NEIGHBORHOOD CANCERIZATION: NEW APPROACHES LINKING SOCIAL AND BIOLOGICAL MECHANISMS OF CANCER

Shannon M. Lynch

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

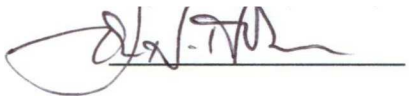
2015

Supervisor of Dissertation

Timothy R. Rebbeck

Professor of Epidemiology

Graduate Group Chairperson



John Holmes, Associate Professor of Medical Informatics and Epidemiology

Dissertation Committee

Timothy Rebbeck, Professor-Epidemiology Nandita Mitra, Associate Professor-Biostatistics

Charles Branas, Professor-Epidemiology Harold Riethman, Associate Professor-Oncogenesis

NEIGHBORHOOD CANCERIZATION: NEW APPROACHES LINKING SOCIAL AND BIOLOGICAL MECHANISMS OF CANCER

Dedication page

My dissertation is dedicated to the memory of my Uncle and my first scientific mentor, Steven Ranjo. Thank you for fostering my love of science. All those museum visits and late night, award winning science fair projects paid off. I am forever grateful to you.

ACKNOWLEDGMENT

I would like to thank my dissertation committee, Tim Rebbeck, Nandita Mitra, Charlie Branas and Harold Riethman. I would also like to thank the following individuals for their contribution to my education while attending the University of Pennsylvania: Carrie Stricker, Peter Kanetsky, David Sarwer, Linda Jacobs, Karen Glanz, Katie Schmitz, Andrea Branas, Michelle Ross, Charnita Zeigler-Johnson, Jonathan Mitchell, Justin Brown, Rebecca Speck. I would especially like to thank my family for their love, dedication, and support during my studies. Special thanks to my husband, John Wittenzellner, parents, Kevin and Terry Lynch, Sisters, Kim McFarland and Steph Lynch, Aunt, Christine Ranjo, Mother-in-law, Susan Wittenzellner, and my children, Ada and Gavin Wittenzellner.

ABSTRACT

NEIGHBORHOOD CANCERIZATION: NEW APPROACHES LINKING SOCIAL AND BIOLOGICAL MECHANISMS OF CANCER

Shannon M. Lynch

Timothy Rebbeck

Novel multidisciplinary and multilevel approaches are required to link biologic and social mechanisms with cancer. We propose a new biosocial concept, “neighborhood cancerization,” which postulates that residents of the same geographically-defined regions can be exposed to common unfavorable circumstances. These common neighborhood-level exposures can in turn have biological consequences that may result in an increased risk of cancer. Just as common “molecular signatures” can differentiate tumor types, “neighborhood signatures” can identify neighborhoods at increased risk for cancers of similar etiologic origins. Under a shared chronic stress hypothesis, we test the neighborhood cancerization theory by first determining the effect of neighborhood circumstances on telomere length (TL), a cellular marker of oxidative stress often implicated in cancer development at the population level. After addressing common methodologic concerns often cited in TL studies, we tested neighborhood and TL associations in a multi-racial, multi-center setting and in the context of individual-level stressors using quantile regression. We then developed and conducted a neighborhood-wide association study (NWAS) using all available U.S Census variables and the Pennsylvania State Cancer Registry in order to empirically identify common neighborhood factors related to prostate cancer. Our novel NWAS approach demonstrates how agnostic, high-dimensional data analyses can be used to identify neighborhoods and people at risk for high grade/high stage, aggressive prostate cancer. Our work has implications for health disparities research, and provides evidence to support the neighborhood cancerization hypothesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

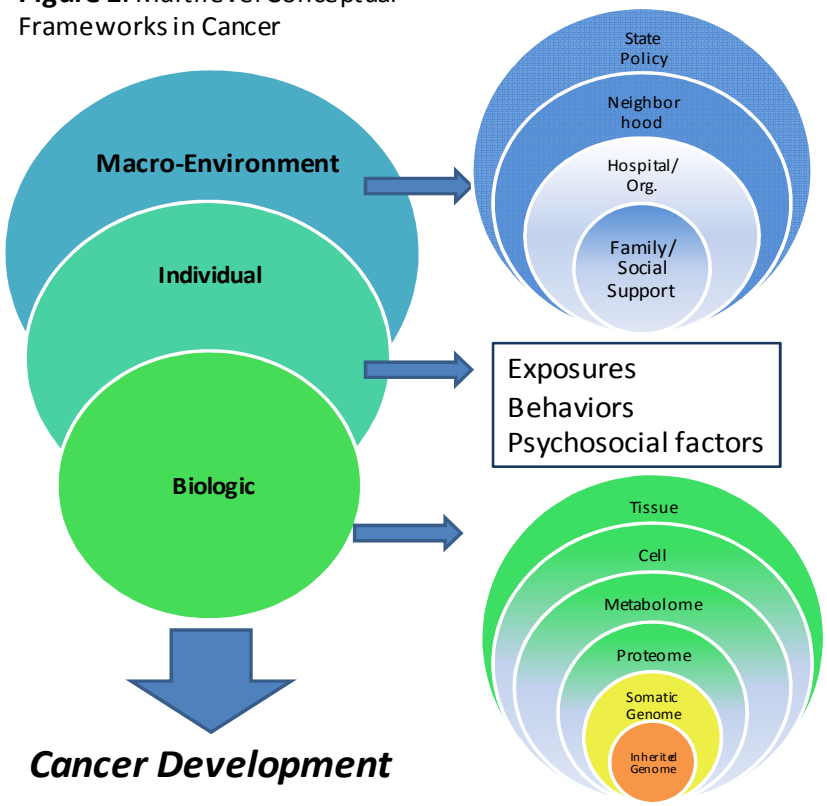
CHAPTER 1

SOCIO-BIOLOGIC CONCEPTS IN CANCER

The Complex, Multilevel Etiology of Cancer

Cancer is a disease of abnormal and dysregulated cell growth[1]. A combination of genetic and environmental factors (e.g. cigarette smoking) can initiate cancer development[2-4], suggesting that cancer is etiologically complex. Thus, cancer research has evolved from focusing on single risk factors to studies of complex interactions between social, behavioral, molecular, and environmental risk factors.

Figure 1. Multilevel Conceptual Frameworks in Cancer



Multilevel conceptual frameworks have been developed to address the complex nature of cancer etiology and allow for the simultaneous assessment of the role of two or more etiological agents within a hierarchical or nested structure [5, 6], [7],

[8],[3, 4, 9-14](Appendix A). Multilevel models are generally characterized by 1) The macro-environment (or “eco-level” [3, 4]); 2) the individual; and 3) biology (Figure 1). Each of these levels is further characterized by sub-levels that define domains of variables involved in cancer etiology or outcomes (Appendix A). This framework supports two main hypotheses: 1) factors at the macro-environment and individual levels can directly affect biological events and result in cancer; 2) factors may interact in a hierarchal fashion, such that biologic-level effects are affected by behaviors or exposures of the individual, and individual level effects are affected by the macro-environment [15]. The ability to address these hypotheses is complex at each level of analysis, and even more complex when crossing levels.

Methodologic challenges in the Multilevel Framework

The biologic level is described here as single biomarkers at the cellular sub-level that have been implicated in cancer development at the population level. Some examples include telomere length and other oxidative stress markers, like cortisol levels[16]. Regardless of biomarkers used for study, inconsistent results (i.e. statistical effects vs no association) are often noted across biomarker studies[17]. Inconsistencies in biomarker research might be due in part to differences among study populations(e.g., age, gender, race, etc.), laboratory approaches(e.g., how blood and DNA is collected, extracted or stored)[18], or statistical methods used [17]. These issues are of particular concern in multicenter research settings where data collection and population demographics often differ, but analyses are combined as if center data comes from a single study. While a number of quality control checks are available to assess validity and reliability of biomarkers prior to associations, there is little consistency in analysis and reporting across study samples[16]. Standardized assessments of biomarker data are lacking, but

would likely improve the acceptability of biomarker findings and lead to a more widespread use of biomarkers in the multilevel context.

Macro-environment is described here in terms of the neighborhood sub-level. Neighborhoods can be defined as the area or place in which a person lives[5], although technical definitions of exact neighborhood polygon boundaries are numerous. Neighborhoods have certain characteristics such as degree of deterioration, urbanization, poverty, educational attainment and percentage of low-income residents that have been correlated with increasing disease rates, risk behaviors, and unfavorable health outcomes [19-23]. Neighborhood effects have been implicated in cancer development[24, 25], even after controlling for individual-level factors[26].

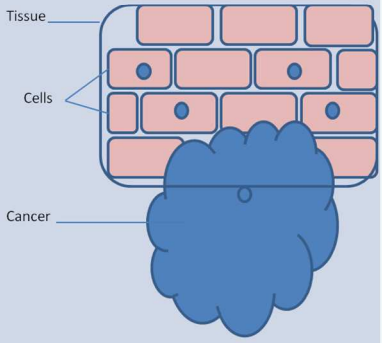
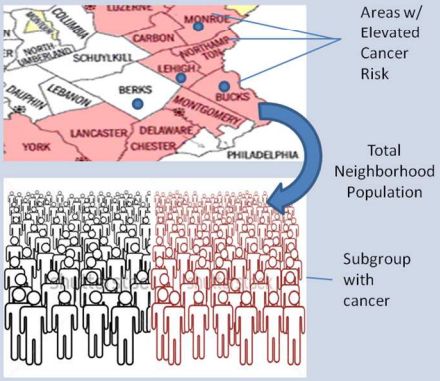
Current methodologic approaches in neighborhood research call for *a priori* selection of neighborhood variables from publically available U.S. Census data or self-administered questionnaires[5, 27]. The problem with this approach is that findings are not easily replicated due to a lack of consistency in defining neighborhood variables and geographic boundaries [27-29]. This has limited the causal inferences that can be made about neighborhood and disease research[27-29] and has likely contributed to fewer studies with a multilevel focus, particularly investigations centered on neighborhood and biology (Appendix B). Animal studies have demonstrated that unfavorable neighborhood exposures, including social isolation from peers, can influence biological parameters related to apoptosis and inflammation[30]. However, few human observational studies have investigated the effects of unfavorable neighborhood characteristics on biological markers related to disease[31]. For example, one multilevel study investigated etiologic effects of all three levels on a cancer outcome[32]; 5 studies investigated macro-environmental effects (namely neighborhood effects) on biomarkers implicated in cancer

at the population level[31, 33-35](Appendix B). Thus, studies linking biologic and social mechanisms in cancer are limited, despite wide-spread acceptance of these multilevel frameworks.

Neighborhood Cancerization: Principles

Current multilevel approaches lack shared pathways and methodologies that can be used to frame the relationship between biologic and macro-environmental factors[5]. Thus, the creation of joint socio-biologic concepts and novel methodologies could open doors for more multidisciplinary cancer investigations. The concept of field cancerization is a pathobiologic theory that was originally proposed to explain synchronous or metachronous tumors of the oral

Figure 2. Socio-Biologic concepts

	Field Cancerization	Neighborhood Cancerization
Levels of investigation	Exposure→Tissue→Cell→ Cancer	Neighborhood→Biologic Effects →Subgroups with cancer
Marker of Cancer Predisposition	Common Cellular Alterations across cells with and without cancer	Common Neighborhood Circumstances (areas with and at risk for cancer)
Model		

mucosa[36]. Field cancerization came to refer to the propensity of a field of tissue to become malignant based on either pathological observations or molecular markers[37]. In theory, field cancerization occurs from simultaneous wide-spread, unfavorable

exposures acting on cells and tissues within organs and organ systems (such as the effect of cigarette smoking on the lung), as well as the migration of “patches” of molecularly altered cells into larger tissue fields or territories that become predisposed to cancer development[37]. Common molecular signatures in pre-malignant cells and primary, malignant tumors provide evidence of the cancer field effect since they can be used to identify tumor types of similar etiologic origins(Figure 2). Molecular markers for field cancerization have been related to epigenetic changes[38-40], tissue micro-environment[41, 42], and telomere function[43] across different tissue types(Figure 2).

Analogous to a field of tissue for which unfavorable exposures may lead to carcinogenesis, residents of neighborhoods may experience common stressors that can lead to unfavorable biological consequences including heightened cancer risks(Figure 2). Under the proposed “neighborhood cancerization” concept, we hypothesize that residents of geographically-defined regions can be exposed to common unfavorable circumstances. These common neighborhood-level exposures can in turn have biological consequences that may result in an increased risk of cancer. Further, just as common “molecular signatures” can differentiate tumor types, “neighborhood signatures” can identify neighborhoods at increased risk for cancers of similar etiologic origins[44]. Thus, neighborhood cancerization concepts at the social level are parallel to field cancerization concepts at the molecular level (Figure 2).

Further, “top-down and bottom-up” mechanisms linking neighborhood and field cancerization processes under similar biologic pathways are possible, but likely complex, involving multilevel approaches, as well as complicated systems approaches[45] (i.e, approaches that incorporate positive and negative feedback loops among multilevel variables, which is outside the scope of this dissertation). Here, we

operate under the hypothesis that unfavorable social and economic environments at the neighborhood-level may act in concert with biological, behavioral or psychosocial factors(i.e., depression[46] and perceived stress[47]) at the individual level [5] to cause chronic stress[48, 49] . In turn, biological processes related to oxidative stress are affected by neighborhood and individual-level exposures in a way that limits the body's ability to fight disease processes like cancer. Thus, chronic stress is related to constant "wearing down" of the body that can affect biological processes, accelerate the rate of decline in physiological functioning, and ultimately affect the body's ability to fight disease processes such as cancer [17, 50-54](Figure 2).

Dissertation Synopsis

To address the issues described above, this investigation focuses on linking biologic mechanisms and social mechanisms (at the macro-level) in cancer by proposing a novel socio-biologic neighborhood cancerization framework. This concept will be tested under a biologic pathway common to both social and biologic sciences, chronic stress. The chronic stress model postulates that constant exposure to unfavorable stressors at the neighborhood level can lead to cellular oxidative stress at the biologic level and ultimately cancer initiation and progression[5] [17, 50-54]. These hypotheses can be tested by applying methodologic approaches from biology to social science and from social science to biology. Shared socio-biologic concepts and methods can serve as a common "language" across disciplines, which can encourage the facilitation of the multicenter, multilevel investigations that are needed to address the complex, multifactorial nature of cancer among individuals who will always be nested within a variety of ambient neighborhood landscapes and risk environments. In directly changing

these environments we may be able to profoundly influence the burden of cancer, perhaps more so than medical care or direct attempts to modify lifestyle[55].

Each chapter of this dissertation addresses a methodologic challenge associated with testing the neighborhood cancerization hypothesis. Chapter 2 focuses on common methodologic concerns often cited in multicenter, biomarker studies. In Chapter 3, we determine the effect of neighborhood circumstances on telomere length (TL), a biological marker of oxidative stress often implicated in cancer development at the population level. We test neighborhood and TL associations in a multi-racial, multi-center setting and in the context of individual-level stressors by applying a social science method, quantile regression, to a biologic outcome. Borrowing methodologies from genome-wide association studies(GWAS), in Chapter 4, we develop and conduct a neighborhood-wide association study (NWAS) using all available U.S Census variables and the Pennsylvania State Cancer Registry in order to empirically identify common neighborhood factors related to prostate cancer. We demonstrate that NWAS can be used to identify common neighborhood signatures that relate to high grade/high stage prostate cancer. The socio-biologic concepts and methods discussed in each chapter have the potential to serve as a common language across disciplines. These concepts and methods can encourage the facilitation of multicenter, multilevel investigations that are needed to address the complex, multifactorial nature of cancer.

References:

1. *What is Cancer*. 2015; Available from: http://www.cancer.gov/cancertopics/what-is-cancer.
2. Morrissey, J.P., et al., *Computational Modeling and Multilevel Cancer Control Interventions.* JNCI Monographs, 2012. **2012**(44): p. 56-66.
3. Gorin, S.S., et al., *Multilevel Interventions and Racial/Ethnic Health Disparities.* JNCI Monographs, 2012. **2012**(44): p. 100-111.

4. Taplin, S.H., et al., *Introduction: Understanding and Influencing Multilevel Factors Across the Cancer Care Continuum.* JNCI Monographs, 2012. **2012**(44): p. 2-10.
5. Lynch, S.M. and T.R. Rebbeck, *Bridging the Gap between Biologic, Individual, and Macroenvironmental Factors in Cancer: A Multilevel Approach.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(4): p. 485-495.
6. Plewis, I., *Statistics in Education*1997, London: Edward Arnold.
7. MacMahon, B. and T. Pugh, *Epidemiology: Principles and Methods.*1970: MacMahon B, Pugh TF. Epidemiology: Principles and Methods. Boston: Little, Brown and Company, 1970 .
8. Judith S. Mausner, J., S. Kamer, and A. Bahn, *Epidemiology: An Introductory Text*1985: Saunders.
9. Galea, S., M. Riddle, and G.A. Kaplan, *Causal thinking and complex system approaches in epidemiology.* Int J Epidemiol, 2010. **39**(1): p. 97-106.
10. Warnecke, R.B., et al., *Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities.* Am J Public Health, 2008. **98**(9): p. 1608-15.
11. Glass, T.A. and M.J. McAtee, *Behavioral science at the crossroads in public health: extending horizons, envisioning the future.* Soc Sci Med, 2006. **62**(7): p. 1650-71.
12. Caporaso, N.E., *Integrative study designs--next step in the evolution of molecular epidemiology?* Cancer Epidemiol Biomarkers Prev, 2007. **16**(3): p. 365-6.
13. Spitz, M.R., N.E. Caporaso, and T.A. Sellers, *Integrative cancer epidemiology--the next generation.* Cancer Discov, 2012. **2**(12): p. 1087-90.
14. Hiatt, R.A. and N. Breen, *The Social Determinants of Cancer: A Challenge for Transdisciplinary Science.* American journal of preventive medicine, 2008. **35**(2): p. S141-S150.
15. Rappaport, S.M., *Implications of the exposome for exposure science.* J Expos Sci Environ Epidemiol, 2011. **21**(1): p. 5-9.
16. Savage, S.A., S.M. Gadalla, and S.J. Chanock, *The Long and Short of Telomeres and Cancer Association Studies.* Journal of the National Cancer Institute, 2013.
17. Sanders, J.L. and A.B. Newman, *Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither?* Epidemiologic Reviews, 2013. **35**(1): p. 112-131.
18. Hofmann, J., Hutchinson, A.A., Cawthon, R., Liu, C.S., Lynch, S.M., et al. , *Telomere Length Varies By DNA Extraction Method: Implications for Epidemiologic Research–Letter.* Unpublished, 2014.
19. Cohen, D.A., et al., *Neighborhood physical conditions and health.* Am J Public Health, 2003. **93**(3): p. 467-71.
20. Wardle, J., et al., *Socioeconomic disparities in cancer-risk behaviors in adolescence: baseline results from the Health and Behaviour in Teenagers Study (HABITS).* Prev Med, 2003. **36**(6): p. 721-30.
21. Borrell, L.N., et al., *Perception of general and oral health in White and African American adults: assessing the effect of neighborhood socioeconomic conditions.* Community Dent Oral Epidemiol, 2004. **32**(5): p. 363-73.
22. Zierler, S., et al., *Economic deprivation and AIDS incidence in Massachusetts.* Am J Public Health, 2000. **90**(7): p. 1064-73.

23. Winkleby, M., K. Sundquist, and C. Cubbin, *Inequities in CHD incidence and case fatality by neighborhood deprivation.* Am J Prev Med, 2007. **32**(2): p. 97-106.
24. Adler, N.E., Ostrove, J.M. , *Socioeconomic status and health: What we know and what we don't.* Ann N Y Acad Sci, 1999. **896**: p. 3-15.
25. Diez Roux, A.V., Mair, C., *Neighborhoods and health.* Ann N Y Acad Sci, 2010. **1186**: p. 125-145.
26. Merkin, S.S., et al., *Neighborhoods and Cumulative Biological Risk Profiles by Race/Ethnicity in a National Sample of U.S. Adults: NHANES III.* Annals of epidemiology, 2009. **19**(3): p. 194-201.
27. Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., *Assessing "Neighborhood Effects": Social Processes and New Directions in Research.* Ann Rev Sociol, 2002. **28**: p. 443-478.
28. Messer, L., . Laraia, B., Kaufman, J., Eyster, J., Holzman, C., Culhane, J., et al. , *The development of a standard neighborhood deprivation index. .* Journal of Urban Health, 2006. **83**(6): p. 1041-1062.
29. Diez Roux, A.V. and C. Mair, *Neighborhoods and health.* Annals of the New York Academy of Sciences, 2010. **1186**(1): p. 125-145.
30. McClintock, M.K., et al., *Mammary cancer and social interactions: identifying multiple environments that regulate gene expression throughout the life span.* J Gerontol B Psychol Sci Soc Sci, 2005. **60 Spec No 1**: p. 32-41.
31. Needham, B.L., et al., *Neighborhood characteristics and leukocyte telomere length: The Multi-Ethnic Study of Atherosclerosis.* Health & Place, 2014. **28**(0): p. 167-172.
32. Rebbeck, T.R., et al., *Context-Dependent Effects of Genome-Wide Association Study Genotypes and Macroenvironment on Time to Biochemical (Prostate Specific Antigen) Failure after Prostatectomy.* Cancer Epidemiology Biomarkers & Prevention, 2010. **19**(9): p. 2115-2123.
33. Barrington, W.E., et al., *Neighborhood socioeconomic deprivation, perceived neighborhood factors, and cortisol responses to induced stress among healthy adults.* Health & Place, 2014. **27**(0): p. 120-126.
34. Epplein, M., et al., *Neighborhood socio-economic characteristics, African ancestry, and Helicobacter pylori sero-prevalence.* Cancer Causes & Control, 2012. **23**(6): p. 897-906.
35. Waggaman, C., P. Julian, and L.M. Niccolai, *Interactive effects of individual and neighborhood race and ethnicity on rates of high-grade cervical lesions.* Cancer Epidemiology. **38**(3): p. 248-252.
36. SLAUGHTER, D.P., H.W. SOUTHWICK, and W. SMEJKAL, *Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin.* Cancer, 1953. **6**(5): p. 963-8.
37. Lochhead, P., et al., *Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression.* Mod Pathol, 2015. **28**(1): p. 14-29.
38. Giovannucci, E. and S. Ogino, *DNA methylation, field effects, and colorectal cancer.* J Natl Cancer Inst, 2005. **97**(18): p. 1317-9.
39. Ushijima, T., *Epigenetic field for cancerization.* J Biochem Mol Biol, 2007. **40**(2): p. 142-50.
40. Ramachandran, K. and R. Singal, *DNA methylation and field cancerization.* Epigenomics, 2012. **4**(3): p. 243-5.

41. Stearman, R.S., et al., *A macrophage gene expression signature defines a field effect in the lung tumor microenvironment.* Cancer Res, 2008. **68**(1): p. 34-43.
42. Martinez-Outschoorn, U.E., et al., *Oxidative stress in cancer associated fibroblasts drives tumor-stroma co-evolution: A new paradigm for understanding tumor metabolism, the field effect and genomic instability in cancer cells.* Cell Cycle, 2010. **9**(16): p. 3256-76.
43. Heaphy, C.M., et al., *Telomere DNA content and allelic imbalance demonstrate field cancerization in histologically normal tissue adjacent to breast tumors.* Int J Cancer, 2006. **119**(1): p. 108-16.
44. Lynch, S.M., Mitra, N., Ravichandran, K., Spangler, E., Zhou, W., Paskett, E., Gehlert, S., et al., *Telomere Length and Neighborhood Cancerization: Evaluating Biological Response to Unfavorable Exposures.* Submitted for Peer Review, 2015.
45. Diez Roux, A.V., *Complex Systems Thinking and Current Impasses in Health Disparities Research.* Framing Health Matters, 2011. **101**: p. 1627-1635.
46. Wolkowitz, O.M., Mellon, S.H., Epel, E.S., Lin, J., Dhabhar, F.S., Su, Y, et al., *Leukocyte Telomere Length in Major Depression: Correlations with Chronicity, Inflammation and Oxidative Stress - Preliminary Findings.* PLoS ONE, 2011. **6**(3): p. e17837.
47. Parks, C.G., et al., *Telomere Length, Current Perceived Stress, and Urinary Stress Hormones in Women.* Cancer Epidemiology Biomarkers & Prevention, 2009. **18**(2): p. 551-560.
48. Geronimus, A.T., Hicken, M., Keene, D., Bound, J., *"Weathering" and Age Patterns of Allostatic Load Scores Among Blacks and Whites in the United States.* Am J Public Health, 2006. **96**: p. 826-833.
49. McEwen, B., *Protective and Damaging effects of stress Mediators.* N Engl J Med, 1998. **338**: p. 171-179.
50. Londoño-Vallejo, J.A., *Telomere instability and cancer.* Biochimie, 2008. **90**(1): p. 73-82.
51. Koorstra, J.B.M., et al., *Pancreatic Carcinogenesis.* Pancreatology, 2008. **8**(2): p. 110-125.
52. Cunningham, J.M., et al., *Telomere Length Varies By DNA Extraction Method: Implications for Epidemiologic Research.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(11): p. 2047-2054.
53. Wentzensen, I.M., et al., *The Association of Telomere Length and Cancer: a Meta-analysis.* Cancer Epidemiology Biomarkers & Prevention, 2011. **20**(6): p. 1238-1250.
54. Butt, H.Z., et al., *Telomere Length Dynamics in Vascular Disease: A Review.* European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery, 2010. **40**(1): p. 17-26.

CHAPTER 2

ADDRESSING METHODOLOGIC CONCERNS IN BIOMARKER RESEARCH: RACE, ETHNICITY, PSYCHOSOCIAL FACTORS, AND TELOMERE LENGTH IN A MULTICENTER SETTING

Shannon M Lynch[1*], M K Peek[2*], Nandita Mitra[1], Krithika Ravichandran[3], Charles Branas[1], Elaine Spangler[1], Wenting Zhou[1], Electra D Paskett[4], Sarah Gehlert[5], Cecilia DeGraffinreid[4], Timothy R Rebbeck[1*], Harold Riethman[3*]

[1]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, USA; [3] Wistar Institute, Philadelphia, PA; [4]The Ohio State University, OH; [2]University of Texas Medical Branch, Galveston, TX; [5] Washington University, St. Louis, MO.

*These authors contributed equally to this work.

'This work was supported by the National Institutes of Health [P50-CA105641, P60-NM006900 and R01-CA85074 to TRR and F31-AG039986 to SML]'. The authors report no conflict of interest related to this work.

Corresponding Author: Shannon M Lynch
Department of Biostatistics and Epidemiology
Perelman School of Medicine at the University of Pennsylvania
243 Blockley Hall, 423 Guardian Drive
Philadelphia, PA 19104-6021
Telephone: 215-898-5300
Fax: 215-573-1050
Email: lynchsh@mail.med.upenn.edu

Abstract

Background: Leukocyte telomere length (LTL) has been associated with age, self-reported race/ethnicity, gender, education, and psychosocial factors, including perceived stress, and depression. However, inconsistencies in associations of LTL with disease and other phenotypes exist across studies. Population characteristics, including race/ethnicity, laboratory methods, and statistical approaches in LTL have not been comprehensively studied and could explain inconsistent LTL associations.

Methods: LTL was measured using terminal restriction fragment assays in 1,510 participants from a multi-center study combining data from 3 centers with different sample characteristics and DNA processing methods. The association between LTL measures, psychosocial factors, and race/ethnicity was evaluated using linear regression and generalized estimating equations, accounting for population characteristics including age, gender, cancer status and center, as well as DNA extraction, a type of laboratory method.

Results: After considering sources of potential bias and confounding in multicenter data, longer LTL was associated with African American race (p-value=0.04) and Hispanic ethnicity (p-value=0.02), and less than a high school education (p-value=0.04). There was an inverse relationship between LTL and perceived stress ($p<0.001$) overall, and between LTL and high school education among African Americans (p-value<0.001).

Conclusions: With proper evaluation and statistical adjustment for center and laboratory effects, combining data from multiple centers is valid and may resolve some inconsistencies in reporting of LTL associations. Biologic effects on LTL may differ under certain psychosocial and racial/ethnic circumstances and could impact future health disparity studies.

Introduction

Telomere DNA consists of long stretches of (TTAGGG) repeat DNA located at the ends of chromosomes[1] and are required for the replication and stability of chromosomes[1]. These repeats naturally shorten with age in all replicating somatic cells[2] due to the inability of the cell to copy the ends of DNA and maintain length over time[3]. Beyond chronological age, telomeres can also shorten prematurely in response to cellular oxidative stress[4, 5]. In normal cells, telomere shortening results in cell senescence or apoptosis[6, 7]. Senescence and apoptosis can function as tumor suppressor mechanisms but can also disrupt normal tissue microenvironments and contribute to aging phenotypes[8-11]. Cells with critically short telomeres that escape apoptosis or senescence [12], and continue to replicate, have unstable genomes and are believed to mark a critical step on the pathway to malignant transformation[2, 13] [14, 15] [4, 16].

Leukocyte telomere length (LTL) has emerged as a potential biomarker of aging, cumulative oxidative stress, and disease, and represents a promising intermediate trait linking chronic cellular stress with disease pathogenesis. Several psychological and social conditions have been associated with both an increase in cellular oxidative stress[5, 17] and subsequent LTL shortening[5, 18-21]. Depression[22], perceived stress[23], and educational attainment[5] are associated with LTL attrition. However, elucidating the complex relationship between psychosocial factors and LTL[24] has been difficult, and inconsistent results have been reported in the literature [5].

While previous studies have demonstrated that membership in certain race/ethnic groups may be associated with a range of socioeconomic and psychosocial factors that could result in shorter LTL [25], namely educational level[5] and perceived stress[23], reports on the effects of race/ethnicity on LTL especially are limited and inconsistent.

Most association studies of LTL have been conducted in female and Non-Hispanic White populations[4, 5]. Studies that include racial/ethnic minorities suggest that Non-Hispanic Whites have shorter LTL compared to African Americans(16, 17) and Hispanics[4]. However, one study suggested that African Americans and Hispanics have shorter LTL than Non-Hispanic Whites[25]. Given the implications for disease prevention, as well as the potential insights into common mechanisms affecting cellular oxidative stress and aging, it is important to better understand both the racial and psychosocial contexts in which changes in telomere biology occur using more diverse samples.

Inconsistencies in telomere research might be due in part to differences among study populations, laboratory approaches, or statistical methods used, sometimes across multiple study sites [4]. Beyond older age and male gender, there is little consensus about the population characteristics that are associated with shorter LTL[4]. The effects of biosample source and laboratory methods on telomere length measurements have been studied, but are still being realized[4, 5, 14, 15, 26]. Differences in cell types used to measure telomere length (i.e. buccal, blood leukocyte, tissue), DNA extraction methods[27], and type of telomere length assay used can affect the validity and reliability of telomere length measurements [4, 5, 14, 15], and ultimately reported LTL associations.

Statistical approaches employed in LTL studies are often chosen based on the characteristics of available study populations and customary laboratory methods. Yet, the type of telomere length assay selected could affect the reporting and statistical analysis of LTL outcome variables. Terminal restriction fragment(TRF) assays, known as the gold standard for measuring LTL[26, 28], measure (TTAGGG)n lengths directly by analysis of Southern blots of restriction digests of genomic DNA with frequently-cutting enzymes; telomere length is reported in terms of the average size of the undigested telomere fragment (which lacks sites for palindrome-dependent restriction enzymes) in

base pairs or kilobases(kb) for each leukocyte DNA sample. Quantitative polymerase chain reaction (qPCR), a high-throughput technique often used in large, population-based studies[4, 29, 30], outputs LTL in terms of T/S ratios. Here, a PCR-generated signal that is dependent upon the total (TTAGGG)n content of the sample (T) is compared to the PCR signal from a known gene present only once in the genome (S). T/S ratios of experimental genomic DNA samples are then each compared with those of a reference genomic DNA sample, determined under identical experimental conditions, to arrive at a value describing the telomere content of each unknown sample, which is assumed to correlate closely with the average telomere length from TRF (24,25). Additionally, some studies account for the potential effects of population characteristics on telomere length outcomes and convert T/S ratios or LTL kb into standardized Z-scores that are adjusted for age and gender[31, 32]. These differences in reported telomere length metrics (e.g. kb, ratios, or Z-scores) can make comparisons across studies difficult, and the implications of using various data transformations and statistical approaches on observed LTL associations has yet to be formally evaluated.

In this study, we use data from a multicenter, multi-racial/ethnic, cross-sectional study designed to investigate the effect of psychosocial factors on cancer-related biomarkers. The study sample includes centers that used different LTL laboratory methods and had different population characteristics. The purpose of this analysis is two-fold. First, we conduct a comprehensive investigation of the collective effect of laboratory procedures, study participant characteristics, and statistical measures in order to better understand telomere length associations and any potential inconsistencies in observed associations across multiple study sites. Second, once these factors have been considered, we evaluate the effect of race/ethnicity on the relationship between psychosocial factors and telomere length.

Methods

Prior to assessing our primary data, we undertook a review of multicenter association studies of LTL in order to ascertain laboratory factors and statistical approaches commonly used in this setting (Supplementary Data File and Supplementary Table 1 in Appendix C). These approaches could serve as methodologic factors that could contribute to inconsistencies in reported associations(4).

Study Sample. Our primary study sample was drawn from three centers: the University of Pennsylvania (Penn), the Ohio State University (OSU), and the University of Texas Medical Branch (UTMB). These centers were originally part of the larger Centers for Population Health and Health Disparities[33] whose main disease focus was on the study of cancer. All study participants were recruited between 2004 and 2012. Each center had its own protocol for recruitment and data collection that has been described previously [34-36], and inclusion/exclusion criteria for each study are listed in Table 1. Study participants agreed to donate a blood sample to extract genomic DNA, and they completed a standardized questionnaire at the time of study enrollment. Study participants were followed-up for cancer status. Informed consent was received from all participants, and study protocols were approved by the Institutional Review Boards of each center.

Covariates. Variables common to all 3 centers included: gender(male/female), age at enrollment (continuous); race/ethnicity(White/Non-Hispanic, African American/Non-Hispanic, and Hispanic), educational status (less than high school or less than 12 years of schooling; high school education or 12 years of schooling/GED); >high school education or >12 years of schooling), disease status (cancer; yes/no), as well as other behavioral factors, including smoking status (ever/never). The psychosocial factors in this study were defined by perceived stress and depression. To evaluate stress, we

used the validated perceived stress scale (PSS)[37, 38]. This is a 10-item global measure of perceived stress where higher scores indicate greater perceived stress(total score range: 1–40). Total PSS was normally distributed in this sample, and we dichotomized this variable to compare high (above median) to low (below median) stress[39, 40]. Questions from the validated Center for Epidemiological Studies-Depression (CES-D) scale[41] and the CES-D revised(R) scale[42] were used to ascertain depressive symptoms. Both the CES-D and CES-DR are 20-item scales (total score range: 0-60). Higher scores, particularly those above 16, suggest more depressive symptoms[41]. The combined total scores from CES-D and CES-DR were positively skewed; we dichotomized at the clinical cut-point of 16[41] to compare those with higher and lower levels of depressive symptoms. PSS, CES-D, and CES-DR scales have been validated in multiethnic studies [43, 44].

Laboratory Methods.

Tissue Source for DNA: Here, All centers followed the same standardized blood draw protocol and used the same tissue source to extract DNA, peripheral blood leukocytes. Twenty milliliters of blood were drawn from each subject by a trained phlebotomist. Samples were centrifuged and buffy coats were stored at -70°C until DNA extraction and telomere assay.

DNA Extraction: Genomic DNA was extracted from each center individually and sent to the Wistar Institute for analysis of LTL. OSU and UTMB samples were processed using the QIAamp DNA Extraction Kit (Valencia, CA). Penn DNA samples were extracted using Chemagen Magnetic Bead technology (n=61) and phenol-chloroform extraction (n=40).

Terminal Restriction Fragment (TRF) assay: TRF length assays were used to measure LTL from extracted DNA on all study samples (using duplicate samples), as described previously by Kimura et al[45] and detailed in Supplementary Laboratory

Methods(Appendix C). Briefly, genomic DNA samples were digested with restriction enzymes *Hinf I* (10 U) and *Rsa I* (10 U; Roche), and mean LTL in kb was determined using Telorun software[45].

Quantitative Telomere PCR (qPCR): For a subset of Cross-Center samples (Penn, n=101 and OSU, n=111), LTL was also measured using the quantitative PCR method developed by Cawthon, modified for compatibility with the Applied Biosystems 7900 HT instrument [30](Supplementary Laboratory Methods-Appendix C). Assays were carried out in triplicate, and center samples were batch analyzed to minimize inter-assay variation. The T/S ratios of each experimental sample relative to the reference sample were generated using the comparative CT (cycle threshold) method[30]**.** T/S ratios and LTL kb were compared for quality control comparisons.

Coefficient of Variation Percentages (CV%): CV% were calculated for duplicate (TRF measurements) or triplicate(qPCR measurements) samples using the pooled standard deviation of the duplicates or triplicates divided by the overall mean of all measurements. The TRF overall CV was 1.25%. The qPCR intra- and inter-plate CV% were 4.9% and 12.9%, respectively.

