IN FOCUS

Integrative Cancer Epidemiology—The Next Generation

Margaret R. Spitz[1], Neil E. Caporaso[2], and Thomas A. Sellers[3]

Summary: We outline an integrative approach to extend the boundaries of molecular cancer epidemiology by integrating modern and rapidly evolving "omics" technologies into state-of-the-art molecular epidemiology. In this way, one can comprehensively explore the mechanistic underpinnings of epidemiologic observations in cancer risk and outcome. We highlight the exciting opportunities to collaborate across large observational studies and to forge new interdisciplinary collaborative ventures. *Cancer Discov; 2(12); 1087–90. ©2012 AACR.*

Epidemiologic studies of all designs have contributed to much of our understanding of cancer etiology (1). Yet the discipline has also received its fair share of criticism (2) with charges that it generates "conflicting results" that tend to "confuse" the public and "disorient" policy makers; and that it is "forever sounding false alarms." The field has matured dramatically since molecular epidemiology first emerged as a defined discipline in the late 1980s as an extension of traditional (classical) epidemiologic research to analyze links between disease and both exposure and biologic risk factors (3). This process mandated incorporating biospecimens into classic epidemiologic study designs and enabled the merging of molecular and biochemical markers of exposure and/or early effect with questionnaire data. The goal was to understand the mechanisms of carcinogenesis and the interplay between lifestyle behaviors, exposure, genes, and cancer etiology. We are now transitioning to an era in which applying advanced technologies (including high-throughput platforms for genotyping and sequencing, omics-based approaches for biomarker discovery and targeted therapies, novel imaging opportunities, and advanced statistical and bioinformatic tools) allows us to dissect the molecular basis of carcinogenesis. Early on in the progression of this new research approach, Sellers (4) was already highlighting the need "to carefully consider the perspectives and expertise of epidemiology and genetics in the design, conduct, and execution" of these large studies that were becoming data driven, complex, and expensive. In their recent commentary, "Bigger, Better, Sooner—Scaling Up for Success," Thun and colleagues (5) also pointed to the value of these large-scale collaborative studies. Kuller (6) noted that the history of epidemiologic advances "is intimately intertwined between epidemiology, pathology, and development of new technologies."

The exciting opportunities to collaborate across large observational studies with the requisite tissue and biospecimens and the need to integrate rapidly evolving high-throughput technologies present the challenge of forging new interdisciplinary collaborative ventures. These must transcend existing dichotomies (e.g., environment vs. gene, risk vs. outcome, somatic vs. germline, genetic vs. epigenetic, and laboratory vs. population based) that do little to advance in-depth understanding but can obscure the complex biologic reality that needs full participation of multidisciplinary teams of scientists to unravel. We therefore present an overarching concept of an integrative approach to merge the boundaries of molecular cancer epidemiology with biobehavioral research, tumor molecular genomics, and systems biology approaches, to explore the mechanistic underpinnings of epidemiologic observations in cancer risk and outcome. We argue that the discipline of molecular epidemiology can provide the framework for the study designs that enable this type of team science.

INTEGRATIVE EPIDEMIOLOGY DEFINITION

Integrative epidemiology was conceived of as a cohesive approach to combine the rigor of epidemiologic study design with the rapid advances in analytic systems and biostatistical and bioinformatic tools mentioned above, using the same populations, biospecimens, and data elements as in case–control or cohort studies of risk to extend to studies of outcome and response to therapy, as well as cancer risk-taking behaviors (e.g., nicotine dependence or physical inactivity). It builds upon the theory that gene discovery and elucidation of broader molecular mechanisms move back and forth between studies of molecular epidemiology and those of tumor molecular genetics, and of intermediate phenotypes, thereby enriching and informing all disciplines (7, 8). Caporaso (8) has argued that such an approach is efficient and although the cost of larger studies is greater, the marginal cost per unit of information is actually lower and the scientific payoff greater. A unifying premise in the concept of integrative epidemiology is that changes in the function of a single gene or pathway can contribute to susceptibility to carcinogenic exposure, predisposition to cancer development, the patient's prognosis, and prediction of response to therapy (7). Integrative cancer epidemiology (7, 8) represents a coalescence of diverse research interests and methodologies relevant across other fields of medicine, for example, cardiovascular epidemiology.

