All the World’s a Stage: Facilitating Discovery Science and Improved Cancer Care through the Global Alliance for Genomics and Health

Mark Lawler¹, Lillian L. Siu², Heidi L. Rehm³, Stephen J. Chanock⁴, Gil Alterovitz⁵, John Burn⁶, Fabien Calvo⁷, Denis Lacombe⁸, Bin Tean Teh⁹, Kathryn N. North¹⁰, and Charles L. Sawyers¹¹, on behalf of the Clinical Working Group of the Global Alliance for Genomics and Health (GA4GH)

Summary: The recent explosion of genetic and clinical data generated from tumor genome analysis presents an unparalleled opportunity to enhance our understanding of cancer, but this opportunity is compromised by the reluctance of many in the scientific community to share datasets and the lack of interoperability between different data platforms. The Global Alliance for Genomics and Health is addressing these barriers and challenges through a cooperative framework that encourages “team science” and responsible data sharing, complemented by the development of a series of application program interfaces that link different data platforms, thus breaking down traditional silos and liberating the data to enable new discoveries and ultimately benefit patients. Cancer Discov; 5(11); 1133-6. ©2015 AACR.

INTRODUCTION

Earlier this year, The Cancer Genome Atlas (TCGA), a U.S. federally funded international genomics project linked to the International Cancer Genome Consortium (ICGC), genetically profiled the last of its 10,000 cancer genomes (1). The ease with which 21st-century molecular technologies have empowered large-scale cancer genomics projects has been impressive, but the unparalleled level of data generated creates its own bioinformatics challenges, making it difficult both to access the data and to conduct new analyses. The full characterization of TCGA is substantive, on the order of multiple petabytes, and it has been released to approved researchers for analysis to be integrated into the wider global cancer community. However, the data-sharing plan for TCGA has not necessarily been embraced universally, a point that continues to inhibit “discovery genomics” and its clinical application. This age of “Big Data” genomics will efficiently advance the development of a series of application program interfaces that link different data platforms, thus breaking down traditional silos and liberating the data to enable new discoveries and ultimately benefit patients. The ease with which 21st-century molecular technologies have empowered large-scale cancer genomics projects has been impressive, but the unparalleled level of data generated creates its own bioinformatics challenges, making it difficult both to access the data and to conduct new analyses. The full characterization of TCGA is substantive, on the order of multiple petabytes, and it has been released to approved researchers for analysis to be integrated into the wider global cancer community. However, the data-sharing plan for TCGA has not necessarily been embraced universally, a point that continues to inhibit “discovery genomics” and its clinical application. This age of “Big Data” genomics will efficiently advance the development of a series of application program interfaces that link different data platforms, thus breaking down traditional silos and liberating the data to enable new discoveries and ultimately benefit patients.

Free the Data: Unlocking the Genomic Silos—the Global Alliance for Genomics and Health

Recognizing the need to unlock “genomic silos” and liberate data for collaborative research/clincial application, a group of like-minded individuals and organizations has created the Global Alliance for Genomics and Health (GA4GH), an international coalition of agencies, universities, and biomedical research institutions, healthcare providers, information technology and life-science companies, patient advocacy organizations, and research funders, dedicated to improving human health by maximizing the potential of genomic medicine through more effective and responsible data sharing (2). Since its inaugural plenary meeting in March 2014 at the Wellcome Trust, GA4GH now has over 360 organizational members from 35 countries throughout the world (3), ranging from global corporations such as Amazon, to Leucine Rich Bio, a bioinformatics company based in Bangalore; from funding agencies such as the Medical Research Council in the United Kingdom, to Zeta Biosciences, a diagnostics start-up in Shanghai.

GA4GH’s mission is to “accelerate progress in human health by helping to establish a common framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data, and by catalyzing data sharing projects that drive and demonstrate the value of data sharing” (4). This will be achieved by (i) convening stakeholders from different sectors; (ii) catalyzing data sharing among members; (iii) creating harmonized approaches with particular analyses of smaller datasets; (ii) the lack of interoperability between old and new datasets; and (iii) the reluctance to share data and provide opportunities for novel investigations. Until the community develops robust guidelines for interoperable data sharing, congruent with local privacy requirements and data protection standards, the field will continue to miss opportunities for genomic discovery and its implementation in clinical care pathways.
emphasis on interoperable tools and methods; (iv) sharing and implementing best practices; (v) fostering a culture of innovation and discovery; and (vi) committing to responsible and secure data sharing, with the overall aim to advance and improve human health.

