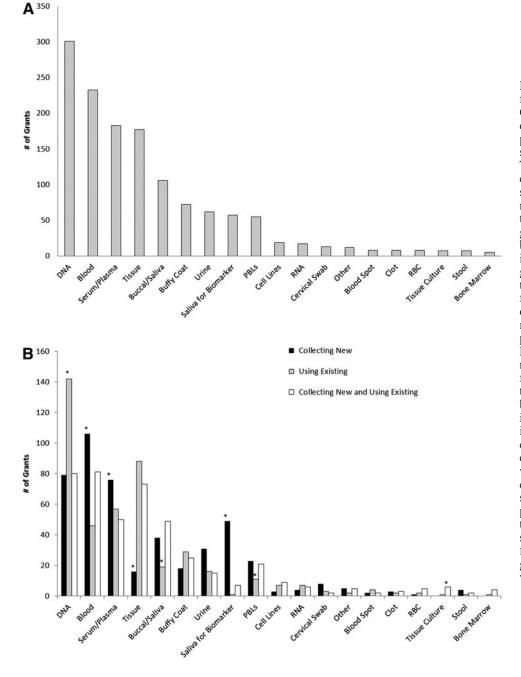


FIG. 2. A. Specimen Types for the 455 Biospecimen Grants. 455 grants were determined to involve biospecimens. The types of biospecimens were recorded. These were not mutually exclusive categories, and as such, a single grant may be represented multiple times in this graph. For example, a grant may propose collecting blood, but then processing it into DNA and serum. That grant would be depicted on this graph three times (once for blood, once for DNA, and once for serum). B. Specimen Types in the 455 Biospecimen Grants by Using Existing/Collecting New Variable. The 455 grants were further subdivided by whether they collected new biospecimens only, used existing biospecimens only, or involved a combination of collecting new and using existing biospecimens. These were not mutually exclusive categories, and as such, a single grant may be represented multiple times in this graph. Asterisks indicate statistically significant difference in the proportion of grants by collection status within a biospecimen type.

95% CI: 1.6–2.6) compared with those only collecting new biospecimens (1.5 publications per year; 95% CI: 0.8–2.2), and 1.9 times higher for studies using a combination of using existing and collecting new (p = 0.0005).

Discussion

We conducted a cross-sectional analysis of the NCI's DCCPS extramural research grant portfolio for biospecimens involved in research projects. While the use of biospecimens in molecular epidemiology studies is common in epidemiology and genomics research, use of biospecimens for behavioral and social research is emerging;^{10–12} this is supported by the observation that only 24% of the behavioral grants (in BRP) involved biospecimens whereas 82% of epidemiologic or genetic grants (in EGRP) involved biospecimens.

DNA was the most common biospecimen included across studies. DNA is relatively easy to obtain from either blood/ blood products and buccal cells/saliva, it does not require stringent storage considerations, and genomic analyses are generally amenable to high-throughput analyses; all of which are important for population-science studies since they typically involve large numbers of participants. Furthermore, significantly more studies used existing DNA compared with those that collected new biospecimens, Since DNA is easily stored, and many analyses require only small amounts of DNA, many researchers were able to leverage pre-existing DNA collections instead of collecting new DNA specimens.

Blood was the second most common biospecimen involved. For the majority of studies, blood was processed into multiple components such as plasma, serum, and buffy coat, and DNA was derived. Studies involving blood or its components tended to involve genetic, protein biomarker, telomere length, vitamin D, lipid, fatty acid and/or immune marker analyses. While most of the blood components were required for those studies, several studies stored one or more of the blood components in a repository for future research use.

It was promising to see the large number of grants collecting tumor tissue (39%), given recent findings from The Cancer Genome Atlas (TCGA) that highlight the importance of examining tissue samples for gaining insights about cancer and subtypes.¹³ Significantly more studies utilized existing tissue compared with collection of new tissue, which is likely a reflection of studies utilizing existing tissue from clinical care. Existing tissue from population-based settings are typically formalin-fixed paraffin embedded (FFPE). In contrast, approximately 50% of studies collecting new tissue were using fresh frozen tissue, which is more laborious to obtain since it is not the standard preservation procedure for many pathology labs. Improved molecular techniques are now permitting expanded uses of FFPE tissue for many molecular assays,^{14–16} which will likely open more opportunities for population-based research using existing tissue resources.

