The Case for a Pre-Cancer Genome Atlas (PCGA)

Joshua D. Campbell1, Sarah A. Mazzilli1, Mary E. Reid2, Samjot S. Dhillon2, Suso Platero3, Jennifer Beane1, and Avrum E. Spira1

ABSTRACT

Understanding the earliest molecular and cellular events associated with cancer initiation remains a key bottleneck to transforming our approach to cancer prevention and detection. While TCGA has provided unprecedented insights into the genomic events associated with advanced stage cancer, there have been few studies comprehensively profiling premalignant and early-stage disease or elucidating the role of the microenvironment in premalignancy and tumor initiation. In this article, we make a call for development of a Pre-Cancer Genome Atlas (PCGA), a concerted initiative to characterize the molecular alterations in premalignant lesions and the corresponding changes in the microenvironment associated with progression to invasive carcinoma. This initiative will require a multicenter coordinated effort to comprehensively profile (cellular and molecular) premalignant lesions and their corresponding “field of injury” collected longitudinally as the lesion progresses towards or regresses from frank malignancy across multiple tumor types. Genomic characterization of alterations in premalignant lesions and their microenvironment, for both bulk tissue and single cells, will enable development of biomarkers for early detection and risk stratification as well as allow for the development of novel targeted cancer interception strategies. The multi-institutional and multidisciplinary collaborative “big-data” effort underlying the PCGA will help usher in a new era of precision medicine for cancer detection and prevention. Cancer Prev Res; 9(2); 119–24. ©2016 AACR.

The Bottleneck for Cancer Prevention and Early Detection

One of the critical barriers to developing new approaches for cancer detection and prevention is the lack of understanding of the molecular and cellular changes that cause cancer initiation and progression. Unlike the extensive work that has been done profiling advanced stage tumors, few studies have comprehensively profiled the genomic alterations found in precancerous tissues. Premalignant lesions are currently characterized by histologic changes that precede the development of invasive carcinoma (1, 2). These lesions can often be identified in regions surrounding an invasive tumor, in biopsies taken from patients undergoing diagnostic evaluation for suspicion of cancer, or in samples acquired during preventive screening. Currently, limited metrics exist to identify lesions that will likely progress to carcinoma and require intervention from those that will naturally regress or remain stable (3, 4). As imaging modalities and screening guidelines advance, the number of lesions identified will grow resulting in a need for more precise risk stratification methods and effective early intervention. Characterization of the molecular alterations in premalignant lesions and the corresponding...
changes in the microenvironment associated with progression would hasten the development of biomarkers for early detection and risk stratification as well as suggest preventive interventions to reverse or delay the development of cancer.

In this article, we make a call for the development of a new collaborative initiative, the "Pre-Cancer Genome Atlas (PCGA)," in which comprehensive genomic profiling of premalignant lesions and their corresponding field of injury is performed longitudinally and combined with clinical data including histology and outcome (progression/regression). Just as The Cancer Genome Atlas (TCGA) has ushered in a new era of precision treatment for advanced stage cancers, we envision the PCGA leading to a new era of personalized approaches for early cancer detection and prevention.

