KAISER
ASSOCIATES

DELIVERING **PRECISION** MEDICINE

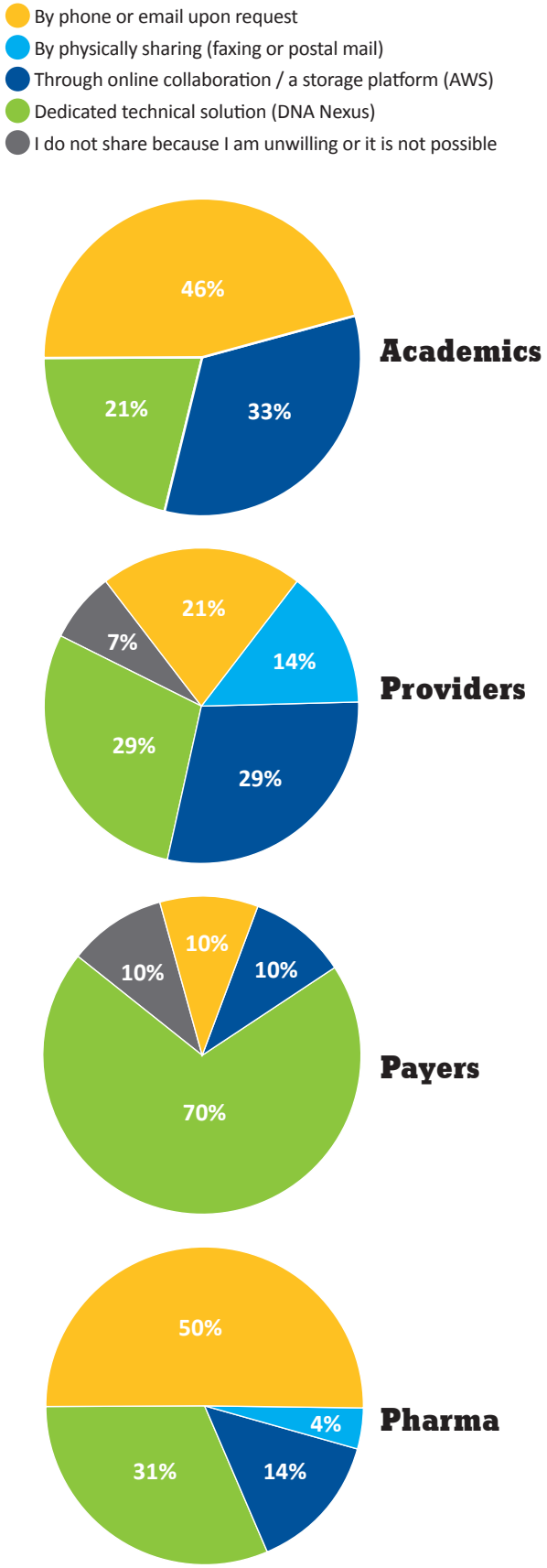
The new challenges and channels for advancing medicine with genomic data

INTRODUCTION

The rapid improvement of genetic sequencing technology has facilitated significant developments in our understanding of human health. Remarkably, the first human genome was sequenced in 2003 by the National Genome Research Institute, the Department of Energy, and their partners in the International Human Genome Sequencing Consortium, and cost $2.7 billion dollars.[1] The cost and innovation curve on DNA sequencing is outpacing Moore's law[2] and creating an unprecedented amount of data that is not only profoundly personal, but also reshaping the way we approach healthcare. Insights from genomes have assisted in uncovering a number of acute genetic differences that can affect how different people respond to treatments, thereby progressing the ability to deliver personalized treatments for patients that best suited for their genetic makeup. By 2015, 13 personalized treatments were approved by the FDA, whereas none were approved in 2010.[3]