Authors' Affiliations: [1]The Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas; [2]Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; and [3]Moffitt Cancer Center, Tampa, Florida

Corresponding Author: Margaret R. Spitz, The Dan L. Duncan Cancer Center, Baylor College of Medicine, One Baylor Plaza, Cullen Building, Suite 450-A8, Mail Stop: BCM-305, Houston, TX 77030. Phone: 713-798-2115; Fax: 713-798-2716; E-mail: spitz@bcm.edu

Since we first wrote about integrative epidemiology in 2005, we have witnessed the development and increasing availability of high-throughput genotyping platforms and massively parallel sequencing that provide orders of magnitude improvement in throughput over our early candidate gene studies using simple TaqMan platforms or Sanger sequencing approaches, enabling "genome-wide" applications that were not possible previously. This integrative concept has been drastically reshaped by the scale of data throughput now possible. The challenge for epidemiologists is to rigorously apply the principles of observational science to these new approaches, including attention to study design, data and sample collection, and marker validation that are hallmarks of high-quality research. Linking of tissue repositories with well-characterized epidemiologic, clinical, phenotypic, and omic data is another requisite. It follows, therefore, that epidemiologists must be reeducated in integrating modern and rapidly evolving "omic" technologies into state-of-the-art molecular epidemiology, as well as becoming facile in incorporating diverse and high-dimensional data. No single scientist can be accomplished in all types of research endeavors, but we all need to appreciate the accelerating pace of new technologies, understand each others' languages, and recognize the potential applications to population studies, to overcome the barriers imposed by our individual disciplines and to effectively communicate and collaborate.

ADVANCES IN INTEGRATIVE EPIDEMIOLOGY

Examples in the literature hint of the broad emergence of these integrative approaches. Ogino and Stampfer (9) urged the incorporation of molecular pathology into traditional epidemiologic studies to examine the relationship between exposures and molecular signatures in the tumor as well as the interactive influences of exposure and molecular features on tumor progression. They termed this approach "molecular pathological epidemiology." Thomas (10) has pointed out the largely untapped potential of genome–environment-wide interaction studies (GEWIS), and stressed the importance of well-designed studies with careful measurement and efficient analysis of both genetic and environmental factors. Khoury and Wacholder (11) also supported the notion that agnostic approaches to interrogating the human genome for genetic risk factors could be extended into a similar approach for gene environment-wide interaction studies. Approaches to evaluate environmental factors associated with disease have not yet yielded the hoped-for technical advances, such as a "chip" or standard bioassays that can broadly survey exposures, although newer metabolomics technologies hold potential. Patel and colleagues (12) have proposed borrowing the genome-wide association study (GWAS) methodology to create a model environmental-wide association study (EWAS) to search for environmental factors associated with disease on a broad scale.

Exome sequencing has proved successful in the identification of genes that cause some rare Mendelian diseases, although many familial cancers remain unexplained after the first wave of exomic exploration, attributable perhaps to how early we still are in applying this technology. Studies are in progress to assess whether a component of the missing heritability of many common disorders resides in rare gene variants of moderate or low penetrance that are potentially tractable by exome sequencing (13), in duplications or deletions (14), or in more subtle changes found in noncoding regulatory regions of the genome. Epigenomic profiling technologies have reached the stage at which large-scale epigenome-wide association studies are also becoming feasible. The correlations that have been observed between genotype and epigenotype (methQTLs) are encouraging for the prospects of further integrated analysis (15).

Garnett and colleagues (16) outlined how systematic pharmacogenomic profiling in cancer cell lines provides a powerful biomarker discovery platform to guide rational cancer therapeutic strategies. As a translational application of this technology, Platz and colleagues (17) integrated data from an efficient, high-throughput *in vitro* screen with available drug use data from a large, prospective cohort study in a successful proof-of-principle study showing that linking biology and epidemiology can inform new indications for existing drugs. Each of the examples cited reinforces the pivotal role that epidemiology can play in bridging basic and clinical research.

No new technology can substitute for careful selection of population samples and refined hypothesis testing (6). We should focus efforts on "smarter" study designs either within existing cohorts or as ancillary studies. For example, in association studies of genetic variation in cancer risk, it can be assumed that rare variants will be enriched at extreme ends of the phenotype being investigated. Therefore, new studies should consider selecting individuals with extreme phenotypes, such as cancer probands from high-risk families or young-onset cases. As a case in point, in a recent editorial, Cirulli and Goldstein (18) have advocated the value of extreme-trait sequencing because variants that contribute to the trait will be enriched in frequency in such groups. Even small sample sizes may suggest candidate variants that can then be replicated in larger samples. Kazma and Bailey (19) also stress how crucial sample selection is in designing these studies. Population-based designs are more suitable for detecting the effect of multiple rare variants, whereas family-based designs enrich for rare variants, for which the effect likely would be concealed at the population level.

