IBM Institute for Business Value

The evolving promise of genomic medicine

How advanced technologies are transforming healthcare and life sciences

IBM Healthcare and Life Sciences

IBM Healthcare & Life Sciences is creating a smarter, more connected healthcare system that delivers better care and empowers people to make better choices. In addition to the company's investment in health technology research and innovation, IBM's healthcare solutions and consulting enable organizations to achieve greater efficiency within their operations, collaborate to improve outcomes, and integrate with new partners for a more sustainable, personalized and patient-centric system focused on value.

By Aditya Pai, Takahiko Koyama and Leonard Lee

After the Human Genome Project was completed in 2003, early successes in genomic medicine fell short of the initial high expectations. But today, a potent mix of influences—including innovation in biology and technology, market demand and consumerism—is furthering an evolution that crosses industries.

Healthcare providers can now personalize care plans thanks to lower sequencing costs that allow genomic data to be combined with electronic health data. Life sciences companies can develop targeted therapies that prevent and alleviate disease symptoms. To capitalize on new advances in science, cognitive computing, analytics and drug discovery, senior leaders across the ecosystem should act quickly to: make genomic medicine a key component of enterprise strategy; address relevant skill gaps; and determine how partnering can bolster critical capabilities.

In three to five years, an individual who interacts with the healthcare system will likely have a genome-based electronic record or "genomic health record."

Sequencing capability will increasingly happen in a cloud-based environment enabling the aggregation and analysis of multiple sources of genomic data.

Transformational opportunity for industries rests in leveraging cognitive computing throughout the entire value chain from pre-clinical discovery to understanding and engaging consumers, and improving health outcomes.

ge·no·mic \ ji-NOH-mik \ medicine

Genomic medicine is the application of the science of genomes, as opposed to genes, to the practice of medicine. The National Institutes of Health describe a genome as "an organism's complete set of *DNA*, including all of its genes. Each genome contains all of the information needed to build and maintain that organism. In humans, a copy of the entire genome — more than 3 billion DNA base pairs — is contained in all cells that have a nucleus."[1]

Note: Please see the Glossary on pages 17 and 18 for definition of important terms and abbreviations

Since 2001, when the first draft of the Human Genome Project was announced, hope surged for genomic medicine to be a panacea for many health conditions from diagnosis to treatment. Soon after the project was complete in 2003, the early tangible examples of success did not match the high expectations of that time. But today, there is renewed confidence in the healthcare and life sciences industries that the hype could indeed match the hope.

Consider what is different now:

- In the healthcare industry, the decreased cost of whole-genome sequencing—coupled with new advances in cognitive computing and drug discovery—has created a new paradigm where genomic data will soon be combined with electronic health data. Care providers are increasingly relying on genomic data to add a unique level of personalization to an individual's care plan. Cognitive computing and other analytics technologies can provide for precision care where decision support enables a reliable diagnosis and care plan, including treatment options.
- In the life sciences industry, the stage is set for a radical transformation. New medical and technological capabilities will increasingly lead to innovative clinical trials, the development of targeted therapies and a focus on health outcomes to prevent and alleviate disease symptoms.

These major changes are positively impacting the three "cornerstones" of genomic medicine: sequencing, translational medicine and personalized healthcare. As genomic medicine continues to proliferate, the importance of a solid privacy, ethical and legal framework to support the complexity of genomic medicine will also become vital.

The field of oncology is already experiencing advances in genomic medicine that are expected to increase on a larger scale through sophisticated genome/proteome/RNA analysis. Coupling such analysis with cognitive computing, oncologists can provide rapid precision oncology decision support.

Introduction

The Human Genome Project was declared complete in April of 2003.[2] Since then, activity has exploded to make sense of our human DNA (Deoxyribo Nucleic Acid) map. Genomic medicine has long had a goal of moving from "one-size-fits-all" to personalized treatments by applying genomic insights at the point of care.

While rational drug design has helped to improve overall drug efficacy levels, many believe that genomic medicine will allow us to better understand the causes and treatment of disease at a molecular level. Consider the following scenario:

> It is 2013 and a 45-year-old woman has been diagnosed with adenocarcinoma of the lung. A PET-CT scan reveals that there is metastasis to the adrenal gland, brain and bone. She meets with her oncologist, who does a lung biopsy of the tumor. Immunohistochemistry confirms that she has Stage IV non-small cell lung cancer.
>
> Her oncologist recommends that she submit her tumor biopsy for molecular testing for EGFR and ALK mutations, which confirms that she has an EGFR3 mutation. This mutation is seen in 30 to 40 percent of women who are non-smokers and of Asian origin, as she is.
>
> The oncologist offers first-line treatment with a particular EGFR inhibitor. Eight weeks later, a PET-CT scan shows shrinkage of her tumors. She does not go through chemotherapy and continues taking a daily pill as part of her treatment regimen.

This scenario is a living example of personalized medicine where genomics research led to the discovery of EGFR mutations and their key role in adenocarcinoma. While EGFR inhibitor drugs have existed in the marketplace for some time, it has been difficult to identify patients for whom a specific drug would have been most effective.

