INTRODUCTION

The Cancer Genome Atlas (TCGA) project has provided many biologic insights through genomic, transcriptomic, epigenomic, and proteomic profiling from a large number of patient samples in many cancer types. However, in-depth clinical tie-ins to TCGA data have yet to be thoroughly performed. A recent study has shown that integrating large-scale molecular profiling data to clinical variables yields statistically significant improvements in prognostic estimates for selected patients with cancer (1). In addition, investigators have found that nearly 90% of TCGA samples harbor alterations in genes linked to predictive, prognostic, or diagnostic associations (1, 2). However, the value of comprehensive genomic tests for medical oncologists planning to optimally guide therapy by matching every possible gene alteration to treatment options when moving from the most relaxed to the strictest criteria to match a gene alteration with a drug. In contrast with previous disease-specific molecular epidemiology efforts, we explore pan-cancer TCGA data and perform a detailed assessment of the direct clinical utility of large-scale molecular profiling by (i) limiting the analysis to “targetable” events—excluding gene variants whose “actionability” is limited to prognostic and/or diagnostic associations; (ii) looking at both mutations and somatic copy-number events across multiple solid tumor types; (iii) carefully assessing whether specific variants in oncogenes have been functionally validated; and (iv) taking into consideration the specific tumor types for which the gene–drug association has been described in the literature.

GENE–DRUG KNOWLEDGE DATABASE AND CLINICAL TARGETABILITY INDEX

To perform a comprehensive clinically oriented analysis of somatic cancer variants, we developed a Gene–Drug
Knowledge Database (GDKD) of predictive genomic biomarkers in oncology—a manually curated database that integrates different layers of annotations: tumor types, genes, variants, and sensitivity/resistance patterns to approved and experimental agents under clinical investigation. All associations are described using standardized terminology in a structured format and linked to PubMed identifiers. Regular updates are made publicly available through Synapse (10). More than 700 variant-specific gene–drug interactions with consensus or emerging therapeutic relevance were curated for this effort, as shown in Fig. 1. We performed an extensive literature search for predictive associations on the most frequent and relevant somatic cancer variants, encompassing 130 genes, 90 drugs, and 40 malignancies. Of note, the GDKD aggregates information from publicly available resources such as My Cancer Genome, Targeted Cancer Care, Personalized Cancer Therapy, and other efforts (2).

For a detailed estimation of targetability, we assessed genomic markers at multiple levels:

1) Gene: Any gene that, when somatically altered in cancer, predicts response or resistance to a specific therapy was considered a targetable gene. We defined three gene lists based on the strength of evidence for predictive associations (10):
   i. Genes whose associations are based on preclinical or clinical studies (n = 122);
   ii. Genes with clinically proven associations based on consensus guidelines, clinical trials, or case reports (n = 50);
   iii. Genes with associations that have been clinically validated and that guided FDA approval of targeted drugs in solid tumors (n = 7).

2) Variant: Two degrees of restrictiveness were defined:
   i. Relaxed: For tumor suppressors (e.g., PTEN), any nonsilent mutation or copy-number loss (GISTIC score +2 or −2).

   For a detailed estimation of targetability, we assessed genomic markers at multiple levels:

   1) Gene: Any gene that, when somatically altered in cancer, predicts response or resistance to a specific therapy was considered a targetable gene. We defined three gene lists based on the strength of evidence for predictive associations (10):
      i. Genes whose associations are based on preclinical or clinical studies (n = 122);
      ii. Genes with clinically proven associations based on consensus guidelines, clinical trials, or case reports (n = 50);
      iii. Genes with associations that have been clinically validated and that guided FDA approval of targeted drugs in solid tumors (n = 7).

   2) Variant: Two degrees of restrictiveness were defined:
      i. Relaxed: For tumor suppressors (e.g., PTEN), any nonsilent mutation or copy-number loss (Genomic Identification of Significant Targets in Cancer (GISTIC) score, −2)

---

**Figure 1.** Workflow of analytic pipeline integrating the Gene–Drug Knowledge Database and the Clinical Targetability Index (CTI) to pan-cancer TCGA data. Statistics of GDKD: distribution of gene variant—drug associations across solid tumor types, most frequent genes, and drugs with predictive associations. Ampl, amplifications; del, deletions; GISTIC, Genomic Identification of Significant Targets in Cancer.
in the gene was considered an event. For oncogenes with targetable amplifications (e.g., MET), only GISTIC scores +2 were considered events. For oncogenes with targetable mutations, any nonsynonymous variant in the gene was considered an event;

ii. Strict: Same as above for tumor suppressors and amplifications/deletions, but for oncogenes with targetable mutations, only functionally validated activating variants were considered events—either preclinical or clinical evidence of sensitivity/resistance to any matched targeted therapy.