Statistical Analysis. Data quality control measures were undertaken to identify any potential measurement errors or inconsistencies. Box plots of LTL measurements were generated to identify outlier points or data errors. LTL is described using means, medians, standard deviations and ranges. The distributions of LTL were not normal, and data transformations were conducted for statistical analysis. Methods used in past multicenter studies were used to investigate inconsistencies and LTL associations (Supplementary Methods/Supplementary Table 1-Appendix C). We evaluated correlations between log-transformed LTL from TRF, and qPCR measurements overall, by center, and by DNA extraction method[27]using[4, 26, 28-30] linear regression.

Relevant study population characteristics overall and by center are summarized by medians and frequencies. Comparisons of population characteristics across center and by LTL were conducted using nonparametric tests (Kruskal-Wallis and Wilcoxon ranked sum) for primary evaluations of population characteristics.

Associations between LTL and age and LTL and psychosocial factors were assessed using the two common telomere length metrics reported in multicenter settings (Supplementary Table 1-Appendix C), log-transformed telomere length(kb) and LTL Z-score. Inverse-weighted variance Z-scores were calculated by subtracting the log-transformed LTL sample mean from the original sample values and then divided by the sample standard deviation[31, 32]. Z-scores were also adjusted for age, gender, and cancer status by estimates within strata and then taking the weighted average across strata[46-51] [52, 53]. Multivariable linear models were used to assess associations with age and psychosocial factors that included relevant population and laboratory factors. Relevant factors (identified from previous multicenter studies; Supplementary Methods/Supplementary Table 1-Appendix C) were chosen for final inclusion in our models using stepwise forward and backward variable selection approaches, with a liberal variable inclusion cut-off of $p<0.25$. GEE models (using an independence correlation structure and robust standard errors) [54]) were also fit in order to account for correlation of observations within centers. Interactions between age, gender and psychosocial factors were evaluated using appropriate cross-product terms within the regression model. Subgroup analyses were further conducted by race/ethnicity and in those without cancer and within the UTMB cohort. All P-values were two-sided. All statistical analyses were conducted using STATA version 9.1.

Results

Laboratory Methods Evaluation. Applying proper quality control measures by accounting for documented laboratory factors known to affect LTL measurements (i.e. DNA extraction protocol, type of telomere assay, etc) ensures lab methods are not prone to error that can subsequently affect LTL associations. In pilot experiments, we found that TRF assay results consistently yielded excellent measurement CVs <2%, so these assays were carried out on all samples. qPCR measurements for the OSU and UPenn samples had higher CVs than the TRF assays(UPenn: qPCR 12.0 CV% and TRF 0.93 CV%; OSU: qPCR 0.12 CV% and TRF 0.01 CV%--consistent with the literature)[14], whereas qPCR measurements of the UTMB samples yielded unacceptably high measurement CVs (27 CV%), possibly due to an unknown analyte affecting the qPCR reaction (24)), and were therefore excluded from the analysis. qPCR and TRF comparisons were made with OSU and UPenn samples[14]. The relationship between log-transformed TRF measures and T/S ratios for 211 samples with both TRF and T/S ratio data showed an overall R^2 of 0.60(Figure 1a). The R^2 within centers was 0.71 for Penn (Figure 1b) and 0.93 for OSU (Figure 1c). Comparing log-transformed TRF to T/S ratios by DNA extraction method, the R^2 for QiAmp DNA extraction was 0.81; for Chemagen, 0.69 and for phenol-chloroform, 0.90. The mean (standard deviation) LTL across all centers was 6.55kb (2.86). Within center, mean LTL was 8.42kb (4.50) for Penn, 6.34kb (1.95) for OSU, and 6.42kb (2.71) for UTMB. Median LTL was significantly different by extraction method (p-value<0.001)(Figure 1d). However, median LTLs were not significantly different between Qiagen and Chemagen methods (p-value=0.48).

Study population evaluation. Baseline characteristics of the study overall (n=1510) and by center were evaluated to determine potential clustering and confounding effects by center (Table 2). The overall study population was 58.8% female. 15.7% had a cancer diagnosis, and 51% had ever smoked cigarettes. The average age was 50.6 years, with a standard deviation of 15.6. All population characteristics, except for

smoking, were significantly different across centers. Baseline study population characteristics of the combined study population (includes all 3 centers) were compared on median LTL and log-transformed LTL(kb) in order to compare our results to literature and to identify factors related to LTL that could be tested in forward and backward regression models with age and LTL(Table 3). Only cancer status had a significant association with median LTL; cancer cases had longer LTL than those without cancer (p-value=0.02). There was no statistical relationship between LTL and gender (median LTL(kb), interquartile range (IQR): men=6.43, 4.14-8.39; women=6.33, 4.39-8.27; Table 3) in the overall cohort and when restricting the population to those without cancer(median TRF(kb), IQR: men=6.01kb, 4.10-8.00; women=6.33, 4.48-8.27, p-value=0.12; Supplementary Table 2-Appendix C).

Association of laboratory factors, population covariates, and LTL. There was no correlation between age and log-transformed (R^2= -0.08, p-value=0.45) and Z-score LTL(R^2= -0.10, p-value=0.49). The best fitting linear regression and GEE model for continuous age and log-transformed LTL or Z-score were the same and included the following: gender (GEE p-value<0.001), cancer status (GEE p-value<0.001), a gender-cancer status interaction (GEE p-value<0.001), and DNA extraction method (GEE p-value<0.001). The gender-cancer status interaction remained when using OSU/UTMB(p-value=0.01) or UTMB data only(p-value=0.02).

Association of Race, Education, Psychosocial Factors and LTL. The distribution of psychosocial factors differed significantly across centers (Table 2). The sample was comprised of 45.6% non-Hispanic Whites, 45.0% Hispanics, and 9.4% African Americans. The Penn and OSU study participants reported higher levels of education than UTMB (p-value<0.001). OSU, which included only females, reported the highest levels of stress and depression (p-value=0.03). There were no statistically significant associations between psychosocial factors and LTL (Table 3). However, African

Americans had the longest median LTL (6.61kb, IQR=4.56-8.82), and Non-Hispanic Whites the shortest (6.11kb, IQR=4.19-8.23). Patterns were consistent when restricting the study population to those without cancer (Supplementary Table 2-Appendix C) and UTMB only. No significant interactions between population characteristics and psychosocial factors were observed.

Associations with log-transformed LTL and Z-scores with race and psychosocial factors were estimated using both using linear regression (Model 1) and GEE models (Model 2)(Table 4). Regardless of LTL measure or statistical model, there was a significant, direct relationship between LTL and race/ ethnicity. For both LTL outcome measures (log-transformed LTL and Z-score), GEE models presented a significant relationship between lower levels of education (less than high school) (log-transformed LTL p-value=0.02) and higher levels of perceived stress(log-transformed p-value<0.001). Associations were similar when limiting the study population to those without cancer and UTMB only (data not shown). Since both LTL measures resulted in similar association results, and we thought it was important to account for cluster effects of center, the final model was chosen to be a GEE models with log-transformed LTL .

No statistically significant associations between log-transformed LTL and psychosocial factors were reported for Caucasians or Hispanics (Table 5). Compared to those with more than a high school education, having only a high school education was significantly related to shorter LTL (p-value<0.001) in African Americans.

Discussion

Inconsistent associations between LTL, race/ethnicity, and psychosocial factors in literature have been reported [5] (16, 17) [23] [25], and few studies have evaluated the association between LTL and psychosocial factors within race/ethnic subgroups[25]. Inconsistencies in literature between socioeconomic and psychosocial factors and LTL

have been attributed to different laboratory and statistical approaches employed in these telomere studies[4], but few studies have evaluated methodological effects. Multi-center studies serve as an ideal opportunity for evaluating methodological effects on LTL associations since they often combine data from centers with heterogenous populations and varying laboratory approaches. Our findings suggest that combining and comparing data from multiple centers is valid and can have little effect on LTL associations, with proper adjustments. We first showed that our telomere measurements in the combined study population were reliable and valid compared to other published studies[4, 5, 14, 15]. More specifically, we evaluated the source of DNA, type of telomere length assays, CV percents, and DNA extraction techniques[4, 5, 14, 15] since they are known to contribute to discrepancies in reporting associations between LTL and cancer and other diseases[3, 14, 27]. Choice of tissue type and assay in our study were consistent with literature(4), and correlations between TRF and T/S ratios for Penn and OSU were within range of other studies (0.60-0.95)[26, 29, 30, 52, 53]. Although not included in TRF and T/S ratio comparisons, UTMB LTL measures were reliable based on a 2.0% TRF CV%[4].

Similar to published studies, phenol-chloroform DNA extraction resulted in longer telomeres than Qiagen methods[14, 27], and Qiagen and Chemagen, both column-based extraction methods, yielded similar median LTL results [14, 27]. The majority of our samples were extracted using Qiagen and Chemagen(97.3%), thus our LTL measurements could be underestimated and result in Type II error. However, the bias is likely nondifferential. Few multicenter studies of LTL report and consider the effects of DNA extraction on study outcomes(Supplementary Table 1-Appendix C), and DNA extraction appears to contribute to inconsistent findings in telomere association studies[14, 27].

Age and male gender have been associated with shorter LTL(4) in many studies. While we see the same trends in our data, we do not observe statistically significant associations(Supplementary Table 2-Appendix C). Although the linear relationship between age and LTL was weaker in the present study for log-transformed LTL(R^2= -0.08) than previously reported(R^2~0.15 [4]), the attenuated association observed between age and LTL when adjusting for other covariates, like gender, is consistent with other studies[53]. Additionally, the rate of telomere attrition may vary over lifespan, with some studies suggesting more rapid attrition in younger ages (childhood) and in later decades of life (over age 70) [55, 56]; the age range of the sample was 26-64 and the median age of the sample was relatively young at 51 years (Table 2). We also found that male cancer cases had longer telomeres compared to non-cancer cases ,and this has been observed in literature[57], although inconsistently[58].

These initial evaluations informed which laboratory and population factors may affect LTL associations in our study. DNA extraction methods, along with age, gender, cancer status, and the interaction of gender and cancer status, were significant confounders. Center-specific study recruitment led to specialized groupings of gender and cancer status by center. Thus, center was a cluster variable, and GEE models, which accounted for the within and between effects of the center cluster variable and include stricter standard errors[59], appeared more appropriate in our analyses. Few multi-center association studies of LTL have accounted for potential cluster effects (Supplementary Table 1-Appendix C). Concern over additional variability in LTL in those w/ cancer and by center prompted us to compare findings when restricting the population to those without cancer and UTMB only. We found that results were robust and that extraneous variability in LTL appeared to be removed with adjustment for relevant population and laboratory methods.

We also evaluated the choice of outcome measure(log-transformed LTL or Z-score). Most multicenter studies of LTL report log-transformed LTL (Supplementary Table 1-Appendix C). However, Z-scores standardize telomeres based on sample distributions and may be more appropriate in instances where the distribution of LTL greatly differs by center or when confounders or model adjustment variables differ by center. Although the magnitude of effects appear different (and often higher with Z-score), they are not comparable. This is because the data transformation associated with each of these measures lends itself to different interpretations. Nevertheless, patterns of association between LTL and race/ethnicity and psychosocial factors were similar regardless of which telomere outcome measure (log-transformed LTL or Z-score) was used.

To our knowledge, this is the first study to evaluate the effect of race/ethnicity on the relationship between socioeconomic and psychosocial factors and LTL, and to more comprehensively investigate the collective effect of laboratory procedures, study population characteristics, and statistical measures on reported LTL associations. We found significant associations between race/ethnicity, low levels of education, perceived stress, and LTL. Associations between high levels of perceived stress and shorter LTL have been reported[5]. We are only the second study to report that both African Americans and Hispanics have longer LTLs than Non-Hispanic Whites[4]. Having less than a high school education was associated with longer LTL, which is an association not typically reported in literature(5)[25]. When stratifying the analysis by race, there was a suggested association between longer LTL and less than a high school education for Hispanics, and a significant association between shorter LTL and having a high school education for African Americans. Thus, the racial, ethnic and educational composition of our sample (including a large number of Hispanics with low education) may have affected our education findings. Studies have found correlations with socioeconomic status (SES) related to education and income, and race, namely lower

SES conditions are associated with African Americans[60]. Being Hispanic is also associated with lower levels of education in literature, as well as improved mortality rates compared to African Americans[60], referred to as the Hispanic paradox[61-63]. Given that shorter LTL is believed to be related to mortality[4], racial composition appears to be an important consideration in LTL studies.

Our study had some limitations. This was a cross-sectional investigation, limiting us to studying variables that were common to all 3 centers. For instance, duration and severity of depression and perceived stress are more consistently associated with shorter LTL[64], and LTL is likely to shorten over time(1). Stratified analyses by race yielded small samples, particularly for African Americans, but findings suggest studies focused on telomere biology by race/ethnicity are warranted. Like most LTL association studies, differences in mean LTL could be influenced by the proportions of different kinds of leukocytes[65]. The average LTL in any given study is considered to be a general average of all the LTLs across all chromosomes and blood leukocytes. Although it is unclear whether differential cell counts are affected by race/ethnicity in a way that would explain the patterns we observed, one previous study found no association between leukocyte type and LTL in African Americans or Non-Hispanic Whites[66].

The large multi-ethnic and multicenter composition of our study allowed for more in depth analysis of the effects of laboratory and statistical approaches on telomere length associations. Our study demonstrated that with proper evaluation and adjustment of center and laboratory effects, combining data from multiple centers, with different laboratory approaches and population characteristics, can be a powerful and valid approach for assessing LTL associations. In addition, evaluating methodologic effects, similar to what we have done here, within and across LTL studies may help resolve inconsistent reports of LTL associations. Our data provide evidence of an association between Hispanics and African Americans and longer LTLs, as well as potential

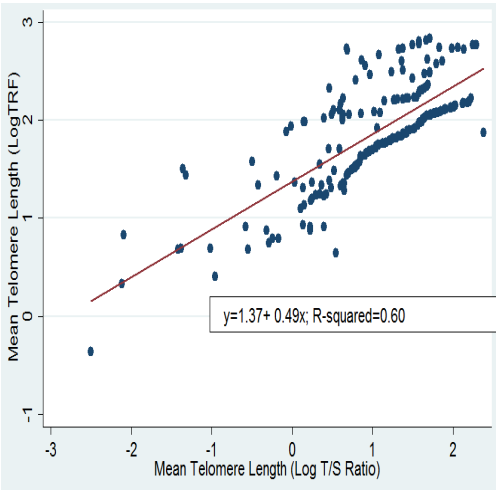
relationships between educational level, perceived stress and LTL for certain racial/ethnic sub-groups. Further study into the effects of socioeconomic and psychosocial factors on LTL by race/ethnicity could have implications for research involving health disparities and disease outcomes.

Table 1: Study Descriptions and Inclusion/Exclusion Criteria

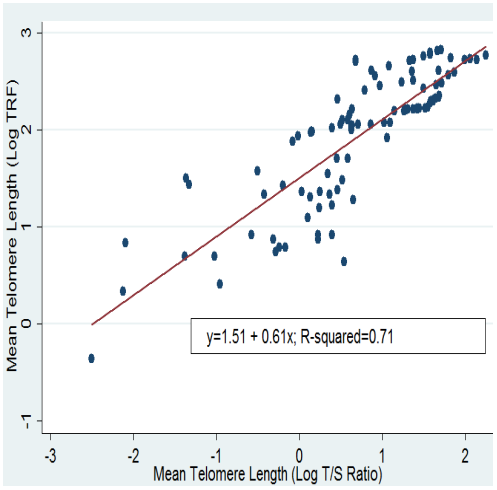
Center	**Original Disease Focus: Primary Race/Gender**	**Sample size(n=1510)**	**Inclusion criteria**
Ohio State University(OSU)[34]	Cervical cancer: Non-Hispanic White/ underserved women	111	Women from Appalachia with an intact uterine cervix and corpus, not pregnant, and no history of cervical cancer recruited at time of routine cervical cytology.
University of Pennsylvania Hospital System (UPenn)[36]	Prostate cancer: Non-Hispanic White and African-American/men	101	Male prostate cancer patients from UPenn urology clinics with blood sample.
University of Texas Medical Branch(UTMB)[35]	Stress effects near oil refineries: Non-Hispanic and Hispanic Whites and African American/men and women	1298	Population-based sample of Non-Hispanic households and a strata sample of Hispanic households in Texas City, TX.

Figure 1. **a-c**. Comparison of Log-transformed TRF to T/S Ratio overall and by center. **d**. Median Telomere Length and R^2 by DNA extraction method.

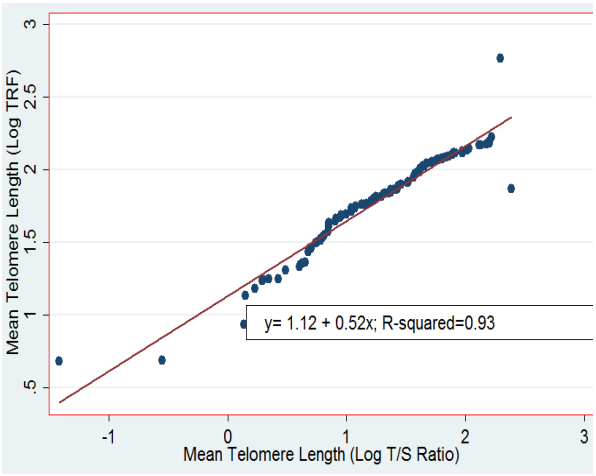
a. Overall



b. UPHS



c. OSU



d. DNA Extraction Method (Center, N, Median, R^2)

Chemagen
(UPHS, n=61, 6.80, R^2=.69) QiAmP
(OSU/UTMB, n=1409, 6.28, R^2=.81)
Phenol-Chloroform
(UPenn, n=40, 9.20,R^2= 0.90)

* R^2 ranges from 0.60-0.95 in literature[26, 29, 30].

Table 2. Study Characteristics

	ALL Centers	**UPenn**	**OSU**	**UTMB**	**p-value[b]**
Population Characteristics					
Total Population (n)	1510	101	111	1298	
Median Telomere length (kb)[a]	6.4 (4.3-8.3)	8.7 (4.2-11.8)	6.3 (5.1-7.9)	6.3 (4.2-8.3)	0.0001
Median Age[a]	51 (38-63)	58 (53-63)	30 (26-43)	51 (38-64)	0.0001
Male Gender (%)	41.2	100	0	40.1	0.0001
Cancer diagnosis(%)	15.7	100	0	10.5	<0.001
Ever Cigarette Smokers(%)	51.0	57.4	42.3	51.2	0.08
Race (%)					
Non-Hispanic White	45.6	89.0	98.2	37.8	
African American	9.4	11.0	1.8	9.9	
Hispanic	45.0	0	0	52.3	<0.001
Education(%)					
> High School	37.0	72.3	64.9	31.8	
High School/GED	29.8	23.7	23.4	30.8	
<High School	33.2	4.0	11.7	37.4	<0.001
Psychosocial Factors					
Median Total Depression Score	6(1-15)	8 (3-13)	12 (7-22)	5 (0-14)	0.0001
Median Total Perceived Stress	19 (17-22)	20 (18-22)	22 (19-24)	19(16-22)	0.0001

*Abbreviations: University of Pennsylvania(UPenn), Ohio State University(OSU), University of Texas Medical Branch(UTMB);

[a] Medians(interquartile range for the median); [b] p-values comparing characteristics across each of the 3 centers using Kruskal Wallis Test or Fisher's Exact Test.

[b] Range from low to high for Depression (0-60) and Perceived Stress Scale (0-40).

Table 3. Unadjusted Median Telomere Length(kb) and Mean Log-Telomere Length(TL in kb) by Study Characteristics(ALL n=1510)

	Median TL (kb) (Interquartile Range)[a]	**p-value**[b]	**Mean logTL(SD)**	**p-value**[c]
Population Characteristics				
Age				
Younger age(<=51)	6.39(4.33-8.29)		1.78(0.44)	
Older age(>51)	6.30(4.20-8.32)	0.68	1.77(0.50)	0.51
Gender				
Female	6.33 (4.39-8.27)		1.77(0.44)	
Male	6.43 (4.14-8.39)	0.58	1.78(0.51)	0.07
Cancer diagnosis				
Yes	7.14 (4.00-9.18)		1.83(0.58)	
No	6.28 (4.30-8.17)	0.02	1.77(0.45)	0.08
Ever Smoker				
Yes	6.45(4.25-8.29)	0.71	1.78(0.47)	
No	6.28(4.31-8.31)		1.77(0.48)	0.89
Race				
Non-Hispanic White	6.11(4.19-8.23)		1.75(0.490	
African American	6.61 (4.56-8.82)		1.83(0.48)	
Hispanic	6.42 (4.42-8.27)	0.12	1.79(0.45)	0.18
Education				
> High School	6.35(4.29-8.30)		1.77(0.50)	
High School/GED	6.13(4.29-8.30)		1.76(0.48)	
<High School	6.43(4.53-8.15)	0.61	1.80(0.43)	0.46
Psychosocial Factors				
Total Perceived Stress				
High Stress (>19)	6.33(4.20-8.32)		1.77(0.48)	

Low Stress (<=19)	6.39(4.33-8.26)	0.60	1.78(0.47)	0.69
Total Depression				
High Depression (>16)	6.45(4.50-8.52)		1.81(0.44)	
Low Depression (<=16)	6.33(4.24-8.23)	0.12	1.76(0.48)	0.19

* Ranges of Median and Mean telomere length are similar to those reported in literature[26, 29, 30].

[a] Medians(interquartile range for the median); [b] p-values comparing characteristics across 3 or more groups using Kruskal Wallis Test, otherwise used Wilcoxon Ranked Sum Test; [c] p-values comparing characteristics across 3 or more groups using ANOVA, otherwise used T-test.

Table 4. Adjusted Regression Estimates(Standard Errors) of Individual Race/Ethnicity, Education, Psychosocial Factors and Log-Transformed and Z-Score Telomere Length(TL in kb).

	Log-Transformed TL		**Z-Score TL**	
	Model 1	**Model 2**	**Model 1**	**Model 2**
Race (compared to Non-Hispanic Whites)				
African-American	0.10(0.04)**	0.09(0.04)**	0.19(0.09)**	0.17(0.08)**
Hispanic	0.07(0.03)**	0.06(0.01)***	0.16(0.06)**	0.13(0.03)***
Education (compared to > High School)				
High School education	0.01(0.03)	0.01(0.03)	0.01(0.06)	0.0004(0.04)
Less than high school	0.06(0.03)*	0.06(0.02)**	0.13(0.06)*	0.12(0.04)**
Perceived Stress				
High Stress(compared to low stress)	-0.02(0.02)	-0.02(0.003)***	-0.05(0.05)	-0.05(0.003)***
Depression				
High Depression (compared to low)	0.04(0.03)	0.04(0.02)*	0.07(0.06)	0.07(0.05)*

Model 1=Linear Regression; **Model 2**=GEE, accounts for clustering of center. LogTL model adjusted by age, gender, cancer status, DNA extraction, and the interaction of gender and cancer status. Z-score adjusted by age, gender and cancer status and model adjusted by DNA extraction.

***p-value $\leq$ 0.001;**p-value >0.001 and <0.05; *suggestion of significance p-value<0.15, but $\geq$0.05.

Table 5. GEE Estimates(Standard Errors) of Individual Socioeconomic and Psychosocial Factors and Log-Transformed Telomere Length(TL in kb) stratified by Race/Ethnicity and adjusted for age, gender, cancer status, gender-cancer status interaction, and DNA extraction method.

	Non-Hispanic Whites (n=688)	**African Americans (n=142)**	**Hispanics[a] (n=688)**
Education			
High School education	0.03(0.04)	-0.11(0.03)***	0.001(0.05)
Less than high school	-0.02(0.05)	0.01(0.02)	0.07(0.04)*
Perceived Stress			
High Stress	0.02(0.03)	-0.001(0.01)	-0.05(0.04)*
Depression			
High Depression	0.09(0.05)	0.02(0.04)	-0.01(0.04)

[a] Linear Regression Model is reported since all Hispanics come from only 1 center. This model is adjusted by age, gender, cancer status and the interaction of gender and cancer status. ***p-value $\leq$ 0.001;**p-value >0.001 and <0.05; *suggestion of significance p-value<0.15, but $\geq$0.05.

References:

1. Cheung, A.L.M. and W. Deng, *Telomere dysfunction, genome instability and cancer.* Frontiers in Bioscience, 2008. **13**(6): p. 2075-2090.
2. Londoño-Vallejo, J.A., *Telomere instability and cancer.* Biochimie, 2008. **90**(1): p. 73-82.
3. Blackburn, E.H., *Telomere states and cell fates.* Nature, 2000. **408**(6808): p. 53-56.
4. Sanders, J.L. and A.B. Newman, *Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither?* Epidemiologic Reviews, 2013. **35**(1): p. 112-131.
5. Starkweather, A.R., et al., *An Integrative Review of Factors Associated with Telomere Length and Implications for Biobehavioral Research.* Nursing Research, 2014. **63**(1): p. 36-50 10.1097/NNR.0000000000000009.
6. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres.* Annu Rev Genet, 2008. **42**: p. 301-34.
7. Muñoz, P., R. Blanco, and M.A. Blasco, *Role of the TRF2 Telomeric Protein in Cancer and Aging.* Cell Cycle, 2006. **5**(7): p. 718-721.
8. Coppé, J.P., Desprez, P.Y., Krtolica, A, Campisi J. , *The senescence-associated secretory phenotype: the dark side of tumor suppression.* Annu Rev Pathol, 2010. **5**: p. 99–118.
9. Davalos, A.R., Coppe, J.P., Campisi, J., Desprez, P.Y. , *Senescent cells as a source of inflammatory factors for tumor progression.* Cancer Metastasis Rev, 2010. **29**: p. 273-83.
10. Jaskelioff, M., Muller, F.L., Paik, J.H., Thomas, E., Jiang, S, Adams, A.C, Sahin, E, Kost-Alimova, M, Protopopov, A, Cadiñanos, J, et al. , *Telomerase*

reactivation reverses tissue degeneration in aged telomerase-deficient mice. Nature, 2010. **469**: p. 102–6. .
11. Sahin, E., Depinho, R.A. , *Linking functional decline of telomeres, mitochondria and stem cells during ageing.* Nature 2010. **464**: p. 520-528.
12. Halvorsen, T.e.a., *Telomerase Activity Is Sufficient To Allow Transformed Cells To Escape from Crisis.* Mol Cell Biol., 1999. **19**(3): p. 1864-1870.
13. Koorstra, J.B.M., et al., *Pancreatic Carcinogenesis.* Pancreatology, 2008. **8**(2): p. 110-125.
14. Cunningham, J.M., et al., *Telomere Length Varies By DNA Extraction Method: Implications for Epidemiologic Research.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(11): p. 2047-2054.
15. Wentzensen, I.M., et al., *The Association of Telomere Length and Cancer: a Meta-analysis.* Cancer Epidemiology Biomarkers & Prevention, 2011. **20**(6): p. 1238-1250.
16. Butt, H.Z., et al., *Telomere Length Dynamics in Vascular Disease: A Review.* European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery, 2010. **40**(1): p. 17-26.
17. Shalev, I., et al., *Stress and telomere biology: A lifespan perspective.* Psychoneuroendocrinology, 2013. **in press**.
18. Epel, E.S., Lin, J., Dhabhar, F.S., Wolkowitz, O.M., Puterman, E., Karan, L., et al., *Dynamics of telomerase activity in response to acute psychological stress.* Brain Behav Immun., 2010. **24**: p. 531-539.
19. O'Donovan A, L.J., Dhabhar FS, Wolkowitz O, Tillie JM, Blackburn E, Epel E *Pessimism correlates with leukocyte telomere shortness and elevated interleukin-6 in post-menopausal women. .* Biol. Psychiatry, 2009. **24**: p. 446-449.
20. O'Donovan, A., Epel, E., Lin, J., Wolkowitz, O., Cohen, B., Maguen, S., Metzler, T., Lenoci, M., Blackburn, E., Neylan, T.C., *Childhood trauma associated with short leukocyte telomere length in posttraumatic stress disorder.* Biol. Psychiatry, 2011a. **70**: p. 465-471.
21. O'Donovan, A., Tomiyama, J., Lin, J., Puterman, E., Adler, N.E., Kemeny, M., Wolkowitz, O.M., Blackburn, E.H., Epel, E.S. , *Stress appraisals and cellular aging: A key role for anticipatory threat in the relationship between psychological stress and telomere length.* Brain, Behavior, and Immun, 2012.
22. Wolkowitz, O.M., Mellon, S.H., Epel, E.S., Lin, J., Dhabhar, F.S., Su, Y, et al., *Leukocyte Telomere Length in Major Depression: Correlations with Chronicity, Inflammation and Oxidative Stress - Preliminary Findings.* PLoS ONE, 2011. **6**(3): p. e17837.
23. Parks, C.G., et al., *Telomere Length, Current Perceived Stress, and Urinary Stress Hormones in Women.* Cancer Epidemiology Biomarkers & Prevention, 2009. **18**(2): p. 551-560.
24. Epel, E.S., Blackburn, E.H., Lin, J., Dhabhar, F.S., Adler, N.E., Morrow, J.D., Cawthon, R.M., *Accelerated telomere shortening in response to life stress.* Proc. Natl. Acad. Sci, 2004. **101**: p. 17312-17315.
25. Diez Roux, A.V., et al., *Race/ethnicity and telomere length in the Multi-Ethnic Study of Atherosclerosis.* Aging Cell, 2009. **8**(3): p. 251-257.
26. Aviv, A., et al., *Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR.* Nucleic Acids Research, 2011. **39**(20): p. e134.
27. Hofmann, J., Hutchinson, A.A., Cawthon, R., Liu, C.S., Lynch, S.M., et al. , *Telomere Length Varies By DNA Extraction Method: Implications for Epidemiologic Research–Letter.* Unpublished, 2014.

28. Elbers, C.C., et al., *Comparison Between Southern Blots and qPCR Analysis of Leukocyte Telomere Length in the Health ABC Study.* The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 2013.
29. Cawthon, R.M., *Telomere measurement by quantitative PCR.* Nucleic Acids Research, 2002. **30**(10): p. e47.
30. Cawthon, R.M., *Telomere length measurement by a novel monochrome multiplex quantitative PCR method.* Nucleic Acids Research, 2009. **37**(3): p. e21.
31. Du, M., et al., *Physical Activity, Sedentary Behavior, and Leukocyte Telomere Length in Women.* American Journal of Epidemiology, 2012. **175**(5): p. 414-422.
32. Codd V., M.M., van der Harst, P., et al. , *Common variants near TERC are associated with mean telomere length.* Nat Genet., 2010. **42**(3): p. 197–199.
33. Warnecke, R.B., et al., *Approaching Health Disparities From a Population Perspective: The National Institutes of Health Centers for Population Health and Health Disparities.* American Journal of Public Health, 2008. **98**(9): p. 1608-1615.
34. Paskett, E.D., et al., *Psychosocial predictors of adherence to risk-appropriate cervical cancer screening guidelines: A cross sectional study of women in Ohio Appalachia participating in the Community Awareness Resources and Education (CARE) project.* Preventive Medicine, 2010. **50**(1–2): p. 74-80.
35. Peek, M.K., et al., *Allostatic Load Among Non-Hispanic Whites, Non-Hispanic Blacks, and People of Mexican Origin: Effects of Ethnicity, Nativity, and Acculturation.* American Journal of Public Health, 2010. **100**(5): p. 940-946.
36. Rebbeck, T.R., et al., *Joint effects of inflammation and androgen metabolism on prostate cancer severity.* International Journal of Cancer, 2008. **123**(6): p. 1385-1389.
37. Krieger, N., Smith, K., Naishadham, D., Harman, C., Barbeau, E.M., *Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health.* Social Science and Medicine, 2005. **61**: p. 1576-1596.
38. Cohen, S., Kessler, R.C., Underwood, L., *Perceived stress scale. Measuring stress: A guide for health and social scientists.* Oxford University Press, 1994. **New York**.
39. Wang, L., et al., *The Effects of Perceived Stress and Life Style Leading to Breast Cancer.* Women & Health, 2012. **53**(1): p. 20-40.
40. Shell, A.M., M.K. Peek, and K. Eschbach, *Neighborhood Hispanic composition and depressive symptoms among Mexican-descent residents of Texas City, Texas.* Social Science & Medicine, 2013. **99**(0): p. 56-63.
41. Radloff, L., *The CES-D scale: a self-report depression scale for research in the general population.* Applied Psychological Measurement, 1977. **1**: p. 385-401.
42. Eaton, W.W., Smith, C., Ybarra, M., Muntaner, C., Tien, A., *Center for Epidemiologic Studies Depression Scale: review and revision (CESD and CESD-R).* In ME Maruish (Ed.). The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, 2004. **3rd Edition**(Volume 3: Instruments for Adults): p. 363-377.
43. Crockett, L.J., Randall, B.., Shen, Y.L., Russell, S.T., Driscoll, A.K., *Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: a national study* Journal of Consulting and Clinical Psychology, 2005. **73**(1): p. 47-58.
44. Flores, E., Tschann, J., Dimas, J., Bachen, E., Pasch, L., De Groat, C., *Perceived discrimination, perceived stress, and mental and physical health*

among Mexican-origin adults. Hispanic Journal of Behavioral Sciences, 2008. **30**(4): p. 401-424.
45. Kimura, M., et al., *Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths.* Nat. Protocols, 2010. **5**(9): p. 1596-1607.
46. Weischer, M., et al., *Short Telomere Length, Cancer Survival, and Cancer Risk in 47102 Individuals.* Journal of the National Cancer Institute, 2013. **105**(7): p. 459-468.
47. Codd, V., et al., *Common variants near TERC are associated with mean telomere length.* Nat Genet, 2010. **42**(3): p. 197-199.
48. Codd, V., et al., *Identification of seven loci affecting mean telomere length and their association with disease.* Nat Genet, 2013. **45**(4): p. 422-427.
49. Maubaret, C.G., Salpea, K.D., Romanoski, C.E., Folkersen, L., Cooper, J.A., et al., *Association of TERC and OBFC1 Haplotypes with Mean Leukocyte Telomere Length and Risk for Coronary Heart Disease* PLoS ONE, 2013. **8**(12): p. e83122. doi:10.1371/journal.pone.0083122.
50. Bojesen, S.E., et al., *Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer.* Nat Genet, 2013. **45**(4): p. 371-384.
51. Nordfjäll, K., et al., *Telomere Length Is Associated With Obesity Parameters but With a Gender Difference.* Obesity, 2008. **16**(12): p. 2682-2689.
52. Levy, D., et al., *Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology.* Proceedings of the National Academy of Sciences, 2010. **107**(20): p. 9293-9298.
53. Hunt, S.C., et al., *Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study.* Aging Cell, 2008. **7**(4): p. 451-458.
54. Hubbard, A.E., et al., *To GEE or Not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health.* Epidemiology, 2010. **21**(4): p. 467-474 10.1097/EDE.0b013e3181caeb90.
55. Aubert, G., , Lansdorp, P. M. , *Telomeres and Aging.* Physiol Rev, 2008. **88**(2): p. 557-579.
56. Engelhardt, M., Kumar, R., Albanell, J., et al., *Telomerase regulation, cell cycle, and telomere stability in primitive hematopoietic cells.* Blood, 1997. **90**(1): p. 182-193.
57. Lynch, S.M., et al., *A prospective analysis of telomere length and pancreatic cancer in the alpha-tocopherol beta-carotene cancer (ATBC) prevention study.* International Journal of Cancer, 2013. **133**(11): p. 2672-2680.
58. Ma, H., et al., *Shortened Telomere Length Is Associated with Increased Risk of Cancer: A Meta-Analysis.* PLoS ONE, 2011. **6**(6): p. e20466.
59. Hubbard, A.E., Ahern, J., Fleischer, N.L., Van der Laan, M., et al., *To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health.* Epidemiology 2010. **21**(4): p. 467-474.
60. National Research Council (US) Panel on Race, E., and Health in Later Life., *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life.*, in *9, Race/Ethnicity, Socioeconomic Status, and Health*, N.B. Anderson, Bulatao, R.A., Cohen B., editors, Editor 2004, National Academies Press (US): Washington (DC).
61. Palloni, A., Arias, E. , *Paradox lost: explaining the Hispanic adult mortality advantage.* Demography, 2004. **41**(385-418).

62. Markides, K.S., Coreil, J., *The health of Hispanics in the Southwestern United States: an epidemiologic paradox. .* Public Health Reports. , 1986. **101**: p. 253-265.
63. Markides, K.S., Eschbach, K., *Hispanic paradox in adult mortality in the United States.* In R. Rogers & E. Crimmins (Eds.), International Handbook of Adult Mortality. New York: Springer, , 2011: p. 227-240.
64. Bleil, M.E., et al., *Depressive symptomatology, psychological stress, and ovarian reserve: a role for psychological factors in ovarian aging?* Menopause, 2012. **19**(11): p. 1176-1185 10.1097/gme.0b013e31825540d8.
65. Weng, N.-p., *Interplay between telomere length and telomerase in human leukocyte differentiation and aging.* Journal of Leukocyte Biology, 2001. **70**(6): p. 861-867.
66. Hunt, S.C., Chen, W., Gardner, J.P., et al., *Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. .* Aging Cell, 2008. **7**(4): p. 451-458.

