Opportunities for Translational Epidemiology: The Important Role of Observational Studies to Advance Precision Oncology

Michael Marrone¹, Richard L. Schilsky², Geoff Liu³, Muin J. Khoury¹,⁴, and Andrew N. Freedman¹

Abstract

Within current oncology practice, several genomic applications are being used to inform treatment decisions with molecularly targeted therapies in breast, lung, colorectal, melanoma, and other cancers. This commentary introduces a conceptual framework connecting the full spectrum of biomedical research disciplines, including fundamental laboratory research, clinical trials, and observational studies in the translation of genomic applications into clinical practice. The conceptual framework illustrates the contribution that well-designed observational epidemiologic studies provide to the successful translation of these applications, and characterizes the role observational epidemiology plays in driving the dynamic and iterative bench-to-bedroom, and bedside-to-bench translation continuum. We also discuss how the principles of this conceptual model, emphasizing integration of multidisciplinary research, can be applied to the evolving paradigm in "precision oncology" focusing on multiplex tumor sequencing, and we identify opportunities for observational studies to contribute to the successful and efficient translation of this paradigm. Cancer Epidemiol Biomarkers Prev; 24(3): 484–9. ©2015 AACR.
Observational Epidemiology in Precision Oncology

Because of the long latency period of cancer development, observational study designs (Table 1) have, by necessity, played an important role in the translation (i.e., development, validation, and evaluation) of effective clinical and public health interventions, especially for cancer screening, early detection, and prevention. The proliferation of high-throughput "-omic" technologies provides tremendous opportunities for observational studies to efficiently develop and evaluate many possible gene-treatment hypotheses applied to cancer treatment and precision oncology. To date, the translation of applications for precision oncology has largely emphasized the laboratory as the starting point. However, observational studies have a unique role in refining and validating these hypotheses in real-world settings.

<table>
<thead>
<tr>
<th>Study design</th>
<th>Description</th>
<th>Inferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td>Longitudinal study following a group of people representing a defined characteristic or experience over time.</td>
<td>Identify risk factors for disease or outcome of interests (i.e., survival). Can identify both prognostic factors° and predictive markers° (i.e., gene-drug interaction).</td>
</tr>
<tr>
<td>Case–cohort</td>
<td>Compare all incident cases from existing longitudinal study (e.g., cohort) to random selection of noncases (e.g., controls) from same cohort.</td>
<td>Identify prognostic factors° associated with outcome of interest.</td>
</tr>
<tr>
<td>Nested case–control</td>
<td>Controls from existing longitudinal study (e.g., cohort) matched to all incident cases from the same cohort on certain factors (i.e., age, gender, race, etc.)</td>
<td>Identify prognostic factors° associated with outcome of interest.</td>
</tr>
<tr>
<td>Case–control</td>
<td>Selection of cases and controls based on disease status.</td>
<td>Identify factors associated with outcome used to select cases and controls.</td>
</tr>
<tr>
<td>Case-series (e.g., clinical series)</td>
<td>A group with a common clinical diagnosis and similar exposure or treatment history.</td>
<td>Identify prognostic factors° associated with clinical outcomes.</td>
</tr>
<tr>
<td>Case-only</td>
<td>All cases selected from a cohort during follow-up (including a RCT).</td>
<td>Identify predictive markers (i.e., gene-drug interaction)°.</td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>Observations made in a representative sample of the larger population of interest at one specific point in time.</td>
<td>Assess prevalence of risk factors and burden of disease at a single point in time.</td>
</tr>
<tr>
<td>Registry</td>
<td>Systematic (often mandated by law) reporting of health outcomes (e.g., disease incidence) by clinical and public health professionals.</td>
<td>Assess temporal trends of disease burden within a defined population.</td>
</tr>
</tbody>
</table>