While these advancements have drastically impacted our understanding of medicine, they have also unlocked new needs from multiple stakeholders. Clinicians, Academics, and Pharmaceutical Researchers seek technology and data solutions to assist and enhance their ability to progress their work. In order to continue to investigate how genetic differences can impact the most effective treatments, it is important that researchers draw statistically significant conclusions about genomic information. Building this kind of scale, increasing the speed at which analysis can be done and results delivered promises to unlock genomics potential as an important industry in our future economy and for our future personal health. This impetus for change provides significant opportunity for technical solutions which will lead to the future success of genomic research. The pinnacle of this opportunity and change is the development of the President's Precision Medicine Initiative and the Vice President's Cancer Moonshot, which seek to leverage public resources to advance medicine and includes initiatives to generate and collaborate on data. Vice President Biden addressed a meeting of the American Association for Cancer Research in April saying, "my job and commitment is to bring together all the human, financial, and knowledge resources we have in the world to seize this moment, to make decades worth of progress in five years."[4] This paper will explore the barriers facing those medical professionals and researchers in rapidly advancing precision medicine, as well as the solutions they have sought to address those needs and to meet the expectations of initiatives like the Cancer Moonshot.

How data is shared today[5]



BARRIERS

Data capabilities for genomics are challenged today in three significant ways: The first is that not all genomic information is collected in the same way or with the same intention. By example, many studies are focused on a specific type of cancer and may sequence tissue samples from a tumor rather than conduct a whole genome sequence