3) Tumor type: Two levels of associations:

i. Unrestricted evidence: Knowledge from one tumor type is translated to all malignancies (e.g., targetability of NRAS codon 61 mutations in melanoma is applied to endometrial cancer);

ii. Restricted evidence: Tumor-type–specific knowledge on targetability of genomic events.

4) Drug: Only associations with agents that are currently in clinical development were considered (we excluded drugs that have not yet been translated to the clinic);

i. Any targeted drug in phase I to III clinical trials or that received regulatory approval;

ii. Genomic markers linked to FDA-approved agents.

Using these criteria, we defined the CTI, with increasing levels of evidence for predictive associations of genomic biomarkers, as summarized in Fig. 1. Briefly, in CTI.1, preclinical studies are taken into consideration when defining a biomarker, such as ERBB3 mutations (11); in CTI.2, we limited the analysis to gene alterations that have clinical associations described in the literature, such as FGFR1 amplifications (12); in CTI.3, we excluded variants in oncogenes that are of uncertain significance; in CTI.4, we focused on predictive evidence derived from studies performed in the same tumor type; and in CTI.5, we considered only associations linked to FDA-approved agents. We then used gene–drug associations from the GKD as “genomic biomarker filters” to assess the prevalence of potentially targetable events at different CTI scenarios. TCGA mutation calls were downloaded from the Synapse TCGA Live Data Portal (13), and copy-number GISTIC scores from the Firehose Broad website (14) on June 12, 2014.

Prevalence of Potentially Targetable Events in Different Scenarios

Table 1 presents the prevalence of potentially targetable events for the different CTIs. The prevalence of potentially targetable events was highest in CTI.5, in which we considered only associations linked to FDA-approved agents. In contrast, the prevalence of potentially targetable events was lowest in CTI.1, in which we considered only preclinical studies. The results show that the prevalence of potentially targetable events varies depending on the CTI.

Figure 2. A, prevalence of targetable events according to different CTIs—CTI.1 to CTI.5—across solid tumors in TCGA. B, mean number of alterations per sample, starting from mutations or copy-number events in Cancer Gene Census (CGC) genes and then looking at potentially targetable events according to different targetability scenarios—CTI.1 to CTI.5. Tumors in graph at right do not have a gene alteration linked to FDA-approved drug (i.e., no CTI.5). C, prevalence of gene variants associated with dramatic responses to targeted therapies (case reports) across solid tumors. For oncogenes, the proportion of variants that have been functionally validated and that remain of uncertain significance is depicted. V.U.S., variants of unknown significance.
alterations per sample. Notably, nearly 75% of the patients still have at least one targetable event according to CTI.3, but only 20% of the tumors have three or more events. This scenario illustrates what medical oncologists working at large research institutions with comprehensive tumor genotyping may face on a daily basis, trying to match many gene alterations that are still of unknown predictive value (emerging evidence derived from early clinical data from a variety of tumor types) with drugs in clinical trials. Surprisingly, a substantial proportion (>50%) of the patients with relatively rare malignancies—bladder, head and neck, stomach, and uterine cancers—would potentially benefit from an expanded mutation/copy-number analysis pipeline to identify alterations in genes that have emerging associations. Examples include genomic events in receptor tyrosine kinases (ERBB2, ERBB3, MET), the PI3K pathway (PIK3CA, PIK3R1, PTEN, MTOR, TSC1), the FGFR pathway (FGFR1, FGFR2, FGFR3), and DNA damage repair signaling (BRCA1, BRCA2, RAD50, ERCC2).