°A measure correlated with a clinical outcome in the setting of natural history or a standard-of-care regimen; it is a variable used to estimate the risk of or time to clinical outcome (5).
°°A measure that identifies patients most likely to be sensitive or resistant to a specific treatment regimen or agent. An effect modifier is particularly useful when that measure can be used to identify the subgroup of patients for whom treatment will have a clinically meaningfully favorable benefit-to-risk profile (5).
point in the bench-to-bedside translational continuum. The discovery of biomarkers able to predict treatment response and the development of molecularly targeted therapies is often portrayed as following a linear, unidirectional process culminating with evaluation in a prospective randomized controlled trial (RCT; ref. 5). Although this might be the prototypical translational model, even this form of translational research is iterative with feedback loops driven by observational epidemiology during the discovery, validation, evaluation, and implementation steps (3). The use of epidemiology and the strengths of observational studies to inform precision oncology have not been generally discussed.

The epidemiologic principles for developing a precision oncology application require explicit description of the population of interest (sociodemographic, clinical, genomic, and treatment characteristics), clearly defining both the genomic and treatment comparisons being made, and capturing the clinical outcomes used to evaluate such comparisons. Attention to these principles, especially the comparisons being made, will help distinguish prognostic factors (predict course of disease regardless of treatment) from predictive factors (able to predict treatment response and outcomes), and will provide a foundation to facilitate further translational research (6). In the development of genomic markers as prognostic and predictive factors, strengths of observational studies include access to diverse patient profiles and cancer types otherwise not necessarily eligible for RCTs designed to address similar questions. Observational studies also allow the examination of large sample sets to interrogate the interaction between a discrete genomic marker and binary treatment variable to predict outcomes. However, observational epidemiology has its limitations in establishing causal effects and is vulnerable to noncausal explanations observed in gene–treatment interactions (e.g., bias and confounding). For example, variation in treatment allocation, dose, and frequency and uneven distribution of risk factors associated with outcomes of interest between treated and untreated groups may allow for potential confounding and selection bias (7).

To overcome such limitations, investigators seeking a more robust measure of effect rely on experimental epidemiology methods, which include the RCT. Many of the inherent qualities of conventional RCTs, including randomized treatment assignments, prespecified participant eligibility criteria, and treatment regimens, rigorous participant follow-up and data collection, and masking participants and trial investigators, substantiate RCTs as the highest levels of evidence to evaluate the efficacy of a clinical or public health intervention. Given their inherent strengths, RCTs have traditionally played a prominent role within the bench-to-bedside continuum. However, conventional RCTs have limited generalizability due to the highly selected patient populations enrolled in trials, and may not efficiently recruit the large number of participants needed to test multiple gene–treatment interactions.

Conceptual Model

To fully realize the potential of precision oncology, it is essential to integrate laboratory discoveries with evidence-based systematic evaluations in both a well-defined clinical setting (e.g., proof-of-principle efficacy studies) and in large heterogeneous populations (e.g., effectiveness study in large populations). This can be done most efficiently by taking advantage of the complementary aspects of laboratory research, clinical trials, and observational studies that use a range of scientific methods to establish a portfolio of evidence. Figure 1 presents our conceptual model, which illustrates the interdependent continuum forming these connections.

The research and development of new cancer therapeutics has historically followed an agnostic approach in preclinical research using in vitro screening of a vast array of molecules followed by animal testing and early-phase clinical trials designed to establish a safe and effective dose. Variation in treatment response may be, in part, due to the heterogeneity of participants based on genetic and other factors. A large proportion of participants with no or suboptimal therapeutic response who share a common molecular profile may produce an overall nonsignificant result for a therapy that is otherwise promising in a different patient subgroup. More recently, development of cancer therapeutics has followed a hypothesis-driven approach focusing on the mechanisms of action to guide drug development. Knowledge of interindividual variation as well as variation in the molecular pathways driving tumorigenesis, observed in clinical research, is fed back into the laboratory as unique genomic and/or molecular targets for future drug development (4, 8, 9). Alternatively, in vitro screening of a library of approved therapeutics used in clinical practice for a variety of noncancer-related indications for growth inhibition could provide new hypotheses for cancer prevention and treatment. Such hypotheses can be evaluated in existing observational cohort studies with long-term follow-up for a large number of participants and many possible cancer outcomes, eventually moving to clinical trials (10, 11).