for the patient. The second is that data is often considered intellectual capital of organizations due to the insights locked within it. 75% of pharmaceutical researchers surveyed in a Kaiser Associates study who do not currently exchange any data stated that legal or security policies of their company would have to be changed in order to increase collaboration.[5] The third is that the cost of generating this type of data, while decreasing, is still significant. Not every institution can resource genomic testing, particularly because genetic testing, in most cases, is not yet reimbursable by health plans.

"The '$1,000 genome' is a long-sought marker of advancement that the genetics industry has seen as 'a tipping point in the pace of genetic discovery.'"

–Aaron Krol, BioITWorld[6]

The commonly reported $1,000 figure, however, does not even begin to cover the costs of running, storing, and maintaining the system and its associated data. Therefore, only well-resourced organizations have the ability to develop internal databases.

Not every component of the data in these data siloes is being used in a competitive way. There are many opportunities in which researchers could work together for both personal and societal benefit, as demonstrated by the multiple partnerships that arise between public and private institutions.

"In an ideal world, [we'd be] without interests and politics, but the current system is very siloed. These silos make sense at points: In clinical applications in which there is managed care or integration with the insurer. In managing a complex case, however, there will be gaps that affect care."

– Medical Director, Managed Care Organization

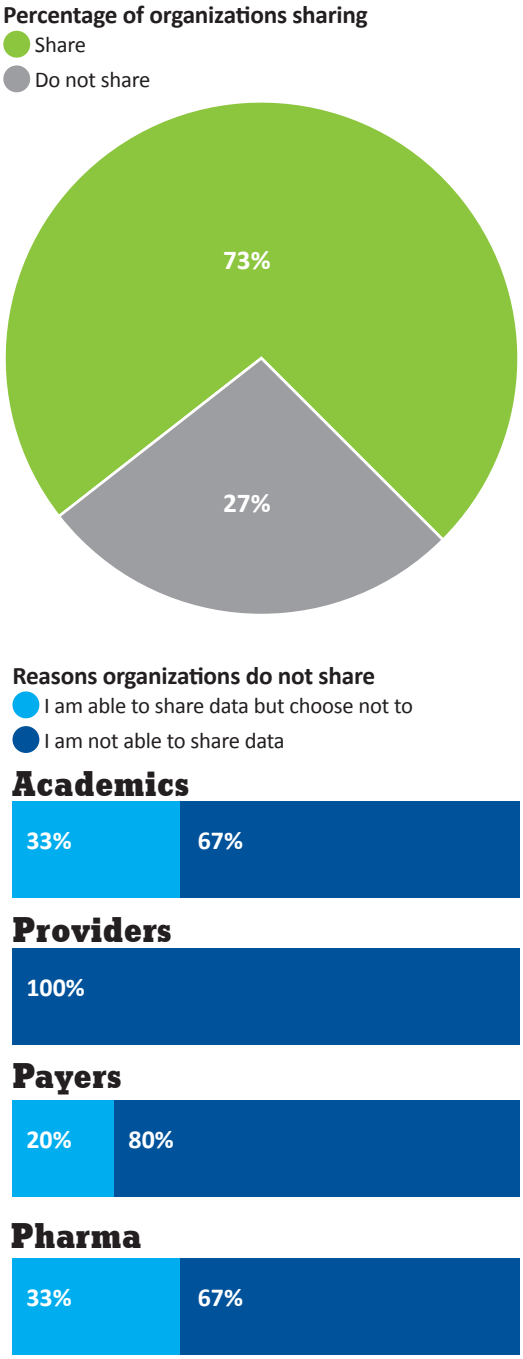
This type of access to a greater sample size ranks among the top two reasons US Academics, Pharmaceutical Researchers and Clinicians are interested in sharing genomics information.[5] Access to a greater sample is particularly important in the case of rare "orphan" diseases, given that there is a smaller universe of potential patients.