Additionally, saliva for biomarkers, blood, and serum/ plasma tended to be collected new; this is possibly due to instability of biomarkers in saliva during storage, and the need to control for pre-analytic variables in serum/plasma and blood studies. There is a recognized need for standardization of blood collection and processing procedures in order to ensure reliable molecular data is obtained.^{2,17,18}

Proportions of cancer sites studied in the biospecimen grants are similar to that overall at NCI.¹⁹ The more common cancers, such as breast, colorectal, prostate, and lung cancers, as well as smoking/substance abuse studies were represented by a large number of grants, while rarer cancers were not studied in a large number of grants. This could be due to the difficulty in amassing sufficient numbers of rare cancer cases, which may be alleviated in part by more openly sharing data and biospecimens among researchers, such as what is being done with cancer consortia.²⁰

Overall, we found that a large number of DCCPS-funded epidemiology studies leveraged established biorepositories. This is consistent with NCI's current emphasis on leveraging existing resources²¹ and is fundamental for facilitating advancements in population-based cancer research. There were observed productivity (i.e., publications) and cost advantages to using existing versus collecting new biospecimens. Some limitations of the cost analysis were: 1) accuracy of the targeted-enrollment data was dependent upon that stated in the grant application and 2) the study design (case-control, cohort, etc.) may affect costs, but this data was not readily available for all grants in a form suitable for analysis. Some limitations for the productivity analysis were: 1) accuracy of the number of publications was dependent upon authors appropriately referencing the correct grant number in publications, 2) publications occurring within the first year of the grant were excluded, and 3) publications that were not in PubMed would have been missed. Despite these limitations, our results are promising, support the known advantages to using existing resources, and demonstrate that many of NCI's grantees are leveraging existing resources. Moreover, these results suggest that a one-time investment could lead to further scientific impact downstream.

While using existing biospecimens is encouraged to facilitate research in more cost and time-efficient manners, as well as to impart a lower burden on participants, we would be remiss not to discuss some caveats. Policies and procedures for evaluating how to balance the preservation of the resource while making biospecimens available to the research community are critical. In addition, researchers must carefully decide whether existing biospecimens are appropriate for the design of the desired study, i.e., are the banked samples "fit for the purpose" of their study. For example, when The Cancer Genome Atlas (TCGA) began, they found only about 30% of existing samples in biobanks were appropriate for TCGA analyses.²² Researchers must consider factors such as study design from where biospecimens already exist and inclusion criteria (including case/control definitions, if pertinent). Researchers must also realize that pre-analytic variables (such as time between blood draw and processing) can greatly affect molecular assay results.²³⁻²⁵ Validation studies should be done to ensure the stored samples are suitable for producing reliable results for a given assay and analyte.⁵ Just as investigators are being encouraged to strengthen the reporting of epidemiology studies (STROBE),²⁶ investigators should be encouraged to standardize and strengthen the reporting of biospecimen collection, processing, storage, and analysis in their molecular epidemiology studies.^{21–24,27}

This analysis serves as a first step at understanding our inventory of DCCPS supported biospecimen collections. However, it also suggests an opportunity to help shape the future of biospecimen research and improve how NCI collects information about biospecimens being used in populationscience studies to foster even greater sharing. Several potential research gaps were identified including the need to expand biospecimen research further into behavioral research studies, record grant-related biospecimen information (biospecimen and study types) more accurately, and to increase collections of some of the rarer specimen types (e.g., stool samples for microbiome projects) and cancer types (e.g., liver cancer). In order to further promote collaborations and sharing biospecimens, we need new and continued funding opportunities to encourage collaborations. We also need investigators to understand the benefits of using existing resources can outweigh the efforts required to design studies that utilize existing biospecimens.

NCI has begun assessing several options for how best to promote collaborations and sharing of biospecimens.^{28,29} NCI DCCPS has developed a biospecimen resources website² and is working on a web-based tool investigators can use to identify potential existing resources. As the scientific community is becoming keenly aware of the importance of sharing data and biospecimens to maximize resources, and biospecimens are being routinely included in population-science research studies, it is crucial to make high-quality data and biospecimens more accessible in responsible ways.

Author Disclosure Statement

The authors declare no conflict of interest.

References

 Khoury MJ, Gwinn M, Clyne M, et al. Genetic epidemiology with a capital E, ten years after. Genet Epidemiol. 2011;35:845–852.