Molecular Characterization of Advanced Stage Cancer via TCGA

As the progression of cancer was initially described pathologically, the molecular processes that guide cancer initiation and development are being continually unraveled. The governing principle guiding cancer development is the same process active in most domains of biology: evolution. Genetic alterations in individual cells occur randomly due to replication errors or as a result of exposure to carcinogens. Mutations, copy number changes, and potentially epigenetic alterations can alter the ability of the cell to proliferate and survive in different environments. A mutation, for example, can confer a selective advantage, allowing the cell and its progeny to proliferate and gradually out-compete other cells lacking the same alteration. Genetic alterations in different molecular pathways can alter various cellular phenotypes or "hallmarks." The acquisition of a hallmark may occur by altering any one of several genes within the same underlying molecular pathway. Many cancer hallmarks have been characterized and include sustained cellular growth and proliferation, resisting cell death, replicative immortality by increasing telomere length, avoiding immune surveillance, as well as others (5). The mechanisms driving invasion can be thought of as a multistep evolution where the cumulative acquisition of driving genetic alterations allows the cells to exhibit multiple hallmarks and invade the proximal tissue.
To characterize the molecular alterations associated with cancer, TCGA consortium performed "multi-omic" profiling on over 11,000 advanced stage tumors from over 33 tumor types including DNA sequencing for mutation detection, SNP microarrays for copy number analysis, RNA sequencing for fusion detection and gene expression analysis, methylation data for epigenomic alterations, reverse protein phase arrays for protein quantification, and small RNA sequencing for miRNA expression analysis over the course of the last 10 years (6). These efforts have had a major impact in two different areas. First, by identifying genes recurrently altered within and across tumors types, the number of putative cancer driver genes has extended from several dozens to several hundred. For example, Lawrence and colleagues examined mutations and indels for 21 tumor types and identified over 250 genes as significantly mutated more than expected by chance (7). Likewise, Zack and colleagues examined focal copy number alterations across 11 tumors types and observed 140 regions recurrently gained or lost, many with novel putative cancer genes (8). Second, TCGA efforts have led to molecular reclassification via multi-omic clustering. Most tumor types had been previously stratified into subgroups using histologic characteristics alone. However, in several tumor types, clustering across global gene expression, copy number, and methylation patterns revealed molecular subgroups largely distinct from their histologic counterparts. In low-grade gliomas (LGG), for example, histology classification suffers from observer variability and does not sufficiently predict clinical outcomes. Clustering of multi-omic data in LGG revealed 3 robust molecular subclasses defined by combinations of IDH1/IDH2 mutation status, 1p/19q codeletion, and TP53 mutation status (9). These molecular subtypes were strongly associated with prognosis and other distinct clinical characteristics beyond standard histology suggesting they should be incorporated into clinical practice.

The Rationale for a PCGA

Despite the significant advances in the genomic characterization of advanced stage disease, a number of critical questions still remain. Given the evolutionary model that a specific sequence of genomic events acquired over many years cause the transition from normal epithelium to invasive carcinoma (10), having a complete catalog of driver genes for each tumor type is only the first step in understanding cancer progression. Indeed, many driver genes are often altered within an advanced stage tumor; however, the order with which the events occurred can be difficult to ascertain. In some circumstances, the predicted clonality of the mutations can be used to infer early events. For example, clonal mutations harbored by all cancer cells occur earlier in the route to frank malignancy compared with subclonal mutations present only in subset of cells (11, 12). This procedure can struggle to reveal the path of progression in some tumor types as mutations in many driver genes are predicted to be clonal. Comprehensive genomic profiling of longitudinally sampled premalignant lesions as they progress toward cancer (as detailed below) will provide critical insights into the sequence of molecular events that drive progression to invasive cancer (Fig. 1). This molecular reclassification of premalignancy will greatly improve our ability to predict which lesions are at higher risk of progressing to invasive carcinoma and allow for the development of novel targeted early interventional and therapeutic strategies.

Another critical component in any evolutionary process is the selective pressure imposed by the environment. As the majority...
of advanced stage tumor profiling has been done on bulk tumor tissue, characterizing the role of immune and stromal cell populations in the process of carcinogenesis has been challenging. In the past, only cell types with specific markers could be systematically identified by either IHC or advanced flow cytometry approaches in combination with gene expression analysis. However, the ability toagnostically characterize all cell populations within a sample is now achievable with the advent of single-cell RNA sequencing. With this technology, the expression state of each individual single cell can be measured and used to determine both the cell type and molecular pathways that may be active within the cell. Recent successes with cancer immunotherapies, such as antibodies blocking PD-1 or PD-L1, that are currently being developed for the treatment of over 30 cancer types (13), underpin the importance of characterizing the contributions of the different cell types in cancer development. Characterization of the immune cell populations in premalignant lesions with progressive versus regressive phenotypes is an opportunity to provide unprecedented insight into the role of the microenvironment in determining cancer initiation and progression, a critical step towards development of immuno-prevention strategies.

Genomic characterization of premalignant lesions using bulk tissue or single cells will help elucidate the mechanisms of disease progression. In turn, these findings can be exploited to develop biomarkers to inform cancer screening/early detection strategies and treatment of cancer at the earliest stages. Identification of premalignant disease processes and their likelihood of progression may help prioritize individuals for cancer screening and dictate the appropriate screening intervals. Biomarker driven cancer screening has the potential to maximize early detection while minimizing false positives that incur added costs as well as increase an individuals' radiation exposure and/or procedure-related complications. Cancer prevention clinical trials could also utilize biomarkers of premalignant disease to select subjects with a high likelihood of progression or response. Molecular selection by adding premalignant biomarkers to the trial entrance criteria has been recently reported to greatly reduce the number of subjects needed to test drug efficacy (14). Biomarkers may also be used in addition to histology to monitor treatment response thereby increasing trial speed and reducing trial cost. Improvements such as these could bring more prevention and targeted agents to the clinic, perhaps reducing the number of people that go on to develop cancer.