"We run trials where we are targeting a small subset of lung cancer. These patients are very, very rare…there are so few patients in the field sometimes the trial is not feasible. Maybe a provider will see one patient in a lifetime that is like a certain patient."

– Scientific Director, Top 10 Pharmaceutical Company

Despite advancements in testing, without greater availability of data, researchers will have a difficult time drawing significant conclusions about any one genetic factor. Similarly, drug trials for rare diseases are often inhibited by a limited number of patients available to conduct a clinical trial on. Furthermore, genomic analysis has deepened scientific understanding about the heterogeneity of cancer tumors from patient to patient. This added accuracy actually poses a similar challenge in designing trials for more prevalent types of cancer by reducing the universe of viable patients. In these cases, greater access to data through data collaboration could serve to inform researchers about available cohorts and vice versa.

Degree of data sharing, and drivers for not sharing[5]



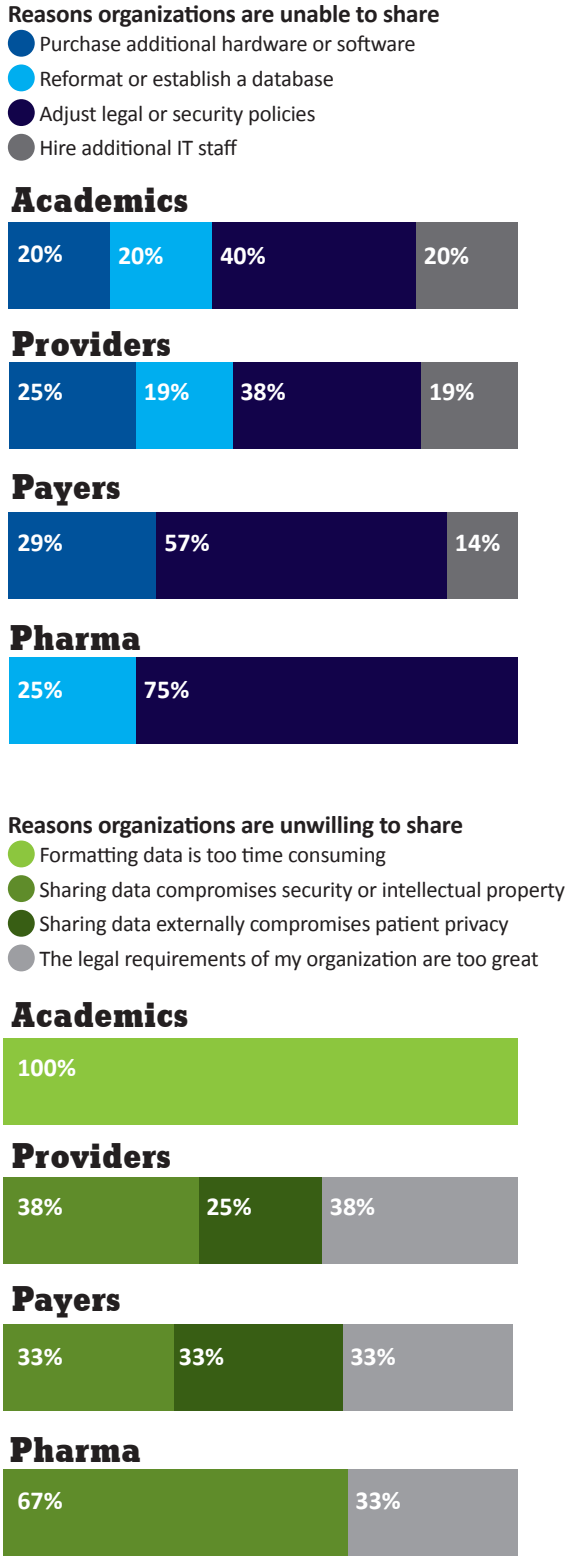
BARRIERS

"Accuracy is more important…at the end of the day I don't want to deploy a team to investigate something that is wrong."
– VP R&D, Specialty Pharmaceutical Company

In order to create the data environment necessary to meet these needs, the burden of handling the data needs to be eased. The critical factor is that genomic data sets are massive—each sequence generates gigabytes of data even after state-of-the-art compression. The National Cancer Institute (NCI) uncovered that if a research organization were to download and house the information in the Genomic Data Commons it would take 23 days to download and about $2M dollars to store.[8] To give perspective, the Genomic Data Commons will be comprised of over 2.5 petabytes of genomic information, which is estimated to be about 10% of genomic data currently unconnected and stored in disparate locations.[9] [10] Transferring these files to alternative databases not only creates inefficiency, but also potential security problems depending on how the information is shared. Currently, information is shared in a variety of inefficient and non-scalable ways: 46% of academics and 50% of pharmaceutical researchers identify phone or email as their primary method for sharing genomic data. The need for effective solutions is compounded by the same researchers (56% of clinicians, 41% of academics, and 40% of pharmaceutical researchers) indicating that data security is a significant problem with their current data sharing tools.

"No matter how big the collaboration, when someone else wants to query, they will be limited to an extremely specific question legally. Exploratory queries are much more difficult. More contracting that is difficult, much more complex because of privacy concerns."
—Medical Director, Top 10 Pharmaceutical Company

Reasons organizations are unable or unwilling to share data[5]



