The GENIE is out of the bottle: landmark cancer genomics dataset released

Kevin Litchfield[1], Samra Turajlic[1,2], Charles Swanton[1,3,4]

1. Translational Cancer Therapeutics Laboratory, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT
2. Renal and Skin Units, The Royal Marsden Hospital, London, SW3 6JJ
3. Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London, WC1E 6BT
4. Department of Medical Oncology, University College London Hospitals, 235 Euston Rd, Fitzrovia, London, United Kingdom, NW1 2BU

Addresses for correspondence:

Prof Charles Swanton,
Translational Cancer Therapeutics Laboratory
The Francis Crick Institute
1 Midland Road
London NW1 1AT

E-mail: Charles.Swanton@crick.ac.uk

Summary: In this issue of *Cancer Discovery* Sawyers, Cerami and colleagues present an overview of the AACR Project GENIE, a landmark study in cancer genomics. The authors summarize the goals and objectives of this ambitious program, together with an analysis of the phase 1 cohort of 19,000 samples. Cancer Discov. ©2017 AACR.

Somatic mutations are a universal feature of cancer, and considered to be a fundamental step in driving tumor growth. Classical oncogenes and tumor suppressor genes, such as *RAS* or *TP53*, while discovered several decades ago remain of relevance in the clinic today. The search for new cancer genes intensified from 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 tumour 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 10's to 100's of genes is commonplace in clinical practice. In this issue of Cancer Discovery Cerami, Sawyers and colleagues (1) present an overview of the AACR Project GENIE , which is set to become the next landmark project in cancer genomics. The consortium members should be highly commended for their efforts in bringing this resource to the scientific community, which will no doubt be utilised extensively by cancer researchers worldwide. 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 ~ten years (2), and for ICGC were >17,000 samples over ~eight years (3). At 19,000 samples in phase 1, increasing to >100,000 in five years, Project GENIE represents a step change in both cohort size and completion time. It should be acknowledged the GENIE project focuses on targeted cancer gene panel data, rather than whole exome or 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, with a very large majority of genes sequenced 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 ~200 genes with known and recurring oncogenic potential. In addition genomic results from the project will meet CLIA/ISO certified processing standards and be accompanied by standardised 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 1 dataset represents a significant step forward, with cohort sizes of >2,000 cases for non-small cell lung, breast and colorectal cancers, immediately rank among the largest publically 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 catalogue 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, primary and metastatic tumor tissue, have all revealed evidence of cancer associated somatic mutations (4,5). While 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 the cBioportal, 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 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 correlate 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 1 cohort the authors make the intriguing observation that cancers of unknown primary are commonly among top 10% highly mutated samples, suggesting checkpoint inhibition therapy may be of potential benefit in 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 sub-groups, 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 1 samples and all used a common large panel with 275+ genes. Concordance in mutation detection frequencies across institutes also appears high in phase 1 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 five 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 1 patients having actionable alterations; longitudinal data on treatment outcome across this large cohort will be of significant interest, and the robust and harmonised clinical annotation across the cohort is key to these interpretations.

Many researchers will remember the poignant address given by Vice President 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 publically 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.

References

1.   Consortium APG. AACR Project GENIE: Powering Precision Medicine Through An International Consortium. Cancer discovery **2017** doi 10.1158/2159-8290.CD-17-0151.
2.   TCGA. 2017 TCGA.   <https://cancergenome.nih.gov/abouttcga/overview>.
3.   ICGC. ICGC (Data Release 24) https://dcc.icgc.org/. **May 18th, 2017**.
4.   Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R*, et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature **2017**;545(7655):446-51 doi 10.1038/nature22364.
5.   Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S*, et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science **2015**;348(6237):880-6 doi 10.1126/science.aaa6806.
6.   Wu L, Qu X. Cancer biomarker detection: recent achievements and challenges. Chemical Society reviews **2015**;44(10):2963-97 doi 10.1039/c4cs00370e.
7.   Manson G, Norwood J, Marabelle A, Kohrt H, Houot R. Biomarkers associated with checkpoint inhibitors. Annals of oncology : official journal of the European Society for Medical Oncology **2016**;27(7):1199-206 doi 10.1093/annonc/mdw181.
8.   Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R*, et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome medicine **2017**;9(1):34 doi 10.1186/s13073-017-0424-2.
9.   Prasad V. Perspective: The precision-oncology illusion. Nature **2016**;537(7619):S63 doi 10.1038/537S63a.
10.   Tannock IF, Hickman JA. Limits to Personalized Cancer Medicine. The New England journal of medicine **2016**;375(13):1289-94 doi 10.1056/NEJMsb1607705.