IN THE SPOTLIGHT

The GENIE Is Out of the Bottle: Landmark Cancer Genomics Dataset Released

Kevin Litchfield¹, Samra Turajlic¹,², and Charles Swanton¹,³,⁴

Summary: In this issue of Cancer Discovery, an overview of the AACR Project GENIE, a landmark study in cancer genomics, is presented by The AACR Project GENIE Consortium. A summary of the goals and objectives of this ambitious program is provided, together with an analysis of the phase I cohort of 19,000 samples. Cancer Discov; 7(8): 796-8. ©2017 AACR.

See related article by The AACR Project GENIE Consortium, p. 818 (1).

Somatic mutations are a universal feature of cancer and considered to be a fundamental step in driving tumor growth. Classic oncogenes and tumor-suppressor genes, such as RAS or TP53, though discovered several decades ago, remain of relevance in the clinic today. The search for new cancer genes has intensified since 2005, through large-scale international initiatives such as The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC). These colossal projects leveraged high-throughput sequencing (HTS) technologies to systematically profile >30 different tumor types across tens of thousands of cases. The technological advancement provided by HTS has heralded previously inconceivably low sequencing costs per mega base, and cancer genome profiling of 10s to 100s of genes is commonplace in clinical practice. In this issue of Cancer Discovery, an overview of AACR Project GENIE, which is set to become the next landmark project in cancer genomics, is presented by The AACR Project GENIE Consortium (1). The consortium members should be highly commended for their efforts in bringing this resource, which will no doubt be utilized extensively by cancer researchers worldwide, to the scientific community. Given the wealth of cancer genomic datasets now available for public access, it is helpful to put into context the scale and rapidity of Project GENIE. For comparison, project sample sizes and timescales for TCGA were >11,000 cases across approximately 10 years (2), and for ICGC were >17,000 samples over approximately 8 years (3). At 19,000 samples in phase I, increasing to >100,000 in 5 years, Project GENIE represents a step change in both cohort size and completion time. It should be acknowledged that the GENIE project focuses on targeted cancer gene panel data, rather than whole-exome or whole-genome profiling. However, in the context of clinically driven research questions, this represents the greatest unmet need. The last decade has been spent sequencing the “long tail” of mutated cancer genes, with a low cost-to-benefit ratio and with a very large majority of sequenced genes found to bear only passenger mutations or rare driver events detected in a handful of cases. Hence, the GENIE dataset represents a pragmatic, well-powered, and efficiently assembled resource, focusing on approximately 200 genes with known and recurring oncogenic potential. In addition, genomic results from the project will meet Clinical Laboratory Improvement Amendments/International Organization for Standardization–certified processing standards and be accompanied by standardized clinical outcome data. The overall goal of Project GENIE is to create a large-scale high-quality cancer genomics database, widely accessible to the global research community, in order to catalyze precision medicine research efforts and improve patient outcome. To this end, the phase I dataset represents a significant step forward, with cohort sizes of >2,000 cases for non–small cell lung, breast, and colorectal cancers, immediately ranking among the largest publicly available HTS genomic datasets for these tumor types.

What stands out most about Project GENIE is its achievements in transcending bureaucratic barriers and galvanizing decisive action across a large group of leading international cancer centers. These efforts should not be underestimated; a large number of influential stakeholders have been successfully engaged and brought on-board with the GENIE program, producing a hugely beneficial end result. This was achieved through careful consideration of both the individual centers’ academic interests and the overall aims of the consortium. Scientifically, genomic profiling of tumor tissue has utility in addressing a number of important research aims, from studying the natural history of disease, to biomarker identification and novel therapeutic discovery. Taking each of these themes in turn, Project GENIE can be put into the following research context. First, regarding basic research into disease natural history, recent large-scale genomics projects have provided considerable insight. Initial efforts to catalog key driver events per tumor type have now been extended to study
the ongoing evolution of cancer across the entire disease course. Molecular profiling of precancerous normal tissue, cell-free circulating nucleic acids, and primary and metastatic tumor tissues have all revealed evidence of cancer-associated somatic mutations (4, 5). Although these observations raise the prospect of novel clinical intervention at various time points, much remains to be understood in terms of basic tumor biology. The GENIE dataset will be well placed to drive forward these basic research efforts, offering a well-powered patient-based resource to cross-validate results from cell-based assays or animal models. In particular, complex functional screens of epistasis, synthetic lethality, and timing/order of mutational events need large sample sizes in order to ensure sufficient numbers in each study group and to correct for multiple testing. Project GENIE will be well powered to address many of these questions. In addition, the accessibility of GENIE results, which are already deposited and available in eBioPortal, is likely to ensure an immediate positive impact to basic research efforts worldwide.