REALIZING THE VISION: ENABLING A DATA-SHARING CULTURE

Global Alliance members are linked by their unity of purpose—to maximize the public benefit of genomics by (i) catalyzing effective integration of molecular and clinical data and (ii) facilitating constructive and effective sharing of these datasets on a global level. Promoting a data-sharing culture has been at the heart of GA4GH activities since its inception. A key early output from the Global Alliance has been the production of a “Framework for the Responsible Sharing of Genomic and Health Related Data” (5), a blueprint that recognizes the principles of privacy and nondiscrimination, but is, nonetheless, rooted in the practicalities that allow the development of harmonized approaches to responsible sharing of genomic and clinical data.

On a practical level, the Global Alliance’s Data Working Group has created a GA4GH Genomics Application Programming Interface (API; refs. 6, 7), which facilitates the exchange of data on DNA sequences and genomic variation on multiple platforms and across multiple organizations. This stable version 0.5.1 API (which is freely available) is designed to overcome the barriers of incompatible infrastructure that currently exist, while taking advantage of common web protocols to make it user friendly; both of these principles echo the Alliance’s vision for maximum sharing of genomic information by both data providers and data “consumers.” The GA4GH’s API provides support to researchers by helping them to manage next-generation sequencing (NGS) “reads” and genomic variants, and other levels of functionality, including genetic variant effect annotations, metadata, and genotype-to-phenotype associations, are in the pipeline. This API has already been widely adopted, in part or whole, by a number of users, including the European Bioinformatics Institute, the U.S. National Center for Biotechnology Information, Google, Genome Savant, and Substitutable Medical Apps Reusable Technologies (SMART) Genomics.

GA4GH is also working with a number of projects/consortia, including Café Variome (8), ClinGen (9), DECIPHER (10), GEM.app (11), GeneMatcher (12), LOVD (13), Phenome Central (14), and the National Human Genome Research Institute’s Undiagnosed Diseases Network (15), in an open collaborative effort called Matchmaker Exchange, which is establishing a federated platform where rare disease cases with specific genetic and phenotypic profiles from different databases can be matched to aggregate evidence in support of novel causes of genetic disease.

EMPOWERING DATA-SHARING INITIATIVES IN CANCER

The initial clinical focus of the Global Alliance, mediated through its Clinical Working Group (cochaired by K.N. North and C.L. Sawyers), has been on rare diseases and cancer, two areas where a collaborative data-sharing approach has the potential to significantly increase the value of individual datasets/databases and translate this value directly to patients for their benefit. Currently, many research centers, clinical laboratories, and rare disease/cancer consortia uncover variants through exome or whole-genome sequencing, but are unable to define the clinical relevance/utility of their findings due to the lack of a similar variant in their own database.

Somatic Tumor Case Matching

In the particular case of cancer-sequencing initiatives, an identical mutation may exist in a different dataset, but fragmentation of data and lack of interoperability between databases make the aggregation of such cases difficult. Initiatives such as Matchmaker Exchange can provide the functionality that could be adopted for somatic tumor case matching in cancer, thus empowering the linking of genomically/phenotypically similar cases from different datasets that can inform both future research questions and help drive potential clinical interventions.

Linking Genotype to Phenotype: The BRCA Challenge

At the inaugural plenary meeting of the Global Alliance, a cancer roundtable (cochaired by M. Lawler and L.L. Siu) led to the generation of the first cancer-specific GA4GH demonstration project. The BRCA Challenge (led by S.J. Chanock and J. Burn) has assembled the world leaders in BRCA research and care and is developing a collaborative environment to pool data from BRCA diagnostic and research sources. Expert-informed clinical annotation of each variant will result in a reliable catalog of BRCA variants, both pathogenic and nonpathogenic with respect to their phenotypic effects. The potential to access and rely upon this resource at a global level is compelling, informing both preventative and therapeutic strategies in cancers in which BRCA has been implicated, such as breast, ovarian, pancreatic, and prostate cancers.