It is now 2015 and whole-genome sequencing coupled with other genomic modalities like RNA-Sequencing and oncology-specific multiplex panels of genes are routinely offered to cancer patients.

The woman, now 47 years old, gets her tumor block analyzed by three such genomic modalities upon expert advice provided by a cognitive system that has been trained in medicine, genomics and the oncology specialty. The cognitive system continues to learn and provides the oncologist with key confidence intervals regarding options.

Upon choosing the option with the highest confidence interval, another mutation is found in the primary tumor which, fortunately, also has a targeted treatment. The woman is now taking two targeted medications and leading a reasonably good quality of life. Her cancer can be treated as a "complex chronic condition" with care that is personalized for her own genetic makeup — targeted care that was not yet available to her in 2013.

Many challenges still exist, but genomic medicine has indeed moved from a vision to a reality. It exists in small, but ever-expanding pockets across the healthcare ecosystem and spans the prediction of drug response to the diagnosis of disease, to the identification of targeted therapies. Major advances have occurred in next-generation sequencing, genome-wide association studies and bioinformatics. In addition, our view of the human genome has expanded to a broader view of t*he human as an ecosystem*, including the bacteria that exist within us, both symbiotic and pathogenic.[3]

As whole-genome sequencing becomes more affordable, personalized medicine will benefit from having EGFR-like regions in the genome analyzed and combined with other clinical and non-clinical attributes — resulting in a truly personalized diagnosis. Various similar examples of genes and corresponding therapies have already been identified for other cancers, such as breast cancer (HER2/neu)[4] and melanoma (BRAF).[5] Similarly, translational medicine is expected to benefit greatly from having more unique and detailed insights into different pathways that explain the molecular basis of disease and provide a sounder rationale for therapies.

The computational challenge in achieving such results on a whole-genome sequence level and combining them with various clinical and non-clinical attributes is not trivial. However, considering the pace at which genomics and its associated "-omics" technologies — which include transcriptomics, proteomics and microbiomics — have evolved, it is only a matter of time before the cancer case scenario has even more specificity and personalization. That said, a concerted effort is required to be able to systematically tackle groups of diseases, each with its own complex mechanisms and relationships, by using analytics and newer advances in computational biology.

Forces impacting genomic medicine in the next five years

Since the completion of the human genome, genomic medicine has been hailed as a promise. Personalized medical treatment, gene therapy, pharmacogenomics, novel biomarkers, and highly specific and sensitive molecular diagnostics attest to the promise.

In the past few years, increased applications of genomic medicine have led to improvements in:

- Cancer detection through blood samples and tumor analysis
- Prenatal screening and diagnosis using fetal cells in maternal blood
- Drug safety labels based on pharmacogenomic data
- Drug dosing, in the case of warfarin,[6] and
- Drug efficacy, for example, related to breast cancer.

While some debate regarding its success continues, we believe that genomic medicine is at a unique inflection point in medicine's history due to three major forces: *innovation in biology and technology; market demand;* and *consumerism*.

Innovation in biology and technology

Today, highly specific panels of genes are being used for cancer testing. Such panels check for the expression of multiple genes known to be associated with specific cancers.[7] The pharmaceutical industry has an increased understanding of complex pathways and corresponding development of targeted drugs. Such drugs include, for example, a list of Tyrokine Kinase inhibitors, as well as drugs that treat resistance to TKIs.[8]

Other innovations include the decreasing cost and increasing capability of high-performance computing platforms that are keeping pace with that of gene discovery, whole-genome sequencing cost and speed. As a result, processing large data sets is becoming easier.

A new facet of computing—cognitive computing—allows a system to make use of natural language processing and machine learning to ingest large and ever-expanding data volumes. It can be trained to provide, with high accuracy and speed, advice which is otherwise manual, time-consuming and prone to missing available evidence.

Market demand

The healthcare and life sciences industries are at a crossroads—characterized by patent expirations, reimbursement pressures and a thirst for new ways of impacting health outcomes through innovative therapies. The increased expectation and market demand for targeted products (such as drugs and tests) can offer significant benefits quickly, versus incremental benefits delivered over time after careful validation and cost benefit analysis.

Genomics is also converging with social media, as evidenced by the sharing of health experiences and questions online. One example is patientslikeme, a health data-sharing platform that includes a blog, where "The future of the personal genome" has been a featured video topic.[9] Increased computing power, imaging capability, the use of wireless sensors and the plethora of accessible digital health information are also part of the technological mix that is helping to push genomics forward. This convergence has led to near-instant dissemination of new discoveries, as well as new forms of social and scientific collaboration.

Consumerism

The growing commercialization of genomics is evident in the increased uptake of direct-to-consumer genomic testing, and in more recent regulatory concerns and recommendations about such technologies.[10] These have led to genomics attracting new media attention and becoming more "mainstream."

Regulatory and legislative precedents in the U.S.—including the 2008 Genetic Non-Discrimination Act (GINA)[11] and the 2013 Supreme Court decision regarding gene patenting with a biotechnology company[12] to protect consumers of genomic data—have led to increased awareness of its sensitive nature.