CHAPTER 3

TELOMERE LENGTH AND NEIGHBORHOOD CANCERIZATION: EVALUATING BIOLOGICAL RESPONSE TO UNFAVORABLE EXPOSURES

Shannon M. Lynch[1], Nandita Mitra[1], Krithika Ravichandran[2], Elaine Spangler[1], Wenting Zhou[1], [3]Electra D. Paskett, [5]Sarah Gehlert, [3]Cecilia DeGraffinreid, Tamara Dubowitz[6], Harold Riethman[2], Charles C. Branas[1], M.K. Peek[4], Timothy R. Rebbeck[1]

[1]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, USA; [2]Wistar Institute, Philadelphia, PA; [3]The Ohio State University, OH; [4]University of Texas Medical Branch, Galveston, TX; [5]Washington University, St. Louis, MO. [6]The RAND Corporation, CA.

This research was supported by grants from the Public Health Service (P50-CA105641, P60-NM006900 and R01-CA85074 to TRR and F31-AG039986 to SML)

Corresponding Author: Shannon M. Lynch
Department of Biostatistics and Epidemiology
Perelman School of Medicine at the University of Pennsylvania
243 Blockley Hall, 423 Guardian Drive
Philadelphia, PA 19104-6021
Telephone: 215-898-5300
Fax: 215-573-1050
Email: lynchsh@mail.med.upenn.edu

Abstract

Background: Analogous to “field cancerization” in pathobiology, “neighborhood cancerization” describes how individuals within areas exposed to common unfavorable circumstances can experience biological consequences that may result in cancer and other diseases. Telomere length (TL) shortening has been associated with exposure to individual-level and neighborhood-level stressors as well as cancer development, yet these relationships are not well understood. This study examined the complex association between neighborhood characteristics and TL to test the neighborhood cancerization hypothesis.

Methods: We studied 1,488 individuals from 127 census tracts in three regions of the US. TL was measured from peripheral blood using terminal restriction fragment(TRF) assays. Multilevel linear models and quantile regression models were fitted adjusting for individual-level characteristics including self-reported race, education, perceived stress and depression. Neighborhood sociodemographic exposures included population density, urban/residential crowding, residential stability/instability, and socioeconomic status (SES).

Results: Neighborhood population density, urban/residential crowding, residential stability/instability, and SES were not significantly associated with TL using standard linear models. Quantile regression revealed significant inverse associations between population density and urban crowding and the 5^{th} (population density, p-value=0.03; urban crowding p-value=0.002), 50^{th}(both p-values <0.001) and 75^{th} percentiles(both p-values<0.001) of the TL distribution. Significant associations between residential stability and TL were seen at the upper tails of the TL distribution (95^{th} percentile-p-value=0.006; 90^{th} percentile-p-value=0.005).

Conclusions: The relationship between neighborhood sociodemographic variables and TL may be nonlinear, with only a portion of the TL distribution being associated with neighborhood-level exposures. Common exposures identified across different neighborhoods can exert biological effects. These results support the neighborhood cancerization hypothesis.

Introduction

The concept of field cancerization was originally proposed over a half century ago to explain synchronous or metachronous tumors of the oral mucosa[1]. Field cancerization came to refer to the propensity of a field of tissue to become malignant based on either pathological observations or molecular markers[2]. In theory, field cancerization occurs from simultaneous wide-spread, unfavorable exposures acting on cells and tissues within organs and organ systems (such as the effect of cigarette smoking on the lung), as well as the migration of patches of molecularly altered cells into larger tissue fields or territories that become predisposed to cancer development[2]. Common molecular signatures in pre-malignant cells and primary, malignant tumors provide evidence of the cancer field effect. The biological basis for field cancerization has been related to epigenetic changes[3-5], tissue micro-environment[6, 7], and telomere function[8] across different tissue types.

Analogous to a field of tissue for which unfavorable exposures may lead to carcinogenesis, residents of neighborhoods may experience common experiences and stressors that can lead to unfavorable biological consequences including heightened cancer risks. Exposure to socioeconomically disadvantaged neighborhoods can lead to poorer health outcomes and a greater chance of death[9, 10], even after controlling for individual-level socioeconomic factors[11]. This suggests that changes to the

neighborhoods themselves, and not necessarily the individuals within those neighborhoods, may be an underutilized yet high-return approach to reducing the population burden of disease outcomes like cancer[12].

Potential mechanisms for explaining this association are complex, but one of the most relevant mechanisms relates to chronic stress[13, 14]. Chronic stress may result from exposure to unfavorable social and economic environments at the neighborhood-level acting in concert with psychosocial factors at the individual-level [15] (i.e., depression[16] and perceived stress[17]) . Thus, chronic stress is related to constant "wearing down" of the body that can affect biological processes, accelerate the rate of decline in physiological functioning, and ultimately affect the body's ability to fight disease processes such as cancer [18-23].

The hypothesis that neighborhood-level characteristics can therefore influence an individual's biological state is supported by numerous multilevel (i.e., accounting for individuals' experiences within environments) conceptual frameworks[15]. For instance, animal models demonstrate that unfavorable macro-level environmental exposures, including social isolation from peers, influence stress-related biological parameters, such as cellular apoptosis and inflammation[24]. Further, human observational studies have shown that unfavorable neighborhood environmental characteristics can affect biological markers related to disease[25]. Telomere length (TL) has emerged as a promising intermediate biological marker along the pathway linking chronic stress and disease pathogenesis. Furthermore, the telomere pathway has been studied as a common molecular signature in field cancerization studies[8, 26], and has been shown to be associated with many cancers on the population level[21].

Telomeres consist of long stretches of (TTAGGG) repeat DNA located at the ends of chromosomes and are designed to protect against DNA degradation[27]. Telomeres naturally shorten with age in all replicating somatic cells [18, 28], but can also shorten prematurely in response to cellular stress[22, 29]. Associations have been reported between shorter blood leukocyte TL attrition and individual-level chronic stress resulting from perceived stress, depression[22, 29], and difficult or stressful life circumstances, like caring for a chronically ill child[26]. Evidence also shows an inverse relationship between shorter LTL and neighborhood socioeconomic status(SES) [15, 25], disadvantage[30, 31] and unfavorable social environment[25], even after adjustment for individual-level factors known to affect telomere length[25] such as biomedical variables (including cardiovascular disease risk factors), lifestyle variables (such as smoking, body mass index, diet, and physical activity) or socioeconomic variables (related to education, income, and employment)[25]. Beyond being a potential biomarker of chronic stress, telomere length is also implicated as a biomarker of aging, cancer, and cardiovascular diseases[18-22, 29]. Thus, understanding the relationship between neighborhood and telomere length could have implications for a range of poor health outcomes [18, 19, 22, 29] [20, 21], and could support the “neighborhood cancerization” hypothesis.

The neighborhood cancerization hypothesis is analogous to the concept of field cancerization in that it considers the biological basis for neighborhood-level effects among groups of commonly exposed individuals from different neighborhoods. We hypothesize that residents of a neighborhood who are (as a group) exposed to stressors or certain neighborhood social or physical characteristics may experience increased cancer risk or unfavorable outcomes via biological responses to these exposures. The goal of this paper is to employ multiple analytic approaches to evaluate how neighborhood-level factors influence telomere length in a diverse, multi-neighborhood

sample, in order to support the hypothesis of neighborhood cancerization. Further, neighborhood factors that are found to be associated with telomere length in this study could be used to create common neighborhood stress signatures, under a shared chronic stress pathway that is currently not well understood.

Methods

Study Population. The study sample included data collected at 3 centers: University of Pennsylvania (Penn), the Ohio State University (OSU), and the University of Texas Medical Branch (UTMB). These 3 centers were originally part of the larger Centers for Population Health and Health Disparities (CPHHD) Initiative[32]. A previous investigation showed that combining data from these 3 centers is a valid approach[33] that increases variation in ethnicity, geography, and neighborhood circumstances. Study participants were recruited between 2004 and 2012. Each center focused on an underserved population from a different geographical area and included protocols for recruitment and data collection that has been described previously[34-36]. Briefly, OSU included Non-Hispanic, White women from rural Appalachia (65% with a high school education or greater) who were not pregnant and without cervical cancer at the time of enrollment[26]. Penn included mostly urban, Non-Hispanic White(89%) and African American(11%) men who were highly educated (72% with greater than high school education) prostate cancer patients from urology clinics within the Penn hospital system[28], and UTMB included a population-based sample of non-Hispanic households and a strata sample of Hispanic households(52% of the study population) in Texas City, TX[27], where 32% of the population had greater than a high school education and 37% had less than a high school education. This geographic and demographic variation allows for comparisons analogous to those in field cancerization that compare molecular alterations in tissue sites within and across organ systems within an individual patient.

Study participants from all centers provided a blood sample, and they completed a standardized questionnaire at the time of study enrollment. Study participants were followed-up for cancer status. Informed consent was received from all participants, and study protocols were approved by the Institutional Review Boards of each center.

Outcome variable. TL was measured from extracted DNA from blood samples. OSU and UTMB samples were processed using the QIAamp DNA Extraction Kit (Valencia, CA). UPHS DNA samples were extracted using Chemagen Magnetic Bead technology (n=57) and phenol-chloroform extraction (n=36). Terminal restriction fragment (TRF) or Southern Blot assays were used to measure TL from extracted DNA on all samples (using duplicate samples), as described previously[37]. The overall coefficient of variation (CV%) was 1.25%, where a CV less than 2% is expected[22, 29]. Mean TRF in kilobases (kb) was determined using Telorun software[37]. TL was reported in kb units.

Neighborhood variables. Census data from the American Community Survey (ACS) were obtained at the census tract level to ascertain sociodemographic neighborhood variables (http://www2.census.gov/acs2009_5yr/summaryfile/). Thus neighborhood is defined here by select census tract social and economic conditions. Census variables were linked to geocoded study data by the Federal Information Processing Standard (**FIPS)** code, which uniquely identifies states, counties, and census tracts[38]. In order to ensure confidentiality, we obtained 5-years of census tract level estimates from the ACS. We used ACS version 2005-2009 since 86% of the study population was accrued between 2004 and 2009. Individuals from the same census tract were assumed to have the same neighborhood characteristics. Our data included 127 unique neighborhood clusters (census tracts) in total. Penn had the most unique number of clusters (n=92), followed by OSU (n=29) and UTMB (n=6).

Variables extracted from the ACS database to represent sociodemographic neighborhood environment were chosen based on literature [28], [39-43] and included population density (overall population/total land area in square miles), urban crowding (housing units/square mile), residential crowding (represented by percent households that have more than one occupant per room), residential stability (percent living in the same house as 1 of year ago) and residential instability (percent who moved their residence within the same State as of 1 year ago). The following variables were used to construct a neighborhood socioeconomic status index (NSES): education (percent of adults over 25 with less than a high school education), employment (percent male unemployment), poverty (percent of households with income below the poverty line, percent of households receiving public assistance, percent of female-headed households with children) and income (median household income) [43]. Briefly, these six variables related to income, education, employment, and poverty were the best representatives of socioeconomic status(SES) in a factor analysis(Cronbach's alpha=0.93)[43]. Each of the 6 variables were summed after being transformed (i.e., higher values corresponded to higher SES). This total score was then standardized to a mean of zero and a standard deviation (sd) of one. Thus, an index score greater than zero denotes a tract with SES above the sample average[43].

Individual-level Covariates. Variables common to all 3 centers and that were found to relate to telomere length in a previous analysis[33] were included as covariates in statistical models: gender (male/female), age at enrollment(continuous); race/ethnicity (White/Non-Hispanic, African American/Non-Hispanic, and Hispanic), educational status (<high school (or less than 12 years of schooling), a high school education (12 years of schooling/GED), or >high school education(>12 years of schooling)), disease status (cancer; yes/no), total perceived stress score dichotomized at the median[44, 45], as

measured from the Perceived Stress Scale [46, 47] (total score range: 1–40); and depression dichotomized at a clinical cutpoint of 16[40] [48, 49], as measured from questions from Center for Epidemiological Studies-Depression (CES-D) scale[50] and the CES-D revised(R) scale[51] (total score range: 0-60).

Statistical Analysis. The distributions of TL and neighborhood variables were examined for non-normality and appropriate data transformations were conducted. Natural log-transformed TL was used as the outcome variable for all analyses. For ease of interpretation, continuous neighborhood variables (with the exception of the NSES index) were scaled by dividing by their standard deviation[25]. We used linear mixed effect models to account for the multilevel nature of the data. This model allowed clustering of individuals within neighborhoods and centers to estimate associations between neighborhood variables and TL before and after adjustment for covariates [25]. Quantile regression was also used to assess associations with neighborhood factors within segments of the TL distribution, accounting for clustering by census tract and confounding by center [52, 53]. Quantile regression coefficients at the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 95^{th} TL percentiles were considered. The coefficients at each TL percentile are interpreted as the change in log-transformed TL, given each unit increase in the neighborhood variable standard deviation. Interactions among covariates were evaluated in stratified analysis and by taking the cross-product terms of each variable in both multilevel linear regression and quantile regression models. Individual-level and neighborhood-level covariates were assessed for multicollinearity before inclusion in statistical models using correlation matrices[54]. Robust standard errors are reported and all tests were two-sided. A p-value<0.05 was considered to be statistically significant. All analyses were performed using Stata 12.1 (StataCorp LP, College Station, TX)[55].

Results

Baseline characteristics of the study sample are presented in Table 1. Of the 1,488 participants, 58.8% were female, 15.7% had a cancer diagnosis, and the median age was 51 years (interquartile range (IQR) 38-63). The sample was comprised of 45.6% non-Hispanic Whites, 45.0% Hispanics, and 9.4% African Americans. The overall sample reported mild levels of stress on the Perceived Stress Scale(median score=19; IQR=17-22) and low levels of depressive symptoms(median score=6, IQR=1-15). The study sample had a relatively low median neighborhood SES index overall(-0.11; IQR=-0.68-0.48), though neighborhood SES was different by region (UPenn=0.80(IQR=0.38-1.11); OSU=-0.35(IQR=-0.61- -0.17); UTMB=-0.10, IQR=-0.68-0.48); data not shown). We report a median population density (total population/total area of land use in square miles) of 3857.3(IQR=1694-5101)). The U.S. reports a population density of 87.5, Galveston, TX reports 1158.2, and Philadelphia county reports 11,379.4. The median percent of households considered to be crowded (i.e., greater than one occupant per room) was 2.6% for the overall study population, which is lower than the national average[56]. The median percent of the population still living in the same house as of 1 year was 86.7%, and the median percent of the population that moved within the same state in the past year was 2.9%.

No statistically significant associations between any neighborhood factor and log-transformed TL in multilevel linear regression models. These findings did not change when adjusting for covariates and psychosocial factors (Table 2). In quantile regression models, significant associations were seen between log-transformed TL and population density and urban crowding at lower tails of the TL distribution (the 5^{th}, 50^{th}, and 75^{th} percentiles), and between residential crowding and TL at the 50^{th} percentile (p-value=0.03) (Table 3). For both population density and urban crowding, magnitudes of

effect were small, but twice as large at the 5th percentile(0.10 for population density; 0.11 for urban crowding) than the 75th percentile of the log-transformed TL distribution(0.05 for population density; 0.04 for urban crowding). For neighborhoods where residents remained in the same house in the past year, there was a significant, positive association between TL and residential stability at the highest levels of the TL distribution (95th percentile-p-value=0.006; 90th percentile-p-value=0.005). For neighborhoods where residents moved within the same state in the past year, there was a significant, inverse relationship between TL and moving within the same state in the past year at the 90th (p-value=0.02) and 95th percentiles (p<0.001).

Discussion

This is the second study in adults to evaluate the relationship between TL and neighborhood, and the first to adjust associations by individual-level psychosocial variables. We report that neighborhood sociodemographic stressors have a complex, possibly non-linear relationship with TL, a biological marker used in the study of age-related disease, cellular and psychosocial stress, and field cancerization [8, 26, 57, 58]. We found that unfavorable neighborhood characteristics, namely urban crowding and population density, were significantly related to shorter TL. This finding is consistent with reports that shorter TLs are associated with unfavorable neighborhood characteristics [25, 30, 31]. Residential stability (remaining in the same house in the past year) and instability (moving within the same state in the past year) were more strongly associated with longer TL at the upper tail of the TL distribution. These findings are consistent with hypotheses that longer TL is associated with more favorable neighborhood circumstances and less chronic stress[30, 31] [25]. Low levels of residential stability are likely to affect groups of individuals differently, depending on their social position or cultural resources[59]. Both impoverished and flourishing

neighborhoods can have a high level of residential stability, yet impoverished neighborhoods often have poorer health[59]. When we adjusted quantile regression models for neighborhood SES, associations of TL with urban crowding and residential stability remained unchanged (data not shown). Further, these relationships persisted even after adjustment for individual-level psychosocial stressors.

Our findings are similar to results from a previous study of neighborhood and TL in adults, using different neighborhood variables and analytic approaches[25]. The previous study investigated the relationship between TL and composite scores of neighborhood SES (from Year 2000 U.S. Census variables) and self-reported, neighborhood social stressors (e.g., social cohesion, aesthetics, and safety)[25]. This study modeled mean TL so that associations at the upper and lower tails of the TL distribution were not distinguished. Given that the lower tail of the TL distribution is of particular interest in disease susceptibility, we examined the association between telomere length and neighborhood SES and socio-demographic environment using quantile regression, an approach that could provide additional insights into neighborhood associations given its focus on the extremes of the telomere length distribution[52, 53].

We examined associations using a composite score of neighborhood SES and neighborhood variables from the U.S. Census related to population density, urban and residential crowding and stability. We chose these variables to represent the sociodemographic stressors because they provide insight into the social norms of a neighborhood, as well as general insights into the neighborhood physical landscape[59, 60]. Additionally, these variables have shown associations with psychosocial stress[59, 60] and premature death[61] and have been considered surrogates for self-reported measures of neighborhood stress investigated in previous TL studies[25]. More specifically, urban and residential crowding are related to increases in social stress[61,

62], can negatively affect family and social relationships[63], and impact social cohesion[59, 60, 62]. Residential stability is related to neighborhood safety (i.e., safer neighborhoods are related to increases in residential stability)[64], and residential stability can affect social cohesion[59, 60]. This is because when a number of well-established residents or families leave a neighborhood, there can be a destabilization of social norms and a disruption of social networks[59, 60]. Thus, our findings show that while U.S. Census variables are generally considered less specific measures of neighborhood phenomena than self-report data, they are readily accessible, reasonable measures of numerous social and environmental phenomena, and can be used to identify potential biologic effects of neighborhood environment and justify more in-depth neighborhood investigations.

Our findings also suggest that when considering the relationship between complex exposures such as neighborhood characteristics and biological variables such as TL, novel statistical modeling tools may be required to obtain relevant insights into the relationship between neighborhood factors and TL. While no statistically significant associations were observed using multilevel, linear regression models, relatively constant associations were observed between shorter TL and population density and urban crowding using quantile regression. Quantile regression allows for the study of predictors across the entire TL distribution, without having to categorize a continuous outcome variable (with concomitant reduction in statistical power), and is better able to evaluate effects at the extremes of a distribution [65]. Our results are consistent with reports that shorter TLs are associated with increases in urban crowding or population density [25, 30, 31]. Our findings demonstrate that using an approach like quantile regression may identify associations that are otherwise missed by modeling simple linear relationships or focusing on the mean of an outcome.

Our analysis supports the concept of neighborhood cancerization. Unfavorable neighborhood circumstances were related to TL, a biomarker indicated in disease development[66], assuming a chronic stress pathway model that extends from macro-environment (neighborhood) exposures to effects occurring at the cellular level[15]. Like field cancerization, the present analysis identified potential common neighborhood characteristics or signatures across different neighborhoods that were associated with biological changes that have been implicated in carcinogenesis[2]. Longitudinal follow-up of TL and changes in residential stability, neighborhood gentrification, and in-out migration of neighborhood residents and families over time could shed light into our findings[59] and could provide further evidence of the neighborhood cancerization effect[67].

More recently, field cancerization theory has been extended to include etiologic field effects that focus not just on complex molecular changes and interactions within a cell, but interactions of the whole host organism with external stimuli, a concept known as the interactome[67]. The interactome includes gene-environment interaction studies[67], although these studies have been almost entirely limited to traditional, often geo-atmospheric measures of environment and much less so the social and population measures of neighborhood environments used here. The neighborhood cancerization effect proposed here extends the etiologic field effect hypothesis and provides some insights into potential gene-neighborhood interactions. Although this has not been tested, it is possible that individuals at the lower tails of the TL distribution in this study(for instance, the 5^{th} and 10^{th} percentiles) could be genetically predisposed to having shorter TL, and those in the upper tails of the TL distribution (the 90^{th} and 95^{th} percentiles) could be genetically predisposed to having longer TL. Thus, based on our findings, neighborhood effects related to residential stability may only be relevant in

those who have inherently longer TL. This suggests that studying the relationship between gene-neighborhood interactions, in the context of other relevant individual and neighborhood-level social factors, is warranted and supports the etiologic field effect concept[67].

Our study had some limitations. This was a cross-sectional study in three U.S. regions, each with its own ascertainment strategy, and it is not a nationally representative sample. This sampling strategy could affect generalizability of results. However, we included data from multiple geographic regions, ethnicities, and disease states to include maximal variation of factors that may influence TL. There is not a standard, agreed upon approach to defining neighborhoods, and while the utility of pre-defined boundaries to define neighborhoods, such as census tract, has been questioned, it is a commonly used approach and has the benefit of allowing for standardized assessments of neighborhoods with readily available data[68, 69]. Future studies could consider so-called boundary-free geographic methods to measure environments that are more complex yet are an improvement upon more commonly polygon-based methods[70]. Despite these few limitations, our findings were similar to a population-based study with a comparable demographic composition[25].

Our findings demonstrate that neighborhood factors exert effects on biology, even after adjustment for psychosocial stressors at the individual-level. The results of this study provide evidence to support the hypotheses that neighborhood circumstances can have biologic consequences, under a concept termed, neighborhood cancerization. TL may therefore be a biomarker of the biological influences of neighborhood circumstances on human health and disease. We conclude that neighborhood-level factors may contribute to TL, chronic stress and disease development[40, 71], but that the relationship of neighborhood on TL is complex.

Table 1. Study Characteristics including neighborhood and individual-level psychosocial factors.

Population Characteristics	
Total Population (n)	1488
Median Telomere length (kb)[a]	6.4 (4.3-8.3)
Median Age[a]	51 (38-63)
Male Gender (%)	41.2
Cancer diagnosis(%)	15.7
Race (%)	
Non-Hispanic White	45.6
African American	9.4
Hispanic	45.0
Education(%)	
> High School	37.0
High School/GED	29.8
<High School	33.2
Median Total Perceived Stress[b]	19 (17-22)
Median Depressive Symptoms[b]	6(1-15)
Neighborhood Factors	
Median Neighborhood SES Index	-0.11(-0.68-0.48)
Median Population Density (total population/sq. mile)	3857.3(1694-5101)
Median Urban Crowding (housing units/sq. mile)	1755.8(695-1833)
Residential Crowding (%)	
Median % households that ARE crowded	2.6 (0.6-11)
Residential Stability(%)	
Median % living in the same house as 1 year ago	86.7(78-91)
Median % moved within the same state(not county)	2.9 (1.2-2.9)

[a] Medians(interquartile range for the median);
[b] Range from low to high for Depression (0-60) and Perceived Stress Scale (0-40)

Table 2. Estimates and standard errors(SE) for associations between log-transformed telomere length (kb) and neighborhood factors(scaled by standard deviation(SD), before and after adjustment for covariates (n=1488).

	Neighborhood Factors scaled by SD			
	Model 1 (Estimate, SE)	p-value	**Model 2** (Estimate, SE)	p-value
Increasing Neighborhood SES Index	-0.01, 0.04	0.90	-0.01, 0.04	0.89
Increasing Population Density (sq. meters)	-0.02, 0.02	0.32	-0.03, 0.02	0.20
Increasing Urban Crowding (housing units/sq. mile)	-0.02, 0.02	0.36	-0.02, 0.02	0.23
Increasing Residential Crowding				
% Crowded households	-0.05, 0.07	0.52	-0.05, 0.08	0.48
Increasing Residential Stability(%)				
%Same house in past year	0.06, 0.05	0.18	0.05, 0.05	0.23
Increasing Residential Instability (%)				
%Moved within same state in past yr	-0.01, 0.04	0.89	-0.001, 0.04	0.98

Model 1 is the crude analysis without covariates. Model 2 includes adjustment for age, gender, cancer status, race/ethnicity, perceived stress(high/low), depression(high/low), education level, and the interactions of gender and cancer status, and race/ethnicity and educational level.

Table 3. Associations with Neighborhood factors (standard deviation adjusted) across the log-transformed telomere length distribution(kb).

	Log-Transformed Telomere Length Distribution					
	5th Percentile	10th Percentile	50th Percentile	75th Percentile	90th Percentile	95th Percentile
Neighborhood Factor	Coeff(SE) p-value	Coeff(SE) p-value	Coeff(SE) p-value	Coeff(SE) p-value	Coeff(SE) p-value	Coeff(SE) p-value
Increasing Neighborhood SES	0.003(0.02) 0.86	0.01(0.03) 0.58	-0.004(0.03) 0.85	0.01(0.02) 0.71	-0.02(0.03) 0.51	0.01(0.03) 0.72
Increasing Population Density (population/sq. mile)	-0.10(0.05) 0.03*	-0.07(0.04) 0.11	-0.05(0.01) <0.001*	-0.05(0.01) <0.001*	-0.05(0.04) 0.27	-0.01(0.02) 0.74
Increasing Urban Crowding (units/sq. mile)	-0.11(0.03) 0.002*	-0.08(0.03) 0.02*	-0.04(0.01) <0.001*	-0.05(0.01) <0.001*	-0.04(0.04) 0.28	-0.02(0.02) 0.45
Increasing Residential Crowding						
% crowded households	-0.01(0.02) 0.70	-0.02(0.01) 0.27	-0.001(0.001) 0.03*	-0.01(0.02) 0.78	0.03(0.03) 0.28	0.002(0.001) 0.09
Increasing Residential Stability(%)						
% in the same house in past year	0.03(0.04) 0.56	0.02(0.04) 0.64	0.03(0.03) 0.27	0.04(0.02) 0.12	0.04(0.02) 0.005*	0.04(0.01) 0.006*
Increasing Residential Instability(%)						
% Moved within same state in past yr	0.03(0.03) 0.32	0.03(0.02) 0.19	-0.01(0.01) 0.70	-0.01(0.01) 0.25	-0.03(0.01) 0.02*	-0.03(0.01) <0.001*

Model includes adjustment for age, gender, cancer status, race/ethnicity, perceived stress(high/low), depression(high/low), center, educational level, and the interactions of gender and cancer status, and race/ethnicity and educational level.

*p-value less than 0.05

References:

1. SLAUGHTER, D.P., H.W. SOUTHWICK, and W. SMEJKAL, *Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin.* Cancer, 1953. **6**(5): p. 963-8.
2. Lochhead, P., et al., *Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression.* Mod Pathol, 2015. **28**(1): p. 14-29.
3. Giovannucci, E. and S. Ogino, *DNA methylation, field effects, and colorectal cancer.* J Natl Cancer Inst, 2005. **97**(18): p. 1317-9.
4. Ushijima, T., *Epigenetic field for cancerization.* J Biochem Mol Biol, 2007. **40**(2): p. 142-50.
5. Ramachandran, K. and R. Singal, *DNA methylation and field cancerization.* Epigenomics, 2012. **4**(3): p. 243-5.
6. Stearman, R.S., et al., *A macrophage gene expression signature defines a field effect in the lung tumor microenvironment.* Cancer Res, 2008. **68**(1): p. 34-43.
7. Martinez-Outschoorn, U.E., et al., *Oxidative stress in cancer associated fibroblasts drives tumor-stroma co-evolution: A new paradigm for understanding tumor metabolism, the field effect and genomic instability in cancer cells.* Cell Cycle, 2010. **9**(16): p. 3256-76.
8. Heaphy, C.M., et al., *Telomere DNA content and allelic imbalance demonstrate field cancerization in histologically normal tissue adjacent to breast tumors.* Int J Cancer, 2006. **119**(1): p. 108-16.
9. Adler, N.E., Ostrove, J.M. , *Socioeconomic status and health: What we know and what we don't.* Ann N Y Acad Sci, 1999. **896**: p. 3-15.
10. Diez Roux, A.V., Mair, C., *Neighborhoods and health.* Ann N Y Acad Sci, 2010. **1186**: p. 125-145.
11. Merkin, S.S., et al., *Neighborhoods and Cumulative Biological Risk Profiles by Race/Ethnicity in a National Sample of U.S. Adults: NHANES III.* Annals of epidemiology, 2009. **19**(3): p. 194-201.
12. Branas, C.C. and J.M. MacDonald, *A Simple Strategy to Transform Health, All Over the Place.* Journal of public health management and practice : JPHMP, 2014. **20**(2): p. 157-159.
13. Geronimus, A.T., Hicken, M., Keene, D., Bound, J., *"Weathering" and Age Patterns of Allostatic Load Scores Among Blacks and Whites in the United States.* Am J Public Health, 2006. **96**: p. 826-833.
14. McEwen, B., *Protective and Damaging effects of stress Mediators.* N Engl J Med, 1998. **338**: p. 171-179.
15. Lynch, S.M. and T.R. Rebbeck, *Bridging the Gap between Biologic, Individual, and Macroenvironmental Factors in Cancer: A Multilevel Approach.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(4): p. 485-495.
16. Wolkowitz, O.M., Mellon, S.H., Epel, E.S., Lin, J., Dhabhar, F.S., Su, Y, et al., *Leukocyte Telomere Length in Major Depression: Correlations with Chronicity, Inflammation and Oxidative Stress - Preliminary Findings.* PLoS ONE, 2011. **6**(3): p. e17837.
17. Parks, C.G., et al., *Telomere Length, Current Perceived Stress, and Urinary Stress Hormones in Women.* Cancer Epidemiology Biomarkers & Prevention, 2009. **18**(2): p. 551-560.
18. Londoño-Vallejo, J.A., *Telomere instability and cancer.* Biochimie, 2008. **90**(1): p. 73-82.
19. Koorstra, J.B.M., et al., *Pancreatic Carcinogenesis.* Pancreatology, 2008. **8**(2): p. 110-125.

20. Cunningham, J.M., et al., *Telomere Length Varies By DNA Extraction Method: Implications for Epidemiologic Research.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(11): p. 2047-2054.
21. Wentzensen, I.M., et al., *The Association of Telomere Length and Cancer: a Meta-analysis.* Cancer Epidemiology Biomarkers & Prevention, 2011. **20**(6): p. 1238-1250.
22. Sanders, J.L. and A.B. Newman, *Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither?* Epidemiologic Reviews, 2013. **35**(1): p. 112-131.
23. Butt, H.Z., et al., *Telomere Length Dynamics in Vascular Disease: A Review.* European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery, 2010. **40**(1): p. 17-26.
24. McClintock, M.K., et al., *Mammary cancer and social interactions: identifying multiple environments that regulate gene expression throughout the life span.* J Gerontol B Psychol Sci Soc Sci, 2005. **60 Spec No 1**: p. 32-41.
25. Needham, B.L., et al., *Neighborhood characteristics and leukocyte telomere length: The Multi-Ethnic Study of Atherosclerosis.* Health & Place, 2014. **28**(0): p. 167-172.
26. Epel, E.S., et al., *Accelerated telomere shortening in response to life stress.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(49): p. 17312-17315.
27. Cheung, A.L.M. and W. Deng, *Telomere dysfunction, genome instability and cancer.* Frontiers in Bioscience, 2008. **13**(6): p. 2075-2090.
28. Blackburn, E.H., *Telomere states and cell fates.* Nature, 2000. **408**(6808): p. 53-56.
29. Starkweather, A.R., et al., *An Integrative Review of Factors Associated with Telomere Length and Implications for Biobehavioral Research.* Nursing Research, 2014. **63**(1): p. 36-50 10.1097/NNR.0000000000000009.
30. Mitchell, C., et al., *Social disadvantage, genetic sensitivity, and children's telomere length.* Proceedings of the National Academy of Sciences, 2014. **111**(16): p. 5944-5949.
31. Theall, K.P., et al., *Neighborhood disorder and telomeres: Connecting children's exposure to community level stress and cellular response.* Social Science & Medicine, 2013. **85**(0): p. 50-58.
32. Health, N.I.o. *NIH Centers for Population Health and Health Disparities*. 2014 [cited 2014 September 25, 2014]; Available from: http://cancercontrol.cancer.gov/populationhealthcenters/cphhd/.
33. Lynch, S.M., Peek, M.K., et al., *Race/ethnicity, Psychosocial Factors, and Telomere Length in a Multicenter Setting.* Unpublished, submitted manuscript, 2015.
34. Paskett, E.D., et al., *Psychosocial predictors of adherence to risk-appropriate cervical cancer screening guidelines: A cross sectional study of women in Ohio Appalachia participating in the Community Awareness Resources and Education (CARE) project.* Preventive Medicine, 2010. **50**(1–2): p. 74-80.
35. Peek, M.K., et al., *Allostatic Load Among Non-Hispanic Whites, Non-Hispanic Blacks, and People of Mexican Origin: Effects of Ethnicity, Nativity, and Acculturation.* American Journal of Public Health, 2010. **100**(5): p. 940-946.
36. Rebbeck, T.R., et al., *Joint effects of inflammation and androgen metabolism on prostate cancer severity.* International Journal of Cancer, 2008. **123**(6): p. 1385-1389.
37. Kimura, M., et al., *Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths.* Nat. Protocols, 2010. **5**(9): p. 1596-1607.
38. Census, U.S. 2014; Available from: http://quickfacts.census.gov/qfd/meta/long_fips.htm.

39. Krieger, N., Williams, D., Moss, N., *Measuring social class in US public health research: concepts, methodologies, and guidelines.* Annu Rev Public Health, 1997. **18**: p. 341-378.
40. Messer, L., . Laraia, B., Kaufman, J., Eyster, J., Holzman, C., Culhane, J., et al. , *The development of a standard neighborhood deprivation index.* . Journal of Urban Health, 2006. **83**(6): p. 1041-1062.
41. Caughy, M., Hayslett-McCall, K., O'Campo, P. , *No neighborhood is an island: incorporating distalneighborhood effects into multilevel studies of child development competence.* Health and Place, 2007. **13**: p. 778-798.
42. Krieger, N., Chen, J., Waterman, P., Rehkopf, D., Subramanian, S. , *Race/ethnicity, gender, and monitoring socioeconomic gradients in health: A comparison of area-based socioeconomic measures – the public health disparities geocoding project.* Am J Public Health, 2003. **93**: p. 1655-71.
43. Dubowitz, T., Heron, M., Bird, C.E., Lurie, N., Finch, B.K., Basurto-Dávila, R., Hale, L., Escarce, J.J. , *Neighborhood socioeconomic status and fruit and vegetable intake among whites, blacks, and Mexican Americans in the United States.* Am J Clin Nutr, 2008. **87**(6): p. 1883-1891.
44. Wang, L., et al., *The Effects of Perceived Stress and Life Style Leading to Breast Cancer.* Women & Health, 2012. **53**(1): p. 20-40.
45. Shell, A.M., M.K. Peek, and K. Eschbach, *Neighborhood Hispanic composition and depressive symptoms among Mexican-descent residents of Texas City, Texas.* Social Science & Medicine, 2013. **99**(0): p. 56-63.
46. Krieger, N., Smith, K., Naishadham, D., Harman, C., Barbeau, E.M., *Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health.* Social Science and Medicine, 2005. **61**: p. 1576-1596.
47. Cohen, S., Kessler, R.C., Underwood, L., *Perceived stress scale. Measuring stress: A guide for health and social scientists.* Oxford University Press, 1994. **New York**.
48. Crockett, L.J., Randall, B.., Shen, Y.L., Russell, S.T., Driscoll, A.K., *Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: a national study*

Journal of Consulting and Clinical Psychology, 2005. **73**(1): p. 47-58.
49. Flores, E., Tschann, J., Dimas, J., Bachen, E., Pasch, L., De Groat, C., *Perceived discrimination, perceived stress, and mental and physical health among Mexican-origin adults.* Hispanic Journal of Behavioral Sciences, 2008. **30**(4): p. 401-424.
50. Radloff, L., *The CES-D scale: a self-report depression scale for research in the general population.* Applied Psychological Measurement, 1977. **1**: p. 385-401.
51. Eaton, W.W., Smith, C., Ybarra, M., Muntaner, C., Tien, A., *Center for Epidemiologic Studies Depression Scale: review and revision (CESD and CESD-R).* In ME Maruish (Ed.). The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, 2004. **3rd Edition**(Volume 3: Instruments for Adults): p. 363-377.
52. Hao, L., Naimen, D.Q., *Quantile Regression*. Quantitative Applications in the Social Sciences, ed. T.F. Liao2007, Thousand Oaks: SAGE Publications.
53. Beyerlein, A., et al., *Genetic markers of obesity risk: stronger associations with body composition in overweight compared to normal-weight children.* PLoS One, 2011. **6**(4): p. e19057.
54. Kumar, T.K., *Multicollinearity in Regression Analysis.* The Review of Economics and Statistics, 1975. **57**(3): p. 365-366.