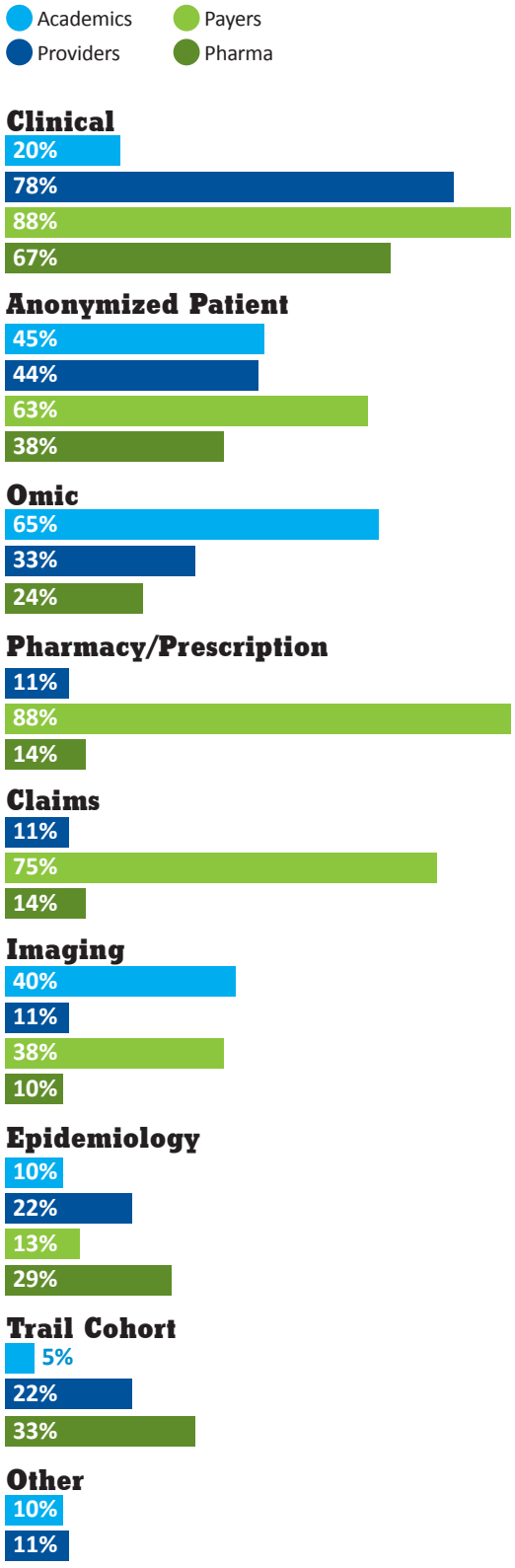
Finally, these barriers also inhibit the development of effective solutions for Clinical Decision Support (CDS) which will enable the real-time application of genomic information to improve patient health. As analytics and electronic medical record systems improve, Clinical Decision Support is becoming an increasingly powerful tool for clinicians as it can provide guidance on a number of factors including patient history, drug interactions, and even treatment plan. Some current CDS systems incorporate a number of diagnostic tests based on genomic information and can assist doctors in determining the course of treatment based on the results of the tests. However, as costs come down, it will become more economically viable to sequence the patient's entire genome rather than conduct diagnostics. The analysis involved is a complicated process, however, and most clinicians in a recent Kaiser Associates study indicated they are very interested in utilizing these results.[5] By integrating and analyzing existing data sources like Electronic Medical Records with Whole Genome Sequences, clinically actionable insights are far more likely to be developed. The integration of patient information and genomic data, however, will assist in discovering patterns in treatment that can serve as evidence for new standards of care and support the everyday utility of genomics in a clinical setting.

EMERGING SOLUTIONS

In an effort to overcome these barriers, a number of solutions are beginning to emerge. Adoption of the solutions will be driven by their ability to deliver greater statistical power through an increased sample size but also have the ability to assist in reducing the analytical burden of massive genomic data sets. Broadly speaking, there are two major arenas for innovation and solutions-building: coordination through alliances and technical solution development. Alliances present the most immediate solutions, given that they generally fit within the frameworks that have been historically employed. However, some technical solutions are offering radically different approaches to solving these challenges.

The American Association of Cancer Research's GENIE Project, the ORIEN Alliance, American Society for Clinical Oncology's CancerLinQ, and the Pistoia Alliance in the European Union rank among the most notable alliances, and each uniquely enables access to larger data sets. By example, the GENIE project combines seven phase one projects into a single registry. Participants include academic medical centers from four countries. The genomic information included cov-

Types of data shared, by stakeholder group[5]