Exploring major changes ahead from genomic medicine

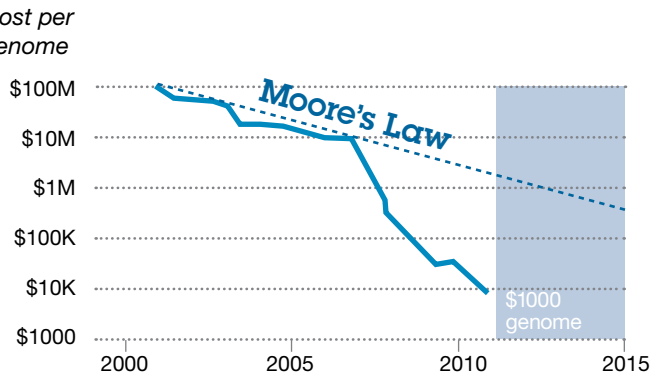
Genomic medicine will be a "game-changer" for important stakeholders that include patients, providers, researchers, payors, diagnostic companies, policy makers, life sciences and governments. Three significant impacts are underway: (1) the growing use of a new genomic health record; (2) greater benefits for stakeholders in the three genomic medicine cornerstones of sequencing, translational medicine and personalized healthcare; and (3) opportunities for radical industry transformation.

1. Growing use of a new genomic health record

In the next three to five years, it is likely that an individual who interacts with the healthcare system will have a genome-based electronic record or "genomic health record." It will be used in tandem with other electronic healthcare tools for decision support, prevention, and customized testing and treatments.[13] At first, the record may not contain the whole genome sequence, but instead perhaps the results from testing specific genes.

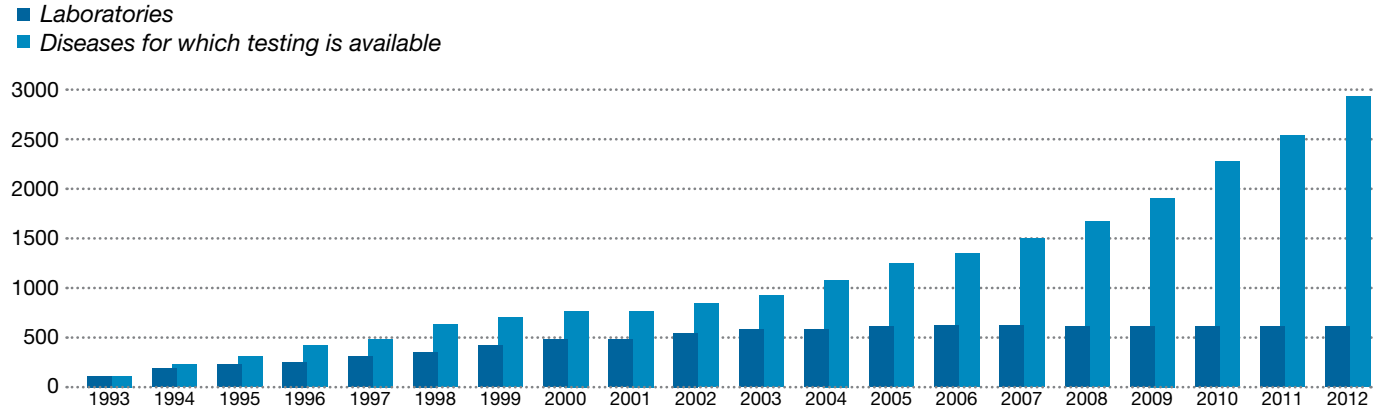
This prediction of such widespread use stems from:

- The decreased costs of sequencing the human genome (see Figure 1),[14]
- The proliferation and availability of genome-based tests in the past five years (see Figure 2),[15]
- The rising adoption of electronic medical records (EMRs, see Figure 3),[16]
- The increased use of genome data to recommend targeted treatments using companion molecular diagnostics, and
- A growing willingness of payors to reimburse payments on some genetic tests today.



Source: National Human Genome Research Institute" http://www.genome.gov/images/content/cost_per_genome.jpg; Adapted by IBM Research.

Figure 1: **Faster processing speeds have reduced the cost of sequencing dramatically.**



Source: Roberta A. Pagon, M.D., Professor of Pediatrics, University of Washington and Editor-in-Chief, GeneReviews www.genereviews.org

Figure 2: **The use of genome-based testing has continued to rise sharply in recent years.**

US EMR Adoption Model℠

Stage	Cumulative Capabilities	2013 Q4	2014 Q1
Stage 7	Complete EMR, CCD transactions to share data; Data warehousing; Data continuity with ED, ambulatory, OP	2.9%	3.1%
Stage 6	Physician documentation (structured templates), full CDSS (variance & compliance), full R-PACS	12.5%	13.3%
Stage 5	Closed loop medication administration	22.0%	24.2%
Stage 4	CPOE, Clinical Decision Support (clinical protocols)	15.5%	15.7%
Stage 3	Nursing/clinical documentation (flow sheets), CDSS (error checking), PACS available outside Radiology	30.3%	27.7%
Stage 2	CDR, Controlled Medical Vocabulary, CDS, may have Document Imaging, HIE capable	7.6%	7.2%
Stage 1	Ancillary — Lab, Rad, Pharmacy — All Installed	3.3%	3.2%
Stage 0	All Three Ancillaries Not Installed	5.8%	5.6%

Note: See Glossary for definitions of acronyms used in this figure.
Source: HIMSS Analytics® EMR Adoption Model℠; 2013 N = 5458; 2014 N = 5445.