55. Stata. *Stata 12 help for sqreg*. 2012 [cited 2012 August 22nd 2012]; Available from: http://www.stata.com/help.cgi?sqreg.
56. *http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF*. Population, Housing Units, Area, and Density: 2010 - United States -- States; and Puerto Rico 2010 Census Summary File 1 2012 [cited 2015; Available from: http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF.
57. Kananen, L., Surakka, I., Pirkola, S., Suvisaari, J., et al., *Childhood adversities are associated with shorter telomere length at adult age both in individuals with an anxiety disorder and controls.* PLoS One, 2010. **5**(e): p. 10826.
58. Tyrka, L.H., Price, H.T., Kao, B. et al, *Childhood maltreatment and telomere shortening: preliminary support for an effect of early stress on cellular aging.* Biol. Psychiatry, 2010. **67**: p. 531-534.
59. Howden-Chapman, P., *Housing standards: a glossary of housing and health.* J Epidemiol Community Health, 2004. **58**: p. 162-168.
60. Entner Wright, B.R., Caspi, A., Moffit, T.E., et al., *Factors associated with doubled-up housing--a common precursor to homelessness.* Social Service Review, 1998. **72**: p. 92-111.
61. Kellet, J.M., *Crowding and Mortality in London Boroughs.* In: Burridge, R., Ormandy, D., eds., 1993. **Unhealthy housing**(London: Chapman and Hall): p. 209-222.
62. Gabe, J., Williams, P., *Women, crowding, and mental health.* In: Burridge, R., Ormandy, D., eds., 1993. **Unhealthy Housing**(London: Chapman and Hall): p. 191-208.
63. Clark, W.A.V., Deurloo, M.C., Dieleman, F.M. , *Housing consumption and residential crowding in US housing markets.* Journal of Urban Affairs, 2000. **22**(1): p. 49-63.
64. Boggess, L. and J. Hipp, *Violent Crime, Residential Instability and Mobility: Does the Relationship Differ in Minority Neighborhoods?* Journal of Quantitative Criminology, 2010. **26**(3): p. 351-370.
65. Altman, D.G. and P. Royston, *The cost of dichotomising continuous variables.* BMJ, 2006. **332**(7549): p. 1080.
66. Lynch, S.M., et al., *A prospective analysis of telomere length and pancreatic cancer in the alpha-tocopherol beta-carotene cancer (ATBC) prevention study.* International Journal of Cancer, 2013. **133**(11): p. 2672-2680.
67. Lochhead, P., Chan, A.T., Nishihara, R., Fuchs, C.S., Beck, A.H., Giovannucci, E.G., Ogino, S., *Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression.* Modern Pathology, 2015. **28**: p. 14-29.
68. Dietz, R.D., *The estimation of neighborhood effects in the social sciences: An interdisciplinary approach.* Social Science Research, 2002. **31**: p. 539-575.
69. Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., *Assessing "Neighborhood Effects": Social Processes and New Directions in Research.* Ann Rev Sociol, 2002. **28**: p. 443-478.
70. Branas, C.C., et al., *A Difference-in-Differences Analysis of Health, Safety, and Greening Vacant Urban Space.* American Journal of Epidemiology, 2011.
71. Diez Roux, A.V. and C. Mair, *Neighborhoods and health.* Annals of the New York Academy of Sciences, 2010. **1186**(1): p. 125-145.

CHAPTER 4

A NEIGHBORHOOD-WIDE ASSOCIATION STUDY (NWAS): IDENTIFICATION OF NEIGHBORHOOD CANCERIZATION SIGNATURES IN PROSTATE CANCER

Shannon M. Lynch[1], Nandita Mitra[1], Michelle Ross[1], Craig Newcomb[1], Karl Dailey[1], Tara Jackson[1], Charnita M. Zeigler-Johnson[2], Harold Riethman[2], Charles Branas[1], Timothy R. Rebbeck[1].

[1] University of Pennsylvania, Perelman School of Medicine, PA 19104

[2] Thomas Jefferson University, Philadelphia, PA

[3]Wistar Institute, Philadelphia, PA 19104

This research was supported by grants from the Public Health Service (P50-CA105641, P60-NM006900 and R01-CA85074 to TRR and F31-AG039986 to SML)

Corresponding Author:

Shannon M. Lynch
235 Blockley Hall
University of Pennsylvania
423 Guardian Drive
Philadelphia, PA 19104
Email: lynchs@mail.med.upenn.edu
Tel: 215-898-1793

Abstract

Background: Cancer is accepted to be the result of complex interactions of multiple variables measured at biological, individual and neighborhood environmental levels. However, systematic approaches to assess neighborhood-level effects are limited. We propose a Neighborhood-Wide Association Study (NWAS), a systematic approach analogous to a genome-wide association study (GWAS), in order to identify neighborhood-level signatures associated with aggressive prostate cancer.

Methods: We empirically evaluated the association between all Year 2000 U.S. Census variables and prostate cancer aggressiveness among White prostate cancer cases reported to the Pennsylvania State Cancer Registry using a multi-phase approach that accounted for age, year of diagnosis, spatial effects, and multiple comparisons. The outcome of interest was aggressive (Stage≥3 and Gleason grade ≥7) vs. non-aggressive (Stage<3 or Gleason grade<7) cancer. Using generalized estimating equations (Phase 1) and Bayesian statistics (Phase 2), we calculated odds ratios (OR) and credible intervals (CI). In Phase 3, principal components analysis was used to account for correlation among variables.

Results: From 14,663 census variables, we identified the top 17 variables associated with prostate cancer aggressiveness. The top two hits related to transportation (OR=1.05; CI=1.001-1.09) and poverty (OR=1.07; CI=1.01-1.12). Our findings also confirm previous associations between poverty, income, housing, employment, immigration, and cancer.

Discussion: This NWAS methodology addresses gaps in neighborhood research by introducing a standardized evaluation of a myriad of complex neighborhood factors on a disease outcome. This approach has implications for health disparities research, and

provides a foundation for multidisciplinary, multilevel research by proposing a common methodological framework for identifying relevant neighborhood variables.

Introduction

Numerous conceptual frameworks suggest that cancer results from a complex interaction of factors at the macro-environmental (e.g. neighborhood), individual, and biologic levels[1, 2]. However, novel approaches that could be used to evaluate the joint effect of these multiple levels have not kept pace with other fields, in which high-dimensional computing approaches have been used to discover etiologic agents. For example, genome-wide association studies (GWAS) have driven population-based cancer research in recent years [3-6]. GWAS studies use high-throughput, low cost technology and readily available genome-mapping data to evaluate the role of millions of genetic markers with a variety of diseases and traits using a unbiased, model-free framework [7, 8].

Applying methods borrowed from GWAS, environmental-wide association studies (EWAS) were subsequently developed to study the effect of exposures at the individual level (e.g., pesticides, cigarette smoking, plastics, air pollution, etc.), and to provide insight for subsequent gene-environment interaction studies[9]. However, EWAS methods have not been applied to cancer outcomes, and the impact of social environment, particularly at the neighborhood level has not been comprehensively studied using this approach. Thus, borrowing concepts from GWAS and EWAS, we propose a novel, empirical approach known as a neighborhood-wide association study (NWAS) to evaluate the effect of multiple neighborhood-level environmental exposures on disease etiology and outcomes. The overarching goal of this method is to

systematically identify “neighborhood signatures” that may be related to disease phenotypes.

In proposing this new method, we also borrow from the pathobiologic concept of field cancerization, which refers to the propensity of a field of tissue to become malignant based on either pathological observations or common molecular markers[10, 11]. Common molecular signatures that provide evidence of the cancer field effect include changes in epigenetics[12-14], the tissue microenvironment[15, 16] and telomere function[17]. In theory, field cancerization can occur from simultaneous, wide-spread, and unfavorable exposures acting on cells and tissues within organs and organ systems (such as the effect of cigarette smoking on the lung), as well as the migration of “patches” of molecularly altered cells into larger tissue fields or territories that become predisposed to cancer development[11].

Neighborhood circumstances and lived conditions also contribute to geospatial effects on populations [18, 19]. Under the proposed “neighborhood cancerization” concept, we hypothesize that residents of a neighborhood who are (as a group) exposed to certain neighborhood characteristics may experience a similar biological or social effect that may relate to their cancer risk as a group, and that common unfavorable neighborhood circumstances identified across different neighborhoods can contribute to tumors of similar etiologic origins[19].

Compared to other cancers, prostate cancer is disproportionally affected by social circumstances. When compared to European American (EA) men, African American (AA) men are more likely to receive differential treatments for prostate cancer, and are twice as likely to die of prostate cancer[20]. This is the largest disparity seen in any cancer site. Studies of neighborhood and prostate cancer show that neighborhoods with poor socioeconomic (SES) circumstances, as measured by single, *a priori* selected U.S. census variables and SES deprivation scores derived from combinations of US census

variables, are related to high-grade prostate cancer[21-24]. These effects are apparent independent of individual level exposures[25-28]. Thus, the study of neighborhood effects on prostate cancer risk is warranted, particularly given that few prostate cancer risk factors at the individual level have been identified [29].

We hypothesize that signatures of factors that influence disease risk and severity may impact groups of residents in neighborhoods, and that these effects can be efficiently identified by using an NWAS approach. In this paper, we introduce the NWAS methodology, and demonstrate how agnostic, high-dimensional data analyses can be used to identify neighborhoods at risk for high grade/high stage, aggressive prostate cancer.

Methods

Study Population. Anonymized data from the Pennsylvania Department of Health Prostate Cancer Registry was provided on prostate cancer patients diagnosed in the Commonwealth of Pennsylvania from 1995 to 2005. Residential addresses of prostate cancer patients were cleaned by trained research staff and geocoded at the census tract level by using Arc GIS software. The registry included variables related to prostate cancer diagnosis (tumor stage and grade), age at diagnosis, year of diagnosis, and race/ethnicity. We focused only on Caucasian prostate cancer cases in this analysis (n=80,575), and excluded those who had only a P.O. Box address (n=112), and those missing tumor grade or stage (n=3371), age (n=2), or year of diagnosis (n=4). A total of 77,086 men were included in the final analysis (Appendix D Supplementary Methods-Phase 0 Data Cleaning).

Neighborhood Variables. All 24,634 census tract variables available in the 2000 U.S. Census Summary File 1 (SF1) and Summary File 3 (SF3) were downloaded from Social Explorer (http://www.socialexplorer.com). Year 2000 U.S. Census data was used

since it served as the midpoint for the prostate cancer registry data from 1995-2005. Social explorer automatically calculates percentages, aggregates, and medians for each variable in the U.S. Census tables. The SF1 form, referred to as the "100%" data, is distributed to every household in the U.S. and asks questions about each person within the housing unit related to demographic information, such as age, gender, and race, as well as general housing questions related to occupancy, tenure, etc[30]. The SF3 form is distributed to 5% of all people and housing units in the U.S and asks more specific questions related to socioeconomic status and physical environment characteristics, such as migration, language ability, disability, veterans status, vehicle availability, kitchen and plumbing facilities, etc. [31] . Both SF1 and SF3 variables are used frequently in social science investigations[24-27, 32-35]. All SF1 and SF3 variables were evaluated for missingness (Appendix D Supplementary methods/ Appendices E and F and Supplementary Digital Files 1-4). Variables with greater than 10% missingness (n=8,092) and modal values that comprised over 95% of the data (n=1,879) were excluded. After these exclusions, 14,663 census variables were included in the analysis.

Data Join of Study Population and Neighborhood Variables. Microsoft Visual Studio 2008 was used to execute the join by importing both the census tract level cancer registry and the SF1 and SF3 census data into a program where the geographic identifier was set to the census tract FIPS code[30]. Individual prostate cancer cases were linked to the neighborhood variable values of the census tract in which they live. Thus, cases arising from the same census tract were assumed to have the same neighborhood characteristics.

Outcome Definition. All incident, Caucasian prostate cancer cases occurring in PA from 1995-2005 were included in this study. Incident prostate cancer cases were identified according to ICD-0-3 site and morphology coding. We assumed complete

case ascertainment, given that medical facilities are required by law to report all diagnosed cases of prostate cancer in PA [36]. Tumor stage is said to measure screening effects, whereas tumor grade is said to measure the biologic composition of the tumor[37], thus we created a combined, "prostate cancer aggressiveness" variable for our primary outcome[38, 39] that was defined by cases with a high tumor stage (stage 3 or 4) and high tumor grade (grade 7+), compared to controls with other combinations of these two variables[38, 39]. Tumor stage and grade were determined by Surveillance, Epidemiology, and End Results (SEER) coding criteria for stage and histology variables, respectively[40, 41]. Subjects with low stage prostate cancer (Stage 1 and 2) were defined by SEER Stages 0 and 1; subjects with high-stage (Stages 3 and 4) were defined by SEER Stages $\geq$2[38, 40]. Low tumor grade (Gleason score of 6 or below) and high-grade (Gleason score of 7 or greater) prostate cancer was determined from the SEER 6th digit coding for histology and differentiation[42]. Sixth digit diagnosis codes that were equal to 6 or 9 were excluded because grade or differentiation could be not determined, was not stated, or was not applicable[42]. Subjects who were missing grade variables also were excluded from our outcome definition (n=3,371) (Appendix D: Supplementary Methods-Phase 0). After these exclusions, we defined two comparison groups for analysis: aggressive prostate cancer cases (n=6,416) and non-aggressive prostate cancer cases (70,670). On average, there were 2 cases of aggressive prostate cancer compared to 23 non-aggressive prostate cancer cases in each census tract.

Statistical Analysis. Data reduction techniques were applied across 3 analytical phases to identify continuous neighborhood variables associated with aggressive prostate cancer. Each phase included progressively more stringent statistical criteria, in order to minimize false positive findings. All models were adjusted for age at diagnosis and year of diagnosis.

In *Phase 1*, we used Generalized Estimating Equations (GEE) with a logit link function, binomial distributions, and robust standard error methods to estimate odds ratios (OR) that describe the relationship between census variables and disease aggressiveness[43]:

$$Logit(p) = \alpha + \beta_0 x_{age} + \beta_1 x_{year\ of\ diagnosis} + \beta_2 x_{neighborhood\ variable\ (i,\ j)} + Corr + Error; \quad (Eq.\ 1),$$

where i= individual prostate cancer cases; j=census tracts and Corraccounts for clustering effects within an exchangeable correlation matrix

P-values were Bonferroni-corrected to an alpha of 0.05 to account for multiple comparisons[44], thus Bonferroni-adjusted p-values less than 0.05 were significant in this analysis. Phase 1 analyses were conducted using SAS 12.0 statistical software.

In *Phase 2*, we used spatial regression to further evaluate those variables that reached statistical significance in Phase 1 and to account for and describe the spatial variability in our data. We specified a Bayesian random effects model in which we allow both global and local smoothing:

$$Logit(p) = \alpha + \beta_0 x_{age} + \beta_1 x_{year\ of\ diagnosis} + \beta_2 x_{neighborhood\ variable\ (i,\ j)} + V_{(j)} + U_{(j)} \quad (Eq.\ 2)$$

where i= individual prostate cancer cases; j=county and V(j) are independent non-spatial random effects and U(j) are spatial random effects. We model the spatial random effects using an intrinsic conditional auto-regressive (ICAR) prior:

$$U_j | U_k, k \in \delta_j \sim N(\overline{U}_J, \omega_U^2 / m_j),$$

where δ_j is the set of neighbors of county *j*, m_j is the number of neighbors, $\overline{U}_J$ is the mean of the spatial random effects of the neighbors, and ω_U^2 is the conditional variance whose magnitude determines the amount of spatial variation[45]. This model imposes smoothing by assuming that the spatial effect in a particular county is similar to the mean of the spatial effects in near-by counties, with the strength of the similarity determined by the number of neighbors (counties with more neighbors will have stronger similarities imposed). We define counties *j* and *k* to be neighbors if they share a common boundary.

Under this model, we must assign distributions to σ_V^2 and ω_U^2. We specify a Gamma(0.05,0.001) prior distribution on $\tau_V = \sigma_V^{-2}$ and on $\tau_U = \omega_U^{-2}$. We conducted sensitivity analyses using different prior distributions (Gamma(1,0.026; 0.05,0.026; 1, 0.001),and results were similar. County-level data was used instead of census tract-level data since each geographic area must include at least 1 case and 1 control. Neighborhood variables were Z-score transformed (subtracting the mean and dividing by the standard deviation) in order to compare odds ratios from many regressions[9]. The proportion of residual variability that is spatial in nature was calculated by dividing the posterior marginal variance of the spatial random effects by the sum of the posterior marginal variances of the spatial and non-spatial random effects. In this phase, there is again a large multiple testing problem and so we adjust our significance threshold and set it to 0.05/n for n the number of variables identified in Phase 1, which corresponds to a Bonferroni corrected threshold of 0.05. Since we are performing a Bayesian analysis, *p*-values are not available to assess significance. Hence, we calculate (100-0.05/n)% credible intervals and neighborhood variables whose credible intervals included zero were not considered significant. Phase 2 analyses were conducted using Integrated Nested Laplace Approximations (INLA) [46] implemented in the INLA package in R statistical software.

In *Phase 3*, we account for correlation among the most significant variables identified in Phase 2 by applying standard principal components analysis (PCA), a technique frequently used in neighborhood research[47, 48]. PCA converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components[47]. PCA analyzes total variance[49], and variable loadings onto components represent the correlation between the variable and that component[50]. PCA, as opposed to factor analysis, was chosen for data reduction in

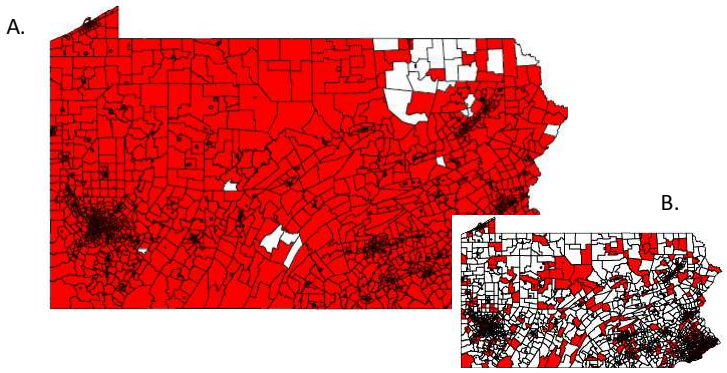
this study because we sought an empirical summary of total neighborhood-level variance explained by the census variables, rather than confirmation of an underlying factor structure comprised of previously identified domains[48, 51]. No independent components emerged following exploratory factor analysis. We hypothesized that variables within each principal component were correlated; this is analogous to linkage disequilibrium in GWAS where single nucleotide polymorphisms within larger gene regions are considered correlated[9]. Additionally, we determined that the most significant variable within each component(i.e. the variable with the tightest credible interval from Phase 2) was the best representation of that principal component for use in future neighborhood and prostate cancer aggressiveness studies. This approach is similar to fine mapping approaches that are often used post-GWAS[52, 53]. Fine mapping methods identify the specific SNP within a gene region that is most relevant to the outcome of interest.

In this analysis, neighborhood variables with a maximum loading of <0.30 on any of the retained principal components were eliminated, given that loadings between 0.3 and 0.5 are considered acceptable[48] [49, 54]. Principal components (and the variables that load on those components) that accounted for up to 90% of the data were retained in order to determine “top hits” in this study. For variables that loaded strongly(>0.30) on more than one component or factor variable, the magnitude of the correlation coefficient for each component, as well as the variable description, were used to determine the most appropriate component placement. Phase 3 was conducted using STATA/SE 12.0 statistical software.

Results

Figure 1a displays the distribution of all prostate cancer study participants in this analysis by census tract, and Figure1b displays the distribution of aggressive prostate cancer cases by census tract. Of the reported 3,135 census tracts in the Commonwealth of Pennsylvania (PA) in Year 2000, 3,037 (97%) census tracts are represented in our study sample (Figure 1a; Appendix D). Most aggressive prostate cancer cases are clustered in urban areas, namely Pittsburgh and Philadelphia (Figure 1b). The average age of the study population was 69.2 (standard deviation (sd) 9.4) and mean year of diagnosis was 2000. The average age of aggressive cases was 69.8 (sd-10.4) and of nonaggressive cases was 68.8 (sd-9.0). Figure 2 summarizes the study methods and findings of Phases 1-3. In Phase 1, we identified 517 census variables that were significantly associated with prostate cancer aggressiveness at Bonferroni significance levels (Supplementary Digital File 5). In Phase 2 we identified 217 variables that were still significant at Bonferroni-adjusted credible intervals after adjustment for spatial variability (Supplementary Digital File 6). The average amount of spatial variation across the 217 models was 0.34 (range: 0.14-0.50), which is considered substantial. In Phase 3, 17 uncorrelated principal components were identified from 217 neighborhood variables, with 76 variables loading on Component 1 and 51 loading on Component 2. Components 1-8 explain 80% of the variance among the top 217 neighborhood variables, and these components relate to poverty level (Component 1), white only characteristics (Component 2), household income and worth (Component 3), male

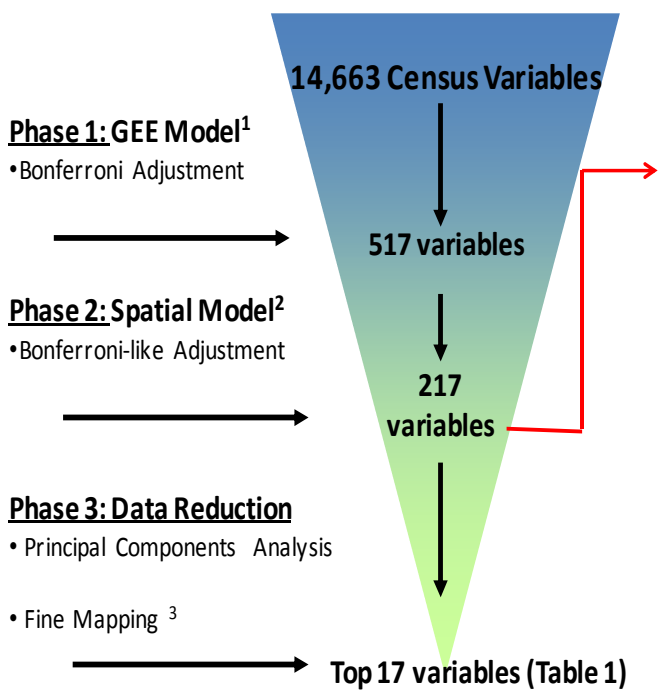
Figure 1a. All Prostate Cancer Cases by Census Tract
b. Aggressive Prostate Cancer Cases by Census Tract



householders living alone component 4), rented houses built before 1939 (Component 5), civilian population (Component 6), household income above $60K (component 7), and immigration (Component 8) (Figure 2).

The top 17 most significant variables within each of the 17 principal components are described in Table 1. The top hit or most significant variable in this analysis (based on significance from Phase 2) loaded on Component 12 (Workers mode of transportation). Specifically, percent workers 16 years and over taking trolley or street car public transportation to work was related to an OR of 1.05 and a credible interval(CI) of 1.001-1.09, which can be interpreted as the odds of having aggressive prostate cancer for each unit increase in the neighborhood percentage of workers 16 years and over taking

Figure 2. Summary of NWAS Methods and Top Neighborhood Components



Top Neighborhood Components (by ranking of variance explained)*	# Variables Loading
1.Poverty (below poverty)	76
2.White Only Family and Housing	51
3.Householder income/worth	31
4.Male Householder Living Alone	11
5.Rented Household built 1939 or earlier	14
6. Civilian Population	4
7. Household income >$60K	4
8.Immigration	1
9. Owner occupied housing unit worth	3
10.Poverty Status Males under 50	5
11.Housing units built 1940-49	3
12. Workers Mode of Transportation	3
13.Rented Household no vehicle	4
14.Males in Protective Services	1
15. Male earnings 7.5K-9,999	2
16. Male over 65 (not family household)	3
17. Male household renter aged 55-64	1

trolley or streetcar public transport to work. Seventeen of the top 30 variables (top 10% of significant variables from Phase 2) loaded on Component 1(Supplementary Digital File 6). Component 1 related to poverty and is best described by variables related to poverty status based on income and female head of households with children variables (Supplementary Digital File 6). The most significant variable from Component 1 represented Non-Hispanic Whites aged 6-11 for whom poverty status was determined (OR=1.07, credible interval=1.01-1.12). In general, 15 of the top hits were socioeconomic variables and two also related to physical environment (% Aggregate income of Occupied Housing units built 1940-1949 and % Renter occupied housing unit built 1939 or earlier with householder aged 15-24 years).

Discussion

Based on GWAS and EWAS frameworks, we propose the neighborhood-wide association study (NWAS). We systematically and comprehensively assessed the association of 14, 663 neighborhood variables measured in the U.S. Census SF1 and SF3 forms with prostate cancer aggressiveness, based on case data from the PA State Cancer Registry. Through a series of progressively more stringent model adjustments and data reduction techniques, we identified the top 17 neighborhood variables associated with aggressive prostate cancer. These findings confirm previous associations between neighborhood and prostate cancer, and they provide new insights into the role of neighborhood in prostate cancer development.

Previous neighborhood studies have focused on associations between prostate cancer and pre-determined variables from the U.S. Census that represent socioeconomic status (SES) domains related to education, income, poverty, and employment (Supplementary Discussion Table 1-[21, 22] [23, 24] [24, 32-35] [25-27]). Housing variables related to vacancy and median home/rental values have also been

used as potential indicators of neighborhood SES (Appendix G). Our findings support that neighborhood income and poverty (Components 1,2,3,7,8, 9, 10, 15), employment (Component 14)) and housing variables(Components 3, 4, 5, 9, 11, 13, 16, 17) are related to prostate cancer aggressiveness; however, neighborhood educational level did not appear to be an important determinant of aggressive prostate cancer in this NWAS. Immigration status (Component 9) has also been studied at the neighborhood and individual-level in prostate cancer. Studies of neighborhoods with higher rates of foreign-born immigrants have shown associations with decreased risk for cancer, despite the socioeconomic status of the individual[55]. Even if individuals are diagnosed with late stage prostate cancer, survival is particularly improved for those who live in high ethnically homogeneous enclaves, suggesting the strong role that social support may play in prostate cancer progression[55, 56]. Other variable components identified in the NWAS have been implicated in neighborhood research or cancer research in general, namely, variables related to mode of transportation(Components 12 and 13) and age of housing units(Components 5, 11). Not having a vehicle or taking public transport to work are measures that are often associated with urban vs rural areas, and they have beenused as measures for access to medical care[57] [58, 59]. Access to care is often cited as cause of disparity in prostate cancer treatment[58] and survival[57] in both urban and rural settings. Higher cancer incidence and mortality rates are often noted in more urban settings, and cases arising from rural environments often are diagnosed at a later stage of disease[60]. Further, previous studies of neighborhood and disease suggest that physical environment (e.g., housing vacancy) is relevant to disease risk[30]. Age of housing unit was associated with aggressive prostate cancer in the NWAS. Assuming that older age of the house is associated with deterioration, this finding supports the hypothesis that poor housing can lead to poor health outcomes[61]. Thus, NWAS findings are biologically plausible, given results support the association

between previously defined socioeconomic domains that were based on existing social theories[62, 63].

While the NWAS top hits represent similar socioeconomic domains/components presented in literature, the variables presented here, within each of the principal components, are potentially more informative for inferring causation and identifying areas in need of intervention. A limitation of previous neighborhood studies has been a lack of consistency in variable selection for analysis[63]. For example, poverty can be defined as the proportion of individuals or households below the federal poverty level, or as a percentage on public assistance[48]. This lack of "common methods" or "common neighborhood variables" for analysis is pervasive across neighborhood research in general[48, 63], making the accumulated evidence of the role of neighborhood in disease difficult to assess systematically for causation[63]. NWAS identified mostly combination variables or variables that represent more than one socioeconomic construct(e.g percentage of male householders not living with family, which represents gender, family, household, and poverty information). Previous studies generally select a single variable that represents fewer socioeconomic parameters (e.g. household income only) and/or develop indices from these variables [24, 32-35] [25-27]. Improper variable selection using a *priori* approaches could bias association results and lead to false negative associations. Thus, NWAS addresses this research gap by empirically identifying neighborhood factors relevant to prostate cancer. Additionally, it is unlikely that previous studies would select to study both variables: percentage of male nonfamily householders living alone AND percentage of male nonfamily householders living alone over 65. Although these two variables appear to represent similar information by name, the NWAS identified these variables as two separate components. Variables related to single resident households have been used as markers of social support[64], and it is possible that they represent separate or potentially dynamic changes in the role of social

support across the lifespan. This is similar to findings in GWAS where top hits from the same region may provide additional information about the function of that genetic locus[65]. Additionally, renter versus owner-occupied housing units appeared to be more strongly associated with aggressive prostate cancer in the NWAS, and could be another indicator of degree of socioeconomic status, beyond income. Given the specificity of top variables from NWAS, groups (in neighborhoods that possess these particular neighborhood characteristics) who are likely to have unfavorable outcomes/aggressive disease can be identified for recruitment in studies and can lead to studies that better characterize the neighborhood environment. This could result in the development of targeted interventions and strategies for addressing these neighborhood-specific factors in high risk areas.

Given that this analysis was restricted to white men with prostate cancer (median age 66), it is not surprising that top hits included variables related to older age ranges(Components 2, 10, 13, 17) and white only neighborhood characteristics (Component 2). Neighborhood investigations often adjust by percent Hispanic or percent African American as a measure of neighborhood segregation (24, 33-35), but segregation or socioeconomic circumstances related to other racial groups did not appear to be associated with aggressive prostate cancer. This could suggest that relevant, race-specific neighborhood characteristics may predict in a manner that is dependent on the individual subject's own race. The conduct of an NWAS in other racial groups is warranted in order to allow for comparisons across other race/ethnicities. Thus, the NWAS approach has implications for health disparities research, particularly teasing apart racial versus socioeconomic effects.

Some of the top variables did not relate to socioeconomic or physical environmental variables, and instead related to population denominators used to calculate percentages in the U.S. Census (i.e., Component 6) or included age ranges that are not typically

associated with cancer development(i.e. Components 1, 5). For example, civilian non-institutionalized population 5 years and older is an imputed variable in the SF3 form that could reflect adjustments for study sampling approaches, given SF3 is based only on a subset, i.e., 5% of the population. Although we made inferences using extremely conservative significance levels, it also is possible that some of these findings could be spurious. In the same token, there is potential for false negatives in this analysis, given that our goal was to minimize Type I error and focus on finding true positives using a multi-phase approach and stringent parameters for statistical significance. Further, 9 of the top 17 variables had mean percent values of less than 1%. Although these variables may be rare, they would still be useful for data reduction purposes, particularly if the goal is to identify census tracts at the upper range of these particular variables for prevention or intervention. Further, some variables, namely taking trolley or street cars to work or living in older, rented housing units, may be a reflection of urban versus rural nature of the U.S. Census data. More prostate cancer cases come from more densely populated urban environments[60], but to address this issue, we account for the spatial nature of the data in Phase 2. Although odds ratio did not change more than 8% when comparing odds ratios from Phase 1 to Phase 2, spatial variance was substantial in our models (greater than 30%), suggesting that spatial events should be considered in neighborhood and cancer studies.

In general, standardized data processing and design approaches often used in administrative datasets like the PA Cancer Registry and U.S. Census, can introduce systematic bias and limit inferences that can be made[30]. For instance, we did not have the ability to adjust for individual level factors beyond race, age, and date of diagnosis. Neighborhood socioeconomic factors are believed to exert separate effects from individual level data[21-24, 38], but conduct of NWAS in populations that allow for individual-level adjustments will be needed in future studies, particularly to make

generalizations about causality[22]. Bias related to data missingness is also a concern in administrative datasets, and could be a factor in this NWAS. The response rate for the census data for the State of Pennsylvania is 70%[66]. Also, we limited our dataset to variables with less than 10% missingness based on GWAS, EWAS, and social epidemiologic studies that have used an acceptable missingness level of $\leq$10%[9] [67, 68]. Based on our assessments of missingness in this study, bias related to reporting or administrative rules appears likely nondifferential (Appendix D); however, future studies that investigate missingness effects in NWAS, using both spatial autocorrelation and imputation techniques[69], are warranted.

There are other study strengths and limitations to note. A hallmark of GWAS studies has been replication of findings in similar study populations. However, with the generally low observed odds ratios in current GWAS studies and that fact that replication in similar study population is not always feasible, many investigators are favoring a single discovery phase adjusted for multiple comparisons[70]. This school of thought applies when using State Cancer Registry and U.S. Census data. Neighborhood characteristics, as well as disease rates, likely differ by State, thus, it's not clear whether other State registries would serve as appropriate replication groups. Clinical populations of prostate cancer cases that give rise to cases within State cancer registries (that also likely include more detailed collections of individual-level factors) and the comparison of cancer registries within the same State from different time periods (i.e. using prostate cancer diagnosed from 1995-2005 and the 2000 US Census versus prostate cancer cases diagnosed from 2005-2015 and the 2010 Census) could potentially serve as appropriate replication groups and warrant further investigation. Focus on appropriate replication groups was outside the scope of this study, and instead we aimed to introduce the NWAS as a nascent, but novel approach focused on discovering potentially new research angles in neighborhood and cancer research.

The use of county and census tract-level data does not come without some criticism[22]. Area-level data analysis assumes that people within the same geographic area experience similar socioeconomic circumstances. In reality, this may not be the case and people may not spend the majority of their time at their self-reported residential addresses. However, use of U.S. Census variables are warranted given that they are widely available, allowing for systematic analysis and consistency across studies. Additionally, previous studies of prostate cancer have inconsistently accounted for the effects of spatial autocorrelation related to U.S. Census administrative boundaries[63], and have focused mostly on socioeconomic, compared to physical environment [22, 63]. This was the first study to comprehensively evaluate the role of all available US census variables, accounting for similarities across administrative census boundaries and including physical and social characteristics, in cancer. We found that spatial variance contributes substantially to model variance (though magnitudes of effect do not change by more than 8% from Phase 1 to Phase 2). Thus, NWAS does contribute new findings to the literature and could provide justification for more precise measurements of neighborhood-level attributes[71]. Additionally, NWAS methodology could be applied to other available, community or national-level databases, which could lead to more relevant neighborhood boundary definitions.

Our prostate cancer outcome was derived from simplified and broader categories of tumor stage and Gleason grade often employed in SEER and State Cancer Registries[41]. Most prostate cancer cases die with and not of the disease, making aggressive prostate cancer the most relevant outcome[20, 72, 73]. It is possible that including more detailed clinical information related to stage, grade, metastasis, and prostate specific antigen(PSA) level[74-76] could improve our outcome definition and the specificity of our findings[77, 78].

Despite limitations, this study will be the first to systematically and empirically evaluate the role of the macro-environment in prostate cancer. We demonstrate for the first time that high dimensional data analysis can be applied to publically available, social datasets. Findings from this study can be used as "neighborhood signatures" in the identification of neighborhoods that possess "high risk" characteristics and are in need of disease intervention and prevention efforts. NWAS results are hypothesis-generating and can lead to studies focused on the etiologic role of neighborhood on prostate cancer and other diseases. The NWAS method has implications for health disparities research and can be applied across a number of health and disease settings. NWAS can serve as a common methodology across disciplines, and thus can facilitate multicenter, multilevel investigations.

Table 1. Summary of Neighborhood Variable "Top Hits" Associated with Aggressive Prostate Cancer by Phase Results.