We then assessed the impact on predictive genomic biomarker frequency when moving from relaxed to stricter targetability criteria. Figure 2B depicts the mean number of alterations per sample in different scenarios, starting with any mutation or high-level amplification or deletion in genes that have been casually implicated in cancer, as catalogued by the Cancer Gene Census (CGC) and other groups (15, 16). On average, solid tumors have nearly 30 cancer-related alterations and only 17% of them are considered targetable according to CTI.1. Thyroid cancer is again an outlier, with only three alterations in cancer genes per sample. Notably, when excluding variants of uncertain significance in oncogenes (scenario CTI.2 to CTI.3), the mean reduction in the number of targetable events is 25%, but with significant differences across solid tumors. As expected, diseases with higher prevalence of events in rarely mutated oncogenes are more likely to be affected (more than 50% reduction in targetability rate), because the majority of non–hotspot events have not been functionally validated, as discussed later. Examples include stomach, uterine, and kidney clear-cell carcinomas with ERBB2, FGFR2, and MTOR mutations. Of note, the largest impact on the prevalence of targetable alterations occurs when we ignore genomic events that have been matched to targeted drugs in different malignancies. Diseases in which the targetability of genomic events has been understudied (with more than a 90% drop when moving from scenario CTI.3 to CTI.4) include bladder cancer, stomach cancer, kidney clear-cell carcinoma, squamous lung cancer, and head and neck cancer. Further preclinical–clinical validation of potential targets is needed in these tumor types. In scenario CTI.4, 39% of the patients have at least one targetable event.

By looking at scenario CTI.5, which represents the strictest criteria to match gene alterations to approved targeted agents, we confirmed that the distributions of TCGA genomic alterations are concordant with the molecular epidemiology of validated biomarkers in cancer: (i) 13% of lung carcinomas with activating EGFR mutations (sensitivity to erlotinib, afatinib); (ii) 48% of melanoma samples with BRAFV600E mutations (sensitivity to vemurafenib, dabrafenib, or trametinib) and 4% with activating KIT mutations (sensitivity to imatinib); (iii) 13% of breast cancer samples with ERBB2 amplifications (sensitivity to trastuzumab, pertuzumab, ado-trastuzumab emtansine, lapatinib)—GISTIC scores ≥2 are likely the best approximation to the FDA-approved biomarkers (in situ hybridization or protein overexpression); and (iv) 45% of colorectal cancer samples with KRAS or NRAS mutations (resistance to cetuximab and panitumumab). These five tumor types have targeted drugs recommended by consensus guidelines and a companion biomarker diagnostic that tests positive in approximately 25% of the samples. If clinicians could add to the therapeutic armamentarium unapproved drugs linked to emerging biomarkers in the same tumor types, this number would increase to 55%. This larger population is indeed eligible for clinical trials testing the predictive value of genomic biomarkers.

Case Reports and Variants of Unknown Significance in Oncogenes

It is already known that validated actionable alterations, such as BRAFV600E in melanoma, can be found in unexpected tumor types and occasionally predict for benefit with matched therapies (4, 17). Case reports of dramatic responses to targeted agents are increasingly being published, not only in the setting of validated biomarkers in one tumor type being translated to other diseases, but also with rare genomic events not previously linked to benefit with a particular drug. Examples include activating mutations in AKT1 (AKT inhibitors; ref. 18), DDR2 (dasatinib plus erlotinib; ref. 19), ERBB2 (anti-HER2 agents; refs. 20, 21), FGFR2 (pazopanib; ref. 6), and MTOR (everolimus plus pazopanib; ref. 22), as well as inactivating events in RAD50 (irinotecan plus CHK1/2 inhibitor; ref. 23) or TSC1 (everolimus; ref. 24). We decided to study the prevalence of these genomic events across solid tumors, as shown in Fig. 2C. When considering only functionally validated mutations in the oncogenes AKTI, DDR2, ERBB2, FGFR2, and MTOR plus deleterious mutations in the tumor suppressors RAD50 and TSC1, on average 6% of cancer samples test positive for a potentially relevant genomic biomarker. Surprisingly, even though none of the cases of spectacular tumor response linked to this set of genes was reported in uterine cancer, translating the information derived from other tumor types could have a direct therapeutic impact in 15% of the patients. Interestingly, within this selected list of rarely mutated oncogenes, 80% of the variants are of uncertain functional significance, reflecting the gap in our understanding of the clinical relevance of a large percentage of alterations in genes potentially linked to clinical actions.
It is important to emphasize that not all alterations in clinically relevant genes described here will act as “drivers” and portend response to therapeutic targeting. In addition, many of the rare oncogene variants, with nearly 2% prevalence across solid tumors, are of unknown functional significance, requiring further experimental validation. In this context, we believe that clinical validation of a potential biomarker (novel mutation in oncogene matched to targeted drug) is acceptable only when mechanisms to annotate response or lack of benefit are available. Ultimately, prospective implementation of global genomic profiling in clinical trials will define its general applicability and whether matching gene alterations to targeted drugs is associated with improved patient outcomes. The Molecular Analysis for Therapy Choice (MATCH) NCI initiative, an umbrella protocol for multiple single-arm trials, plans to assign over 1,000 patients that progress after one line of standard therapy to matched agents based on next-generation sequencing. Other examples include the Michigan Oncology Sequencing Project (MI-ONCOSEQ), Canadian IMPACT protocol, and the French MOSCATO trial, with systematic use of comprehensive genomic profiling to guide early clinical trial enrollment. We believe that to best inform clinical practice, reports of next-generation sequencing tests should include standardized classifications of predictive genomic biomarkers, with multilevel associations to define targetability.