Figure 3: **EMR adoption in the U.S. and Canada is maturing.**

A genomic health record would extract and integrate relevant electronic health data with a person's genome data. To use such data effectively, new types of decision support will be required to personalize risk, prevention and follow-up treatment. Owing to the complex uses of and needs for such data, the genomic health record is likely to be distinct from the traditional EMR often generated from an EMR system.

The healthcare and life sciences industries are poised for greater use of clinical decision support. This is supported by recent EMR adoption data.[17] In Q1 2014, approximately 40 percent of U.S. hospitals were at Stage 5-7 on the HIMSS (Healthcare Information and Management Systems Society) EMR adoption Model℠.[18]

As EMRs adopt more features, the overall integration and interoperability of health information will increase and with it, bring a new demand for clinical decision support. Such integration is only the start of truly leveraging genomic information in personalizing health data.

With the rising use of EMRs, healthcare integration and interoperability is finally maturing. Advances in cognitive computing—such as machine learning and natural language processing—will build on this trend to accelerate the adoption of genomic medicine and its integration with electronic health data into this new genomic health record.

Adoption of genomic health records can provide a completely different level of decision support to users of genomic data through the application of cognitive computing. As defined by IBM Research, "Cognitive computing systems learn and interact naturally with people to extend what either humans or machine could do on their own. Cognitive computing systems are not based on programs that predetermine every answer or action needed to perform a function or set of tasks; rather, they are trained using artificial intelligence and machine learning algorithms to sense, predict, infer and, in some ways, think. Cognitive computing systems get better over time as they build knowledge and learn a domain—its language and terminology, its processes and its preferred methods of interacting."[19]

Cognitive computing has the potential to bring unique advances to translational medicine and to clinical decision support at the point of care through two capabilities: understanding natural language text or speech during diagnosis and treatment; and interrogating vast volumes of medical information to provide an accurate, probabilistic answer.

Based on the availability of accurate molecular diagnostic tests, targeted treatments and a greater understanding of pathways and genes that lead to specific cancers, a select group of cancers would be excellent candidates to be used as an initial step. With the massive expansion of knowledge across various health conditions, a person's genomic health record could provide the ultimate source of personalization for a care provider—creating an entirely new approach to care delivery that is personalized using molecular information on individuals and combined with their most current healthcare status.

2. Greater benefits for stakeholders in each genomic medicine cornerstone

As the genomic health record evolves, each of the three cornerstones of genomic medicine stands to benefit substantially:

- **Sequencing**—Processing raw data into usable form.
- **Translational medicine**—Finding relationship between genome and phenotypes and discovering/developing treatments.
- **Personalized healthcare**—Applying useful clinical insights to patients.

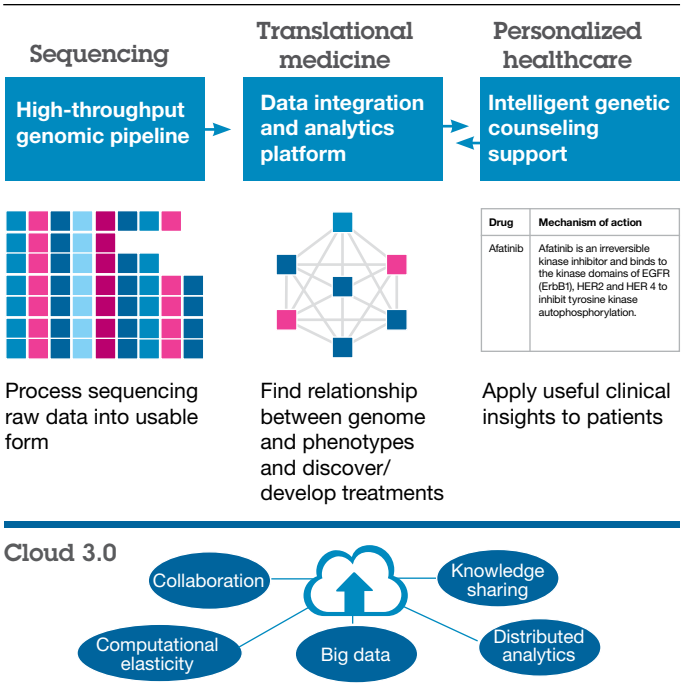
All of these cornerstones will work in a cloud-based model that adds new dimensions of data sharing, collaboration and efficiency (see Figure 4). A secure cloud offers benefits that include rapid access of data and greater collaboration of large data sets that can be exchanged and simultaneously accessed and worked on.