			Phase 1				Phase 2			Phase 3
	Mean (sd)	**Range**	**Odds Ratio**	**CI**	**p-value**	**Rank**	**Odds Ratio**	**CI**	**Rank**	**Component Load**
Census Variable										
%White alone population for whom poverty status is determined age 6-11 years (pct_sf3_pct075a006)	0.38 (0.53)	0-20.6	1.09	1.05-1.12	0.03	463	1.07	1.01-1.12	12	1
%White, Non-hispanics where poverty status determined aged 18-64 below poverty level in 1999 (pct_SF3_p159i007)	4.4 (4.7)	0-100	1.01	1.01-1.02	.003	147	1.06	1.02-1.12	11	2
% Male Nonfamily households below poverty level (pct_sf3_p092021)	1.6 (1.9)	0-35.4	1.03	1.02-1.04	0.03	461	1.06	1.01-1.11	14	3
%Male householder living alone (nonfamily household) (pct_SF3_h019093)	4.69 (3.9)	0-39.5	1.02	1.01-1.02	0.03	466	1.07	1.01-1.12	61	4
% Renter occupied housing unit built 1939 or earlier with householder aged	0.75 (1.4)	0-31.8	1.05	1.03-1.06	0.003	136	1.07	1.02-1.11	2	5

15-24 years (pct_SF3_hct005083)										
Imputed civilian non-institutionalized population 5 years and older (pct_sf3_p120002)	6.39 (2.6)	0-63.2	1.02	1.01-1.03	0.02	415	1.06	1.01-1.11	91	6
%Household income $60K-74,999 (pct_SF3_p052012)	10.9 (3.6)	0-25.1	0.98	(0.97-0.99)	0.048	516	0.95	0.89-0.99	202	7
%Foreign born naturalized citizen at or above poverty level (pct_SF3_pct051020)	2.0 (2.1)	0-18.9	0.96	0.94-0.97	8.8 X 10-6	10	0.93	0.87-0.99	217	8
%Household income of $10K-19,999 with owner-occupied housing unit value of $10K-19,999 (pct_SF3_hct017019)	0.34 (1.1)	0-23.0	1.06	1.04-1.08	2.7 X 10-5	15	1.05	1.00-1.10	3	9
% Population for whom poverty status is determined aged 45-54 years, under 0.50(pct_sf3_pct050102)	0.33 (0.41)	0-12.6	1.18	1.12-1.25	6X10-5	19	1.08	1.03-1.13	10	10
% Aggregate income of Occupied Housing units built 1940-1949 (pct_sf3_hct015042)	0.67 (1.1)	0-28.3	1.06	1.03-1.08	0.001	94	1.06	1.01-1.11	7	11
%Workers 16 years and over taking public transportation, namely trolley or street cars, to work (pct_SF3_p030007)	0.12 (0.64)	0-14.6	1.10	1.06-1.13	0.0001	25	1.05	1.001-1.09	1	12
Renter occupied housing unit with householder aged 55-64 with no vehicle available (pct_SF3_h045025)	0.54 (0.99)	0-15.1	1.06	1.04-1.09	0.003	155	1.07	1.02-1.12	8	13
%Male Protective Service Occupations: fire fighting, prevention, and law enforcement workers (pct_SF3_p050026)	0.89 (0.93)	0-16.7	0.93	0.90-0.96	0.04	499	0.94	0.89-0.99	213	14
%Males with earnings of $7500-9,999 in 1999 (pct_SF3_p084006)	1.50 (0.96)	0-37.1	1.07	1.04-1.10	0.01	323	1.05	1.001-1.10	41	15
Male householder over 65 living alone in nonfamily household (pct_SF1_p030012)	7.2 (2.5)	0-34.5	1.03	1.02-1.04	.005	196	1.07	1.02-1.13	100	16
% Household Renters aged 55-64 years (pct_sf3_hct004093)	0.74 (0.75)	0-9.3	1.09	1.05-1.12	0.03	462	1.06	1.01-1.12	53	17

References:

1. Lynch, S.M. and T.R. Rebbeck, *Bridging the Gap between Biologic, Individual, and Macroenvironmental Factors in Cancer: A Multilevel Approach.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(4): p. 485-495.
2. Warnecke, R.B., et al., *Approaching Health Disparities From a Population Perspective: The National Institutes of Health Centers for Population Health and Health Disparities.* American Journal of Public Health, 2008. **98**(9): p. 1608-1615.
3. Yeager, M., et al., *Genome-wide association study of prostate cancer identifies a second risk locus at 8q24.* Nat Genet, 2007. **39**(5): p. 645-9.
4. Eeles, R.A., et al., *Multiple newly identified loci associated with prostate cancer susceptibility.* Nat Genet, 2008. **40**(3): p. 316-21.
5. Gudmundsson, J., et al., *Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24.* Nat Genet, 2007. **39**(5): p. 631-7.
6. Kibel, A.S., *A common variant associated with prostate cancer in European and African populations Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorleifsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, Johansson JE, Balter K, Wiklund F, Montie JE, Yu X, Suarez BK, Ober C, Cooney KA, Gronberg H, Catalona WJ, Einarsson GV, Barkardottir RB, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K, deCODE Genetics Inc., Reykjavik, Iceland.* Urol Oncol, 2007. **25**(5): p. 446-7.
7. Varghese, J.S. and D.F. Easton, *Genome-wide association studies in common cancers—what have we learnt?* Current Opinion in Genetics & Development, 2010. **20**(3): p. 201-209.
8. Chen, G., P. Wei, and A.L. DeStefano, *Incorporating biological information into association studies of sequencing data.* Genetic Epidemiology, 2011. **35**(S1): p. S29-S34.
9. Patel, C.J., Bhattacharya, J., Butte, A.J., *An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus.* PLoS One, 2010. **5**(5): p. e10746.
10. SLAUGHTER, D.P., H.W. SOUTHWICK, and W. SMEJKAL, *Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin.* Cancer, 1953. **6**(5): p. 963-8.
11. Lochhead, P., et al., *Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression.* Mod Pathol, 2015. **28**(1): p. 14-29.
12. Giovannucci, E. and S. Ogino, *DNA methylation, field effects, and colorectal cancer.* J Natl Cancer Inst, 2005. **97**(18): p. 1317-9.
13. Ushijima, T., *Epigenetic field for cancerization.* J Biochem Mol Biol, 2007. **40**(2): p. 142-50.
14. Ramachandran, K. and R. Singal, *DNA methylation and field cancerization.* Epigenomics, 2012. **4**(3): p. 243-5.
15. Stearman, R.S., et al., *A macrophage gene expression signature defines a field effect in the lung tumor microenvironment.* Cancer Res, 2008. **68**(1): p. 34-43.

16. Martinez-Outschoorn, U.E., et al., *Oxidative stress in cancer associated fibroblasts drives tumor-stroma co-evolution: A new paradigm for understanding tumor metabolism, the field effect and genomic instability in cancer cells.* Cell Cycle, 2010. **9**(16): p. 3256-76.
17. Heaphy, C.M., et al., *Telomere DNA content and allelic imbalance demonstrate field cancerization in histologically normal tissue adjacent to breast tumors.* Int J Cancer, 2006. **119**(1): p. 108-16.
18. Lochhead, P., Chan, A.T., Nishihara, R., Fuchs, C.S., Beck, A.H., Giovannucci, E.G., Ogino, S., *Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression.* Modern Pathology, 2015. **28**: p. 14-29.
19. Lynch, S.M., Mitra, N., Ravichandran, K., Spangler, E., Zhou, W., Paskett, E., Gehlert, S., et al., *Telomere Length and Neighborhood Cancerization: Evaluating Biological Response to Unfavorable Exposures.* Submitted for Peer Review, 2015.
20. American Cancer Society, *Cancer Facts and Figures*, 2009, American Cancer Society: Atlanta.
21. Messer, L., et al., *The Development of a Standardized Neighborhood Deprivation Index.* Journal of Urban Health, 2006. **83**(6): p. 1041-1062.
22. Diez Roux, A.V. and C. Mair, *Neighborhoods and health.* Annals of the New York Academy of Sciences, 2010. **1186**(1): p. 125-145.
23. Wardle, J., et al., *Socioeconomic disparities in cancer-risk behaviors in adolescence: baseline results from the Health and Behaviour in Teenagers Study (HABITS).* Prev Med, 2003. **36**(6): p. 721-30.
24. Borrell, L.N., et al., *Perception of general and oral health in White and African American adults: assessing the effect of neighborhood socioeconomic conditions.* Community Dent Oral Epidemiol, 2004. **32**(5): p. 363-73.
25. Ziegler-Johnson, C., et al., *Prostate Cancer Severity Associations with Neighborhood Deprivation.* Prostate Cancer, 2011. **2011**.
26. Byers, T.E., et al., *The impact of socioeconomic status on survival after cancer in the United States.* Cancer, 2008. **113**(3): p. 582-591.
27. Carpenter, W., et al., *Racial differences in PSA screening interval and stage at diagnosis.* Cancer Causes and Control, 2010. **21**(7): p. 1071-1080.
28. Lyratzopoulos, G., et al., *Population based time trends and socioeconomic variation in use of radiotherapy and radical surgery for prostate cancer in a UK region: continuous survey.* BMJ, 2010. **340**.
29. Crawford, E.D., *Epidemiology of Prostate Cancer.* Urology, 2003. **62**: p. 3-12.
30. Bureau, U.S.C., *2000 Census Technical Documentation for SF 3.* United States Department of Commerce, 2007.
31. Bureau, U.S.C., *2000 Census Technical Documentation for SF 3 Data.* US Department of Commerce, 2007.
32. Diez Roux, A.V., D.R. Jacobs, and C.I. Kiefe, *Neighborhood characteristics and components of the insulin resistance syndrome in young adults: the coronary artery risk development in young adults (CARDIA) study.* Diabetes Care, 2002. **25**(11): p. 1976-82.
33. Stoll, B.A., *Affluence, Obesity, and Breast Cancer.* Breast J, 2000. **6**(2): p. 146-149.
34. Robert, S.A. and E.N. Reither, *A multilevel analysis of race, community disadvantage, and body mass index among adults in the US.* Soc Sci Med, 2004. **59**(12): p. 2421-34.
35. Robert, S.A., et al., *Socioeconomic risk factors for breast cancer: distinguishing individual- and community-level effects.* Epidemiology, 2004. **15**(4): p. 442-50.
36. *Part III, Chapter 27 Communical and Noncommunicable Diseases, Section 27.31*, in *28*, P.S. Legislature, Editor 1980: Harrisburg, Pennsylvania.

37. Penney, K.L., et al., *Gleason Grade Progression Is Uncommon.* Cancer Research, 2013. **73**(16): p. 5163-5168.
38. Zeigler-Johnson, C., Tierney, A., Rebbeck, T.R., Rundle, A., *Prostate Cancer Severity Associations with Neighborhood Deprivation.* Prostate Cancer 2011.
39. Eeles, R.A., et al., *Multiple newly identified loci associated with prostate cancer susceptibility.* Nat Genet, 2008. **40**(3): p. 316-321.
40. *SEER Program Coding and Staging Manual 2000*, E. National Cancer Institute. Surveillance, and Endpoints Research (SEER). Editor 2000: Bethesda, MD.
41. SEER., *SEER Training Modules*, in *Prostate Cancer* N.I.o.H. U.S. Department of Health and Human Services, National Cancer Institute, Editor 2014: Bethesda, MD.
42. *Rule G* SEER Training Modules 2012 October 30, 2013]; Available from: http://training.seer.cancer.gov/coding/guidelines/rule_g.html.
43. Hubbard, A.E., et al., *To GEE or Not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health.* Epidemiology, 2010. **21**(4): p. 467-474 10.1097/EDE.0b013e3181caeb90.
44. Aickin, M., Gensler, H. , *Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods.* Am J Public Health, 1996. **86**: p. 726-728.
45. Besag, J., York, J., Mollie, A., *Bayesian Image Restoration, with Two Applications in Spatial Statistics.* Ann Inst Statist Math, 1991. **43**(1): p. 1-59.
46. Ru, H., Martino, S., *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.* J.R. Statist. Soc. B, 2008. **71**(2): p. 319-392.
47. Hatcher, L., *Principle Component Analysis.* SAS Institute Inc, 1994.
48. Messer, L., . Laraia, B., Kaufman, J., Eyster, J., Holzman, C., Culhane, J., et al. , *The development of a standard neighborhood deprivation index. .* Journal of Urban Health, 2006. **83**(6): p. 1041-1062.
49. Tabachnick, B.G., Fidell, L.S., *Chapter 13: Principal Components and Factor Analysis. Using Multivariate Statistics. 3.* Northridge, California: California State University, Harper Collins College, 1996. **635-708**.
50. Wilson, W.T., *The Truly Disadvantaged: The Inner City, The Underclass, and Public Policy*, C.U. Press, Editor 1987: Chicago, IL.
51. Kim, J.-O., Mueller, C.W. , Vol 07-014. Newbury Park, California: Sage; 1978., *Factor Analysis: Statistical Methods and Practical Issues.* Newbury Park, CA, 1978. **07-014**(Sage).
52. *Fine Mapping and GWAS*. Nature/Nurture: Genetic epideiology and statistics-how tos and wherefores blog 2014; Available from: https://blogs.sph.harvard.edu/peter-kraft/2013/07/03/fine-mapping-and-gwas/.
53. Meuwissen, T.H. and M.E. Goddard, *Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci.* Genetics, 2000. **155**(1): p. 421-430.
54. *Principal Components Analysis*. http://www.unt.edu/rss/class/mike/6810/Principal%20Components%20Analysis.pdf University of Texas (accessed 2015).
55. Schupp, C.W., Press, D.J., Gomez, S.L., *Immigration factors and prostate cancer survival among Hispanic men in California: does neighborhood matter?* Cancer, 2014. **120**(9): p. 1401-1408.
56. Carriere, G.M., Sanmartin, C., Bryant, H., Lockwood, G., *Rates of cancer incidence across terciles of the foreign-born population in Canada from 2001-2006.* Can J Public Health, 2013. **104**(7): p. e443-449.

57. Guidry, J.J., Aday, L.A., Zhang, D., Winn, R.J., *Transportation as a barrier to cancer treatment.* Cancer Pract, 1997. **5**(6): p. 361-366.
58. Patel, K., Kenerson, D., Wang, H., Brown, B., Pinkerton, H., Burress, M., Cooper, L., Canto, M., Ukoli, F., Hargreaves, M., *Factors influencing prostate cancer screening in low income African Americans in Tennessee.* J Health Care Poor Underserved, 2010. **21**(1 Suppl): p. 114-126.
59. Goodwin, J.S., Hunt, W.C., Samet, J.M., *Determinants of Cancer Therapy in Elderly Patients.* Cancer, 1993. **72**(594-601).
60. Monroe, A.C., Ricketts, T.C., Savitz, L.A., *Cancer in Rural versus Urban Populations: A Review.* J Rural Health, 1992. **8**(3): p. 212-220.
61. Thomson, H., Thomas, S., Sellstrom, E., Petticrew, M., *Housing Improvements for Health and Associated Socio-economic Outcomes.* Cochrane Database Syst Reve, 2013.
62. Dietz, R.D., *The estimation of neighborhood effects in the social sciences: An interdisciplinary approach.* Social Science Research, 2002. **31**: p. 539-575.
63. Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., *Assessing "Neighborhood Effects": Social Processes and New Directions in Research.* Ann Rev Sociol, 2002. **28**: p. 443-478.
64. Thompson, E.E., Krause, N., *Living Alone and Neighborhood Characteristics as Predictors of Social Support in Late Life.* J Gerontol B Psychol Sci Soc Sci, 1998. **53**(6): p. S354-64.
65. Salinas, C.A., et al., *Multiple Independent Genetic Variants in the 8q24 Region Are Associated with Prostate Cancer Risk.* Cancer Epidemiology Biomarkers & Prevention, 2008. **17**(5): p. 1203-1213.
66. United States Census Bureau, U., *Response Rates by State.* 2000.
67. Thomas, G., et al., *Multiple loci identified in a genome-wide association study of prostate cancer.* Nat Genet, 2008. **40**(3): p. 310-315.
68. Lynch, S.M. *Missing Data (Soc 504)*. 2003.
69. Paternoster, L., et al., *Genome-wide Association Study of Three-Dimensional Facial Morphology Identifies a Variant in PAX3 Associated with Nasion Position.* The American Journal of Human Genetics. **90**(3): p. 478-485.
70. Thomas, D., Casey, G, Conti, DV, Haile, RW, Lewinger, JP, Stram, DO., *Methodological Issues in Multistage Genome-wide Association Studies.* Stat Sci, 2009. **24**(4): p. 414-429.
71. Basta, L.A., T.S. Richmond, and D.J. Wiebe, *Neighborhoods, daily activities, and measuring health risks experienced in urban environments.* Social Science & Medicine, 2010. **71**(11): p. 1943-1950.
72. Society, A.C., *Cancer Facts and Figures.* 2011.
73. D'Amico, A.V., Whittington, R., Malkowicz, S., Fondurulia, J., Chen, M. , *The combination of preoperative prostate specific antigen and postoperative pathological findings to predict prostate specific antigen outcome in clinically localized prostate cancer.* Journal of Urology, 1998. **160**(6): p. 2096–2101.
74. *Prostate Cancer Staging*, in *Cancer Staging References*, American Joint Committee on Cancer, Editor 2013.
75. Chodak, G.W., et al., *Results of Conservative Management of Clinically Localized Prostate Cancer.* New England Journal of Medicine, 1994. **330**(4): p. 242-248.
76. Gittes, R.F., *Carcinoma of the Prostate.* New England Journal of Medicine, 1991. **324**(4): p. 236-245.
77. Stephenson, A.J., et al., *Preoperative Nomogram Predicting the 10-Year Probability of Prostate Cancer Recurrence After Radical Prostatectomy.* Journal of the National Cancer Institute, 2006. **98**(10): p. 715-717.

78. Kattan, M.W., T.M. Wheeler, and P.T. Scardino, *Postoperative Nomogram for Disease Recurrence After Radical Prostatectomy for Prostate Cancer.* Journal of Clinical Oncology, 1999. **17**(5): p. 1499.

APPENDIX *A*

Bridging the Gap between Biological, Individual and Macro-Environmental Factors in Cancer: A Multi-Level Approach

Shannon M. Lynch, Timothy R. Rebbeck

Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology and Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA 19104

Published in *Cancer Epidemiology, Biomarkers and Prevention*.

This research was supported by grants from the Public Health Service (P50-CA105641, P60-NM006900 and R01-CA85074 to TRR and F31-AG039986 to SML)

Corresponding Author: Shannon M. Lynch

Department of Biostatistics and Epidemiology

University of Pennsylvania School of Medicine

243 Blockley Hall, 423 Guardian Drive

Philadelphia, PA 19104-6021

Telephone: 215-898-5300

Fax: 215-573-1050

Email: lynchsh@mail.med.upenn.edu

Abstract

To address the complex nature of cancer occurrence and outcomes, approaches have been developed to simultaneously assess the role of two or more etiological agents within hierarchical levels including the: 1) macro-environment level (e.g., health care policy, neighborhood, or family structure); 2) individual level (e.g., behaviors, carcinogenic exposures, socioeconomic factors and psychological responses); 3) biological level (e.g., cellular biomarkers and inherited susceptibility variants). Prior multilevel approaches tend to focus on social and environmental hypotheses, and are thus limited in their ability to integrate biological factors into a multilevel framework. This limited integration may be related to the limited translation of research findings into the clinic. We propose a "Multi-level Biological And Social Integrative Construct" (MBASIC) to integrate macro-environment and individual factors with biology. The goal of this framework is to help researchers identify relationships among factors that may be involved in the multifactorial, complex nature of cancer etiology, to aid in appropriate study design, to guide the development of statistical or mechanistic models to study these relationships, and to position the results of these studies for improved intervention, translation, and implementation. MBASIC allows researchers from diverse fields to develop hypotheses of interest under a common conceptual framework, to guide transdisciplinary collaborations, and to optimize the value of multilevel studies for clinical and public health activities.

Motivation

Cancer is etiologically complex and its causes are multifactorial. Risk factors associated with cancer development have been identified that represent a variety of levels of influence on health and disease (**Table 1**). Macro-environment factors including health system, neighborhood or community characteristics, have increasingly been linked to cancer incidence and mortality [199, 200]. In addition, social determinants and processes [14, 199, 201] have been identified as cancer risk factors, including

socioeconomic status or self-reported race [202-205]. Environmental exposures at the level of the individual [202] including cigarette smoking [206], radon [207], asbestos [208], diet [209], and physical activity [210] are causally associated with some cancers. Applied and fundamental investigations have identified a wide array of biologic factors mechanistically involved in carcinogenesis including those of the tumor microenvironment, metabolome, proteome, transcriptome, and genome. For example, hundreds of novel genetic susceptibility loci have been identified through candidate and genome-wide association studies (GWAS [144]).

Studies of factors at a single level have provided a great deal of insight into the etiology of disease. Despite successes in identifying cancer risk factors, these approaches are limited and at some point the information obtained from these single-level studies reach a saturation point, and have provided as much information as they can. It is clear that the factors reported to date do not fully explain cancer incidence in the general population. For example, while smoking is strongly associated with lung cancer [211], most smokers will not be diagnosed with lung cancer, whereas some non-smokers will [212]. While *BRCA1* or *BRCA2* mutation carriers have a greatly increased lifetime risk of developing breast cancer [213], some *BRCA1/2* mutation carriers are never diagnosed with breast or ovarian cancer, even at an advanced age. GWAS have identified a wealth of susceptibility genes, but the identification of novel genes using this approach is unlikely to continue *ad infinitum*. Therefore, risk factors studied in isolation and identified by standard approaches are unable to fully explain the complex, multifactorial causes of cancer. For this reason, cancer research has evolved from focusing on single factors to studies of complex relationships between social, behavioral, molecular, and environmental factors.

Overview of Current Multilevel Approaches

To address the complex nature of cancer etiology, multilevel approaches have been developed to simultaneously assess the role of two or more etiological agents within a hierarchical or nested structure [6]. A number of conceptual frameworks have been proposed that integrate information across levels of disease etiology, including the "web of disease" of MacMahon and Pugh [7], the "wheel" of Mausner et al. [8], "systems epidemiology"[9], and more recent models of multifactorial etiology [3, 4, 10-13]. Multilevel approaches are generally characterized by three main levels: 1) macro-environment, referred to elsewhere as "eco-level" [3, 4]; 2) individual; and 3) biology **(Table 1).** Each of these levels is further characterized by sub-levels (**Table 1**) that define domains of variables involved in cancer etiology or outcomes. Multi-level conceptual frameworks are based on the premise that factors affecting disease act within and across levels to collectively affect disease. These approaches generally hypothesized that cancer outcomes can result from the complex relationship of factors at multiple levels in at least two ways **(Table 1).** First, factors at the macro-environment and individual levels can directly affect the biological events and result in cancer. Second, factors may confer risk in a hierarchal fashion, such that biologic-level effects are affected by behaviors or exposures of the individual, and individual level effects are affected by the macro-environment [15].

The relationships described above in the context of a multilevel model refer to both statistical and biological interactions. Here, we use the term "interaction" generically to refer to any non-additive statistical structure that can be constructed between two or more factors. This concept includes that of effect modification, mediation[214], as well

as biological structures that may be defined between two or more factors (e.g., epistasis among genetic loci). The goal of Multi-level Biological And Social Integrative Construct" (MBASIC) we propose below is not to define a specific form for interaction. A variety of statistical methods have been developed to guide the implementation of hierarchal, longitudinal or multilevel models [6, 215-217]. Instead, MBASIC provides a framework around which a researcher can generate hypotheses about the relationship among etiological agents in a consistent manner. When results of these hypothesis tests are known, investigators using this framework may be better able to compare and combine their results to form more coherent multilevel inferences.

Most multilevel approaches lack a detailed focus on mechanisms that can be used to frame the relationships between macro-environment or individual- level factors. In part, the limited incorporation of mechanistic hypotheses stems from the early multilevel frameworks having evolved from research focused on social factors. Thus, multi-level conceptual approaches have tended to take a "top-down" approach that is focused on the role of social determinants at the macro-environment level (**Table 2**). Only more recently has a detailed consideration of the biological level been included in multilevel studies. For instance, the model of Warnecke et al. [10] centers on health disparities as the outcome of interest and defines macro-environment level factors by policies, institutions, and social or physical factors. They also include a single level including biological factors. Similarly, the models of Taplin et al. [4] and Gorin et al. [3] focus on improved cancer care, sub-dividing the macro-environment level by national and state health policy, local community environment, organization or practice setting, health care provider teams, and family/social support. Across proposed multi-level frameworks, the traditional individual level risk factors for cancer (e.g., smoking, race, diet) are also

considered, whereas biological factors in these constructs remain broadly defined by genes, proteins, enzymes and other somatic changes in the cellular environment [2-4]. In these models, all biological processes are treated in a manner similar to those of other levels without accounting for the extensive knowledge of biological processes, pathways, and etiologic relationships that are involved in carcinogenesis.

Approaches that focus on macro-environment have had an impact on the conceptual advancement of our understanding of disease etiology. While the National Institutes of Health has increasingly recognized and encouraged the use of multilevel approaches to go beyond investigating individual level factors to include macro-environment level exposures [218, 219], many of the current approaches[218, 219]come from the perspective of social and environmental research, and the full integration of biological level factors has yet to be realized. A search of PubMed for the term “multilevel analysis” or “multilevel model” and “cancer” resulted in 55 articles published between 2002 and 2012, although the majority of these (26 of 55, 47%) were published since 2010. Most of these studies focused on individual-level and macro-environmental factors, and few incorporated biological factors. Thus, work is needed to improve the understanding of which factors at each level are relevant to the disease, the hierarchical nature of the relationship of those factors, and the effective application of integrative multilevel approaches to achieve meaningful etiological inferences.

Multi-level Biological and Social Integrative Construct (MBASIC).

To the degree that a researcher has knowledge of biological mechanisms of human cancer, multilevel models could be used to harness this information and to generate

hypotheses that link macro-environment or individual level factors with mechanisms of carcinogenesis. Current multi-level conceptual approaches, while created to promote multi-disciplinary research, often lack detailed descriptions of the biological level that could be used to unite traditionally distinct fields (e.g., molecular biology and social epidemiology). MBASIC defines the multilevel framework (construct) to include three main hierarchal levels that contribute to cancer etiology and levels of carcinogenesis (i.e., macro-environment, individual, and biological factors; **Figure 1**), where the biological level is more specifically defined. This multilevel etiological model is then placed in the context of interventions, and translation/implementation (i.e., T0-T4; [220, 221]; **Figure 1**) Thus, this framework allows researchers from the fields of public health, health policy, prevention, behavioral sciences, sociology, epidemiology, biology, clinical medicine, and others to test hypotheses of interest under a common conceptual framework, to address the dynamic nature of carcinogenesis, to facilitate translation of multilevel studies to clinical and public health strategies, and to support multi-disciplinary collaborations.

The primary goal of MBASIC is to consistently and systematically frame complex hypotheses about cancer etiology, that once tested, can expeditiously inform intervention and implementation levels, under the umbrella of a common framework. As may be expected with any comprehensive conceptual framework, the full range of MBASIC components is not meant to be implemented in any one study. Instead, MBASIC is meant to aid the researcher in stating hypotheses for individual studies that address a part of the complete framework. Thus, MBASIC provides the framework for hypotheses that allow comparison and compilation of individual study results using formal (e.g., meta-analytic) or *ad hoc* means. Individual studies built around the

MBASIC framework could also motivate multidisciplinary collaborations and could rationalize single, large-scale multilevel studies in the future.

Predictive and Mechanistic Links Between and Among Hierarchical Levels of Etiology

An important goal of the MBASIC is to guide researchers to consistently and systematically incorporate biological mechanisms into a multilevel framework. Despite the substantial limitations in our ability to generate meaningful statistical or epidemiological models of mechanism and biological events [222, 223], knowledge of existing biological pathways emerging from animal, tumor, and other *in vivo* studies can be employed to improve generation of hypotheses about how each of the three hierarchal levels relates with the others in order to frame questions about the complexity of cancer etiology[224]. The well-known molecular epidemiology paradigm [225-229] provides a useful structure into which existing biological knowledge can be incorporated into a multilevel framework. As shown in **Figure 2** and defined below, the effect of exposures can be measured by biomarkers of biologically effective doses (BED), early biological effects (EBE), and altered structure and function (ASF) that are predictive of disease [226-229]. The formation of these biomarkers can be influenced by inherited genotypes (IG). These factors can give rise to somatic genomic (SG) changes involved in carcinogenesis. Note that while prior constructs include markers of internal dose, which have great value as biomarkers for research, clinical or screening purposes, we exclude these in the present framework to emphasize biological and mechanistic effects in the multilevel etiology of cancer. While spontaneous mutation may give rise to the biomarkers of disease and effect shown in **Figure 2**, the multilevel construct assumes

that each of the biomarkers occur in response to an initial macro-environment or individual level exposure, even though that exposure may not be known or measurable.

We adapt the traditional molecular epidemiology approach [226-229] in two ways: by considering the nested hierarchical nature of the multilevel model (**Figure 2**); and by expanding the definition of "exposure" to include both macro-environment level and individual level exposures. As noted in **Table 1**, relevant etiological factors can be measured by biomarkers (i.e., BED, EBE, ASF) of exposure or disease at the biological level. These biomarkers reflect somatic changes and are often measured at the tissue or cellular level. For example, biomarkers of exposure to cigarette smoking at the individual level can be measured by exposure biomarkers such as DNA adducts [226-229] in blood; prostate specific antigen (PSA) levels or chromosomal instability [229] measured in blood can serve as markers of disease. Thus, these factors may be framed as both processes leading to disease and as intermediates reflecting the relationship between macro-environmental and individual factors, separately, and disease (**Table 1**). For instance, macro-environment level variables can induce a psychological response, which can be directly measured at the biological level. Witnessing a crime in a neighborhood environment can lead to flight or fight cellular responses that cause increases in cortisol levels. Thus, cortisol is a biomarker of a macro-environment exposure. An example of a linkage between the individual level and the biological level is that of the human exposome [230]. The exposome is defined by environmental exposures (including lifestyle factors) that represent combined exposures from all sources, from the prenatal period onwards [230]. The exposome can be measured by biomarkers at the cellular level via bodily fluids or tissue that can serve as surrogates for exogenous or endogenous environmental exposures. For instance, exposure to

organophosphate pesticides can be measured by certain metabolites, and dietary factors, like vitamin intake, can be measured by antioxidant metabolites. Like the GWAS approach, epidemiology has employed environment-wide association studies (EWAS) [230, 231] that use an agnostic approach to identifying environmental factors involved in disease. Future EWAS studies in cancer are warranted to provide practical evidence for a link between individual level exposures and the biological level. While EWAS and GWAS share some conceptual similarities, there are numerous methodological differences between the two approaches[232]. However, the results of each can provide information that may promote the development of multilevel hypotheses in cancer etiology.

While the examples above demonstrate how macro-environment and the individual level factors can each separately affect the biological level as an exposure, we can also demonstrate the hierarchal effect among exposures at multiple levels on the biologic level. For example, exposure to a group of friends who smoke cigarettes could prompt an individual to change her behavior and also start to smoke cigarettes. This change in behavior at the individual level influences molecular carcinogenesis at the biologic level (i.e. DNA adducts; BED) and chromosomal damage (ASF). Despite symptoms of decreased lung function over the course of 15-20 years, the individual is genetically predisposed to nicotine dependence, is unable to quit smoking, and ultimately ends up developing lung cancer. Here, the behavior change served as an intermediate between the macro-environment and biological events involved in carcinogenesis. Thus, this example demonstrates the biological plausibility of how a macro-environmental factor can impact an individual, affecting her biological environment, ultimately resulting in disease. When the macro-environment, individual, and biological factors are collectively

considered in order to predict or explain a cancer outcome, statistical methods will be needed to determine which levels or which risk factors within each level are most relevant to the cancer outcome under study. Thus, it is possible for intermediates to serve as surrogates of exposure and disease, but the importance of each level and each factor within each level will need to be determined statistically based on available methods.

Biology in a Multilevel Framework

Starting with the Levels of Etiology (**Figure 1**), the biological level can be subdivided into sub-levels with a hierarchical order based on our knowledge of biology and carcinogenesis: tissues are comprised of cells, which contain genes. Somatic mutations and cellular events (e.g. DNA replication) may be involved in carcinogenesis. In the following sections, we build the framework around which the biological level can be optimally incorporated into multilevel analysis (**Figure 2**).

Tissues: Tissues warrant consideration as a unique biological sub-level in a multilevel framework for two reasons. First, cellular markers and processes that are measured in normal tissue, pre-neoplasia, or malignant tumors could serve as potential markers of exposure, disease, or prognosis. Second, tumors occur at the tissue level. Most cancers are diagnosed and staged using tissue samples or by imaging techniques that may identify lesions in a particular organ. A growing area of research is focused on the tumor microenvironment, defined by normal cells, signaling molecules, matrices and blood vessels that surround and feed a tumor cell [233]. A tumor can alter its microenvironment (as defined by cellular and genomic sub-level factors), and the

microenvironment can affect how a tumor grows and spreads. Data about the role of the tumor microenvironment are rapidly becoming available via initiatives such as The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov).

Cells: The cellular sub-level is characterized by proteins, enzymes, and other biomarkers that can be detected in bodily fluids and tissues. In the context of our model, the cellular level includes the transcriptome, proteome, and metabolome, where biomarkers of exposure and disease can be measured **(Figure 2).** The transcriptome includes the various forms of RNA in the cell that affect gene expression and cellular function [234-236]. The proteome includes the total set of proteins expressed in a given cell at a given time [235]. Examples of factors measured in the proteome include prostate specific antigen (PSA) and CA-125 [229, 234, 237]. Complex protein interactions are referred to as the metabolome [235, 238]. Therefore, even within the cellular sub-level, there is an emerging hierarchy [235]. Many approaches for disease biomarker discovery focus on a single biomarker at the cellular level, despite an emerging expectation that panels of biomarker analytes will be needed to provide sufficient sensitivity and specificity for cancer screening, diagnosis or prognosis [234, 238]. Therefore, there is a shifting focus to the role of the pathway-based and statistical interactions among cellular factors, but progress in this area is limited by available, high-throughput technologies that can detect and organize the millions of proteins obtained from a given biological sample.

Somatic Genomics (SG): The SG sub-level (**Figures 1** and **2**) is defined by acquired somatic genomic changes over the course of a person's lifetime. The SG level is

defined by factors that can be both markers of disease and markers of exposure. SG examples include mutations, copy number variants, and epigenetic changes occurring in DNA [234, 239]. Early studies of SG used methods that identify potential susceptibly loci *a priori*, but this approach used a small number of genetic markers, rarely identified robust associations between candidate genes and cancer, and most findings were not replicable in other studies [144].

Inherited Genomics (IG): The IG sub-level (**Figures 1** and **2**) is comprised of inherited susceptibility loci that serve as markers of disease risk and outcome. IG includes hereditary cancer syndromes [240], which confer a high risk of cancer development. IG research may use family-based linkage methods to identify important inherited, high-penetrance genes, such as *BRCA1* and *BRCA2*. However, the mutations in these genes are rare in the general population [144, 201, 213], and only explain a small fraction of familial aggregation and cancer risk. GWAS have identified many dozens of cancer susceptibility loci [227, 241], most of which were not previously hypothesized to be involved in cancer susceptibility [144]. Despite this success, genetic risk variants identified from GWAS, alone and in combination, explain a relatively minor proportion of disease risk, and have had limited translational value to the clinic. This has led to a focus on the identification of rare variants that may account for larger proportions of cancer genetic risk[145].

Given the limited clinical utility of SG and IG findings focused on single disease loci and statistical interactions thereof, there has been a renewed interest in studying epistasis, defined as genes at two or more loci that produce phenotype effects that are different

than the expected effects of the individual loci [242]. At both the SG and IG sub-levels, gene-gene interaction studies are being conducted to ascertain the independent and joint effects of risk loci on cancer outcomes [243]. These studies may use multiple cancer risk susceptibility loci based on pathway or shared biological function, or be combined using statistical predictive models independent of biological knowledge.

Non-Hierarchical Effects Within and Across Levels of Etiology

Mechanisms and example methodologies have been proposed to build on the definition of the biological level and to illustrate how interactions between and among factors at each level relates to one another, assuming a hierarchal structure for levels of etiology (**Figure 1**). Hypotheses that consider the hierarchical framework of MBASIC are readily constructed from the discussion provided above. However, the effects of factors within each of these levels need not follow a strict hierarchy. In the context of predictive (as opposed to mechanistic) models, each level can dynamically affect another. Thus, statistical (causal) inferences need not be constrained in a linear hierarchical fashion [244]. Concepts in social science and genetics support this assertion. According to the Social Ecological Perspective [245, 246], human health results from the complex interaction of personal factors (e.g., behaviors, biology, psychology) as well as physical and social environments (e.g., geography, built environment, culture, economics, politics, and social relationships) [245]. For instance, a combination of geography, psychology, and behavior without a clear hierarchal or biological link could interact (statistically) and affect disease outcomes. Additionally, changes in eating habits at the individual level may affect social relationships at the macro-environment level since a person who is more conscious of their eating habits may prefer to be around other

healthy eaters; the effect of each level on the other is not necessarily linear, top-down, or bottom-up. In the field of genetics, penetrance [247] is defined by the probability of a phenotype given genotype. Even though a person is born with a disease genotype, lack of exposure to harmful environmental factors or carcinogens may prevent the disease from occurring. While it is likely that the disease genotype and exposure have some biological link, in the absence of this knowledge, specific methodologies aimed at analyzing gene-environment interactions, more recently, gene-environment interaction-wide association studies (GEIWAS) [248], can be developed to help elucidate statistical interactions across levels. Since it is clear that biological, social, and environmental factors interact in some way in cancer etiology, a multilevel framework is needed to both organize and guide traditionally separate fields of cancer research; however, these frameworks should also account for the dynamic nature of the disease.

Expanding MBASIC: Levels of Intervention, Implementation and Evaluation

MBASIC expands the utility of the multilevel approach by including levels of etiology and carcinogenesis with levels of intervention and implementation/evaluation, all of which can influence one another in a nonlinear manner. Levels of intervention are characterized by primary, secondary, and tertiary prevention strategies and survivorship that range from risk assessment to detection to diagnosis and treatment (**Figure 1**). Previous multilevel studies have focused on assessing factors within the levels of intervention [14], particularly cancer care outcomes like detection or screening at the individual level or practice setting sub-level [3, 4].

The implementation/evaluation level is characterized by changes made through the application and translation of relevant interventions [249]. Implementation/evaluation may occur through national, state, or local policy or health care systems changes, and the impact of interventions and implementation will ultimately be seen in changes to the health status of a population. The levels of implementation/evaluation are based on the translational model of Khoury et al. [220, 221], which describes five translational phases (**Figure 1**): T0/T1 (determination of mechanisms, etiology and development of interventional strategies); T2 (development of evidence-based policy and practice); T3 (implementing evidence-based guidelines to elicit health care system changes); and T4 (surveillance and monitoring the effect of changes on health outcomes in populations). Appropriate consideration of the dissemination, implementation, and evaluation of research findings into health systems is critical if the potential of multilevel models is to be realized. For instance, knowledge of the role of macro-environmental factors (e.g., residential location, social environment) in individuals with specific biological characteristics and risk factor profile could provide a resource-efficient approach to early detection or screening for cancer.