We have shown here the magnitude of change of targetability when applying increasingly stringent criteria for a predictive genomic biomarker. Overall, 85% of potentially targetable events are rejected with the strictest definition. Thyroid cancer is an outlier, with only a 28% overall drop in the targetability rate, most likely related to its lower complexity at the genomic level as compared with the other solid tumors included in this report. One clear limitation of genomic biomarker identification and characterization in our study is the lack of fusion genes in the final list of targetable events (e.g., ALK, ROS1, and NTRK1, known sensitizers to crizotinib in lung adenocarcinomas; refs. 26–28). However, the rarity of these events indicates that the results presented here should not be largely affected (29). Defining the prevalence of these events across rare solid tumors and providing easy access to fusion calls through TCGA data portal would be highly beneficial to the cancer research community.

Finally, biologic and clinical interpretation of genomic events is primarily a challenge in data standards and management. We were able to aggregate in our structured GDKD the fragmented knowledge of uncoordinated and overlapping efforts from multiple institutions. Given that these efforts do not share a common set of standards, the products of each group could not be easily integrated. As proposed by Good and colleagues (30), the only alternative moving forward is the creation of a community-based open resource to connect cancer genomic events with the necessary evidence to evaluate their biologic and clinical significance. The information should be released in an interactive web-based tool, subjected to editing, validation, and critique from the medical community. Most importantly, knowing that our collective understanding of predictive genomic events grows on a daily basis, the database should aim to reincorporate these advances in a timely manner. Finally, we believe that integrating this resource to clinical trial databases, with expansion of ClinicalTrials.gov, will substantially increase therapeutic options for patients with cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank Obi Griffith for GDKD support and Larson Omberg for technical help in data acquisition.

Grant Support

R. Dienstmann is a recipient of “La Caixa International Program for Cancer Research & Education.”

Published online February 5, 2015.

REFERENCES

# Database of Genomic Biomarkers for Cancer Drugs and Clinical Targetability in Solid Tumors

Rodrigo Dienstmann, In Sock Jang, Brian Bot, et al.

*Cancer Discovery* 2015;5:118-123.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: <a href="http://cancerdiscovery.aacrjournals.org/content/5/2/118">http://cancerdiscovery.aacrjournals.org/content/5/2/118</a></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 26 articles, 11 of which you can access for free at: <a href="http://cancerdiscovery.aacrjournals.org/content/5/2/118.full#ref-list-1">http://cancerdiscovery.aacrjournals.org/content/5/2/118.full#ref-list-1</a></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Citing articles</th>
<th>This article has been cited by 5 HighWire-hosted articles. Access the articles at: <a href="http://cancerdiscovery.aacrjournals.org/content/5/2/118.full#related-urls">http://cancerdiscovery.aacrjournals.org/content/5/2/118.full#related-urls</a></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>E-mail alerts</th>
<th>Sign up to receive free email-alerts related to this article or journal.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reprints and Subscriptions</td>
<td>To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a>.</td>
</tr>
<tr>
<td>Permissions</td>
<td>To request permission to re-use all or part of this article, contact the AACR Publications Department at <a href="mailto:permissions@aacr.org">permissions@aacr.org</a>.</td>
</tr>
</tbody>
</table>