Although clinical trials are the primary source of information on the utility of therapeutic agents, for a large number of rare cancers, and for certain population subgroups, clinical trial data are often unavailable to assess the utility of a treatment in a population of interest. In addition, older established anticancer drugs were typically assessed at a time when biologic specimens were not collected within the clinical trials that established their clinical utility. In such instances, observational clinical studies can provide information on drug effectiveness in particular subpopulations as well as the incidence of rare acute adverse events and long-term toxicity in particular risk subgroups. This concept can also be extended to nonpharmacologic and technology-driven therapies, such as radiation therapy. This is particularly important for patient subgroups with particular genomic profiles as well as patients typically not represented in trials such as the elderly and those with comorbidities. Discoveries arising from post-marketing surveillance and observational studies can guide researchers in cultivating new hypotheses that can inform basic molecular research. Using this feedback loop, new prospective clinical studies using innovative trial designs (e.g., FOCUS and Lung-MAP trials) can be initiated that incorporate genomic components into their design to establish both the utility and safety of precision therapy based on a person’s genomic and clinical risk profile.

Examples of Observational Epidemiology Driving Translation of Precision Oncology

Discovery of a prognostic marker to advance drug development

There is perhaps no better example of using observational epidemiologic clinical research to translate laboratory discoveries into improved population health outcomes than the development

Cancer Epidemiology, Biomarkers & Prevention
of trastuzumab for treating HER2/neu-positive breast cancer. In the mid 1980s several independent laboratory investigations provided evidence of the biologic mechanisms by which HER2/neu promotes cancer proliferation, and demonstrated increased levels of this marker in a number of different tumor types (12–14). Two key observational studies in women with lymph node-positive breast cancer showed that increased HER2/neu expression was associated with lower disease-free survival and overall survival compared with normal HER2/neu expression (15, 16). The contributions from observational epidemiologic studies, demonstrating the prognostic importance of HER2/neu overexpression (15–17), provided sufficient evidence to drive the development of the first ever molecularly targeted cancer therapy—trastuzumab, a monoclonal antibody targeting tumor cells expressing HER2/neu. Eventually, a phase III trial investigating the efficacy of trastuzumab in participants with HER2/neu amplification confirmed the ability of HER2/neu expression to predict treatment outcomes with trastuzumab (18). During this period, additional observational studies helped to establish the proportion of patients carrying this molecular change in their breast cancer tumors, leading to a clearer understanding of the potential clinical impact and costs associated with this new biomarker-driven therapy that shaped subsequent health care policy (19). The progression through the translational continuum using observational epidemiology methods to build on prior knowledge (laboratory and observational) to inform clinical trial design highlights the reciprocal nature of translational epidemiology (Fig. 1).

Marker discovery and confirmation of predictive value
Since the introduction of anti-EGFR therapy, additional molecular discoveries have led to a greater understanding of the genes involved in the EGFR pathway in colorectal carcinogenesis. After cetuximab (in 2006) and panitumumab (in 2007) received approval by the FDA for treatment of metastatic colorectal cancer, investigators began to wonder whether treatment outcomes observed in the trials were mediated by variation in EGFR expression. Once it became clear that EGFR expression did not predict treatment outcomes with either EGFR-targeted therapy, there was great interest in whether mutations in the downstream coding genes in the EGFR pathway (i.e., KRAS, BRAF, and PI3KCA) could influence outcomes with anti-EGFR therapy (20). Based on the evidence from two observational clinical studies in metastatic colorectal cancer patients undergoing cetuximab therapy, investigators were able to show that individuals whose tumors had a KRAS mutation did not achieve the same clinical response compared with those with wild-type KRAS tumors (21, 22). Further hypothesis-driven “prospective-retrospective” analyses, comparing KRAS wild-type with KRAS-mutant tumors in archived tumor samples from the RCTs that investigated the efficacy and safety of cetuximab and panitumumab, confirmed the ability of KRAS mutation status to predict treatment outcomes (23, 24). Leveraging knowledge of the molecular landscape of the EGFR pathway with the observational data, researchers were able to demonstrate the prognostic characteristics of KRAS, and when combined with the retrospective analyses of trial data validating the gene–treatment interaction (i.e., prediction), developed and validated a new hypothesis of the molecular mechanism mediating treatment response to EGFR-targeted therapy.