Cornerstone 1: Sequencing

A sequencing project typically has four phases: experimental design and sample collection; sequencing; data processing and downstream analysis.

High-performance computing and scalable storage solutions are required to process the vast quantities of data from whole-genome sequencing. Among the most significant challenges are:

- To process the data fast enough to keep pace with the output speed of sequencer(s)
- To manage and store the massive amount of data ("really big data") generated by sequencers and analysis.



Source: Adapted from IBM Global Technology Outlook, 2014.

Figure 4: **Three cornerstones of genomic medicine can operate in the cloud: sequencing, translational medicine and personalized healthcare.**

Sequencing has become both more economical and much faster than in recent years. However, clinical reliability of such large-scale sequencing efforts remains in question, especially when whole-genome sequencing information could be used for translational medicine and clinical decision making. New advances in whole-genome sequencing technology have led to greater throughput of sequencing data at a decreasing cost. And when faster, cheaper machines operate in parallel, thousands of whole genome sequences could be generated per year.

In the data processing phase, raw sequencing reads are mapped to a known reference genome or assembled *de novo*. The sequence alignment information is stored and compressed. The differences (variants) between the reference and sequence genome are identified. The applications and algorithms are typically CPU-intensive and input/output-intensive. The *de novo* assembly algorithms have a very large memory footprint, while the reference-mapping applications and variant-calling algorithms have much smaller memory requirements.

Traditional generation of raw sequence will not only become faster, cheaper and more accurate, but also infused with analytics capability that renders more than just raw data results.[20] Many vendors now offer sequence data analysis, in addition to software-as-a-service analysis platforms for scientists in research and clinical settings. Such services aggregate and analyze multiple sources of information including gene expression, proteomic and sequence data.

It is crucial for stakeholders in the sequencing data collection cornerstone to have a high-performance computing IT solution that can scale up appropriately and accelerate the analysis of the sequencing data. An HPC solution could help manage the data deluge resulting from the much higher data generation rate of future sequencers.

Raw sequence data infused with analytics capability can increase that data's value to downstream users, including researchers and the pharmaceutical industry. Sequencing capability is increasingly going to include analytics capability offered in a cloud-based environment with the ability to aggregate and analyze multiple sources of genomic, proteomic data.

Cornerstone 2: Translational medicine

Translational medicine can broadly encompass the need for new methodologies and solutions to understand disease mechanisms. It includes:

- The discovery of biomarkers used for discovery of drug targets, diagnostics, prognosis and risk factors,
- Genome-specific data models, and
- The integration of genomic and non-genomic analytics capabilities to discover clinically useful insights.

Translational medicine thus has many diverse data inputs, such as gene sequence data from the first cornerstone, and phenotypic and environmental information.

Although the capability to rapidly and cheaply sequence the human genome is important, translational medicine analyzes genome data with corresponding phenotype information—such as those reported in medical records or environmental information—to provide the real value to researchers who can then obtain insights into human biology and health.

Yet, to find such genotype-phenotype relationships, there are two requirements: large populations with accurate medical records; and large computational capacity with sophisticated algorithms (see sidebar, "Rizzoli Orthopedic Institute").

Phenotypes are subjective in many cases, and vary among physicians and hospitals. In the translational medicine cornerstone, data from many sources—such as patients' medical records, public databases, literature and in-house experimental data—is heterogeneous, multidimensional and mostly unstructured. The integration of such large and diverse healthcare data with genomic data is highly challenging for researchers who now also face unprecedented data volumes.

Most translational medicine platforms in the last few years have focused on data processing and data integration. This step is essential to derive insights from the information captured in genomic records starting from Single Nucleotide Polymorphism data and next-generation sequencing of raw information to fully annotated whole genomes.

It is also critical to link genomic data to additional sources, such as phenotypic and environmental data in the case of related patient studies, and to available textual information in patents and scientific literature. Linking externally available open databases of genomics, transcriptomics, proteomics, metabolomics and microbiomics requires another step to bring all the information together to allow advanced data mining and deep analysis.

Rizzoli Orthopedic Institute: Analyzing multiple disease factors to gain insights that allow individualized care for patients with rare genetic conditions[21]

Rizzoli Orthopedic Institute aspired to create more evidence-based treatment protocols for rare bone diseases to guide more effective treatment and reduce unnecessary testing and procedures. Integrating clinical, genetic and medical imaging data from hundreds of patients with a rare inherited bone disorder is arming doctors at Rizzoli Orthopedic Institute with new insights into better predicting how disease symptoms will likely progress in an individual person. Advanced algorithms identify patterns within the data, guiding research programs to further understand the underlying disease. With a greater understanding of the disease, doctors are able to match treatment strategies to the individual, enhancing patients' care and improving their quality of life. The personalized treatment approach also means that patients with less severe symptoms are avoiding the pain of unnecessary surgery and the inconvenience of repeatedly going to the center for imaging procedures.

To bring together such diverse data, researchers around the world often build their own integrated data set from internal data and some external available data. Data subsets are thus replicated and integrated repeatedly in various ways, each time to get the desired answer to a different research question of a specific study.