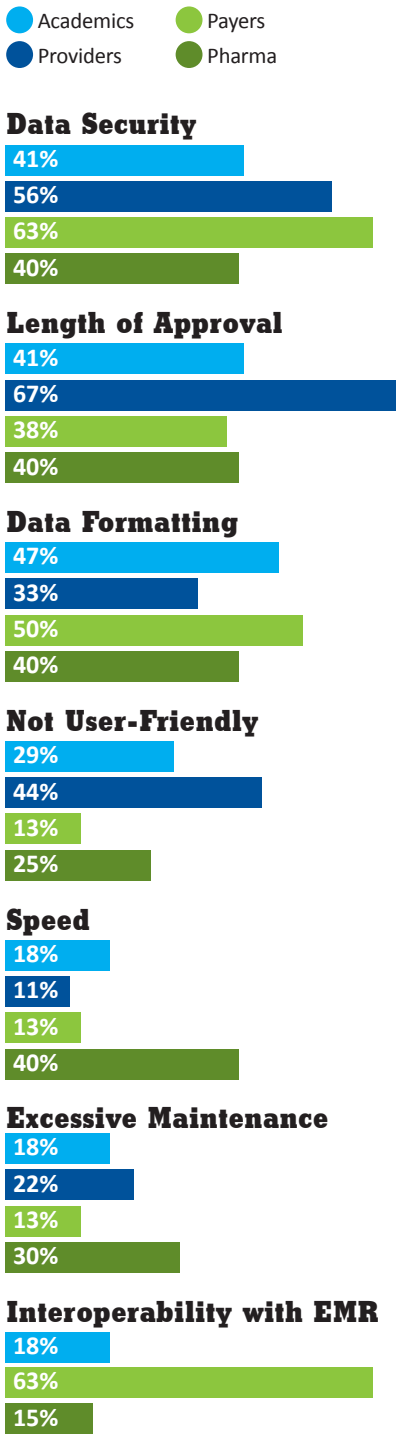
ers multiple cancer types, including rare cancers, and is accessible for researchers to conduct in depth analyses.[12] While this project makes important progress, researchers would still benefit from having access to data outside of a limited number of clinical trials to incorporate other samples from past research projects within large medical organizations.

Alliances are often limited to one type of stakeholder group, such as Academic Medical Centers or Pharmaceutical companies. There is significant opportunity to improve data access for a diverse sets of stakeholders at the same time. This is what many Technical solutions are seeking to accomplish, and meeting the needs of such a diverse set of stakeholders is a monumental task. In reaction, Healthcare IT and analytics providers are taking an equally diverse set of approaches to generating solutions. By example, Google and Microsoft have developed Cloud Storage options with computing, analytical, and sharing offerings. IBM's Watson, ADAM (a Hadoop platform), and Intel's Collaborative Cancer Cloud (CCC) have taken alternative approaches to enabling data access and analytics. Intel's CCC takes a different spin on the cloud that enables collaboration without requiring data to leave the data owner's data center. CCC is enabled by technology that supports secure collaboration by querying across multiple sites facilitated by a central execution engine. These queries return rapid results through Intel's GenomicsDB technology that have the capability to accelerate this analysis process 11x.

"The time it now takes to perform the variant discovery process went from eight days to 18 hours, however, that's with 100 whole genomes. We routinely process projects with thousands of samples, so that speedup itself is truly transformative. We recently needed to abandon our attempt to run variant discovery on an eight thousand sample project because we estimated it would take 90 days without GenomicsDB. With GenomicsDB, however, it should take under a week. This means we can say 'yes' to our researchers far more often, on far more ambitious projects."

– Eric Banks, Senior Director of Data Sciences and Data Engineering at Broad[13]

Challenges with current data sharing approaches, by stakeholder group[5]



Solutions like Intel's CCC have the opportunity to solve many of the needs emerging in the advancement of genomic research and personalized medicine.

"The ability to process a lot of data from many patients to figure out who has the markers for a certain drug would be good [because it would] expedite the whole process. This is important for some of these cancers like acute leukemia, [where] you need to treat that patient immediately. Hours matter when the blood stream is getting blocked off. You need to get the right drug in immediately."

– Surgeon & Department Chair, Top Tier Academic Medical Center

Access to greater amounts of data through secure channels will enable new technical solutions to meet the needs of a diverse set of stakeholders. If met, these solutions have the ability to facilitate the capture of a tremendous business opportunity, and a great change in the way that healthcare is delivered.

METHODOLOGY

Kaiser Associates' research methodology for this white paper focused on current practices, future state, pain points, adoption drivers, and barriers. In addition to a comprehensive literature review, Kaiser conducted 30 in-depth interviews and a 76-respondent online study with senior / executive representatives from four stakeholder groups – Academics, Pharmaceutical Researchers, Payers, and Clinicians.