Hypothesis generation and discovery of pharmacokinetic mechanisms
Despite the variability in response to remission maintenance therapy with 6-mercaptopurine (6-MP) in children with acute lymphoblastic leukemia, neither the exact mechanism by which 6-MP is able to maintain remission was immediately known, nor were the causes of serious treatment-related toxicities. In a series of observational epidemiologic investigations, researchers linked genetic variability in thiopurinemethyltransferase (TPMT), an important enzyme involved in 6-MP metabolism (25) to levels of 6-thioguanine nucleotides (6-TGN), the active metabolites of 6-MP (26, 27). Additional observational studies reported the association between 6-TGN and 6-MP treatment-related toxicity (i.e., myelosuppression; refs. 28, 29). The sequential accumulation of observational epidemiologic evidence including: (i) the candidate gene analysis; (ii) the correlation between 6-TGN and toxicity; and (iii) the negative correlation between TPMT activity and 6-TGN concentration, allowed researchers to develop a chain of evidence linking the ability of TPMT to predict 6-MP treatment-related toxicity. Despite the absence of direct evidence of the association between TPMT genotype and 6-MP-related toxicities from RCTs, the chain of evidence influenced the Clinical Pharmacogenetics Implementation Consortium Guidelines to advocate for customized 6-MP dosing based on TPMT genotype to limit potential toxicities (30).

Opportunities and Challenges for Observational Epidemiology in the Era of Multiplex Tumor Evaluations
Knowledge gained from high-throughput interrogations of genomic alterations has led to the characterization of dynamic pathways and driver mutations associated with tumorigenesis (4, 31), and has revealed great molecular diversity within histologically similar cancer types. For the first time in the post–GWAS era, genomic information has the potential to become integrated into everyday clinical decision making. At the vanguard of genomic medicine, the current precision oncology paradigm has the technological capabilities to address the single most important question in biomedical research—what is the best treatment for each individual?

Much of the evidence driving the paradigm shift is rooted in the scientific theory of cancer biology (i.e., mechanistic pathways and driver mutations) with supporting evidence from the traditional single-marker-single-drug applications (e.g., companion diagnostics). A series of hypothesis-driven and “discovery” clinical trials, forging a consilience between genomic alterations in key pathways and the molecular targets of therapies in development, individually, may validate a single a priori gene–treatment interaction hypothesis in a small subgroup of molecularly defined patients. However, there remains a need to consider further observational study designs to investigate a growing number of molecularly targeted therapies in an ever-shrinking pool of patients with molecularly defined tumor subtypes, and in the setting where more and more tumors are found to contain multiple targetable molecular aberrations, many of which will require treatment with combinations of drugs (32).

The systematic accumulation of individual patient experiences in the real-time practice of precision oncology possesses the necessary elements to advance evidence-based practice while simultaneously driving further translational research. Integrating
epidemiologic principles in this type of active-learning system, anchored by observational studies, will allow for efficient knowledge integration through rapid synthesis and evaluation of precision oncology while ensuring continuous investigation, discovery, and evidence implementation (33, 34). Several contemporary examples of an active learning system that collect comprehensive genomic and treatment data from unique ‘N of 1’ encounters include CancerLinQ (35) and CancerCommons (36). While these model systems rely on different architectural principles in the construction of an active-learning knowledge base (complete electronic health records from participating institutions vs. donated clinical data from individual patients), they are designed to provide a reliable evidence base, generated across diverse institutional and health systems environments, to be used in real-time clinical decision making (37).