Empowering Cancer Care Coordination—the Precision Cancer Medicine App

Cancer-specific apps are now in development and being deployed to empower clinical/molecular data-sharing approaches that can underpin clinical benefit. For instance, the SMART Genomics Precision Cancer Medicine (PCM) app (16) integrates cancer genomic and clinical information to compare an individual patient to others with similar phenotypes and genotypes, using electronic clinical information sharing using Fast Healthcare Interoperability Resources, a new language that promotes clinical data interoperability. PCM is used as a point-of-care mobile app at the Vanderbilt Ingram Cancer Centre (Nashville, TN) that integrates patient, clinical, and somatic mutation data and is intended to aid cancer care coordination for both the physician and the patient.
data generated from different clinical sequencing activities in the global cancer community, in order to maximize the value of these datasets to inform future patient care. A number of interinstitutional and/or pan-country activities are ongoing.

**SPECTA**

SPECTA (Screening Patients for Efficient Clinical Trial Access; ref. 17) is the European Organization for Research and Treatment of Cancer (EORTC)'s molecular screening platform that aims to identify a patient's individual molecular profile and use this information to match the patient with an appropriate molecularly guided/targeted clinical trial. SPECTA is a unique pan-European multitumor longitudinal clinically annotated dynamic biobank, while also providing real-life benchmarking cohorts of patient-level detailed biologic information. This collaborative data-driven healthcare approach, involving over 40 clinical centers from more than 15 European countries, has the potential to accelerate access to novel treatments for patients with cancer in a variety of tumor types. In colorectal cancer, SPECTAcolore has recruited over 670 patients, facilitated through its collaborative clinical centers across the pan-European EORTC Gastrointestinal Cancer Group; SPECTAlung opened for recruitment in June 2015, and similar “matchmaking” approaches are at various stages of development in melanoma, brain, prostate, and rare cancers.

**NCI-MATCH**

NCI-MATCH (the NCI’s Molecular Analysis for Therapy Choice; ref. 18) is a precision medicine clinical trial that includes multiple arms that assess the effectiveness of treating different cancers according to the specific molecular abnormalities detected in the tumor. The trial design is extremely flexible, with the ability to add or drop different treatment arms based on efficacy over the course of the study. NCI-MATCH will commence shortly, initially with 10 different arms targeting different “actionable” mutations. Approximately 3,000 different tumors will be screened by NGS with a predicted 1,000 of these patients matched to new or existing therapies, with plans to expand the study to 20 to 25 treatment arms, over 2,400 sites within the United States, emphasizing the power and scope of this approach.

**Cancer Core Europe**

Cancer Core Europe (19) is a collaborative approach between six European institutions [Institut Gustave Roussy Cancer Campus (IGRCC), Paris, France; Cancer Research UK Cambridge Institute, Cambridge, UK; Netherlands Cancer Institute, Amsterdam, the Netherlands; German Cancer Research Centre (DKFZ), Heidelberg, Germany; Karolinska Institute, Stockholm, Sweden; and Vall d’Hebron Institute of Oncology, Barcelona, Spain] to develop a European Virtual eCancer Hospital with intercompatible clinical molecular profiling laboratories. Sharing of laboratory and IT protocols permits aggregation of genomics and associated clinical data from each of these six leading European institutions, representing an unrivalled discovery resource for identifying new targets for clinical intervention.

**GENIE**

The GENIE project (Genomics Evidence Neoplasia Information Exchange; led by C.L. Sawyers) is a pilot collaboration initiated by the American Association for Cancer Research (AACR) to integrate genomic profiles and clinical outcome of tens of thousands of patients sequenced locally at eight different cancer centers in the United States, Canada, and Europe [Centre for Personalized Cancer Treatment, the Netherlands; Dana Farber Cancer Institute (DFCI), Boston, MA; IGRCC, Paris, France; Memorial Sloan Kettering Cancer Center, New York, NY; The University of Texas MD Anderson Cancer Center, Houston, TX; Princess Margaret Cancer Centre, Toronto, Canada; University of California San Francisco Helen Diller Family Comprehensive Cancer Center, San Francisco, CA; Vanderbilt Ingram Cancer Center, Nashville, TN]. Similar to Cancer Core Europe, it aims to leverage aggregated clinical and genomic data to underpin new treatment modalities that can yield significant clinical benefit.

In addition, CancerLinQ (Learning Intelligence Network for Quality; ref. 20), an American Society of Clinical Oncology (ASCO)-led health information technology approach, will aggregate relevant data to create a learning health system for personalized cancer care. Aggregation and analysis of these “real-world” clinical data and delivering real-time feedback to clinical providers will enable oncology practices to measure their performance against peer-informed standards, will provide individual clinicians with quality data that can underpin personalized cancer care decisions, and will uncover novel patterns in aggregated data that can inform outcomes research, benefits, costs, and new treatment options for cancer patients.