Simultaneous consideration of multiple levels in the MBASIC framework may impact a number of cancer outcomes. The levels or sublevels of inference (etiology, carcinogenesis, intervention, or implementation/evaluation; **Figure 1**) could serve as the outcome or exposure of interest. For instance, healthcare system changes (e.g., insurance coverage) can affect individual level behavior (e.g., participation in smoking cessation programs), which can affect the cellular environment (e.g., carcinogen levels and formation of DNA adducts). Therefore, interactions within and across levels can be modeled in a variety of ways, and extent to which the four levels of inference impact the

trait of interest will vary depending on the etiological setting. For instance, during cancer initiation, living in a community that promotes cancer screening and having access to primary care may play a prominent role in cancer early detection(ref). After cancer is diagnosed, the oncology provider and social support may become a predominant influence on clinical and psychosocial outcomes. Both of these scenarios may be imposed on a common biological context (e.g., a cancer having a specific mechanistic cause), but the relevant individual and macro-level factors may differ substantially.

MBASIC Example: Prostate Cancer

In the United States, prostate cancer (CaP) is the second leading cause of cancer death in men [250]. CaP is of public health concern because it disproportionately affects different races. African American men are more likely to be diagnosed with and die from CaP than any other racial group, and this disparity is the largest observed for any cancer [251]. Despite the burden of CaP, particularly for African American men, little is known about the etiology and predictors of poor prognosis for the disease. At present, the only widely agreed-upon risk factors for CaP are at the individual level: race, age, and family history of CaP [153]. Tumor and patient characteristics used to identify men with a poor prognosis include tumor stage, Gleason score or grade, and prostate-specific antigen (PSA) level at diagnosis. However, these clinical characteristics are imperfect in their ability to determine long-term prognosis and appropriate treatment options. Thus, CaP is a good example of the potential value of the MBASIC framework.

PSA Screening: From the Cellular Level to T4 Implementation

In the 1980s, studies on the cellular level demonstrated that PSA levels could serve as markers for CaP recurrence [252]. The use of PSA screening for patients undergoing treatment was approved in 1986 [252-254]. Despite studies in the late 1980s suggesting that PSA might not be an ideal biomarker for screening and early detection of CaP [252-254], the FDA also approved PSA as an early detection screening test in 1994, and PSA became one of the first FDA approved early detection biomarkers for cancer [252]. The FDA based its approval on a large clinic study whose results suggested that men with PSA values above 4.0mg/mL could be biopsied for cancer[255]. As a result of this bench to beside clinical translation (T1 phase), screening guidelines with often conflicting recommendations from different organizations like the United States Preventive Services Task Force [256], the American Cancer Society [257], and the National Comprehensive Cancer Network [258], started to emerge. These guidelines affected clinical practice (T2 phase), resulting in more men being screened for and diagnosed with CaP [252]. The health care system was also affected by these guidelines: insurance companies, particularly the Veterans Association and Medicare, incurred large costs covering routine PSA screening (T3 phase) [259]. Continued research at both the population level (T4 phase) and levels of causation (cellular and individual levels) in more recent years have shown that PSA screening may not improve CaP mortality rates, that early detection of CaP can often lead to unnecessary treatment for some and insufficient treatment in others [252, 260, 261]. As a result, researchers continue to developed enhanced PSA screening tests that are more sensitive and specific [252]. Guidelines for routine PSA screening are continually being revised in the context of individual level factors. These include questioning the utility of screening for men under the age of 75 [256], focusing on screening high risk groups [202], and recommending baseline PSA measures in men under age 50 [145]. Despite its limitations and controversies, PSA screening illustrates how cellular and individual levels

of causation, resulting biomarker interventions, and health care implementation (**Figure 1**) can inform one another to optimize the early detection of CaP.

The PSA scenario also suggests that a comprehensive evaluation of PSA in early detection of CaP may benefit from the use of MBASIC to frame the hypotheses and approaches needed to improve screening and treatment for CaP. While macro-level factors have yet to be widely used in the context of PSA screening, it is not hard to imagine that screening strategies may be optimized by having a better understanding of those men who are most likely to have unfavorable CaP outcomes based on their socio-economic situation, access to health care, or other macro-environmental factors. The role of macro-environmental factors in CaP risk and mortality are beginning to emerge from the health disparity and PSA screening literature. Screening behaviors can be affected by economic, physical, and social characteristics of residential neighborhoods [262]. Neighborhoods considered to be disadvantaged or low-income have been correlated with higher levels of pollutants, overcrowding, violence, less social cohesion, and less access to services [263].[263]. Screening practices can affect CaP incidence, and low-income neighborhoods often have fewer medical facilities that are overburdened with indigent care to provide optimal screening[262]. This can lead to differential screening practices by neighborhood [264] and differences in both the diagnosis and treatment of CaP, particularly among Caucasian versus African American men [151, 265]. Therefore, neighborhood measures could serve as a surrogate for access to care in CaP, and appear to be a relevant macro-environment level measure to investigate for

this cancer outcome. In the setting of an MBASIC approach, men with known biological risk profiles may therefore benefit from targeted intervention if they also reside in defined disadvantaged neighborhoods. For instance, this concept is further illustrated by a multilevel analysis that investigated the role of individual level characteristics and census-tract neighborhood variables and stage of prostate cancer. Consistent with data showing an association between race, stage, and socioeconomic circumstances like living in a low income area, using geographical information systems technology, Xiao et al. [266] went beyond identifying factors associated with prostate cancer stage and suggest community education and outreach in areas with unfavorable neighborhood characteristics. In the context of MBASIC, discovery and early translation can be leveraged in a single study and can provide additional insights that would not be as readily apparent in studies focusing on a single etiological level.

Prostate Cancer Disparities: Piecing Together Studies on Biology and Neighborhood

Because of the complex etiology of CaP, an understanding of CaP disparities may benefit from a multilevel approach. A growing body of literature supports this hypothesis. Rundle et al. [267] reported an association between neighborhood SES (based on median income level of a census tract) modifies the association between individual smoking status and PAH-DNA adduct levels in prostate tissue (BED). We reported an interaction between CaP genetic susceptibility loci identified in GWAS and census-tract level neighborhood variables on time to PSA failure in men who had undergone radical prostatectomy [32]. We identified no main effects of the genetic

variants or neighborhood factors on PSA failure by themselves, but found statistically significant interactions between neighborhood variables and the susceptibility loci. Specifically, genotypes at *MSMB* and *HNF1B/TCF2* predicted time to PSA failure in men from disadvantaged neighborhoods. This suggests that context-specific effects of genotype should be explored and may improve the ability to identify groups that may experience poor CaP outcomes. It is important to note that these studies represent predictive models that may have implications for implementation or translation, but themselves do not provide direct mechanistic conclusions. In general, these studies may motivate a continued focus on multi-level approaches and provide rationale for the utility of multilevel models in cancers like CaP, where typical single disciplinary approaches provide limited insight into disease etiology.

Charge to the Scientific Community

We have proposed a unifying conceptual framework that allows researchers from public health, policy, oncology, health services research, behavioral science, epidemiology, and the biomedical sciences to test hypotheses of interest under a common framework. The MBASIC framework allows researchers to generate common inferences from otherwise disparate individual research findings by using a common conceptual model. As illustrated by the prostate cancer example, taking a multilevel approach can help to expedite translation of etiologic findings into translational efforts, more than would occur in studies focused on single levels of etiology alone. By providing a stronger basis for inclusion of biological factors in a multilevel hierarchy, MBASIC bridges the gap between social science and biology in order to foster multidisciplinary collaboration and streamline intervention, implementation, and translation efforts. Emerging biomedical

technologies enable population-based studies to include biomarker data such that the landscape of cancer research is changing and the lines between disciplines are increasingly blurring. MBASIC can serve as a road map for hypothesis generation and the development of emerging multidisciplinary teams. The MBASIC framework allows individual studies to more effectively piece together individual research findings under a common conceptual model. Knowledge gained from this integration can be used to rationalize the costs of future, large-scale, multilevel studies. Finally, MBASIC represents a framework around which transdisciplinary research (i.e., research that generates new fields of inquiry) can be built.

References

1. Krieger N. Defining and investigating social disparities in cancer: critical issues. Cancer Causes and Control. 2005;16:5-14.

2. Krieger N, Quesenberry C, Peng T, Horn-Ross P, Stewart S, Brown S, et al. Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, black, Hispanic, and white residents of the San Francisco Bay Area, 1988–92 (United States). Cancer Causes and Control. 1999;10:525-37.

3. Hiatt RA, Breen N. The Social Determinants of Cancer: A Challenge for Transdisciplinary Science. American journal of preventive medicine. 2008;35:S141-S50.

4. Holmes JH, Lehman A, Hade E, Ferketich AK, Gehlert S, Rauscher GH, et al. Challenges for Multilevel Health Disparities Research in a Transdisciplinary Environment. American journal of preventive medicine. 2008;35:S182-S92.

5. Sciences NCIaNIfEH. Cancer and the Environment. National Institutes of Heatlh; 2003.

6. Aarts MJ, Lemmens VEPP, Louwman MWJ, Kunst AE, Coebergh JWW. Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. European journal of cancer (Oxford, England : 1990). 2010;46:2681-95.

7. Bradley CJ, Given CW, Roberts C. Race, Socioeconomic Status, and Breast Cancer Treatment and Survival. Journal of the National Cancer Institute. 2002;94:490-6.

8. Dayal HH, Polissar L, Dahlberg S. Race, socioeconomic status, and other prognostic factors for survival from prostate cancer. Journal of the National Cancer Institute. 1985;74:1001-6.

9. The Health Consequences of Smoking: A Report of the Surgeon General. In: Services USDoHaH, editor.2004.

10. Lubin Jay H. Environmental Factors in Cancer: Radon. Reviews on Environmental Health2010. p. 33.

11. Orenstein MR, Schenker MB. Environmental asbestos exposure and mesothelioma. Current Opinion in Pulmonary Medicine. 2000;6:371-7.

12. Gibson TM, Ferrucci LM, Tangrea JA, Schatzkin A. Epidemiological and Clinical Studies of Nutrition. Seminars in oncology. 2010;37:282-96.

13. Na H-K, Oliynyk S. Effects of physical activity on cancer prevention. Annals of the New York Academy of Sciences. 2011;1229:176-83.

14. Varghese JS, Easton DF. Genome-wide association studies in common cancers—what have we learnt? Current Opinion in Genetics & Development. 2010;20:201-9.

15. Hecht SS. Tobacco Smoke Carcinogens and Lung Cancer. Journal of the National Cancer Institute. 1999;91:1194-210.

16. Thun MJ, Henley SJ, Burns D, Jemal A, Shanks TG, Calle EE. Lung Cancer Death Rates in Lifelong Nonsmokers. Journal of the National Cancer Institute. 2006;98:691-9.

17. Rebbeck TR, Kauff ND, Domchek SM. Meta-analysis of Risk Reduction Estimates Associated With Risk-Reducing Salpingo-oophorectomy in BRCA1 or BRCA2 Mutation Carriers. Journal of the National Cancer Institute. 2009;101:80-7.

18. Plewis I. *Statistics in Education*. London: Edward Arnold; 1997.

19. MacMahon B, Pugh T. Epidemiology: Principles and Methods.: MacMahon B, Pugh TF. Epidemiology: Principles and Methods. Boston: Little, Brown and Company, 1970 . 1970.

20. Judith S. Mausner J, Kamer S, Bahn A. Epidemiology: An Introductory Text: Saunders; 1985.

21. Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol. 2010;39:97-106.

22. Warnecke RB, Oh A, Breen N, Gehlert S, Paskett E, Tucker KL, et al. Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. Am J Public Health. 2008;98:1608-15.

23. Taplin SH, Anhang Price R, Edwards HM, Foster MK, Breslau ES, Chollette V, et al. Introduction: Understanding and Influencing Multilevel Factors Across the Cancer Care Continuum. JNCI Monographs. 2012;2012:2-10.

24. Gorin SS, Badr H, Krebs P, Das IP. Multilevel Interventions and Racial/Ethnic Health Disparities. JNCI Monographs. 2012;2012:100-11.

25. Glass TA, McAtee MJ. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. Soc Sci Med. 2006;62:1650-71.

26. Caporaso NE. Integrative study designs--next step in the evolution of molecular epidemiology? Cancer Epidemiol Biomarkers Prev. 2007;16:365-6.

27. Spitz MR, Caporaso NE, Sellers TA. Integrative cancer epidemiology--the next generation. Cancer Discov. 2012;2:1087-90.

28. Rappaport SM. Implications of the exposome for exposure science. J Expos Sci Environ Epidemiol. 2011;21:5-9.

29. Gelfand LA, Mensinger JL, Tenhave T. Mediation analysis: a retrospective snapshot of practice and more recent directions. J Gen Psychol. 2009;136:153-76.

30. Hox JJ. *Applied Multilevel Analysis*. Amsterdam: TT-Publikaties; 1994.

31. Kreft I, De Leeuw J. *Introducing Multilevel Modelling*. London: Sage; 1998.

32. Raudenbush SW, Willms JDE. *Schools, Classrooms and Pupils*. San Diego: Academic Press.; 1991.

33. Morrissey JP, Lich KH, Price RA, Mandelblatt J. Computational Modeling and Multilevel Cancer Control Interventions. JNCI Monographs. 2012;2012:56-66.

34. Office of Behavioral and Social Sciences Research NIoH. Toward Higher Levels of Analysis: Progress and Promise in Research on Social and Cultural Dimensions of Health. Bethesda, MD2000.

35. Lam TK, Spitz, M, Schully, SD, Khoury, MJ. Drivers of Translational Cancer Epidemiology in the 21st Century: Needs and Opportunities. Accepted article. 2012.

36. Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? Genet Med. 2007;9:665-74.

37. Khoury MJ, Gwinn M, Bowen MS, Dotson WD. Beyond Base Pairs to Bedside: A Population Perspective on How Genomics Can Improve Health. American Journal of Public Health. 2011;102:34-7.

38. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. Int J Epidemiol. 1981;10:383-7.

39. Thompson WD. Effect modification and the limits of biological inference from epidemiologic data. J Clin Epidemiol. 1991;44:221-32.

40. Cortessis V, Thomas DC. Toxicokinetic genetics: an approach to gene-environment and gene-gene interactions in complex metabolic pathways. IARC Sci Publ. 2004:127-50.

41. Perera FP. Molecular Epidemiology: Principles and Practices. In: Schulte PA, Perera, F.P., editor. San Diego: Academic Press; 1993.

42. Perera F, Weinstein, IB. Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. J Chronic Dis. 1982;35:581-600.

43. Guy M, et al. Identification of new genetic risk variants for prostate cancer. Asian J Androl. 2009;11:49-55.

44. Perera FPF. Molecular cancer epidemiology: a new tool in cancer prevention. J Natl Cancer Inst. 1987;78:887-98.

45. Raymond SK. Molecular epidemiology: Principles and practices. Edited by Paul A. Schulte and Frederica P. Perera, 588 pp. San Diego: Academic Press, 1993, $80. Hepatology. 2005;19:547-.

46. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. PLoS ONE. 2010;5:e10746.

47. Thomas DC, Siemiatycki J, Dewar R, Robins J, Goldberg M, Armstrong BG. The problem of multiple inference in studies designed to generate hypotheses. Am J Epidemiol. 1985;122:1080-95.

48. Ioannidis JP, Loy EY, Poulton R, Chia KS. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. Sci Transl Med. 2009;1:7ps8.

49. Swartz MA, Iida N, Roberts EW, Sangaletti S, Wong MH, Yull FE, et al. Tumor Microenvironment Complexity: Emerging Roles in Cancer Therapy. Cancer Research. 2012;72:2473-80.

50. Tainsky MA. Genomic and proteomic biomarkers for cancer: A multitude of opportunities. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2009;1796:176-93.

51. Dove A. Proteomics: Translating Genomics into products. Nature America Inc. 1999.

52. Meltzer PS. Cancer genomics: Small RNAs with big impacts. Nature. 2005;435:745-6.

53. Andriole GL, Crawford ED, Grubb RL, Buys SS, Chia D, Church TR, et al. Prostate Cancer Screening in the Randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: Mortality Results after 13 Years of Follow-up. Journal of the National Cancer Institute. 2012;104:125-32.

54. Spratlin JL, Serkova NJ, Eckhardt SG. Clinical Applications of Metabolomics in Oncology: A Review. Clinical Cancer Research. 2009;15:431-40.

55. McDermott U, Downing JR, Stratton MR. Genomics and the Continuum of Cancer Care. New England Journal of Medicine. 2011;364:340-50.

56. Foulkes WD. Inherited Susceptibility to Common Cancers. New England Journal of Medicine. 2008;359:2143-53.

57. Project TIHH.

58. Chen G, Wei P, DeStefano AL. Incorporating biological information into association studies of sequencing data. Genetic Epidemiology. 2011;35:S29-S34.

59. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Human Molecular Genetics. 2002;11:2463-8.

60. Fehringer G, Liu G, Briollais L, Brennan P, Amos CI, Spitz MR, et al. Comparison of Pathway Analysis Approaches Using Lung Cancer GWAS Data Sets. PLoS ONE. 2012;7:e31816.

61. Weiner BJ, Lewis MA, Clauser SB, Stitzenberg KB. In Search of Synergy: Strategies for Combining Interventions at Multiple Levels. JNCI Monographs. 2012;2012:34-41.

62. Stokols D. Establishing and maintaining healthy environments: Toward a social ecology of health promotion. American Psychologist. 1992;47:6-22.

63. Stokols D. Translating social ecological theory into guidelines for community health promotion. Am J Health Promot. 1996;10:282-98.

64. Lobo I. Same genetic mutation, different genetic disease phenotype. Nature Education. 2008;1.

65. Cornelis MC, Tchetgen Tchetgen EJ, Liang L, Qi L, Chatterjee N, Hu FB, et al. Gene-Environment Interactions in Genome-Wide Association Studies: A Comparative Study of Tests Applied to Empirical Studies of Type 2 Diabetes. American Journal of Epidemiology. 2012;175:191-202.

66. Yano EM, Green LW, Glanz K, Ayanian JZ, Mittman BS, Chollette V, et al. Implementation and Spread of Interventions Into the Multilevel Context of Routine Practice and Policy: Implications for the Cancer Care Continuum. JNCI Monographs. 2012;2012:86-99.

67. Society AC. Cancer Facts and Figures 2010 Tables and Figures. 2010 [cited 2011 Retrieved April 19 2011];

68. Lansdorp PM. Stress, Social Rank, and leukocyte telomere length. Aging Cell. 2006;5:583-4.

69. Crawford ED. Epidemiology of Prostate Cancer. Urology. 2003;62:3-12.

70. De Angelis G, et al. Twenty Years of PSA: From Prostate Antigen to Tumor Marker. Rev Urol. 2007;9:113-23.

71. Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. Prostate-Specific Antigen as a Serum Marker for Adenocarcinoma of the Prostate. New England Journal of Medicine. 1987;317:909-16.

72. Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJJ, et al. Measurement of Prostate-Specific Antigen in Serum as a Screening Test for Prostate Cancer. New England Journal of Medicine. 1991;324:1156-61.

73. Catalona W, et al. Selection of optimal prostate specific antigen cutoffs for early dection of prostate cancer: receiver operating characteristic curves. J Urol. 1994;152:2037-42.

74. Force USPST. Screening for Prostate Cancer. 2012; Available from: http://www.uspreventiveservicestaskforce.org/uspstf/uspsprca.htm

75. American Cancer Society A. Guidelines for the Early Detection of Cancer (prostate cancer). In: Society AC, editor.2012.

76. National Comprehensive Cancer Network N. Prostate Cancer Early Detection V.2. NCCN Clinical Practice Guidelines in Oncology. 2007.

77. Vickers AJ, Roobol MJ, Lilja H. Screening for Prostate Cancer: Early Detection or Overdetection? Annual Review of Medicine. 2012;63:161-70.

78. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360:1320-8.

79. Andriole GL, Grubb RL, 3rd, Buys SS, Chia D, Church TR, Fouad MN, et al. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009;360:1310-9.

80. Zeigler-Johnson C. Prostate Cancer Severity Associations with Neighborhood Deprivation. Prostate Cancer. 2011;2011.

81. Subramanian SV, et al. Neighborhood effects on the self-rated health of elders: uncovering the relative importance of structural and service-related neighborhood environments. Journals of Gerontology-Series B. 2006;61:S153-S60.

82. Fowke J, et al. Prostate cancer screening between low-income African-American and Caucasian Men. Urologic Oncology: Seminars and Original Investigations. 2005;23:333-40.

83. Berglund A, Garmo H, Robinson D, Tishelman C, Holmberg L, Bratt O, et al. Differences according to socioeconomic status in the management and mortality in men with high risk prostate cancer. European journal of cancer (Oxford, England : 1990). 2012;48:75-84.

84. Carpenter W, Howard D, Taylor Y, Ross L, Wobker S, Godley P. Racial differences in PSA screening interval and stage at diagnosis. Cancer Causes and Control. 2010;21:1071-80.

85. Xiao H, Gwede CK, Kiros G, Milla K. Analysis of prostate cancer incidence using geographic information system and multilevel modeling. J Natl Med Assoc. 2007;99:218-25.

86. Rundle A, Richards C, Neslund-Dudas C, Tang D, Rybicki BA. Neighborhood socioeconomic status modifies the association between individual smoking status and PAH-DNA adduct levels in prostate tissue. Environmental and Molecular Mutagenesis. 2012:n/a-n/a.

87. Rebbeck TR, Weber AL, Walker AH, Stefflova K, Tran TV, Spangler E, et al. Context-Dependent Effects of Genome-Wide Association Study Genotypes and Macroenvironment on Time to Biochemical (Prostate Specific Antigen) Failure after Prostatectomy. Cancer Epidemiology Biomarkers & Prevention. 2010;19:2115-23.

Table 1: Hierarchical Level Definitions

Level	Sub-Level	Factors at this Level Can Serve As:
Macro-Environment	• Health Policy (National, State, Local) • Community, Neighborhood • Social and Built Environment • Practice Setting and Health Care Providers • Family and Social Support	• Exposures that affect individual risk factors • Exposures that affect biological processes • Contextual variables [32]
Individual	• Behaviors • Exposures • Psychological Determinants • Socioeconomic Factors	• Exposures leading to disease • Intermediates between the macro-environment and disease
Biological	• Tissue • Cell • Somatic Genome (SG) • Inherited Genome (IG)	• Processes leading to disease • Intermediates and biomarkers reflecting the relationship between macro-environmental and individual factors

Table 2: Multilevel Framework Examples

Level	Hiatt and Breen [14]	Warnecke et al. [10]	Taplin et al. [4] Gorin et al. [3]	Morrissey et al. [26]
Macro-Environ-ment	Defined by Factors: Social determinants and Health Care Systems	Defined by sub-levels (from largest to smallest): Social conditions (e.g., discrimination), Institutions (e.g., Families), Neighbor-hood, Social Relationships	Defined by sub-levels: National/state policy, local community, organization or practice setting, health care providers, family/ social support	Defined by sub-levels from Taplin et al. [4] Gorin et al. [3]
Individual	Defined by Factors: Social Determinants, Behavioral/ Psychological factors	Defined by Factors: Age, Socioeconomic status, Education, Obesity, Tobacco Use, Acculturation, Diet, Race	Defined by Factors: Biological Factors, Sociodemographics, insurance coverage, risk status, comorbidities, knowledge, attitudes, beliefs, decision-making preferences, psychological reaction/coping	Similar to Taplin et al. [4] Gorin et al. [3] where individual is described as the patient.
Biologic	Defined by Factors: genes and biomarkers	Defined by Factors: Allostatic Load (e.g., combination of stress markers or other biomarkers), Metabolic Processes, Genetic Mechanisms		Defined by sub-levels (largest to smallest level): Organ, Tissue, Cell, Gene, Molecule, Atom
Primary Outcome of Interest	Cancer Control Continuum (Pre-disease, pre-clinical, incidence, morbidity/	Cancer Health Disparities	Cancer Care Continuum (Risk assessment, primary prevention,	Cancer Care Continuum

	survival, mortality) Interventions (Prevention, early detection, diagnosis/treat-ment, quality of life)		detection, diagnosis, treatment, survivorship, end of life)	

Figure Legends

Figure 1: Multi-level Biological And Social Integrative Construct (MBASIC). This framework includes levels of etiology, carcinogenesis, intervention, and implementation/evaluation, as well as previously defined phases of translation (i.e., T0-T4; [220, 221]. Biological levels include inherited genome (IG), somatic genome (SG), as well as related biomarkers of biologically effective dose (BED), biomarkers of early biological effect (EBE), and biomarkers of altered structure and function (ASF) [225].

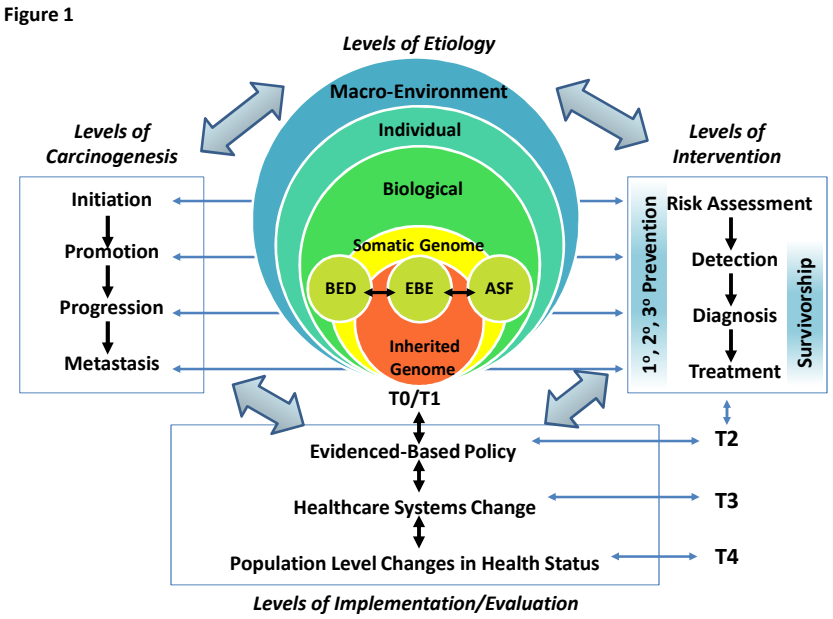
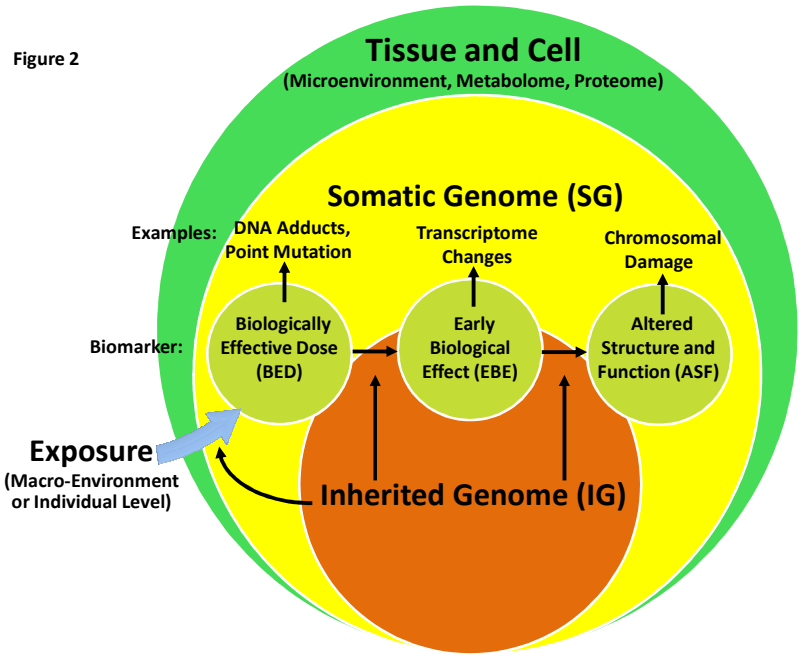


Figure 2: Incorporating Molecular Epidemiology and Biomarkers in the Multilevel Framework.

Biological levels include inherited genome (IG), somatic genome (SG), as well as related biomarkers of biologically effective dose (BED), biomarkers of early biological effect (EBE), and biomarkers of altered structure and function (ASF) [225].



APPENDIX B

Overview of Systematic Review of Multilevel Studies in Cancer Etiologic Research

I. Background. A number of multilevel conceptual models exist in cancer research(see Lynch and Rebbeck, CEBP, 2013). In general, these models state that macro-environment, defined by neighborhood, families, hospitals; individual factors, like behaviors, exposures and social circumstances; and biologic factors, like genes and other molecular markers, can each individually and/or collectively affect cancer outcomes.

II. Problem/hypothesis. A number of these conceptual models exist, but they have not been tested in literature. Very few cancer studies include factors from all 3 levels: macro-environment, individual and biology, or consider macro-environment and biology together, when studying cancer outcomes.

III. Study Aims. The goal is to conduct a systematic review to identify the number of cancer studies that include factors from all 3 levels (or at least the macro and biologic levels) and summarize their findings in order to begin to identify research gaps, evaluate the utility of multilevel conceptual frameworks, and suggest next steps.

IV. Study Methods. A manuscript review was conducted to identify multilevel cancer epidemiology research studies. A multilevel study is defined here as a study that includes the analysis of an association between a cancer outcome and variables at ALL three hierarchal levels, macro-, individual, and biologic levels OR at least the macro- and biologic levels in adult populations. Cancer outcomes could include, incidence, prevalence, mortality, case-control status, or prognosis outcomes. Biologic outcomes could include biomarkers related to oxidative stress, given these markers have been implicated in cancer at the biologic level. Otherwise, articles were excluded if only one or two levels (namely individual and macro-level or individual and biologic level) were represented in the analysis. Gene-environment interaction studies where the environment was measured at the individual level were excluded since they only represented one level outside the individual. An advanced search of the electronic database, PubMed/Medline, was conducted between May-July 2014. Studies from 2002 to the present were identified by entering required terms: “cancer” and “epidemiology”, with alternating key words focused on identifying macro-environmental studies including: family relations (n=63), health disparities (n=276), macroenvironment (n=3), multilevel (n=268), neighborhood (n=284) then combining the following biologic level terms with selected macro-environmental terms listed above, “genotype” and “macroenvironment”(n=1), “biomarker” and “neighborhood” (n=0), and : “gene environment interaction” (n=245). Key words were also paired individually with the terms “biomarker” and “genes” including: “family practice” and “genes” (n=6) and “family practice” and “biomarker” (n=0), “health services” and “genes” (n=27) and “health services” and “biomarker” (n=11), “healthcare disparities” and “genes” (n=1) and healthcare disparities” and “biomarker” (n=1), “poverty” “gene” (n=3) and “poverty” “biomarkers” (n=23), “psychology” and “genes” (n=87) and “psychology” and “biomarker” (n=8), “social perception” and

"genes" (n=1) and "social perception" and "biomarker" (n=0), "social support" and "genes" (n=6) and "social support" and "biomarker" (n=1), "socioeconomic factors" and "genes" (n=31) and "socioeconomic factors" and "biomarker" (n=10), and "state medicine" and "genes" (n=0) and "state medicine" and "biomarker" (n=0). Citations in articles were cross-referenced to obtain additional sources.

V. Results: Studies that met inclusion criteria are summarized in Table 1, and include descriptions of methodological approaches, cancer/biomarker outcomes, and study findings. One study includedfactors at all 3 levels and 4 studies included investigations of macro-environment or neighborhood on biomarkers indicated in cancer.

Appendix B Table 1. Evaluation and Summary of Multilevel Cancer Studies from 2002-Present

	Study Characteristics			**Multi-Level Approaches**				**Method**	
	Study Design	**Sample Size**	**Race/ Gender**	**Outcome**	**Macro-environmental Predictor /con-founders**	**Individual Level Factor(s) Predictors /con-founders**	**Biologic Factors/ con-founders**	**Statistics**	**Findings**
Studies by author, year									
Rebbeck et al, 2010[32]	Cohort Survival Analysis	444	White males	Prostate cancer bio-chemical failure (BF)	**Pre-dictors:** U.S. Census Tract variables: Aging/ social isolation; education ;Housing quality; SES:	**Con-founder:** Age	**Con-founder:** Tumor Stage/ Gleason Grade **Pre-dictors:** 3p12, multiple regions at 8q24, 11q13, 17q24; and candidate genes including *CTBP2* , *HNF1B/TCF2*, *JAZF1*, *LMTK2*, *MSMB*, and *NUDT* , Xp11, *KLK3*, *OATP1B1 RNASEL, MSR1*	Cox models	Significant association found between BF and MSMB and older single heads of households and HNF1B and income.
Barrington et al, 2014[33]	Cross-sectional	543	White men and women	Cortisol reactivity	**Pre-dictors:** Neighbor hood Dep-rivation	**Pre-dictors:** Age, gender, work, education,		Multi-level Growth Curve Model	Significant relationship between Neighbor-hood Deprivation

					Index	fear of crime, social control			and Cortisol Level in women
Waggaman et al, 2014[35]	Cross-sectional	669	Black, His-panic, White women	Pre-cancer cervical lesions	**Pre-dictors:** Pro-portion female black or Hispanic or living below poverty at census tract level	**Con-founders:** Age, race		Poisson Model	Found a a marginally significant interaction ($P < 0.05$) between individual race/ ethnicity and area race
Epplein et al 2012[34]	Cohort	665	336 Black men and woman; 329 White men and women	Sero-negativity to H. pylori and CagA	**Pre-dictors:** US Census 2000 used to find income or wealth; education ; work ;crowding	**Con-founders:** Race (African American ancestry) and Marital status.		Poly-tomous logistic re-gression	Neighbor-hood-level measures of education, work and house values are associated with CagA+ H. pylori sero-prevalence
Need-ham et al., 2014[31]	Cross-sectional	973	Hispanic, Black, White men and women	Telomere Length	Social environ-ment and Neighbor hood Dis-advantage	**Con-founders:** Age, race, lifestyle factors, biomedical factors, socio-economic factors		Linear Multi-level Models	Neighbor-hood social environment was associated with Telomere Length

References:

1. *What is Cancer*. 2015; Available from: http://www.cancer.gov/cancertopics/what-is-cancer.
2. Morrissey, J.P., et al., *Computational Modeling and Multilevel Cancer Control Interventions.* JNCI Monographs, 2012. **2012**(44): p. 56-66.
3. Gorin, S.S., et al., *Multilevel Interventions and Racial/Ethnic Health Disparities.* JNCI Monographs, 2012. **2012**(44): p. 100-111.
4. Taplin, S.H., et al., *Introduction: Understanding and Influencing Multilevel Factors Across the Cancer Care Continuum.* JNCI Monographs, 2012. **2012**(44): p. 2-10.
5. Lynch, S.M. and T.R. Rebbeck, *Bridging the Gap between Biologic, Individual, and Macroenvironmental Factors in Cancer: A Multilevel Approach.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(4): p. 485-495.
6. Plewis, I., *Statistics in Education*1997, London: Edward Arnold.
7. MacMahon, B. and T. Pugh, *Epidemiology: Principles and Methods.*1970: MacMahon B, Pugh TF. Epidemiology: Principles and Methods. Boston: Little, Brown and Company, 1970 .

8. Judith S. Mausner, J., S. Kamer, and A. Bahn, *Epidemiology: An Introductory Text*1985: Saunders.
9. Galea, S., M. Riddle, and G.A. Kaplan, *Causal thinking and complex system approaches in epidemiology.* Int J Epidemiol, 2010. **39**(1): p. 97-106.
10. Warnecke, R.B., et al., *Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities.* Am J Public Health, 2008. **98**(9): p. 1608-15.
11. Glass, T.A. and M.J. McAtee, *Behavioral science at the crossroads in public health: extending horizons, envisioning the future.* Soc Sci Med, 2006. **62**(7): p. 1650-71.
12. Caporaso, N.E., *Integrative study designs--next step in the evolution of molecular epidemiology?* Cancer Epidemiol Biomarkers Prev, 2007. **16**(3): p. 365-6.
13. Spitz, M.R., N.E. Caporaso, and T.A. Sellers, *Integrative cancer epidemiology--the next generation.* Cancer Discov, 2012. **2**(12): p. 1087-90.
14. Hiatt, R.A. and N. Breen, *The Social Determinants of Cancer: A Challenge for Transdisciplinary Science.* American journal of preventive medicine, 2008. **35**(2): p. S141-S150.
15. Rappaport, S.M., *Implications of the exposome for exposure science.* J Expos Sci Environ Epidemiol, 2011. **21**(1): p. 5-9.
16. Savage, S.A., S.M. Gadalla, and S.J. Chanock, *The Long and Short of Telomeres and Cancer Association Studies.* Journal of the National Cancer Institute, 2013.
17. Sanders, J.L. and A.B. Newman, *Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither?* Epidemiologic Reviews, 2013. **35**(1): p. 112-131.
18. Hofmann, J., Hutchinson, A.A., Cawthon, R., Liu, C.S., Lynch, S.M., et al. , *Telomere Length Varies By DNA Extraction Method: Implications for Epidemiologic Research–Letter.* Unpublished, 2014.
19. Cohen, D.A., et al., *Neighborhood physical conditions and health.* Am J Public Health, 2003. **93**(3): p. 467-71.
20. Wardle, J., et al., *Socioeconomic disparities in cancer-risk behaviors in adolescence: baseline results from the Health and Behaviour in Teenagers Study (HABITS).* Prev Med, 2003. **36**(6): p. 721-30.
21. Borrell, L.N., et al., *Perception of general and oral health in White and African American adults: assessing the effect of neighborhood socioeconomic conditions.* Community Dent Oral Epidemiol, 2004. **32**(5): p. 363-73.
22. Zierler, S., et al., *Economic deprivation and AIDS incidence in Massachusetts.* Am J Public Health, 2000. **90**(7): p. 1064-73.
23. Winkleby, M., K. Sundquist, and C. Cubbin, *Inequities in CHD incidence and case fatality by neighborhood deprivation.* Am J Prev Med, 2007. **32**(2): p. 97-106.
24. Adler, N.E., Ostrove, J.M. , *Socioeconomic status and health: What we know and what we don't.* Ann N Y Acad Sci, 1999. **896**: p. 3-15.
25. Diez Roux, A.V., Mair, C., *Neighborhoods and health.* Ann N Y Acad Sci, 2010. **1186**: p. 125-145.
26. Merkin, S.S., et al., *Neighborhoods and Cumulative Biological Risk Profiles by Race/Ethnicity in a National Sample of U.S. Adults: NHANES III.* Annals of epidemiology, 2009. **19**(3): p. 194-201.
27. Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., *Assessing "Neighborhood Effects": Social Processes and New Directions in Research.* Ann Rev Sociol, 2002. **28**: p. 443-478.