The development of an active-learning system potentially provides a robust research infrastructure for national outcome-based cohorts and consortiums (histology-based or histology-independent) as well as cancer genome databases, like My Cancer Genome, that can efficiently integrate findings into clinical practice while also generating new gene–treatment interaction hypotheses. The Lung Cancer Mutation Consortium (LCMC) recently published results from a multi-institution observational study that used multiplex genotyping methods to systematically identify 10 driver mutations in 1,007 participants with lung adenocarcinoma and to use this information to guide treatment decisions (38). The large effects observed in smaller, observational investigations of therapies targeting specific genomic alterations in molecularly homogeneous populations may provide sufficient evidence of clinical validity and utility otherwise not readily available from traditional RCTs and may provide the groundwork for clinical trials to validate the treatment effects of promising targeted therapies (4, 39).

As a growing evidence-base emerges from the everyday practice of precision oncology, it is important to maintain the proper epidemiologic perspective when designing future experimental trials to test the gene–treatment hypotheses that arise from an active-learning environment. The introduction of an active-learning system for precision oncology within our conceptual model illustrates where observational studies can add more certainty to the underlying treatment hypotheses while simultaneously generating evidence of clinical validity and utility of the potential benefits and harms in precision oncology. The dynamic process of evaluating rational combinations of targeted therapies through an active-learning system will improve our understanding of cancer biology, enhance our understanding of tumor evolution and the emergence of drug-resistance, and contribute to mechanistic approaches to drug development. In doing so, the bidirectional flow of information between laboratory science, clinical trials, and observational studies can be efficiently linked with clinical and population outcomes across cancer types.

Conclusion

We have demonstrated the unique opportunities that observational epidemiology has to offer to advance the application of precision oncology. Study designs, hypotheses, and inferences drawn from observational epidemiology help define the complementary and iterative bench-to-bedside translation of laboratory science, clinical trials, and observational epidemiology (Fig. 1). All epidemiologic study designs, including clinical trials, have limitations and different designs answer different questions, so an understanding of the advantages and limitations of each design at each point within the dynamic translational continuum is important. Failure to recognize the context-specific strengths and weaknesses can jeopardize the development of a promising application and contribute to slowing the pace of translation.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: M. Marrone, M.J. Khoury, A.N. Freedman
Development of methodology: M. Marrone, M.J. Khoury, A.N. Freedman
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J.A.N. Freedman
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.N. Freedman
Writing, review, and/or revision of the manuscript: M. Marrone, R.L. Schilsky, G. Liu, M.J. Khoury, A.N. Freedman
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): G. Liu, A.N. Freedman
Study supervision: M.J. Khoury, A.N. Freedman

Acknowledgments

The authors thank Dr. Margaret Spitz for her helpful comments on an early version of this commentary.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 19, 2014; revised November 5, 2014; accepted November 22, 2014; published online March 6, 2015.

References

# Opportunities for Translational Epidemiology: The Important Role of Observational Studies to Advance Precision Oncology


*Cancer Epidemiol Biomarkers Prev* 2015;24:484-489.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: <a href="http://cebp.aacrjournals.org/content/24/3/484">http://cebp.aacrjournals.org/content/24/3/484</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cited articles</td>
<td>This article cites 37 articles, 14 of which you can access for free at: <a href="http://cebp.aacrjournals.org/content/24/3/484.full#ref-list-1">http://cebp.aacrjournals.org/content/24/3/484.full#ref-list-1</a></td>
</tr>
<tr>
<td>Citing articles</td>
<td>This article has been cited by 3 HighWire-hosted articles. Access the articles at: <a href="http://cebp.aacrjournals.org/content/24/3/484.full#related-urls">http://cebp.aacrjournals.org/content/24/3/484.full#related-urls</a></td>
</tr>
</tbody>
</table>

**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://cebp.aacrjournals.org/content/24/3/484](http://cebp.aacrjournals.org/content/24/3/484). Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.