THE FUTURE

Khoury and colleagues (20) recently challenged the cancer epidemiology community to reflect on the critical scientific priorities they will be confronting in the near future. We cite a few opportunities below. Beyond analysis of germline DNA, great value can be added by collection of tumor tissues for integration of somatic genetic alterations and extraction of RNA species for profiling. New opportunities are available for analysis of blood samples for measurement of circulating microRNAs or tumor cells, and the specific requirements to ensure valid measurement of these species must be defined at the start of the study by scientists with expertise in their measurement. Advances in imaging play a major role in early detection of cancer, and such data have proved to be quite valuable in our understanding of breast cancer (through quantification of mammographic density; ref. 21). Indeed, Kumar and colleagues (22) have proposed that the comprehensive interrogation of radiologic images ("radiomics") can reveal and refine early detection by cancer

imaging studies, and doing so within the framework of an epidemiologic study, complemented by other types of biologic, risk factor, and tissue data, may prove to be a powerful approach.

The value and contributions of team science are now recognized as essential to ensure the successful application of advances in technical capabilities to the understanding of underlying biologic complexities (23). To fully empower such collaboration across disciplines, education at many levels is needed. Senior scientists need exposure to new disciplines. Newer scientists need immersion in informatics and emerging technologies, always with the caveat that the ever more rapid march of technology will make continual reeducation mandatory for all. Kuller (6) has pointed out that epidemiologists require a solid background in biologic sciences and an understanding of the new tools being developed, in addition to their solid quantitative skills. Recognizing the evolving challenges being faced in computational resources and data management, the National Cancer Institute sponsored a workshop in 2011 titled "Next Generation Analytic Tools for Large-Scale Genetic Epidemiology Studies of Complex Diseases" (24), highlighting, among other needs, those related to annotation and curation of biologic pathway databases, tools for data visualization, and new open-source, user-friendly analytic tools. The group also recommended improved computational training and support for graduate students and postdoctoral fellows, as such skills are critical to properly leverage and interpret increasingly dense datasets across multiple sources and platforms. It is likely that medical schools and schools of public health will need to develop integrated programs and that grant review committees and funding agencies will have to be reoriented to this new research reality. Multiple discussions will be needed to establish standard research guidelines for this evolving area of research, as suggested by Ogino and colleagues (25), along the lines of STROBE (strengthening the reporting of observational epidemiology; ref. 26). Implementing genetic risk prediction in clinical practice will require a comprehensive evaluation of risk prediction models and recommendations for the reporting of genetic risk prediction studies (GRIPS) that have been proposed to maximize the synthesis of data across multiple studies (27).

We will also be facing challenges in translating scientific discoveries into meaningful interventions at the population level. Khoury and colleagues (28) have discussed the role of "translational epidemiology" along the multidisciplinary research continuum, from basic discovery through evidence guidelines to implementation in practice, and in assessing population health outcomes. Greater efforts in knowledge integration at all phases of the research continuum will be required so that findings will be translated to inform treatment and prevention trials, thereby filling the "translational gap" from discovery to global impact.

In summary, a need exists for rapid and efficient integration of the emerging wealth of genomic, epigenomic, and transcriptomic information for prediction of risk and improvements in disease outcomes. Inevitably, this process will mandate an integrated philosophy and a growing emphasis on molecular epidemiology research and the application of research approaches intrinsic to observational science to all aspects of translational research. We advocate this approach, as it offers unprecedented opportunities for discovery of causes, mechanisms, and outcomes of cancer while being attentive to the rigor of study design, careful population selection, and pristine data collection.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