The result is duplication and inefficiency in advanced data mining and deep analysis. Once a researcher devises a subset of integrated data, it is often analyzed separately in a "quick and dirty" mode. This methodology has limitations including inefficiencies in the use of big data, lower replicability of studies, the need for more detailed technical requirements and a lack of widespread use.

Future translational medicine platforms will have to overcome these limitations—and include not only the integrated data layers, but additional layers—to automate and speed up preprocessing. Such platforms will also need a robust methodology and pipeline for performing advanced analytics and enabling repeatable research results. Data analytics is thus at the core of translational medicine (see sidebar, "A leading medical and research center").

It is expected that any big data platform for translational research will include the required layers for effective analysis, including cognitive computing capability. Genomic data and analytics can also be combined with scientific literature through the use of cognitive computing, the power of which was demonstrated by Watson in the Jeopardy! Grand Challenge in 2011.[22] It can help scientists and clinical research doctors obtain insights from vast amounts of data and publications. In addition, machine learning methods can be employed to elucidate insights automatically. Such insights can help construct disease progression models for a variety of diseases to assess risk factors and treatment planning.

A leading medical and research center: Aiming to grow translational research capabilities

A cloud-based HPC solution designed and developed with a business partner will serve as the foundation for a growing translational research capability at a leading international medical and research center. The solution will provide researchers and administrators with the input/output performance, scalability and flexibility they will need to adjust to rapidly changing tools and capability demands inherent in computationally intensive life science research.

The solution is designed to enable two key research capabilities: 1) the efficient creation and execution of complex workflows for genomic analysis, and 2) the secure integration of genomic data with de-identified clinical data in a common, easily accessible repository for downstream analytics. Through this effort, the center aims to become an international leader in translational research — a center of excellence where top biomedical scientists can work collaboratively to develop novel clinical therapies that improve both local and global health outcomes.

Translational medicine is a key innovation lever for the discovery of targeted treatments that are specific to key biochemical pathways regulated by genes which—when awry—can create or predispose a disease state.

Such insights can be achieved with high efficacy and low side effects, especially toxicity. Treatment insights can then be shared and validated among translational medicine communities to improve healthcare and advance personalized healthcare.

Cornerstone 3: Personalized healthcare

Personalized healthcare focuses on an individual (that is, a patient) and the ability of a clinician to provide unique treatment based on the individual's personal characteristics. Genomic medicine will add even more complexity for geneticists, genetic counselors and specialists, such as medical oncologists who must link a person's genomic sequence to appropriate evidence-based personalized treatments. Three areas in personalized healthcare are expected to stand out: even more advances in cognitive computing; greater use of disease progression models; and a growing impact of genomic medicine on oncology.

Advances in cognitive computing—such as machine learning and natural language processing—are expected to increase as clinicians seek machine-based "second opinions" which are made with high confidence and rigor, and at a pace far greater than is possible manually. Information sources can include dbSNP (SNP database), HGMD (Human Genome Mutation Database), COSMIC (Catalogue Of Somatic Mutations In Cancer), UMD (UNC Microarray Database), LOVD (Leiden Open Variation Database), MutDB (database that associates protein structural information with mutations and polymorphisms in gene sequences), HVP (Human Variome Project) and programs like SIFT (Sorting Intolerant From Tolerant) and Polyphen2.

Such information will comprise the genomic health record at the point of care. Clinical decision support tools will allow physicians and genetic counselors to understand and navigate the massive amount of genomic and drug information in literature and databases to provide personalized diagnosis, prevention advice and the most effective treatment plans.

Disease progression models that include historical paths of past patients, with various information including genomics and other "-omics information" will guide physicians to evaluate a patient. Patient similarity will be performed on the models to define similarities between patients using multi-dimensional and complex data.

Oncology is the first area in medicine where the impact of genomic medicine has already been documented from predisposition testing to diagnosis, to precision treatment with a targeted drug.[23] And this appears to be only the tip of the iceberg.[24] With a reduction in sequencing costs, the availability of multiple oncology diagnostic panels, the rapid proliferation of targeted treatments and a new paradigm of cognitive computing, the oncology field is one that should quickly experience the benefits from personalized/precision care and a new genomic health record (see sidebar, "New York Genome Center").

With the help of cognitive computing and other computational techniques, the landscape of risk factor profiling will change dramatically as we process more human genomes and translational medicine advances. Certain considerations will be important in this era of genomic medicine, building upon strong principles which have been tested with more traditional genetic disorders (see sidebar, "Privacy, legal, social and ethical aspects of genomic medicine").

New York Genome Center: Accelerating personalized, life saving treatments for cancer patients[25]

The New York Genome Center (NYGC) and IBM are partnering in a first-of-a-kind program to accelerate personalized, life-saving treatment for cancer patients. IBM and NYGC will test a unique Watson prototype designed specifically for genomic research as a tool to help oncologists deliver more personalized care to cancer patients. NYGC and its medical partner institutions plan to initially evaluate Watson's ability to help oncologists develop more personalized care to patients with glioblastoma, an aggressive and malignant brain cancer that kills more than 13,000 people in the U.S. each year.