Recent Genomic Studies of Premalignant Disease

Studies demonstrating the ability of genomic profiling to provide insights into the biology of premalignancy have been recently reviewed (15). We reviewed several recent examples that set the stage for a larger PCGA initiative outlined below. Stachler and colleagues characterized two major paths of esophageal adenocarcinoma development by performing whole exome sequencing on tumor and adjacent Barrett esophagus within the same patient (16). In contrast to previous hypotheses, the majority of esophageal adenocarcinomas developed by first obtaining a TP53 mutation, followed by whole genome doubling and genomic instability, and finally oncogene amplification resulting in frank malignancy. The remainder of tumors displayed progressive inactivation of tumor suppressors such as CDKN2A and SMAD4, followed by oncogene activation, and genome instability. Interestingly, some patients had lesions with different sets of somatic alterations suggesting that they had formed independently (i.e., were clonally unrelated). These results suggest that extensive sampling of suspect areas is necessary to accurately capture the diversity of alterations and that comprehensive methodologies capable of detecting complex events such as whole genome doubling are needed.

Shain and colleagues utilized targeted sequencing of cancer genes in primary melanomas and adjacent precursor lesions to uncover the order of key driver events (17). The well-characterized and targetable mutation BRAF V600E in addition to other mutations in the MAPK pathway was substantially enriched in benign lesions, suggesting these are early events in melanoma carcinogenesis. Mutations in other common driver genes were observed only in intermediate or later stages of disease such as CDKN2A loss, TERT promoter mutations, or the SWI/SNF chomatin modifiers ARID1A, ARID2, or SMARCAD1. In the PCGA initiative, it will be important to sample various premalignant histologies as was done by Shain and colleagues to inform our understanding of cancer pathogenesis and risk of cancer development or how close to invasive malignancy a lesion may be.

Somatic mutations have been observed in tissue without clear histologic evidence of cancer. Mutations in genes such as DMN3TA, TET, and ASXL1 were found in the blood of subjects without any appreciable hematologic abnormalities at the time of sample collection (18). However, the presence of somatic mutations was associated with an increased risk for developing a hematologic malignancy as well as an increase in overall mortality as determined by longitudinal follow-up. These results support the notion that many people may already have a “first hit” that can produce a premalignant clonal expansion of cells. This raises the question of when clinical intervention should be applied in the premalignant setting. The number of hits that are needed to warrant clinical intervention may be different for various cancer types and may be dependent on a concurrent understanding of the interactions between precancerous cells and the immune system. We propose that answers to these questions can be most readily answered with a combination of longitudinal sampling, molecular profiling, and thorough clinical characterization ideally throughout the entire process of cancer development.

A Roadmap for the PCGA

Advances in genomic profiling pioneered by TCGA and related studies have largely overcome many of the technical profiling challenges. The largest obstacle impeding the understanding of cancer initiation and progression and development of early detection tools is the lack of systematic collection, annotation, and profiling of premalignant lesions. We propose a concerted multi-institutional effort to collect premalignant tissue across multiple tumor types followed by comprehensive genomic profiling to enhance our understanding of early-stage disease and build upon the foundation created by TCGA. Like TCGA, this will likely require more than 20 medical centers coordinating effort to collect and annotate the relevant clinical specimens. A number of recent examples of this type of coordinated effort include an NCI initiative to collect >1,000 surgically resected pancreatic cyst samples from 5 medical centers as well as a United States Department of Defense-funded consortium collecting...
airway samples via bronchoscopies from >1,000 smokers at risk for lung cancer at 11 military and Veteran hospitals.

Premalignant lesions can be identified and collected in a variety of ways. In tissues that can only be accessed via invasive surgery, premalignant lesions may be identified and sampled by histologic review of fresh or banked tumor specimens and their resection margins. Comparing the overlap of genomic alterations in the premalignant lesions to those found in the invasive tumor can identify early events in the process of carcinogenesis (19). However, cross-sectional studies like these may have limitations due to formalin fixation or the small quantity of tissue available after laser capture microdissection (LCM). These challenges may require researchers to use less comprehensive targeted approaches such as those utilized in the study of AAH lesions adjacent to lung adenocarcinoma and may preclude the use of some genomic technologies such as RNA sequencing (20). Additional limitations include the requirement that the tumor to be evolutionary related to some of the profiled premalignant lesions to infer early versus late events. Many lesions may in fact arise independently and will be clonally distinct from the tumor.