This study was sponsored by Intel Corporation.

SOURCES

[1] **Human Genome Project. The Human Genome Project Completion: Frequently Asked Questions**
Available at: https://www.genome.gov/11006943/human-genome-project-completion-frequently-asked-questions.
Accessed July 8, 2016.

[2] **Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)**
Available at: www.genome.gov/sequencingcostsdata.
Accessed July 8, 2016.

[3] **Pritchard D. Seizing Momentum: Reflecting on the 'Year of Personalized Medicine'**
Available at: http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/Personalized-Medicine-in-Brief1.pdf.
Accessed July 8, 2016.

[4] **Lowy DR. A Shared Commitment for Accelerating Progress with the Cancer Moonshot**
Available at: http://www.cancer.gov/news-events/cancer-currents-blog/2016/moonshot-biden-aacr.
Accessed July 8, 2016.

[5] **Kaiser Associates Study**

[6] **Krol A. What You Need to Know About Illumina's New Sequencers.**
Available at: http://www.bio-itworld.com/BioIT_Article.aspx?id=133900.
Accessed July 8, 2016.

[8] **Davidsen T. National Cancer Institutes: NCI Cancer Genomics Cloud Pilots (and Genomic Data Commons).**
Available at: https://www.genome.gov/multimedia/slides/tcga4/23_davidsen.pdf.
Accessed July 8, 2016.

[9] **Davidsen T. National Cancer Institutes: NCI Cancer Genomics Cloud Pilots (and Genomic Data Commons).**
Available at: https://www.genome.gov/multimedia/slides/tcga4/23_davidsen.pdf.
Accessed July 8, 2016.

[10] **Gebelhoff R. Sequencing the genome creates so much data we don't know what to do with it.**
Available at: https://www.washingtonpost.com/news/speaking-of-science/wp/2015/07/07/sequencing-the-genome-creates-so-much-data-we-dont-know-what-to-do-with-it.
Accessed July 8, 2016.

[12] **American Association for Cancer Research. Project GENIE Infographic.**
Available at: http://www.aacr.org/Documents/GENIE_InfoGraph.pdf.
Accessed July 8, 2016.

[13] **Broad Institute. Broad Institute, Intel work together to develop tools to accelerate biomedical research.**
Available at: http://genomicinfo.broadinstitute.org/acton/media/13431/broad-intel-collaboration.
Accessed July 8, 2016.

ABOUT KAISER ASSOCIATES

ABOUT KAISER

Founded in 1981, Kaiser Associates is an international strategy consulting firm that serves as a key advisor to the world's leading companies. We provide our clients with the unique insight to drive critical decision-making and solve their most pressing problems.

Kaiser's Healthcare Practice engages with executives at leading Life Sciences companies, including pharmaceutical, medical device, clinical diagnostics, consumer health, and health IT. We work with our clients to identify new growth markets, develop long-term portfolio strategies, and maximize commercial success.

Kaiser's Technology Practice advises clients whose businesses span the whole stack from infrastructure and operating systems to gaming and analytics using a unique blend of quantitative analytics fueled by a best-in-class primary research capability. Kaiser's projects typically center on market opportunity assessments and competitive landscape understanding in hardware, software, and services, with substantial concentration in IoT and cloud topics.

ABOUT THE AUTHORS

Bob Serrano

Bob is a Vice President in Kaiser Associates' Global Healthcare Practice based in Washington, D.C.

You may contact him by email at bserrano@kaiserassociates.com

Jeremy Williams

Jeremy is a Principal in Kaiser Associates' Technology Practice based in Washington, D.C.

You may contact him by email at jwilliams@kaiserassociates.com

Lily Carpenter-Israel

Lily is an Analyst in Kaiser Associates' Global Healthcare Practice based in Washington, D.C.

You may contact her by email at lcarpenterisrael@kaiserassociates.com

Design and Graphics by Kelly Martin Design