Second, regarding biomarker identification, this is a field that has faced significant challenges, with numerous findings failing to validate and commonplace reporting of conflicting results (6). A major limitation has been insufficient statistical power, with smaller cohort sizes leading to type I error, an issue particularly acute when stratifying patient groups based on individual gene mutational status. When 100,000 cases are reached, Project GENIE will provide sample sizes of approximately an order of magnitude higher than TCGA/ICGC datasets, which should finally enable robust validation of many genomic biomarkers. As well as being relevant for molecularly targeted therapies, genomic biomarkers are also likely to be of relevance for immunotherapy, with increasing evidence associating mutational burden with checkpoint inhibitor response rate (7). Recent data have shown mutational burden estimates from panel data closely correlated with true whole-exome mutation counts (8); hence, the Project GENIE resource has potential to support researchers in this area of urgent clinical need. In addition, the search for validated prognostic biomarkers, to aid patient stratification for surveillance and/or adjuvant therapies, will be greatly furthered by the GENIE resource.

Third, regarding novel therapeutic discovery, several opportunities are likely to arise through expansion of existing FDA-approved therapies to new indications. Indeed, in the phase I cohort, the authors make the intriguing observation that cancers of unknown primary are enriched within the top 10% most highly mutated samples, suggesting that checkpoint inhibition therapy may be of potential benefit for a subset of these hard-to-treat patients. In addition, the inclusion of data from sponsored research agreements, such as >2,000 rare breast cancer samples with either mutant ERBB2 or AKT E17K, will accelerate approval of novel therapies for these rare breast cancer subtypes. At a strategic level, Project GENIE will be well placed to participate in further such studies, with its ability to identify genetically defined patient subgroups, which are often rare and difficult to recruit in sufficient numbers in other contexts.

In terms of challenges, Project GENIE will need to manage the potential inter-site variability across the eight contributing centers, with a combination of both amplicon- and hybrid-capture–based panels, with differing gene content. However, across all panels, there is a core 44-gene overlap, and the three largest contributing centers (Dana-Farber, Memorial Sloan Kettering, and Vanderbilt- Ingram Cancer Centers) collectively submitted >14,000 of the 19,000 phase I samples, and all used a common large panel with 275+ genes. Concordance in mutation detection frequencies across institutes also appears high in phase I data.

Finally, the most pertinent question that Project GENIE can address is the general utility of genetic profiling and therapy matching in a routine clinical context. Several recent reports (9, 10) have questioned the clinical value of precision medicine approaches, and this remains a difficult but important question that must be addressed. With 19,000 samples, increasing to 100,000 within 5 years, Project GENIE will offer unparalleled insights into the applicability of cancer genomic profiling at a routine population level. The final results from this project are likely to have long-lasting and broad implications on the implementation of cancer genomics within routine clinical care. The initial results appear positive, with >30% of phase I patients having potentially actionable alterations; longitudinal data on treatment outcome across this large cohort will be of significant interest, and the robust and harmonized clinical annotation across the cohort is key to these interpretations.

Many researchers will remember the poignant address given by then Vice President Joe Biden at the 2016 AACR meeting, which conveyed a renewed sense of urgency and importance to the cancer research community. To deliver such a direct response to this, by publicly releasing genomic data from 19,000 samples within 12 months of his address, serves as a powerful example of what can and should be achieved through collaboration.

Disclosure of Potential Conflicts of Interest

C. Swanton has received consulting and speaker fees from Boehringer Ingelheim, Eli Lilly, Novartis, and Roche; has received speaker fees from Celgene, GlaxoSmithKline, Pfizer, and Servier; has ownership interest in Achilles Therapeutics, Apogen Biotechnologies, Epic Biosciences, and GRAIL; and has patents for immune checkpoint intervention in cancer (PCT/EP2016/071471), a method of detecting tumor recurrence (1618485.5), and a method for treating cancer (PCT/EP2016/059401); and is a consultant/advisory board member for Apogen, Epic Biosciences, and GRAIL. No potential conflicts of interest were disclosed by the other authors.

Grant Support

K. Litchfield is supported by a UK Medical Research Council Skills Development Fellowship Award (grant reference number MR/P014712/1). S. Turaligic is a Cancer Research UK clinician scientist and is funded by Cancer Research UK (grant reference number C50947/A18176) and the National Institute for Health Research (NIHR) Biomedical Research Centre at the Royal Marsden Hospital and Institute of Cancer Research (grant reference number A109). C. Swanton is a senior Cancer Research UK clinical research fellow and is funded by Cancer Research UK (TRACERx), the Rosetrees Trust, NovoNordisk Foundation (ID 16584), EU FP7 (projects PREDICT and RESPONSIFY, ID: 259303), the Prostate Cancer Foundation, the Breast Cancer Research Foundation, the European Research Council (THESEUS), and the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Published online August 1, 2017.
REFERENCES

The GENIE Is Out of the Bottle: Landmark Cancer Genomics Dataset Released

Kevin Litchfield, Samra Turajlic and Charles Swanton

*Cancer Discov* 2017;7:796-798.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><a href="http://cancerdiscovery.aacrjournals.org/content/7/8/796">http://cancerdiscovery.aacrjournals.org/content/7/8/796</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 8 articles, 1 of which you can access for free at:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><a href="http://cancerdiscovery.aacrjournals.org/content/7/8/796.full#ref-list-1">http://cancerdiscovery.aacrjournals.org/content/7/8/796.full#ref-list-1</a></td>
</tr>
</tbody>
</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>