In contrast, other tissues accessible via relatively noninvasive procedures may be a useful starting point for PCGA studies, including bronchoscopy for the respiratory track, endoscopy for the upper gastrointestinal track, colonoscopy for the large intestine, the Papanicolaou test for the cervix, or visual examination of the skin and oral cavity. In fact, some of the earliest studies in cancer progression started with surveying genomic alterations in polyps in the colon (21). Importantly, the ability to identify suspect regions by visual examination or other facilities enables specialized tools to collect samples, albeit often small in size, to be collected and stored in conjunction with formalin fixed samples for histologic review. Fresh frozen tissue is more amenable to genomic profiling and is critical for technologies such as single-cell RNA-seq that currently require tissue dissociation and cell sorting soon after sample procurement. Similarly, single-cell DNA-seq can be achieved by performing single nuclei sorting from fresh frozen tissue. This approach can have even higher resolution to unravel evolutionary relationships among subclonal populations (22, 23).

Sample collection from accessible tissues also lends itself to repeated sampling of the same site over time. As screening studies become more common and are increasingly implemented as standard of care, the opportunities for sample collection of premalignant tissue will become more prevalent. There is also the potential to leverage recent advances in “liquid biopsy” technology to longitudinally follow genomic alterations in circulating blood that may reflect alterations found within premalignant lesions (19). Despite the technical feasibility of collecting premalignant tissue detailed above, significant barriers remain in having relatively healthy patients contribute research samples that increase risk and procedural time, as well as the challenge to physicians in collecting, annotating, and banking these additional research specimens. This type of tissue sampling will take careful thought and organization by the participating organizations and will likely proceed at a slower pace than with tumor samples. However, the return on investment will be significant: longitudinal profiling of premalignant lesions will allow us to better elucidate both the order of somatic alterations as well as the corresponding changes specific to the premalignant microenvironment that enable the transformation and ultimate invasion leading to frank carcinoma.

Conclusion
Recent advances in cancer screening and next-generation sequencing technology have set the stage for an unprecedented opportunity to characterize the genomic alterations associated with premalignant disease progression. While TCGA has provided us with a comprehensive catalog of driver genes for each tumor type, the sequence of these genomic events that characterize the progression of premalignant lesions to invasive cancer remains to be unraveled. In addition, we know little about how changes in the immune cells and premalignant microenvironment contribute to disease initiation and progression. Comprehensive profiling of genomic and microenvironment changes that occur longitudinally in premalignant lesions as they progress towards (or regress away from) invasive cancer, a “Pre-Cancer Genome Atlas (PCGA),” will provide novel targets for disease interception that can be used to both develop early detection biomarkers as well as enable personalized therapeutic approaches. Creation of this PCGA will require a multi-institutional and multidisciplinary collaborative big-data “pre-cancer moonshot” effort (consistent and aligned with the recent Obama/Biden initiative) to collect, annotate, and profile premalignant lesions across multiple tumor types. This initiative will also require development of novel high-throughput functional screens in the premalignant in vitro setting as well as in vivo models of premalignancy to test the functional role of candidate genes and immune cell types. Ultimately, the PCGA will help usher in a new era of precision medicine for cancer detection and prevention.

Disclosure of Potential Conflicts of Interest
All authors except S. Platero have received a commercial research grant from Janssen Pharmaceuticals. S. Platero is an employee of Janssen Pharmaceuticals.

A. Spira is a consultant to Varacyte Inc. No potential conflicts of interest were disclosed by the other authors.

Grant Support
This work was supported by Janssen Pharmaceuticals.

Received January 25, 2016; accepted January 25, 2016; published online February 2, 2016.

References


# The Case for a Pre-Cancer Genome Atlas (PCGA)

Joshua D. Campbell, Sarah A. Mazzilli, Mary E. Reid, et al.


### Updated version

Access the most recent version of this article at:

http://cancerpreventionresearch.aacrjournals.org/content/9/2/119

### Cited articles

This article cites 21 articles, 5 of which you can access for free at:

http://cancerpreventionresearch.aacrjournals.org/content/9/2/119.full#ref-list-1

### Citing articles

This article has been cited by 16 HighWire-hosted articles. Access the articles at:

http://cancerpreventionresearch.aacrjournals.org/content/9/2/119.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://cancerpreventionresearch.aacrjournals.org/content/9/2/119. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.