Despite groundbreaking discoveries into the genetic drivers of cancers like glioblastoma, few patients benefit from personalized treatment that is tailored to their individual cancer mutations. Clinicians lack the tools and time required to bring DNA-based treatment options to their patients and to do so, they must correlate data from genome sequencing to reams of medical journals, new studies and clinical records — at a time when medical information is doubling every five years.

This joint NYGC-Watson initiative aims to speed up this complex process, identifying patterns in genome sequencing and medical data to unlock insights that will help clinicians bring the promise of genomic medicine to their patients. The combination of NYGC's genomic and clinical expertise coupled with the power of IBM's Watson system will enable further development and refinement of the Watson tool with the shared goal of helping medical professionals develop personalized cancer care.

The new cloud-based Watson system will be designed to analyze genetic data along with comprehensive biomedical literature and drug databases. Watson can continually "learn" as it encounters new patient scenarios, and as more information becomes available through new medical research, journal articles and clinical studies. Given the depth and speed of Watson's ability to review massive databases, the goal of the collaboration is to increase the number of patients who have access to care options tailored to their disease's DNA.

Watson will complement rapid genome sequencing and is expected to dramatically reduce the time it takes to correlate an individual's genetic mutations with reams of medical literature, study findings and therapeutic indications that may be relevant. The intention is to provide comprehensive information to enable clinicians to consider a variety of treatment options that the clinician can tailor to a patient's genetic mutations. It will also help NYGC scientists understand the data detailing gene sequence variations between normal and cancerous biopsies of brain tumors.

Privacy, legal, social and ethical aspects of genomic medicine

The importance of keeping genetic information private is spurring legal actions to protect patients and citizens. The most noteworthy legal protection for U.S. consumers came from the GINA (Genetic Information Nondiscrimination Act) in May 2008. The GINA legislation protects Americans from discrimination based on the results derived from a genetic test. It also forbids insurance companies from reducing coverage or increasing price, and prevents companies from making employment decisions based on an individual's genetic information. GINA also forbids employers and insurance companies from requesting or demanding a genetic test.

Genomic information elicits two important privacy considerations: informed consent and confidentiality.

Informed consent is vital and requires policy measures to be in place, both at the research and healthcare provider levels. From a research perspective, citizens or patients from whom DNA is obtained must be informed about what it is likely to be used for. In some cases, a genetic test that is ordered by a provider on a patient may also include the option for additional research on the DNA sample.

In such cases, separate consent must be obtained for both: the genetic test being ordered; and the research that would be conducted on the patient's DNA sample. For example, the topic of incidental findings during the course of genome-based investigations has raised numerous ethical discussions about genomic testing and its impact on individuals and their families.

Confidentiality is a serious consideration in genomics as it provides confidence to a given citizen or patient that his or her genomic information will be treated as confidential and will only be seen by those in the circle who are providing care. The security of the information is critical, as genomic information needs to be protected and safeguarded by appropriate industry standard security software and encryption.

In the case of more standard genetic tests, these may be part of a patient's electronic health/medical record and thus need to be treated with appropriate security measures. Confidentiality also requires appropriate measures to "anonymize" data in a research setting so that a given sample is not readily associated with a specific individual. In addition, both privacy and security measures must be in place to protect data from malicious access.

Social and ethical aspects of genomic medicine will also require ongoing attention and the future implications have yet to be defined. For example, managing a diagnosis of Huntington disease or needing to provide a nurturing environment for a child with Down syndrome could clearly both have social and ethical implications that extend beyond the families that are directly affected.

3. Opportunities for radical industry transformation

The transformational opportunity for industries rests in leveraging cognitive computing throughout the entire value chain; from pre-clinical discovery to understanding and engaging consumers, and improving health outcomes.

The pharmaceutical industry, in particular, faces numerous challenges with decreased R&D productivity, high costs of bringing drugs to market, more stringent regulatory measures and payors that are not willing to pay for "me-too drugs" or for drugs that offer incremental benefits.[26] By embracing *genomic* approaches to drug discovery, new and radical transformation opportunities will become available.

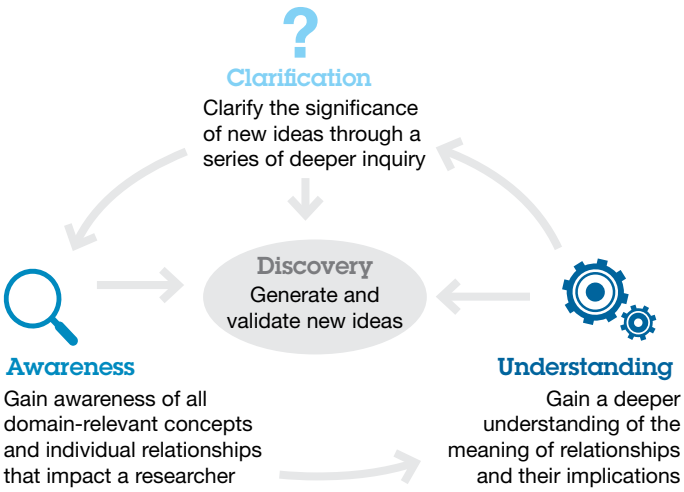
The finding of new biomarkers, for example, can be used for a wide range of molecular diagnostics and drug target identification. Such biomarkers can also provide more specificity for drug development. Innovative clinical trials can lead to accelerated approval of targeted drugs that meet criteria of efficacy, utilization and cost-effectiveness. With recent advances in genomic medicine, it is not uncommon for patients with specific cancers to undergo genetic testing.