28. Messer, L., . Laraia, B., Kaufman, J., Eyster, J., Holzman, C., Culhane, J., et al. , *The development of a standard neighborhood deprivation index. .* Journal of Urban Health, 2006. **83**(6): p. 1041-1062.
29. Diez Roux, A.V. and C. Mair, *Neighborhoods and health.* Annals of the New York Academy of Sciences, 2010. **1186**(1): p. 125-145.
30. McClintock, M.K., et al., *Mammary cancer and social interactions: identifying multiple environments that regulate gene expression throughout the life span.* J Gerontol B Psychol Sci Soc Sci, 2005. **60 Spec No 1**: p. 32-41.
31. Needham, B.L., et al., *Neighborhood characteristics and leukocyte telomere length: The Multi-Ethnic Study of Atherosclerosis.* Health & Place, 2014. **28**(0): p. 167-172.
32. Rebbeck, T.R., et al., *Context-Dependent Effects of Genome-Wide Association Study Genotypes and Macroenvironment on Time to Biochemical (Prostate Specific Antigen) Failure after Prostatectomy.* Cancer Epidemiology Biomarkers & Prevention, 2010. **19**(9): p. 2115-2123.
33. Barrington, W.E., et al., *Neighborhood socioeconomic deprivation, perceived neighborhood factors, and cortisol responses to induced stress among healthy adults.* Health & Place, 2014. **27**(0): p. 120-126.
34. Epplein, M., et al., *Neighborhood socio-economic characteristics, African ancestry, and Helicobacter pylori sero-prevalence.* Cancer Causes & Control, 2012. **23**(6): p. 897-906.
35. Waggaman, C., P. Julian, and L.M. Niccolai, *Interactive effects of individual and neighborhood race and ethnicity on rates of high-grade cervical lesions.* Cancer Epidemiology. **38**(3): p. 248-252.
36. SLAUGHTER, D.P., H.W. SOUTHWICK, and W. SMEJKAL, *Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin.* Cancer, 1953. **6**(5): p. 963-8.
37. Lochhead, P., et al., *Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression.* Mod Pathol, 2015. **28**(1): p. 14-29.
38. Giovannucci, E. and S. Ogino, *DNA methylation, field effects, and colorectal cancer.* J Natl Cancer Inst, 2005. **97**(18): p. 1317-9.
39. Ushijima, T., *Epigenetic field for cancerization.* J Biochem Mol Biol, 2007. **40**(2): p. 142-50.
40. Ramachandran, K. and R. Singal, *DNA methylation and field cancerization.* Epigenomics, 2012. **4**(3): p. 243-5.
41. Stearman, R.S., et al., *A macrophage gene expression signature defines a field effect in the lung tumor microenvironment.* Cancer Res, 2008. **68**(1): p. 34-43.
42. Martinez-Outschoorn, U.E., et al., *Oxidative stress in cancer associated fibroblasts drives tumor-stroma co-evolution: A new paradigm for understanding tumor metabolism, the field effect and genomic instability in cancer cells.* Cell Cycle, 2010. **9**(16): p. 3256-76.
43. Heaphy, C.M., et al., *Telomere DNA content and allelic imbalance demonstrate field cancerization in histologically normal tissue adjacent to breast tumors.* Int J Cancer, 2006. **119**(1): p. 108-16.
44. Lynch, S.M., Mitra, N., Ravichandran, K., Spangler, E., Zhou, W., Paskett, E., Gehlert, S., et al., *Telomere Length and Neighborhood Cancerization: Evaluating Biological Response to Unfavorable Exposures.* Submitted for Peer Review, 2015.
45. Diez Roux, A.V., *Complex Systems Thinking and Current Impasses in Health Disparities Research.* Framing Health Matters, 2011. **101**: p. 1627-1635.

46. Wolkowitz, O.M., Mellon, S.H., Epel, E.S., Lin, J., Dhabhar, F.S., Su, Y, et al., *Leukocyte Telomere Length in Major Depression: Correlations with Chronicity, Inflammation and Oxidative Stress - Preliminary Findings.* PLoS ONE, 2011. **6**(3): p. e17837.
47. Parks, C.G., et al., *Telomere Length, Current Perceived Stress, and Urinary Stress Hormones in Women.* Cancer Epidemiology Biomarkers & Prevention, 2009. **18**(2): p. 551-560.
48. Geronimus, A.T., Hicken, M., Keene, D., Bound, J., *"Weathering" and Age Patterns of Allostatic Load Scores Among Blacks and Whites in the United States.* Am J Public Health, 2006. **96**: p. 826-833.
49. McEwen, B., *Protective and Damaging effects of stress Mediators.* N Engl J Med, 1998. **338**: p. 171-179.
50. Londoño-Vallejo, J.A., *Telomere instability and cancer.* Biochimie, 2008. **90**(1): p. 73-82.
51. Koorstra, J.B.M., et al., *Pancreatic Carcinogenesis.* Pancreatology, 2008. **8**(2): p. 110-125.
52. Cunningham, J.M., et al., *Telomere Length Varies By DNA Extraction Method: Implications for Epidemiologic Research.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(11): p. 2047-2054.
53. Wentzensen, I.M., et al., *The Association of Telomere Length and Cancer: a Meta-analysis.* Cancer Epidemiology Biomarkers & Prevention, 2011. **20**(6): p. 1238-1250.
54. Butt, H.Z., et al., *Telomere Length Dynamics in Vascular Disease: A Review.* European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery, 2010. **40**(1): p. 17-26.
55. Branas, C.C. and J.M. MacDonald, *A Simple Strategy to Transform Health, All Over the Place.* Journal of public health management and practice : JPHMP, 2014. **20**(2): p. 157-159.
56. Cheung, A.L.M. and W. Deng, *Telomere dysfunction, genome instability and cancer.* Frontiers in Bioscience, 2008. **13**(6): p. 2075-2090.
57. Blackburn, E.H., *Telomere states and cell fates.* Nature, 2000. **408**(6808): p. 53-56.
58. Starkweather, A.R., et al., *An Integrative Review of Factors Associated with Telomere Length and Implications for Biobehavioral Research.* Nursing Research, 2014. **63**(1): p. 36-50 10.1097/NNR.0000000000000009.
59. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres.* Annu Rev Genet, 2008. **42**: p. 301-34.
60. Muñoz, P., R. Blanco, and M.A. Blasco, *Role of the TRF2 Telomeric Protein in Cancer and Aging.* Cell Cycle, 2006. **5**(7): p. 718-721.
61. Coppé, J.P., Desprez, P.Y., Krtolica, A, Campisi J. , *The senescence-associated secretory phenotype: the dark side of tumor suppression.* Annu Rev Pathol, 2010. **5**: p. 99–118.
62. Davalos, A.R., Coppe, J.P., Campisi, J., Desprez, P.Y. , *Senescent cells as a source of inflammatory factors for tumor progression.* Cancer Metastasis Rev, 2010. **29**: p. 273-83.
63. Jaskelioff, M., Muller, F.L., Paik, J.H., Thomas, E., Jiang, S, Adams, A.C, Sahin, E, Kost-Alimova, M, Protopopov, A, Cadiñanos, J, et al. , *Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice.* Nature, 2010. **469**: p. 102–6. .
64. Sahin, E., Depinho, R.A. , *Linking functional decline of telomeres, mitochondria and stem cells during ageing.* Nature 2010. **464**: p. 520-528.
65. Halvorsen, T.e.a., *Telomerase Activity Is Sufficient To Allow Transformed Cells To Escape from Crisis.* Mol Cell Biol., 1999. **19**(3): p. 1864-1870.

66. Shalev, I., et al., *Stress and telomere biology: A lifespan perspective.* Psychoneuroendocrinology, 2013. **in press**.
67. Epel, E.S., Lin, J., Dhabhar, F.S., Wolkowitz, O.M., Puterman, E., Karan, L., et al., *Dynamics of telomerase activity in response to acute psychological stress.* Brain Behav Immun., 2010. **24**: p. 531-539.
68. O'Donovan A, L.J., Dhabhar FS, Wolkowitz O, Tillie JM, Blackburn E, Epel E *Pessimism correlates with leukocyte telomere shortness and elevated interleukin-6 in post-menopausal women.* . Biol. Psychiatry, 2009. **24**: p. 446-449.
69. O'Donovan, A., Epel, E., Lin, J., Wolkowitz, O., Cohen, B., Maguen, S., Metzler, T., Lenoci, M., Blackburn, E., Neylan, T.C., *Childhood trauma associated with short leukocyte telomere length in posttraumatic stress disorder.* Biol. Psychiatry, 2011a. **70**: p. 465-471.
70. O'Donovan, A., Tomiyama, J., Lin, J., Puterman, E., Adler, N.E., Kemeny, M., Wolkowitz, O.M., Blackburn, E.H., Epel, E.S. , *Stress appraisals and cellular aging: A key role for anticipatory threat in the relationship between psychological stress and telomere length.* Brain, Behavior, and Immun, 2012.
71. Epel, E.S., Blackburn, E.H., Lin, J., Dhabhar, F.S., Adler, N.E., Morrow, J.D., Cawthon, R.M., *Accelerated telomere shortening in response to life stress.* Proc. Natl. Acad. Sci, 2004. **101**: p. 17312-17315.
72. Diez Roux, A.V., et al., *Race/ethnicity and telomere length in the Multi-Ethnic Study of Atherosclerosis.* Aging Cell, 2009. **8**(3): p. 251-257.
73. Aviv, A., et al., *Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR.* Nucleic Acids Research, 2011. **39**(20): p. e134.
74. Elbers, C.C., et al., *Comparison Between Southern Blots and qPCR Analysis of Leukocyte Telomere Length in the Health ABC Study.* The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 2013.
75. Cawthon, R.M., *Telomere measurement by quantitative PCR.* Nucleic Acids Research, 2002. **30**(10): p. e47.
76. Cawthon, R.M., *Telomere length measurement by a novel monochrome multiplex quantitative PCR method.* Nucleic Acids Research, 2009. **37**(3): p. e21.
77. Du, M., et al., *Physical Activity, Sedentary Behavior, and Leukocyte Telomere Length in Women.* American Journal of Epidemiology, 2012. **175**(5): p. 414-422.
78. Codd V., M.M., van der Harst, P., et al. , *Common variants near TERC are associated with mean telomere length.* Nat Genet., 2010. **42**(3): p. 197–199.
79. Warnecke, R.B., et al., *Approaching Health Disparities From a Population Perspective: The National Institutes of Health Centers for Population Health and Health Disparities.* American Journal of Public Health, 2008. **98**(9): p. 1608-1615.
80. Paskett, E.D., et al., *Psychosocial predictors of adherence to risk-appropriate cervical cancer screening guidelines: A cross sectional study of women in Ohio Appalachia participating in the Community Awareness Resources and Education (CARE) project.* Preventive Medicine, 2010. **50**(1–2): p. 74-80.
81. Peek, M.K., et al., *Allostatic Load Among Non-Hispanic Whites, Non-Hispanic Blacks, and People of Mexican Origin: Effects of Ethnicity, Nativity, and Acculturation.* American Journal of Public Health, 2010. **100**(5): p. 940-946.
82. Rebbeck, T.R., et al., *Joint effects of inflammation and androgen metabolism on prostate cancer severity.* International Journal of Cancer, 2008. **123**(6): p. 1385-1389.

83. Krieger, N., Smith, K., Naishadham, D., Harman, C., Barbeau, E.M., *Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health.* Social Science and Medicine, 2005. **61**: p. 1576-1596.
84. Cohen, S., Kessler, R.C., Underwood, L., *Perceived stress scale. Measuring stress: A guide for health and social scientists.* Oxford University Press, 1994. **New York**.
85. Wang, L., et al., *The Effects of Perceived Stress and Life Style Leading to Breast Cancer.* Women & Health, 2012. **53**(1): p. 20-40.
86. Shell, A.M., M.K. Peek, and K. Eschbach, *Neighborhood Hispanic composition and depressive symptoms among Mexican-descent residents of Texas City, Texas.* Social Science & Medicine, 2013. **99**(0): p. 56-63.
87. Radloff, L., *The CES-D scale: a self-report depression scale for research in the general population.* Applied Psychological Measurement, 1977. **1**: p. 385-401.
88. Eaton, W.W., Smith, C., Ybarra, M., Muntaner, C., Tien, A., *Center for Epidemiologic Studies Depression Scale: review and revision (CESD and CESD-R).* In ME Maruish (Ed.). The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, 2004. **3rd Edition**(Volume 3: Instruments for Adults): p. 363-377.
89. Crockett, L.J., Randall, B.., Shen, Y.L., Russell, S.T., Driscoll, A.K., *Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: a national study.* Journal of Consulting and Clinical Psychology, 2005. **73**(1): p. 47-58.
90. Flores, E., Tschann, J., Dimas, J., Bachen, E., Pasch, L., De Groat, C., *Perceived discrimination, perceived stress, and mental and physical health among Mexican-origin adults.* Hispanic Journal of Behavioral Sciences, 2008. **30**(4): p. 401-424.
91. Kimura, M., et al., *Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths.* Nat. Protocols, 2010. **5**(9): p. 1596-1607.
92. Weischer, M., et al., *Short Telomere Length, Cancer Survival, and Cancer Risk in 47102 Individuals.* Journal of the National Cancer Institute, 2013. **105**(7): p. 459-468.
93. Codd, V., et al., *Common variants near TERC are associated with mean telomere length.* Nat Genet, 2010. **42**(3): p. 197-199.
94. Codd, V., et al., *Identification of seven loci affecting mean telomere length and their association with disease.* Nat Genet, 2013. **45**(4): p. 422-427.
95. Maubaret, C.G., Salpea, K.D., Romanoski, C.E., Folkersen, L., Cooper, J.A., et al., *Association of TERC and OBFC1 Haplotypes with Mean Leukocyte Telomere Length and Risk for Coronary Heart Disease* PLoS ONE, 2013. **8**(12): p. e83122. doi:10.1371/journal.pone.0083122.
96. Bojesen, S.E., et al., *Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer.* Nat Genet, 2013. **45**(4): p. 371-384.
97. Nordfjäll, K., et al., *Telomere Length Is Associated With Obesity Parameters but With a Gender Difference.* Obesity, 2008. **16**(12): p. 2682-2689.
98. Levy, D., et al., *Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology.* Proceedings of the National Academy of Sciences, 2010. **107**(20): p. 9293-9298.

99. Hunt, S.C., et al., *Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study.* Aging Cell, 2008. **7**(4): p. 451-458.
100. Hubbard, A.E., et al., *To GEE or Not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health.* Epidemiology, 2010. **21**(4): p. 467-474 10.1097/EDE.0b013e3181caeb90.
101. Aubert, G., , Lansdorp, P. M. , *Telomeres and Aging.* Physiol Rev, 2008. **88**(2): p. 557-579.
102. Engelhardt, M., Kumar, R., Albanell, J., et al., *Telomerase regulation, cell cycle, and telomere stability in primitive hematopoietic cells.* Blood, 1997. **90**(1): p. 182-193.
103. Lynch, S.M., et al., *A prospective analysis of telomere length and pancreatic cancer in the alpha-tocopherol beta-carotene cancer (ATBC) prevention study.* International Journal of Cancer, 2013. **133**(11): p. 2672-2680.
104. Ma, H., et al., *Shortened Telomere Length Is Associated with Increased Risk of Cancer: A Meta-Analysis.* PLoS ONE, 2011. **6**(6): p. e20466.
105. Hubbard, A.E., Ahern, J., Fleischer, N.L., Van der Laan, M., et al., *To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health.* Epidemiology 2010. **21**(4): p. 467-474.
106. National Research Council (US) Panel on Race, E., and Health in Later Life., *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life.*, in *9, Race/Ethnicity, Socioeconomic Status, and Health*, N.B. Anderson, Bulatao, R.A., Cohen B., editors, Editor 2004, National Academies Press (US): Washington (DC).
107. Palloni, A., Arias, E. , *Paradox lost: explaining the Hispanic adult mortality advantage.* Demography, 2004. **41**(385-418).
108. Markides, K.S., Coreil, J., *The health of Hispanics in the Southwestern United States: an epidemiologic paradox. .* Public Health Reports. , 1986. **101**: p. 253-265.
109. Markides, K.S., Eschbach, K., *Hispanic paradox in adult mortality in the United States.* In R. Rogers & E. Crimmins (Eds.), International Handbook of Adult Mortality. New York: Springer, , 2011: p. 227-240.
110. Bleil, M.E., et al., *Depressive symptomatology, psychological stress, and ovarian reserve: a role for psychological factors in ovarian aging?* Menopause, 2012. **19**(11): p. 1176-1185 10.1097/gme.0b013e31825540d8.
111. Weng, N.-p., *Interplay between telomere length and telomerase in human leukocyte differentiation and aging.* Journal of Leukocyte Biology, 2001. **70**(6): p. 861-867.
112. Hunt, S.C., Chen, W., Gardner, J.P., et al., *Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. .* Aging Cell, 2008. **7**(4): p. 451-458.
113. Epel, E.S., et al., *Accelerated telomere shortening in response to life stress.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(49): p. 17312-17315.
114. Mitchell, C., et al., *Social disadvantage, genetic sensitivity, and children's telomere length.* Proceedings of the National Academy of Sciences, 2014. **111**(16): p. 5944-5949.
115. Theall, K.P., et al., *Neighborhood disorder and telomeres: Connecting children's exposure to community level stress and cellular response.* Social Science & Medicine, 2013. **85**(0): p. 50-58.

116. Health, N.I.o. *NIH Centers for Population Health and Health Disparities*. 2014 [cited 2014 September 25, 2014]; Available from: http://cancercontrol.cancer.gov/populationhealthcenters/cphhd/.
117. Lynch, S.M., Peek, M.K., et al., *Race/ethnicity, Psychosocial Factors, and Telomere Length in a Multicenter Setting.* Unpublished, submitted manuscript, 2015.
118. Census, U.S. 2014; Available from: http://quickfacts.census.gov/qfd/meta/long_fips.htm.
119. Krieger, N., Williams, D., Moss, N., *Measuring social class in US public health research: concepts, methodologies, and guidelines.* Annu Rev Public Health, 1997. **18**: p. 341-378.
120. Caughy, M., Hayslett-McCall, K., O'Campo, P. , *No neighborhood is an island: incorporating distalneighborhood effects into multilevel studies of child development competence.* Health and Place, 2007. **13**: p. 778-798.
121. Krieger, N., Chen, J., Waterman, P., Rehkopf, D., Subramanian, S. , *Race/ethnicity, gender, and monitoring socioeconomic gradients in health: A comparison of area-based socioeconomic measures – the public health disparities geocoding project.* Am J Public Health, 2003. **93**: p. 1655-71.
122. Dubowitz, T., Heron, M., Bird, C.E., Lurie, N., Finch, B.K., Basurto-Dávila, R., Hale, L., Escarce, J.J. , *Neighborhood socioeconomic status and fruit and vegetable intake among whites, blacks, and Mexican Americans in the United States.* Am J Clin Nutr, 2008. **87**(6): p. 1883-1891.
123. Hao, L., Naimen, D.Q., *Quantile Regression*. Quantitative Applications in the Social Sciences, ed. T.F. Liao2007, Thousand Oaks: SAGE Publications.
124. Beyerlein, A., et al., *Genetic markers of obesity risk: stronger associations with body composition in overweight compared to normal-weight children.* PLoS One, 2011. **6**(4): p. e19057.
125. Kumar, T.K., *Multicollinearity in Regression Analysis.* The Review of Economics and Statistics, 1975. **57**(3): p. 365-366.
126. Stata. *Stata 12 help for sqreg*. 2012 [cited 2012 August 22nd 2012]; Available from: http://www.stata.com/help.cgi?sqreg.
127. *http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF*. Population, Housing Units, Area, and Density: 2010 - United States -- States; and Puerto Rico 2010 Census Summary File 1 2012 [cited 2015; Available from: http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF.
128. Kananen, L., Surakka, I., Pirkola, S., Suvisaari, J., et al., *Childhood adversities are associated with shorter telomere length at adult age both in individuals with an anxiety disorder and controls.* PLoS One, 2010. **5**(e): p. 10826.
129. Tyrka, L.H., Price, H.T., Kao, B. et al, *Childhood maltreatment and telomere shortening: preliminary support for an effect of early stress on cellular aging.* Biol. Psychiatry, 2010. **67**: p. 531-534.
130. Howden-Chapman, P., *Housing standards: a glossary of housing and health.* J Epidemiol Community Health, 2004. **58**: p. 162-168.
131. Entner Wright, B.R., Caspi, A., Moffit, T.E., et al., *Factors associated with doubled-up housing--a common precursor to homelessness.* Social Service Review, 1998. **72**: p. 92-111.
132. Kellet, J.M., *Crowding and Mortality in London Boroughs.* In: Burridge, R., Ormandy, D., eds., 1993. **Unhealthy housing**(London: Chapman and Hall): p. 209-222.

133. Gabe, J., Williams, P., *Women, crowding, and mental health.* In: Burridge, R., Ormandy, D., eds., 1993. **Unhealthy Housing**(London: Chapman and Hall): p. 191-208.
134. Clark, W.A.V., Deurloo, M.C., Dieleman, F.M. , *Housing consumption and residential crowding in US housing markets.* Journal of Urban Affairs, 2000. **22**(1): p. 49-63.
135. Boggess, L. and J. Hipp, *Violent Crime, Residential Instability and Mobility: Does the Relationship Differ in Minority Neighborhoods?* Journal of Quantitative Criminology, 2010. **26**(3): p. 351-370.
136. Altman, D.G. and P. Royston, *The cost of dichotomising continuous variables.* BMJ, 2006. **332**(7549): p. 1080.
137. Lochhead, P., Chan, A.T., Nishihara, R., Fuchs, C.S., Beck, A.H., Giovannucci, E.G., Ogino, S., *Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression.* Modern Pathology, 2015. **28**: p. 14-29.
138. Dietz, R.D., *The estimation of neighborhood effects in the social sciences: An interdisciplinary approach.* Social Science Research, 2002. **31**: p. 539-575.
139. Branas, C.C., et al., *A Difference-in-Differences Analysis of Health, Safety, and Greening Vacant Urban Space.* American Journal of Epidemiology, 2011.
140. Yeager, M., et al., *Genome-wide association study of prostate cancer identifies a second risk locus at 8q24.* Nat Genet, 2007. **39**(5): p. 645-9.
141. Eeles, R.A., et al., *Multiple newly identified loci associated with prostate cancer susceptibility.* Nat Genet, 2008. **40**(3): p. 316-21.
142. Gudmundsson, J., et al., *Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24.* Nat Genet, 2007. **39**(5): p. 631-7.
143. Kibel, A.S., *A common variant associated with prostate cancer in European and African populations Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorleifsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, Johansson JE, Balter K, Wiklund F, Montie JE, Yu X, Suarez BK, Ober C, Cooney KA, Gronberg H, Catalona WJ, Einarsson GV, Barkardottir RB, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K, deCODE Genetics Inc., Reykjavik, Iceland.* Urol Oncol, 2007. **25**(5): p. 446-7.
144. Varghese, J.S. and D.F. Easton, *Genome-wide association studies in common cancers—what have we learnt?* Current Opinion in Genetics & Development, 2010. **20**(3): p. 201-209.
145. Chen, G., P. Wei, and A.L. DeStefano, *Incorporating biological information into association studies of sequencing data.* Genetic Epidemiology, 2011. **35**(S1): p. S29-S34.
146. Patel, C.J., Bhattacharya, J., Butte, A.J., *An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus.* PLoS One, 2010. **5**(5): p. e10746.
147. American Cancer Society, *Cancer Facts and Figures*, 2009, American Cancer Society: Atlanta.
148. Messer, L., et al., *The Development of a Standardized Neighborhood Deprivation Index.* Journal of Urban Health, 2006. **83**(6): p. 1041-1062.
149. Ziegler-Johnson, C., et al., *Prostate Cancer Severity Associations with Neighborhood Deprivation.* Prostate Cancer, 2011. **2011**.

150. Byers, T.E., et al., *The impact of socioeconomic status on survival after cancer in the United States.* Cancer, 2008. **113**(3): p. 582-591.
151. Carpenter, W., et al., *Racial differences in PSA screening interval and stage at diagnosis.* Cancer Causes and Control, 2010. **21**(7): p. 1071-1080.
152. Lyratzopoulos, G., et al., *Population based time trends and socioeconomic variation in use of radiotherapy and radical surgery for prostate cancer in a UK region: continuous survey.* BMJ, 2010. **340**.
153. Crawford, E.D., *Epidemiology of Prostate Cancer.* Urology, 2003. **62**: p. 3-12.
154. Bureau, U.S.C., *2000 Census Technical Documentation for SF 3.* United States Department of Commerce, 2007.
155. Bureau, U.S.C., *2000 Census Technical Documentation for SF 3 Data.* US Department of Commerce, 2007.
156. Diez Roux, A.V., D.R. Jacobs, and C.I. Kiefe, *Neighborhood characteristics and components of the insulin resistance syndrome in young adults: the coronary artery risk development in young adults (CARDIA) study.* Diabetes Care, 2002. **25**(11): p. 1976-82.
157. Stoll, B.A., *Affluence, Obesity, and Breast Cancer.* Breast J, 2000. **6**(2): p. 146-149.
158. Robert, S.A. and E.N. Reither, *A multilevel analysis of race, community disadvantage, and body mass index among adults in the US.* Soc Sci Med, 2004. **59**(12): p. 2421-34.
159. Robert, S.A., et al., *Socioeconomic risk factors for breast cancer: distinguishing individual- and community-level effects.* Epidemiology, 2004. **15**(4): p. 442-50.
160. *Part III, Chapter 27 Communical and Noncommunicable Diseases, Section 27.31*, in *28*, P.S. Legislature, Editor 1980: Harrisburg, Pennsylvania.
161. Penney, K.L., et al., *Gleason Grade Progression Is Uncommon.* Cancer Research, 2013. **73**(16): p. 5163-5168.
162. Zeigler-Johnson, C., Tierney, A., Rebbeck, T.R., Rundle, A., *Prostate Cancer Severity Associations with Neighborhood Deprivation.* Prostate Cancer 2011.
163. Eeles, R.A., et al., *Multiple newly identified loci associated with prostate cancer susceptibility.* Nat Genet, 2008. **40**(3): p. 316-321.
164. *SEER Program Coding and Staging Manual 2000*, E. National Cancer Institute. Surveillance, and Endpoints Research (SEER). Editor 2000: Bethesda, MD.
165. SEER., *SEER Training Modules*, in *Prostate Cancer* N.I.o.H. U.S. Department of Health and Human Services, National Cancer Institute, Editor 2014: Bethesda, MD.
166. *Rule G* SEER Training Modules 2012 October 30, 2013]; Available from: http://training.seer.cancer.gov/coding/guidelines/rule_g.html.
167. Aickin, M., Gensler, H. , *Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods.* Am J Public Health, 1996. **86**: p. 726-728.
168. Besag, J., York, J., Mollie, A., *Bayesian Image Restoration, with Two Applications in Spatial Statistics.* Ann Inst Statist Math, 1991. **43**(1): p. 1-59.
169. Ru, H., Martino, S., *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.* J.R. Statist. Soc. B, 2008. **71**(2): p. 319-392.
170. Hatcher, L., *Principle Component Analysis.* SAS Institute Inc, 1994.
171. Tabachnick, B.G., Fidell, L.S., *Chapter 13: Principal Components and Factor Analysis. Using Multivariate Statistics. 3.* Northridge, California: California State University, Harper Collins College, 1996. **635-708**.
172. Wilson, W.T., *The Truly Disadvantaged: The Inner City, The Underclass, and Public Policy*, C.U. Press, Editor 1987: Chicago, IL.

173. Kim, J.-O., Mueller, C.W. , Vol 07-014. Newbury Park, California: Sage; 1978., *Factor Analysis: Statistical Methods and Practical Issues.* Newbury Park, CA, 1978. **07-014**(Sage).
174. *Fine Mapping and GWAS*. Nature/Nurture: Genetic epideiology and statistics-how tos and wherefores blog 2014; Available from: https://blogs.sph.harvard.edu/peter-kraft/2013/07/03/fine-mapping-and-gwas/.
175. Meuwissen, T.H. and M.E. Goddard, *Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci.* Genetics, 2000. **155**(1): p. 421-430.
176. *Principal Components Analysis*. http://www.unt.edu/rss/class/mike/6810/Principal%20Components%20Analysis.pdf University of Texas (accessed 2015).
177. Schupp, C.W., Press, D.J., Gomez, S.L., *Immigration factors and prostate cancer survival among Hispanic men in California: does neighborhood matter?* Cancer, 2014. **120**(9): p. 1401-1408.
178. Carriere, G.M., Sanmartin, C., Bryant, H., Lockwood, G., *Rates of cancer incidence across terciles of the foreign-born population in Canada from 2001-2006.* Can J Public Health, 2013. **104**(7): p. e443-449.
179. Guidry, J.J., Aday, L.A., Zhang, D., Winn, R.J., *Transportation as a barrier to cancer treatment.* Cancer Pract, 1997. **5**(6): p. 361-366.
180. Patel, K., Kenerson, D., Wang, H., Brown, B., Pinkerton, H., Burress, M., Cooper, L., Canto, M., Ukoli, F., Hargreaves, M., *Factors influencing prostate cancer screening in low income African Americans in Tennessee.* J Health Care Poor Underserved, 2010. **21**(1 Suppl): p. 114-126.
181. Goodwin, J.S., Hunt, W.C., Samet, J.M., *Determinants of Cancer Therapy in Elderly Patients.* Cancer, 1993. **72**(594-601).
182. Monroe, A.C., Ricketts, T.C., Savitz, L.A., *Cancer in Rural versus Urban Populations: A Review.* J Rural Health, 1992. **8**(3): p. 212-220.
183. Thomson, H., Thomas, S., Sellstrom, E., Petticrew, M., *Housing Improvements for Health and Associated Socio-economic Outcomes.* Cochrane Database Syst Reve, 2013.
184. Thompson, E.E., Krause, N., *Living Alone and Neighborhood Characteristics as Predictors of Social Support in Late Life.* J Gerontol B Psychol Sci Soc Sci, 1998. **53**(6): p. S354-64.
185. Salinas, C.A., et al., *Multiple Independent Genetic Variants in the 8q24 Region Are Associated with Prostate Cancer Risk.* Cancer Epidemiology Biomarkers & Prevention, 2008. **17**(5): p. 1203-1213.
186. United States Census Bureau, U., *Response Rates by State.* 2000.
187. Thomas, G., et al., *Multiple loci identified in a genome-wide association study of prostate cancer.* Nat Genet, 2008. **40**(3): p. 310-315.
188. Lynch, S.M. *Missing Data (Soc 504)*. 2003.
189. Paternoster, L., et al., *Genome-wide Association Study of Three-Dimensional Facial Morphology Identifies a Variant in PAX3 Associated with Nasion Position.* The American Journal of Human Genetics. **90**(3): p. 478-485.
190. Thomas, D., Casey, G, Conti, DV, Haile, RW, Lewinger, JP, Stram, DO., *Methodological Issues in Multistage Genome-wide Association Studies.* Stat Sci, 2009. **24**(4): p. 414-429.
191. Basta, L.A., T.S. Richmond, and D.J. Wiebe, *Neighborhoods, daily activities, and measuring health risks experienced in urban environments.* Social Science & Medicine, 2010. **71**(11): p. 1943-1950.

192. Society, A.C., *Cancer Facts and Figures.* 2011.
193. D'Amico, A.V., Whittington, R., Malkowicz, S., Fondurulia, J., Chen, M. , *The combination of preoperative prostate specific antigen and postoperative pathological findings to predict prostate specific antigen outcome in clinically localized prostate cancer.* Journal of Urology, 1998. **160**(6): p. 2096–2101.
194. *Prostate Cancer Staging*, in *Cancer Staging References*, American Joint Committee on Cancer, Editor 2013.
195. Chodak, G.W., et al., *Results of Conservative Management of Clinically Localized Prostate Cancer.* New England Journal of Medicine, 1994. **330**(4): p. 242-248.
196. Gittes, R.F., *Carcinoma of the Prostate.* New England Journal of Medicine, 1991. **324**(4): p. 236-245.
197. Stephenson, A.J., et al., *Preoperative Nomogram Predicting the 10-Year Probability of Prostate Cancer Recurrence After Radical Prostatectomy.* Journal of the National Cancer Institute, 2006. **98**(10): p. 715-717.
198. Kattan, M.W., T.M. Wheeler, and P.T. Scardino, *Postoperative Nomogram for Disease Recurrence After Radical Prostatectomy for Prostate Cancer.* Journal of Clinical Oncology, 1999. **17**(5): p. 1499.
199. Krieger, N., *Defining and investigating social disparities in cancer: critical issues.* Cancer Causes and Control, 2005. **16**(1): p. 5-14.
200. Krieger, N., et al., *Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, black, Hispanic, and white residents of the San Francisco Bay Area, 1988–92 (United States).* Cancer Causes and Control, 1999. **10**(6): p. 525-537.
201. Holmes, J.H., et al., *Challenges for Multilevel Health Disparities Research in a Transdisciplinary Environment.* American journal of preventive medicine, 2008. **35**(2): p. S182-S192.
202. Sciences, N.C.I.a.N.I.f.E.H., *Cancer and the Environment*, 2003, National Institutes of Heatlh.
203. Aarts, M.J., et al., *Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome.* European journal of cancer (Oxford, England : 1990), 2010. **46**(15): p. 2681-2695.
204. Bradley, C.J., C.W. Given, and C. Roberts, *Race, Socioeconomic Status, and Breast Cancer Treatment and Survival.* Journal of the National Cancer Institute, 2002. **94**(7): p. 490-496.
205. Dayal, H.H., L. Polissar, and S. Dahlberg, *Race, socioeconomic status, and other prognostic factors for survival from prostate cancer.* Journal of the National Cancer Institute, 1985. **74**(5): p. 1001-6.
206. *The Health Consequences of Smoking: A Report of the Surgeon General*, U.S.D.o.H.a.H. Services, Editor 2004.
207. Lubin Jay, H., *Environmental Factors in Cancer: Radon*, in *Reviews on Environmental Health*2010. p. 33.
208. Orenstein, M.R. and M.B. Schenker, *Environmental asbestos exposure and mesothelioma.* Current Opinion in Pulmonary Medicine, 2000. **6**(4): p. 371-377.
209. Gibson, T.M., et al., *Epidemiological and Clinical Studies of Nutrition.* Seminars in oncology, 2010. **37**(3): p. 282-296.