Developing common data schemas and interoperability approaches for the diverse datasets generated by all of these approaches can help inform precision medicine strategies that deliver true therapeutic benefit for the patient with cancer.

In addition to the molecular and clinical data-sharing collaborations outlined above, the cancer genomics community has also responded to the recognized need to develop optimal standards and guidance for NGS approaches, underpinning the identification of gene mutations in different cancers. Some of these challenges are not insignificant, particularly if different institutions perform sequencing at different depths and use different mutation-calling algorithms. A number of initiatives are looking at ways to harmonize these approaches. For example, academic institutions including DFCI, Fred Hutchinson Cancer Center, Memorial Sloan Kettering Cancer Center, and MD Anderson Cancer Center, have been sharing best practices for sample preparation, NGS analysis, and clinical reporting of actionable cancer mutations. Similar efforts are occurring within ClinGen, the NIH-funded resource that is primarily focusing on germline alterations across human disease but is now expanding into somatic cancer mutations. GA4GH is the natural home for integrating these diverse but overlapping approaches, providing an overarching framework that can maximize the “added-value” benefit and ensure inclusive, transparent, and responsible data sharing between these different activities. All of the approaches outlined above represent ways in which the cancer community is responding to the clinico-genomic...
data challenge. In order to harmonize these efforts, GA4GH is fulfilling the “honest broker” role in developing, nurturing, and implementing a global cancer genomic and clinical data-sharing blueprint through this newly established Actionable Cancer Genome Initiative.

HARNESSING THE POWER OF GLOBAL CLINICO-GENOMIC DATABASES FOR IMPROVED CANCER OUTCOMES

The activities highlighted above emphasize the Global Alliance’s momentum—its ability to convene relevant stakeholders from across the globe, but also to harness disparate expertise to deliver workable solutions to data-sharing challenges. This is particularly relevant in cancer as we move away from the “one-size-fits-all” therapeutic blockbuster to a more tailored solution (through either targeted therapy or drug repurposing) for molecularly defined subgroups of patients. In this scenario, the number of patients in these segmented subgroups may be small, challenging our ability to deliver innovative but inexpensive cancer care. However, by aggregating different databases at a global level, we expand the cohort of patients who would benefit from this precision medicine approach. Initiatives such as GENIE, with its intercontinental axis, will provide a blueprint for global partnership and implementation in this rapidly evolving domain.

Implementing the scale of therapeutic intervention envisaged by some of the approaches outlined above is extremely challenging under current health delivery models, not least for the pharmaceutical industry, which tends to focus on specific geographic regions/pricing levels. There are also significant ethical and regulatory challenges, but developing a manageable strategy to harness the enormous potential of global clinico-genomic databases can potentially deliver a more cost-effective therapeutic solution that would benefit healthcare providers, healthcare payers, and industry, but most importantly patients with cancer.

In conclusion, as industrialist Henry Ford once said, “Coming together is a beginning; keeping together is progress; working together is success.” This philosophy of a collaborative workforce with different skills working together toward a common goal that yielded such success in the transportation industry resonates well with the work of the Global Alliance. In the 2 years since its inception, the culture that GA4GH has nurtured within the cancer community has underpinned a tangible willingness to collaborate in an altruistic and productive fashion. A number of practical solutions to the data-sharing challenge have been created. GA4GH continues to develop momentum, catalyzing a data-sharing ethics that is anticipated to enhance global cancer research efforts and champion their translation into meaningful benefits for patients with cancer.

Disclosure of Potential Conflicts of Interest

J. Burn reports receiving commercial research support from AstraZeneca and has received speakers bureau honoraria from the same.

C.L. Sawyers is a director at Novartis. No potential conflicts of interest were disclosed by the other authors.

Published online November 2, 2015.

REFERENCES

All the World's a Stage: Facilitating Discovery Science and Improved Cancer Care through the Global Alliance for Genomics and Health

Mark Lawler, Lillian L. Siu, Heidi L. Rehm, et al.

Cancer Discov 2015;5:1133-1136.

Updated version
Access the most recent version of this article at:
http://cancerdiscovery.aacrjournals.org/content/5/11/1133

Cited articles
This article cites 10 articles, 1 of which you can access for free at:
http://cancerdiscovery.aacrjournals.org/content/5/11/1133.full#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://cancerdiscovery.aacrjournals.org/content/5/11/1133.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, contact the AACR Publications Department at permissions@aacr.org.