REFERENCES

1. Greenwald P, Dunn BK. Landmarks in the history of cancer epidemiology. Cancer Res 2009;69:2151–62.
2. Taubes G. Epidemiology faces its limits. Science 1995;269:164–9.
3. Perera FP. Molecular cancer epidemiology: a new tool in cancer prevention. J Natl Cancer Inst 1987;78:887–98.
4. Sellers TA. The beginning of the end for the epidemiologic focus on gene-environment interactions? Cancer Epidemiol Biomarkers Prev 2006;15:1059–60.
5. Thun MJ, Hoover RN, Hunter DJ. Bigger, better, sooner—scaling up for success. Cancer Epidemiol Biomarkers Prev 2012;21:571–5.
6. Kuller LH. Invited commentary. The twenty-first century epidemiologist. A need for different training? Am J Epidemiol 2012;176:668–71.
7. Spitz MR, Wu X, Mills G. Integrative epidemiology: from risk assessment to outcome prediction. J Clin Oncol 2005;23:267–75.
8. Caporaso NE. Integrative study designs—next step in the evolution of molecular epidemiology. Cancer Epidemiol Biomarkers Prev 2007;16:365–6.
9. Ogino S, Stampfer M. Lifestyle factors and microsatellite instability in colorectal cancer: the evolving field of molecular pathological epidemiology. J Natl Cancer Inst 2010;102:365–7.
10. Thomas D. Gene–environment-wide association studies: emerging approaches. Nat Rev Genet 2010;11:259–72.
11. Khoury MJ, Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies–challenges and opportunities. Am J Epidemiol 2009;169:227–30; discussion 234–5.
12. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PLoS ONE 2010;5:e10746.
13. Snape K, Ruark E, Tarpey P, Renwick A, Turnbull C, Seal S, et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. Breast Cancer Res Treat 2012;134:429–33.
14. Yang XR, Ng D, Alcorta DA, Liebsch NJ, Sheridan E, Li S, et al. T (brachyury) gene duplication confers major susceptibility to familial chordoma. Nat Genet 2009;41:1176–8.
15. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet 2011;12:529–41.
16. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 2012;483:570–5.
17. Platz EA, Yegnasubramanian S, Liu JO, Chong CR, Shim JS, Kenfield SA, et al. A novel two-stage transdisciplinary study indentifies digoxin as a possible drug for prostate cancer treatment. Cancer Discov 2011;1:68–77.
18. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 2010;11:415–25.

19. Kazma R, Bailey JN. Population-based and family-based designs to analyze rare variants in complex diseases. Genet Epidemiol 2011;35 Suppl 1:S41–7.

20. Khoury MJ, Freedman AN, Gillanders EM, Harvey CE, Kaefer C, Reid BC, et al. Frontiers in cancer epidemiology: a challenge to the research community from the epidemiology and genomics research program at the national cancer institute. Cancer Epidemiol Biomarkers Prev 2012;21:999–1000.

21. Kelemen LE, Sellers TA, Vachon CM. Can genes for mammographic density inform cancer etiology? Nat Rev Cancer 2008;8:812–23.

22. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. J Magn Reson Imaging 2012;30:1234–48.

23. Sellers TA, Caporaso N, Lapidus S, Petersen GM, Trent J. Opportunities and barriers in the age of team science. Cancer Causes Control 2006;17:229–37.

24. Mechanic LE, Chen HS, Amos CI, Chatterjee N, Cox NJ, Divi RL, et al. Next generation analytic tools for large scale genetic epide-

miology studies of complex diseases. Genet Epidemiol 2012;36: 22–35.

25. Ogino S, King EE, Beck AH, Sherman ME, Milner DA, Giovannucci E. Interdisciplinary education to integrate pathology and epidemiology: towards molecular and population-level health science. Am J Epidemiol 2012;176:659–67.

26. Gallo V, Egger M, McCormack V, Farmer PB, Ioannidis JP, Kirsch-Volders M, et al. STROBE Statement. Strengthening the reporting of observational studies in Epidemiology–Molecular Epidemiology (STROBE-ME): an extension of the STROBE statement. PLoS Med 2011;8:e1001117.

27. Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJGRIPS Group. Strengthening the reporting of genetic risk prediction studies: the GRIPS statement. PLoS Med 2011;8:e1000420.

28. Khoury MJ, Gwinn M, Ioannidis JP. The emergence of translational epidemiology: from scientific discovery to population health impact. Am J Epidemiol 2010;172:517–24.

CANCER DISCOVERY

Integrative Cancer Epidemiology––The Next Generation

Margaret R. Spitz, Neil E. Caporaso and Thomas A. Sellers

Updated version	Access the most recent version of this article at: http://cancerdiscovery.aacrjournals.org/content/2/12/1087

Cited articles	This article cites 28 articles, 8 of which you can access for free at: http://cancerdiscovery.aacrjournals.org/content/2/12/1087.full#ref-list-1
Citing articles	This article has been cited by 5 HighWire-hosted articles. Access the articles at: http://cancerdiscovery.aacrjournals.org/content/2/12/1087.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerdiscovery.aacrjournals.org/content/2/12/1087. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.