210. Na, H.-K. and S. Oliynyk, *Effects of physical activity on cancer prevention.* Annals of the New York Academy of Sciences, 2011. **1229**(1): p. 176-183.
211. Hecht, S.S., *Tobacco Smoke Carcinogens and Lung Cancer.* Journal of the National Cancer Institute, 1999. **91**(14): p. 1194-1210.
212. Thun, M.J., et al., *Lung Cancer Death Rates in Lifelong Nonsmokers.* Journal of the National Cancer Institute, 2006. **98**(10): p. 691-699.
213. Rebbeck, T.R., N.D. Kauff, and S.M. Domchek, *Meta-analysis of Risk Reduction Estimates Associated With Risk-Reducing Salpingo-oophorectomy in BRCA1 or BRCA2 Mutation Carriers.* Journal of the National Cancer Institute, 2009. **101**(2): p. 80-87.
214. Gelfand, L.A., J.L. Mensinger, and T. Tenhave, *Mediation analysis: a retrospective snapshot of practice and more recent directions.* J Gen Psychol, 2009. **136**(2): p. 153-76.
215. Hox, J.J., *Applied Multilevel Analysis*1994, Amsterdam: TT-Publikaties.
216. Kreft, I. and J. De Leeuw, *Introducing Multilevel Modelling*1998, London: Sage.
217. Raudenbush, S.W. and J.D.E. Willms, *Schools, Classrooms and Pupils*1991, San Diego: Academic Press.
218. Office of Behavioral and Social Sciences Research, N.I.o.H., *Toward Higher Levels of Analysis: Progress and Promise in Research on Social and Cultural Dimensions of Health*, 2000: Bethesda, MD.
219. Lam, T.K., Spitz, M, Schully, SD, Khoury, MJ, *Drivers of Translational Cancer Epidemiology in the 21st Century: Needs and Opportunities.* Accepted article, 2012.
220. Khoury, M.J., et al., *The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention?* Genet Med, 2007. **9**(10): p. 665-674.
221. Khoury, M.J., et al., *Beyond Base Pairs to Bedside: A Population Perspective on How Genomics Can Improve Health.* American Journal of Public Health, 2011. **102**(1): p. 34-37.
222. Siemiatycki, J. and D.C. Thomas, *Biological models and statistical interactions: an example from multistage carcinogenesis.* Int J Epidemiol, 1981. **10**(4): p. 383-7.
223. Thompson, W.D., *Effect modification and the limits of biological inference from epidemiologic data.* J Clin Epidemiol, 1991. **44**(3): p. 221-32.
224. Cortessis, V. and D.C. Thomas, *Toxicokinetic genetics: an approach to gene-environment and gene-gene interactions in complex metabolic pathways.* IARC Sci Publ, 2004(157): p. 127-50.
225. Perera, F.P., *Molecular Epidemiology: Principles and Practices*, P.A. Schulte, Perera, F.P., Editor 1993, Academic Press: San Diego.
226. Perera, F., Weinstein, IB., *Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation.* J Chronic Dis, 1982. **35**(7): p. 581-600.
227. Guy, M., et al., *Identification of new genetic risk variants for prostate cancer.* Asian J Androl, 2009. **11**(1): p. 49-55.
228. Perera, F.P.F., *Molecular cancer epidemiology: a new tool in cancer prevention.* J Natl Cancer Inst, 1987. **78**(5): p. 887-898.
229. Raymond, S.K., *Molecular epidemiology: Principles and practices. Edited by Paul A. Schulte and Frederica P. Perera, 588 pp. San Diego: Academic Press, 1993, $80.* Hepatology, 2005. **19**(2): p. 547-547.

230. Patel, C.J., J. Bhattacharya, and A.J. Butte, *An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus.* PLoS ONE, 2010. **5**(5): p. e10746.
231. Thomas, D.C., et al., *The problem of multiple inference in studies designed to generate hypotheses.* Am J Epidemiol, 1985. **122**(6): p. 1080-95.
232. Ioannidis, J.P., et al., *Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification.* Sci Transl Med, 2009. **1**(7): p. 7ps8.
233. Swartz, M.A., et al., *Tumor Microenvironment Complexity: Emerging Roles in Cancer Therapy.* Cancer Research, 2012. **72**(10): p. 2473-2480.
234. Tainsky, M.A., *Genomic and proteomic biomarkers for cancer: A multitude of opportunities.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2009. **1796**(2): p. 176-193.
235. Dove, A., *Proteomics: Translating Genomics into products.* Nature America Inc., 1999.
236. Meltzer, P.S., *Cancer genomics: Small RNAs with big impacts.* Nature, 2005. **435**(7043): p. 745-746.
237. Andriole, G.L., et al., *Prostate Cancer Screening in the Randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: Mortality Results after 13 Years of Follow-up.* Journal of the National Cancer Institute, 2012. **104**(2): p. 125-132.
238. Spratlin, J.L., N.J. Serkova, and S.G. Eckhardt, *Clinical Applications of Metabolomics in Oncology: A Review.* Clinical Cancer Research, 2009. **15**(2): p. 431-440.
239. McDermott, U., J.R. Downing, and M.R. Stratton, *Genomics and the Continuum of Cancer Care.* New England Journal of Medicine, 2011. **364**(4): p. 340-350.
240. Foulkes, W.D., *Inherited Susceptibility to Common Cancers.* New England Journal of Medicine, 2008. **359**(20): p. 2143-2153.
241. Project, T.I.H.H.
242. Cordell, H.J., *Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans.* Human Molecular Genetics, 2002. **11**(20): p. 2463-2468.
243. Fehringer, G., et al., *Comparison of Pathway Analysis Approaches Using Lung Cancer GWAS Data Sets.* PLoS ONE, 2012. **7**(2): p. e31816.
244. Weiner, B.J., et al., *In Search of Synergy: Strategies for Combining Interventions at Multiple Levels.* JNCI Monographs, 2012. **2012**(44): p. 34-41.
245. Stokols, D., *Establishing and maintaining healthy environments: Toward a social ecology of health promotion.* American Psychologist, 1992. **47**(1): p. 6-22.
246. Stokols, D., *Translating social ecological theory into guidelines for community health promotion.* Am J Health Promot, 1996. **10**(4): p. 282-298.
247. Lobo, I., *Same genetic mutation, different genetic disease phenotype.* Nature Education, 2008. **1**(1).
248. Cornelis, M.C., et al., *Gene-Environment Interactions in Genome-Wide Association Studies: A Comparative Study of Tests Applied to Empirical Studies of Type 2 Diabetes.* American Journal of Epidemiology, 2012. **175**(3): p. 191-202.
249. Yano, E.M., et al., *Implementation and Spread of Interventions Into the Multilevel Context of Routine Practice and Policy: Implications for the Cancer Care Continuum.* JNCI Monographs, 2012. **2012**(44): p. 86-99.
250. Society, A.C. *Cancer Facts and Figures 2010 Tables and Figures*. 2010 [cited 2011 Retrieved April 19 2011].
251. Lansdorp, P.M., *Stress, Social Rank, and leukocyte telomere length.* Aging Cell, 2006. **5**(6): p. 583-584.

252. De Angelis, G., et al., *Twenty Years of PSA: From Prostate Antigen to Tumor Marker.* Rev Urol, 2007. **9**(3): p. 113-123.
253. Stamey, T.A., et al., *Prostate-Specific Antigen as a Serum Marker for Adenocarcinoma of the Prostate.* New England Journal of Medicine, 1987. **317**(15): p. 909-916.
254. Catalona, W.J., et al., *Measurement of Prostate-Specific Antigen in Serum as a Screening Test for Prostate Cancer.* New England Journal of Medicine, 1991. **324**(17): p. 1156-1161.
255. Catalona, W., et al., *Selection of optimal prostate specific antigen cutoffs for early dection of prostate cancer: receiver operating characteristic curves.* J Urol, 1994. **152**(6 Pt 1): p. 2037-2042.
256. Force, U.S.P.S.T. *Screening for Prostate Cancer*. 2012; Available from: http://www.uspreventiveservicestaskforce.org/uspstf/uspsprca.htm.
257. McDonald, J.A., Weathers, B., Barg, F.K., Troxel, A.B., Shea, J.A., Bowen, D., Guerra, C.E., *Donation Intentions for Cancer Genetics Research Among African Americans.* Genet Test Mol Biomarkers, 2012. **16**(4): p. 252-258.
258. National Comprehensive Cancer Network, N., *Prostate Cancer Early Detection V.2.* NCCN Clinical Practice Guidelines in Oncology, 2007.
259. Vickers, A.J., M.J. Roobol, and H. Lilja, *Screening for Prostate Cancer: Early Detection or Overdetection?* Annual Review of Medicine, 2012. **63**(1): p. 161-170.
260. Schroder, F.H., et al., *Screening and prostate-cancer mortality in a randomized European study.* N Engl J Med, 2009. **360**(13): p. 1320-8.
261. Andriole, G.L., et al., *Mortality results from a randomized prostate-cancer screening trial.* N Engl J Med, 2009. **360**(13): p. 1310-9.
262. Zeigler-Johnson, C., *Prostate Cancer Severity Associations with Neighborhood Deprivation.* Prostate Cancer, 2011. **2011**.
263. Subramanian, S.V., et al., *Neighborhood effects on the self-rated health of elders: uncovering the relative importance of structural and service-related neighborhood environments.* Journals of Gerontology-Series B, 2006. **61**(3): p. S153-S160.
264. Fowke, J., et al., *Prostate cancer screening between low-income African-American and Caucasian Men.* Urologic Oncology: Seminars and Original Investigations, 2005. **23**(5): p. 333-340.
265. Berglund, A., et al., *Differences according to socioeconomic status in the management and mortality in men with high risk prostate cancer.* European journal of cancer (Oxford, England : 1990), 2012. **48**(1): p. 75-84.
266. Xiao, H., et al., *Analysis of prostate cancer incidence using geographic information system and multilevel modeling.* J Natl Med Assoc, 2007. **99**(3): p. 218-25.
267. Rundle, A., et al., *Neighborhood socioeconomic status modifies the association between individual smoking status and PAH-DNA adduct levels in prostate tissue.* Environmental and Molecular Mutagenesis, 2012: p. n/a-n/a.
268. Hellenthal N.J., e.a., *Men of Higher Socioeconomic Status Have Improved Outcomes After Radical Prostatectomy for Localized Prostate Cancer.* Urology, 2010. **76**(6): p. 1409-1413.
269. *Prostate Cancer Severity Associations with Neighborhood Deprivation.* Prostate Cancer, 2011. **2011**.

270. Cheng, I., et al., *Socioeconomic status and prostate cancer incidence and mortality rates among the diverse populations of California.* Cancer Causes and Control, 2009. **20**(8): p. 1431-1440.
271. Schymura, M., et al., *Factors associated with initial treatment and survival for clinically localized prostate cancer: results from the CDC-NPCR Patterns of Care Study (PoC1).* BMC Cancer, 2010. **10**(1): p. 152.
272. Marlow, N., et al., *Disparities Associated with Advanced Prostate Cancer Stage at Diagnosis.* Journal of Health Care for the Poor and Underserved, 2010. **21.1**: p. 112-131.

APPENDIX C

Systematic Review of Multilevel studies in Telomere Length.

A manuscript review was conducted to identify epidemiologic, multicenter studies focused on factors affecting telomere length in adult populations. A multicenter study is defined here as a study comprised of individual research projects that collected data independently, then collaborated with other projects, and merged data to analyze telomere length as an overall, combined study outcome. Articles were excluded if only one university or academic center oversaw and was responsible for original data collection under a single study protocol, or if the main association analysis remained stratified by project or center. This is because the methodological concerns associated with varying population characteristics and laboratory methods are often minimized under these circumstances, and these studies have been reviewed previously [17, 52, 53, 58] (**See supplementary methods**).

A search of the electronic database, PubMed/Medline, was conducted in 2014. Studies from 2002 to the present were identified using the key words, “telomere length” paired with “multicenter(n=16)” or “consortium(n=12)” or “registry”(n=12). Citations in articles were cross-referenced to obtain additional sources. Study design, type of risk factor, or disease under study were not selection factors since the focus of this investigation is on methodological considerations in multicenter studies, not overall association findings. Eleven articles were retrieved and 9 met the criteria for inclusion listed below:

1. Research studies that involved adult, human participants and combined data from multiple centers with a primary aim of assessing the relationship between a risk factor and telomere length measured in blood.

2. The study reported clear methodologies for measuring telomere length.

3. It was written in English.

We summarized methodological approaches for each study based on laboratory, population factors, and statistical approaches known to affect telomere length measurements in literature[17, 52, 53, 58, 73] (**Supplementary Table 1**).

Supplementary Table 1. Evaluation and Summary of Methodologies employed in Multi-center,Telomere Length(TL) Association Studies from 2002-Present

	LABORATORY CONSIDERATIONS					POPULATION FACTORS				STATISTICAL METHODS				
	# Labs	**DNA**	**DNA Extract Method**	**Assay[a]**	**Coeff-icient of Vari-ation**	**Mean Ages**	**Race**	**% men**	**Dis-ease**	**Corr-elate w/ Age**	**Telo-mere Out-come**	**TL diff-erence across centers/ labs**	**Statistical Approach**	**Findings**
Studies by author														
Codd et al. March 2010 *(5 Europecohorts)*	5	Blood Leukocyte	Pure-gene; Phenol/ Chloro-form; Qiagen[b]	q-PCR (n=4)[c] / TRF(n=1)	Intra-assay: 1.5-2% Inter-assay: 3.5-3.9%	39.3-60.8	White	43	Heart/ Renal Disease/ no dis-ease	Neg. and Signif-linear regress-ion	Mean TL (T/S ratio or bp); Z-Score adjusts by age and sex	Yes (Mean TL range: 0.82-6.98)	1. Linear Re-gression adjusted by age, center, sex; 2. Fisher Method; 3. Meta-analysis	Same across studies/ various methods that 3q26 gene region (rs12696304) ass-ociates w/ mean TL.
Codd et al. April 2013 *(21 EuropeCohorts)*	5	Blood Leukocyte	Puregene; Phenol/ Chlorofo rm; Qiagen[b]	q-PCR [c]	Intra-assay: 2-5% Inter-assay: 2.7-24.8%	24-71.7	White	42	Heart, no dis-ease	Neg-ative and Sig-nif-icant	Mean TL (T/S ratio or log-transformed T/S ratio); Z-Score adjustsby age and sex	Yes (Mean TL range: .004-3.71)	1. Linear Regression adjusted by age, sex, family; 2. Meta-analysis	Study findings vary slightly by cohort, but loci associated with *TERC*, *TERT*, *NAF1*, *OBFC1* and *RTEL1 are* associated w/ mean TL.
Mau-baret et al. 2013 *(5 EuropeCase-control studies)*	1	Blood Leukocyte	Salting out	q-PCR some TRF assay	Intra-assay: 5%	22.7-69.3	White	85	Heart, no dis-ease	No report	Mean TL (bp from T/S ratio vs TRF plots	Yes (Mean TL range: 6.80-9.82)	1. Linear regression adjusted for age, center, gender, and physical activity; 2. Meta-analysis	Findings vary by study, but overall the previous association between *OBFC1* and TL confirmed
Bojesen et al. 2013 *(Combined 3 European cohorts)*	2	Blood Leukocyte	Phenol/ Chloro-form; Qiagen[b]	q-PCR	No report	52-62	White	0	No dis-ease	No report	Mean TL from Cycle adjust-ments and base pair calculations from the T/S ratio)	Yes. TL adjust by plate and TL as a CT outcome. No center differ-ence	1.Linear regression adjusted for age, plate, sex, principal component and study. 2. Meta-	Findings vary by study in meta-analysis; TERT loci associate w/ longer TL (rs2736108 and rs770552

											and relative change in mean TL per minor allele.		analysis	
Cunn-ingham et al., 2013 *(inter-national biobank, registry, and SPORE studies)*	1	Blood Leukocyte	Phenol-Chloro-form, Pure-gene, Qiagen	q-PCR	CV: 6%	47 - 51	Not re-ported	52	Colon cancer and health y contro l	Not repor-ted	T/S ratio or log-trans-formed T/S ratio and diff-erences in ratios bt DNA ex-traction	Yes, based on DNA ex-traction	ANCOV A	RTL measured by Qiagen extracted DNA is smaller than other methods which could influence incon-sistency across studies
Weis-her et al., 2012 *(2 EuropeCoh orts)*	1	Blood Leukocyte	Phenol/ Chloro-form; Qiagen[b]	q-PCR	CV: 9.3%	56 - 68	White	46	Health y contro l	Neg-ative and Sig-nifi-cant linear re-gressi on	Quartile of TL base pairs (derived from Cycle (CT)) adjust-ments and base pair calcu-lations from the T/S ratio).	Yes. Lab differ-ence assumeba sed on adjust-ment by plate Center differ-ence no report	Chi-square, Kruskal Wallis, p-trend treating TL as an ordinal variable.	TL associate w/ age, bmi, male gender, smoking, alcohol intake.
Levy et al. 2010 *(Combined 4 US Cohorts and 1 European Twin Study)*	1	Blood Leukocyte	Qiagen, Phenol-chloro-form, Pure-gene[b]	TRF assay	No report	35 - 75	White[e] White/ Black[f]	46	Heart dis-ease	report	Mean TL (bp)	No report	Meta-analysis using Linear Regress ion and linear mixed effects regressi on (family studies) adjusted for age, age^2, sex, bmi, smoking .	TERC and OBFC1 associate d with TL; Findings not reported by study.
Hunt et al. 2008 *(2 U.S. studies, one family and one pop-ulation-based)*	1	Blood Leukocyte ; T-cells and neutrophil	Qiagen, Phenol-chloro-form, Pure-gene[b]	TRF and PCR	CV: 1.43% CV: 2.40% q-PCR: 6.40%	19 - 93	White and Black	41	Heart dis-ease	Sign-ificant and neg-ative ass-ociati on with age in whites and blacks , but not by sex	Mean TL age and bmi adjust; Differ-ence in mean TL by race and gender	Not reported	Generali ze estimati ng equatio ns and ex-chang-eable correlati on matrix compari ng race-specific associat ion of age with TL adjusted for sex/BMI	Sex- and BMI-adjusted TL became shorter with age at a steeper slope in blacks than in whites; Findings not reported by study.
Nord-fjall et al. 2008 (2 EuropeCoh orts)	1	Blood Leukocyte	Qiagen, phenol-chloro-form, Pure-gene, Chemag en[b]	q-PCR	CV: 3.96%	44.5 - 60.5	White	52	Can-cer and heart dis-ease	Sig-nifi-cant and Neg-ative	Log-trans-formed T/S ratio adjust for age, sex, and center	No center effet after adjust for age.	Analysis of covarian ce and correlati on	TL associate with an "obesity-phenotyp e" but only in women

[a] Quantitative Polymerase Chain Reaction(q-PCR) is often reported as a ratio of telomere length repeat length (T) to copy number of a single-copy gene or standard(S) DNA, called a T/S ratio; Southern blot assays report telomere length in terms of base pairs(bp).
[b] Had to consult source manuscripts to identify potential sources of DNA extraction and did not find this information for each cohort in some instances. Also, for some cohorts, DNA extraction methods were different in two separate source manuscripts and it's unclear which extraction method was used for the telomere length study.
[c] Different DNA standards(S) were used to generate T/S ratios and were used to explain differences in mean T/S ratios across studies or cohorts.
[e] In the genome-wide association study(GWAS), all were white participants(n=3417).
[f] In the replication of the GWAS findings(n=1893), the study population included whites and blacks and findings differed slightly by race, which could have been due to the small sample size of blacks(n=574).

Supplementary Laboratory Methods

Terminal Restriction Fragment (TRF) assay: Genomic DNA samples were digested with restriction enzymes *Hinf I* (10 U) and *Rsa I* (10 U; Roche), then the digested DNA samples (1-5 µg each) along with molecular weight DNA markers (1-kb DNA ladder plus λ DNA/Hind III fragments; Invitrogen, Carlsbad, CA) were resolved on 0.8% agarose gels and transferred to nitrocellulose membranes by southern blotting. Membranes were hybridized overnight using radioactively-labeled (TTAGGG) probes, and the radioactive signal was detected and digitized using a phosphorImaging system. The phosphorImager signals (adjusted for background) versus DNA migration distances were determined for each sample[91], and mean TRF in kb was determined using Telorun software[91]. Each sample was run in duplicates on separate gels, and the average value in kb was used.

Quantitative Telomere PCR (qPCR): The overall telomere lengths for each experimental sample are determined relative to the reference DNA by comparing the difference in their ratios of the telomere copy number (T) to the single copy gene copy number (S) using quantitative PCR. This ratio has been found to be proportional to

average telomere length[2]. The qPCR reactions are set up as 10ul reactions in a 384 well plate compatible with the Applied Biosystems, 7900 HT. The final DNA concentration for each experimental sample is approximately 20ng diluted in 5 ul of water. Each plate also contains a set of standards (using the reference DNA) spanning an 81-fold range prepared by serial dilution and analyzed in triplicate[1]. These reactions generate the standard curves used for relative quantitation. The multiplex qPCR assay from Cawthon[1] was modified to make it compatible with the ABI 7900 HT. Two master mixes of PCR reagents were prepared, one with the telomere primers (telc and telg) and the other with either the albumin pair (albd, and albu) or the beta-globin pair (hgbu, and hgbd). LTL did not vary by reference gene primer. Five micro-liters of each master mix was added into the appropriate wells. The final concentrations in each PCR reaction were 0.8X SYBR Green I Master Mix (Agilent Technologies), and 900nM of the telomere pair, or 900nM of the albumin pair, or 500nM of the beta-globin pair.

The thermal cycling profile used was 15min at 95°C, 2 cycles of 15s at 94°C, 15s at 49°C, followed by 32 cycles of 15s at 94°C, 10s at 62°C, and 15s at 74°C with data acquisition. The plates were read at 74°C to minimize the interference from the telomere primer-dimers. The ABI software SDS version 2.0 was used to generate two standard curves from each plate, one for the telomere amplification, and the other for the single copy gene. The ratio (T/S) of the telomere copy number (T) to the single gene copy number (S) was generated for each experimental sample, and the value was averaged across the triplicates. The T/S ratios relative to the reference sample were generated using the comparative CT(cycle threshold) method. Center samples were batched analyzed to minimize inter-assay variation.

Supplementary Table 2. Unadjusted Median and Mean LogTelomere Length(TL)by Study Characteristics(No Cancer, n=1261)

	Median TL (kb) (Interquartile Range)[a]	**p-value**[b]	**Mean logTL(SD)**	**p-value**[c]
Age				
Younger age(<=51)	6.35(4.33-8.17)		1.78(0.44)	
Older age(>51)	6.16(4.26-8.17)	0.52	1.76(0.46)	0.49
Gender				
Female	6.33 (4.48-8.27)		1.78(0.44)	
Male	6.01 (4.10-8.00)	0.12	1.74(0.46)	0.13
Race				
Non-Hispanic White	5.96(4.24-7.92)		1.73(0.44)	
African American	6.48(4.37-8.32)		1.77(0.48)	
Hispanic	6.44 (4.42-8.32)	0.03	1.79(0.45)	0.08
Education				
> High School	6.18(4.21-8.06)		1.75(0.45)	
High School/GED	6.05(4.18-8.19)		1.74(0.47)	
<High School	6.44(4.59-8.16)	0.13	1.81(0.42)	0.07
Perceived Stress				
High Stress	6.16(4.23-8.17)		1.76(0.45)	
Low Stress	6.38(4.37-8.15)	0.29	1.78(0.45)	0.72
Depression				
High Depression	6.45(4.50-8.44)		1.79(0.44)	
Low Depression	6.26(4.26-8.03)	0.20	1.75(0.45)	0.22

[a] Medians(interquartile range for the median); [b] p-values comparing characteristics across 3 or more groups using Kruskal Wallis Test, otherwise used Wilcoxon Ranked Sum Test; [c] p-values comparing characteristics across 3 or more groups using ANOVA, otherwise used T-test.

REFERENCES:

1. Cawthon, R.M (2009). Telomere length measurement by novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res*., **37**, e21
2. Cawthon, R.M (2002). Telomere measurement by quantitative PCR. *Nucleic Acids Res*., **30**, e47

APPENDIX D
NWAS Supplementary Material

Phase 0. Data Cleaning

I. Characterization of Data Missingness by Outcome and Neighborhood Variables

Prostate Cancer Outcome. Overall, there were 93,308 incident cases of prostate cancer between 1995-2005 in the State of Pennsylvania. We focused this analysis on Caucasians only (prostate cancer cases=80,575). 112 prostate cancer cases were dropped because their home address was a non-descript P.O. Box address, 6 more were dropped because of missing age and date of diagnosis, leaving 80,457 cases. The main outcome variable, prostate cancer aggressiveness, is a combination variable comprised of tumor stage and grade data. The stage variable is never missing, and the grade variable is missing for 3,371 prostate cancer patients(4.2% of total cases). Thus, we had 77, 086 Caucasian cases left for an analysis with our primary outcome variable; 6, 416 were classified as aggressive cancer and 70,670 cases were classified as "non-aggressive."

Neighborhood variables. Selection of Year 2000 U.S. Census variables for inclusion in this analysis from SF1 and SF3 is summarized in Appendices E and F, respectively. We assessed percent missingness for each census or neighborhood variable from the 2000 US Census SF1(n=8113) and SF3 forms(n=15,521) for just the variables from the US Census and after the join that linked the cancer registry to the US census data. Percentage available are reported in SF1 Supplementary Digital File 1(variables included in analysis) and Digital File 2(variables excluded from analysis), columns B(census file availability) and C(join file availability), and SF3 Supplementary Digital File 3(variables included in analysis) and 3(variables excluded from analysis), and columns B(census file availability) and C(join file availability). Columns B and C correspond to the percentage of census tracts reporting a value(either a percentage, mean, or median) for each census variable, which equates to percent non-missingness for that variable. SF1 and SF3 Supplementary Digital file 1 and 3, respectively, show which variables were included in the final NWAS analysis (SF1 n=5,943; SF3 n=10,599), based on having less than 10% missing data; SF1 and SF3 Digital Files 2 and 4, respectively, show which variables had missingness greater than 10% and were excluded from the analysis. Using SF1 Supplementary Digital File 1 as an example, for each census variable, %nonmissingness in column C was slightly higher than %nonmissingness in column B, but the difference(Column D) was generally not more than 3-6.5%. Thus, census or neighborhood variable missingness did not appear to be majorly affected by the incorporation of case status after joining the 2000 U.S. Census data SF1 and SF3 forms to the Pennsylvania State Prostate Cancer Registry (1995-2005).

Census Tract Level. Individual prostate cancer cases were linked to Year 2000 Census SF1 and SF3 forms at the census tract level. Of the reported 3, 135 census tracts in the entire State of Pennsylvania in Year 2000, 3,037 census tracts are represented in our Caucasian population. Thus, 97% of PA census tracts are covered in this analysis. On

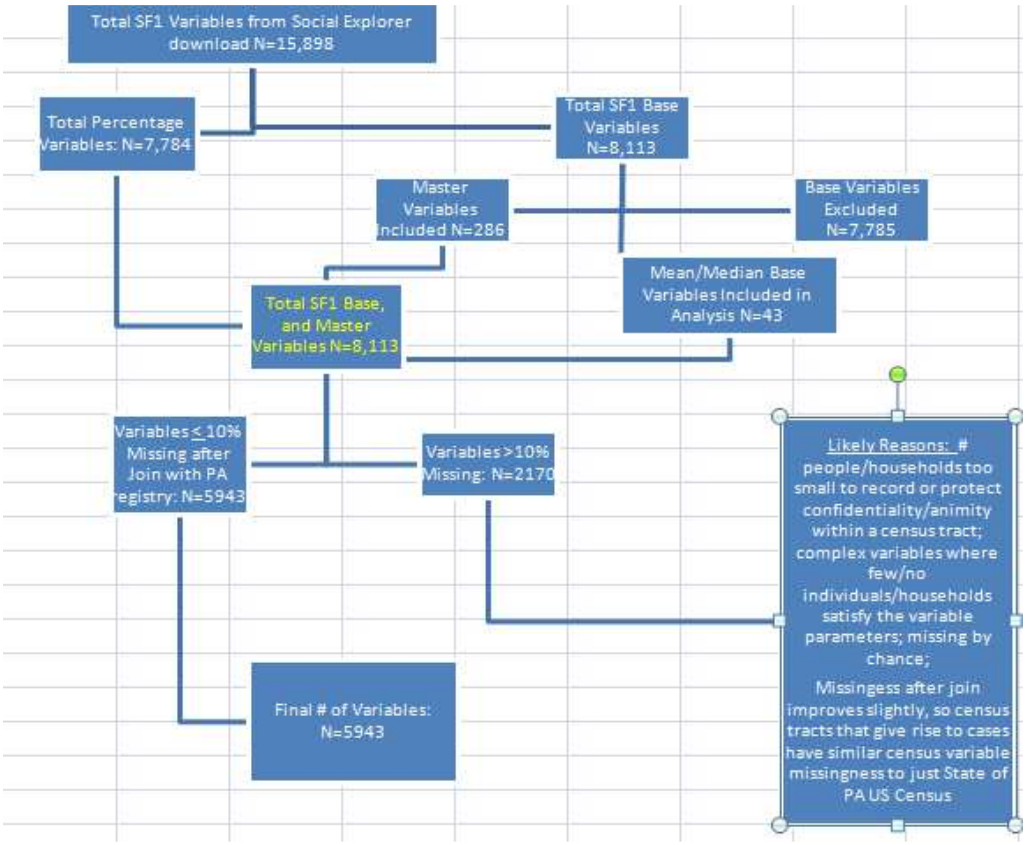
average, there were 2 cases of aggressive prostate cancer compared to 23 controls with non-aggressive prostate cancer in each census tract. We also assessed case status by census block group, but there were <1 cases on average and 8 controls in each census block group. Prostate cancer grade missingness did not appear systematic by census tract (in the grade variable, 3,033 census were represented). Thus, missingness is likely at random.

Statistical Analysis. Using the joined (combined registry and census) final dataset, 1346 (SF1=748; SF3=598) neighborhood variables(8% of the final analytic set) would have been excluded based on census only missingness (missingness>10% in column B in Appendix A for both SF1 and SF3), but were included in the analysis because missingness improved to <10% after the join(column B). After running Phase 1, only 1 of these variables were included in the top 517 hits, and none of these variables were in the top hits after Phase 2, thus findings did not change by this inclusion criteria.

APPENDIX E

Year 2000 SF1 Variable Missingness Descriptions for the State of Pennsylvania
Year 2000 SF1 Variables Pulled from Social Explorer(http://www.socialexplorer.com)
Year 2000 SF1 Variables Linked to Pennsylvania State Prostate Cancer Registry 1995-2005

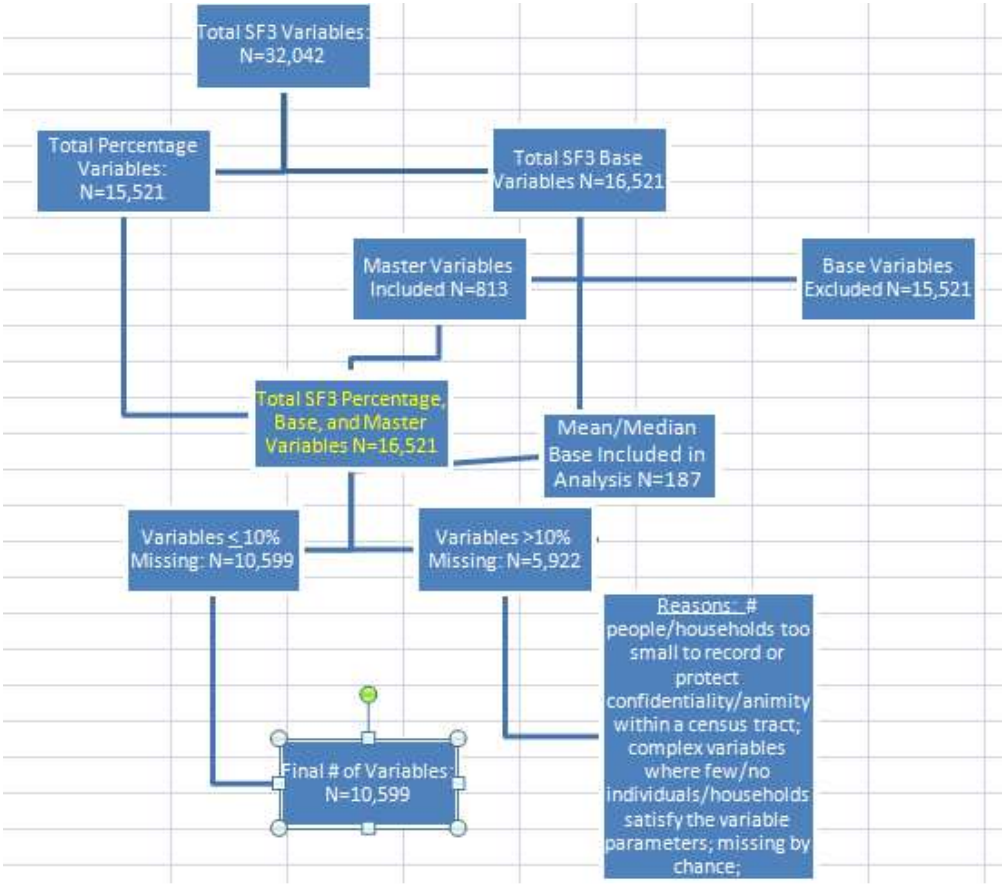
Census Tract Level



Key: Percentage Variables take the Census base variables from which it originates and divides percentage variables by base variables. Social Explorer calculates these variables. Master Variables often serve as denominator variables, such as Total Number of Households. In some cases, it is relevant to include in statistical models, therefore, all master variables with appropriate levels of missingness (i.e. less than 10%) are included in the statistical models (n=286) Some Census Base variables were not represented as percentages, but are relevant because they report means or medians, thus, are included in the analysis (n=43)

APPENDIX F

Year 2000 SF3 Variable Missingness Descriptions for the State of Pennsylvania
Year 2000 SF3 Variables Pulled from Social Explorer(http://www.socialexplorer.com)
Year 2000 SF3 Variables Linked to Pennsylvania State Prostate Cancer Registry 1995-2005



Key: Percentage Variables take the Census base variables from each census table and divide it by the appropriate master variable from the table with which it originates. Social Explorer calculates these variables. Master Variables often serve as denominator variables, such as Total Number of Households. In some cases, it is relevant to include in statistical models, therefore, all master variables with appropriate levels of missingness (i.e. less than 10%) are included in the statistical models

APPENDIX G

Supplementary Discussion Table 1. Examples of Neighborhood Methods used in Prostate Cancer Research

Study (data source)	Neighborhood variables	Outcome	Results	Ref
Method: Neighborhood Indices				
Hellenthal et al. 2010 (CA Cancer Registry)	Principal components analysis used to create an SES score(1-5, 5 being the highest), including median household income, education level, proportion below 200% poverty, and median house value.	PCa Treatment/Survival	Men of lower SES are less likely to undergo radical prostatectomy (RP) or radiation (XRT) for the management of localized prostate cancer. After RP or XRT, men of lower SES have a decreased cancer-specific survival compared with men of higher SES.	[39]

Zeigler-Johnson et al. 2011 (PA Cancer Registry)	1. Analyzed individual SES variables from Census. 2. Calculated a deprivation index (-1 to 1, with 1 being the highest deprivation index) using a principle components analysis (PCA), including: (1) % of households with income <$30,000/year (2) % poverty; (3) % households on public assistance (4) % female head of household with dependent children (5) % households with no car.	Prostate cancer	The highest quartile of neighborhood deprivation was also associated with high Gleason score. For both Caucasians and African-Americans, the highest quartile of neighborhood deprivation was associated with high Gleason score at diagnosis (OR=1.27, 95% CI=1.11-1.44; OR=1.61, 95% CI=1.15-2.25, respectively.) Using a neighborhood deprivation index, associations between prostate cancer severity and neighborhood deprivation across ethnic groups was observed.	[40]
Cheng et al. 2009 (CA Study)	Principal component analysis to develop single SES index from seven census-based indicator variables of SES: 1. mean years of education; 2. median household income; 3. percent living 200% below poverty level; 4.percent blue-collar workers; 5. percent older than 16 years in workforce without	Prostate Cancer risk and mortality	Higher levels of SES were associated with lower mortality rates of prostate cancer deaths (SES Q1 vs. Q5: RR = 0.88; 95% CI: 0.92–0.94). African-Americans had a twofold to fivefold increased risk of prostate cancer deaths in comparison to non-Hispanic Whites across all levels of SES.	[41]

	job; 6. median rent; 7. median house value . This index was used to assign a standardized score to each census block group, which was then categorized into quintile levels.			
Lyratzopoulous et al. 2010 (United Kingdom)	The United Kingdom 2004 Indices of Deprivation.	Prostate Cancer Treatment/Survival	After a diagnosis of prostate cancer, men from lower socioeconomic groups were substantially less likely to be treated with radical surgery or radiotherapy. The causes and impact on survival of such differences remain uncertain.	[38]
Byers et al. 2008 (NPCR POC Study)	Both education and income were classified into 2 Levels (<25% vs 25% of adults aged 25 years with less than a high school education and <20% vs 20% of households with incomes below the Federal Poverty Level). Each patient was then classified as living in a census tract with neither low education nor low income (65% of cases), with only 1 of those indicators	Advanced Prostate, Breast, colon cancer	Low SES was associated with more advanced disease stage and with less aggressive treatment for all 3 cancers.	[36]

	of low SES (20% of cases), or with both of those indicators (15% of cases).			
Schymura et al. 2010 (CDC-NPCR PoC1)	Records were linked by census tract. The following variables were analyzed individually: 1. poverty (<20% versus 20%+ of residents below the 2000 poverty level); 2. education (<25% versus 25%+ of residents age twenty-five and over with less than a high school education); 3.working class status (<66% versus 66%+ working class occupations); 4. urban-rural residence (totally urban, totally rural, urban-rural mix, or unknown).	Prostate Cancer survival	No neighborhood variables were associated with survival from localized prostate cancer State of residence was a significant predictor of treatment type and overall survival.	[42]
Marlow et al. 2010 (national cancer database)	Socioeconomic status was classified using the median household income and proportion of population with a high school diploma from patient's ZIP code of residence.	Advanced Prostate Cancer	Patients residing in areas with lower socioeconomic characteristics have significantly increased odds of advanced PCa.	[43]