- Office of Biorepositories and Biospecimen Research, National Cancer Institute, NIH. NCI Best Practices for Biospecimen Resources, 2011.
- 3. PubMed, Search conducted using the terms: (biospecimens OR blood OR serum OR plasma OR tissue OR urine OR saliva OR DNA OR buccal) AND (epidemiology) AND cancer, May 22, 2013.
- 4. Henderson GE, Cadigan RJ, Edwards TP, et al. Characterizing biobank organizations in the U.S.: Results from a national survey. Genome Med. 2013;5:3.
- 5. Elliott P, Peakman TC, Biobank UK. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. Int J Epidemiol. 2008;37:234–244.
- NCI, Biospecimen Resources for Population Scientists, July 2013.
- 7. Collins R. What makes UK Biobank special? Lancet. 2012;379:1173–1174.
- 8. NIH, NIH Office of Extramural Research, Grants & Funding: Types of Grant Programs, 2014.
- 9. NIH, NIH Office of Extramural Research, Grants & Funding: Developing Your Budget, 2014.
- Ballard-Barbash R, Friedenreich CM, Courneya KS, et al. Physical activity, biomarkers, and disease outcomes in cancer survivors: A systematic review. J Natl Cancer Inst. 2012;104:815–840.
- Granger DA, Fortunato CK, Beltzer EK, et al. Focus on methodology: salivary bioscience and research on adolescence: An integrated perspective. J Adolesc. 2012;35:1081– 1095.
- Piazza JR, Almeida DM, Dmitrieva NO, et al. Frontiers in the use of biomarkers of health in research on stress and aging. J Gerontol B Psychol Sci Soc Sci. 2010;65:513–525.
- 13. TCGA Research Briefs.
- 14. Wagle N, Berger MF, Davis MJ, et al. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. Cancer Discov. 2012;2:82–93.
- 15. Adams MD, Veigl ML, Wang Z, et al. Global mutational profiling of formalin-fixed human colon cancers from a pathology archive. Mod Pathol. 2012;25:1599–1608.
- Weng L, Wu X, Gao H, et al. MicroRNA profiling of clear cell renal cell carcinoma by whole-genome small RNA deep sequencing of paired frozen and formalin-fixed, paraffinembedded tissue specimens. J Pathol. 2010;222:41–51.
- Vaught JB, Henderson MK. Biological sample collection, processing, storage and information management. IARC Sci Publ. 2011;23–42.

- ISBER. 2012 Best Practices for Repositories: Collection, Storage, Retrieval, and Distribution of Biological Materials for Research. Biopreserv Biobank. 2012;10:79–161.
- 19. NCI, FY 2012 Research Funding by Cancer Type.
- Burgio MR, Jr., Ioannidis JP, Kaminski BM, et al. Collaborative Cancer Epidemiology in the 21st Century: The Model of Cancer Consortia. Cancer Epidemiol Biomarkers Prev. 2013.
- Green T, Caga-anan C, Hutter C, Share and Share Alike: How You Too Could Benefit from NCI Data Sharing Resources, Cancer Epidemiology Matters Blog (http:// blog-epi.grants.cancer.gov/2013/03/20/nci-data-sharingresources/), 2013.
- 22. Barker A. Q&A: Anna Barker on the cancer genome atlas. Cancer Discov. 2011;1:193.
- Vaught JB. Blood collection, shipment, processing, and storage. Cancer Epidemiol Biomarkers Prev. 2006;15:1582– 1584.
- 24. Lippi G, Becan-McBride K, Behulova D, et al. Preanalytical quality improvement: In quality we trust. Clin Chem Lab Med. 2013;51:229–241.
- Tworoger SS, Hankinson SE. Collection, processing, and storage of biological samples in epidemiologic studies: Sex hormones, carotenoids, inflammatory markers, and proteomics as examples. Cancer Epidemiol Biomarkers Prev. 2006;15:1578–1581.
- 26. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. PLoS Med. 2007;4:e297.
- 27. Moore HM, Kelly A, Jewell SD, et al. Biospecimen reporting for improved study quality. Biopreserv Biobank. 2011;9:57–70.
- Massett HA, Atkinson NL, Weber D, et al. Assessing the need for a standardized cancer HUman Biobank (caHUB): Findings from a national survey with cancer researchers. J Natl Cancer Inst Monogr. 2011;2011:8–15.
- 29. Vaught J, Rogers J, Myers K, et al. An NCI perspective on creating sustainable biospecimen resources. J Natl Cancer Inst Monogr. 2011;2011:1–7.

Address correspondence to: Danielle M. Carrick, PhD, MHS EGRP, DCCPS, NCI, NIH 9609 Medical Center Drive, Room 4E244 Rockville, MD 20850

E-mail: Danielle.Carrick@nih.gov