Improved discovery capabilities can, however, lead to the detection of rare mutations that only impact a few patients. Rarity makes it more difficult to study a drug which has a targeted impact on a specific mutation. In addition, targeting a mutation with one drug may not be sufficient. Thus the older method of testing one drug at a time per patient paired with a genetic test is inefficient and complicates clinical trials that typically have stringent study criteria and expected end points. In the case of a stage IV cancer patient, time is of the essence and so the conventional clinical trial process does not always work.

To address these challenges, the "Master Protocol" is being used specifically for squamous cell lung cancer at first.[27] The protocol aims to solve a fundamental problem of cancer research using genomic medicine by using a standard trial design for studying many drugs simultaneously. Instead of one genetic test for each drug, the trial will use high-throughput DNA sequencing to scan tumors for 16 different genetic changes (mutations) that could get treated by more than a dozen drugs.

Based on trial results, patients will be put into separate studies for each new drug. Those for whom DNA sequencing doesn't turn up a potential medicine will get a new type of immune system drug. The Master Protocol began in March 2014. It could offer a radically different approach for cancer diagnosis, treatment and follow-up. It could also provide new opportunities to partner with healthcare stakeholders such as life sciences companies, academic medical research centers and payors.

Advances in cognitive computing will continue to evolve, allowing pharmaceutical companies to use tools and systems that can ingest and continuously learn. New awareness, understanding and clarification will set the stage for the discovery and generation of new ideas (see Figure 5).



Source: IBM Global Business Services 2014.

Figure 5: **Discovery and new idea creation in the healthcare and life sciences industries.**

The combination of awareness, understanding and clarification enables discovery by integrating information related to problems that previously required large amounts of manual work and whose solution attempts were potentially prone to missing some of the evidence. The implications of improving discovery are highly relevant for the life sciences industry as it seeks ways to reduce the time and cost of developing drugs that are efficacious and highly targeted.

The healthcare and life sciences industries now have an opportunity to radically transform. Genomic medicine can be a foundation in the development of innovative clinical trial protocols and the search for a new cadre of precision treatments that focus on preventing disease or alleviating symptoms.

Ask yourself: Considerations for industry executives

As a first step, answer the following questions to help your healthcare or life sciences organization develop its genomic medicine strategies:

- ***How have you incorporated genomic medicine into your enterprise vision and strategy at the clinical, research and IT levels?*** Genomic data will continue to proliferate into mainstream healthcare. A data deluge is expected, as well as serious questions about using genomic data for clinical decision support. A new genomic health record will become a reality along with increasing EMR maturity.
- ***Which types of capabilities, skills and personnel will your organization need so you can use genomic medicine?*** Which do you already possess and which do you currently lack? How will you develop a plan to fill any identified gaps? Cognitive computing and analytics leveraging cloud computing will play a crucial role in genomic medicine, as advances in these areas will lead to more accurate genome-based clinical decision support via a system that learns continuously. Understand the requirements for your organization to leverage such advances that enable patient care to be of higher quality, more accurate and safer. Competencies to comply with evolving regulatory standards, privacy and security requirements (for example, HIPAA) will be crucial in an increasingly personalized world where data breaches can be very damaging.
- ***What is your approach to deciding who you should partner with to build complementary capabilities and skills in genomic medicine?*** To more fully exploit the revolutionary technologies and processes for studying and treating diseases, you will likely need to partner with other organizations in new ways. For example, as seen in the Master Protocol, academic research centers are partnering differently with pharmaceutical companies and primary care facilities.

Conclusion

A new genomic health record will become a reality as genomic information gets combined with relevant data extracted from the traditional EMR. Rapid, precision oncology decision support is expected to expand on a larger scale by performing sophisticated genome/proteome/RNA analysis coupled with cognitive computing. The capabilities and technologies associated with cognitive computing are critical to the ongoing genomic medicine evolution by enabling much more sophisticated decision support, innovative clinical trials, new targeted therapies, and a focus on health outcomes to prevent and alleviate disease symptoms.

To benefit from the far-reaching industry transformation that has begun, forward-thinking executives can: verify that genomic medicine is part of their enterprise vision and strategy; assess and plan to fill existing and future skill gaps; and look closely at how and when partnering will help their organizations succeed in meeting stakeholder needs.

To learn more about this IBM Institute for Business Value study, please contact us at iibv@us.ibm.com. For a full catalog of our research, visit: **ibm.com/**iibv

Access IBM Institute for Business Value executive reports on your tablet by downloading the free "IBM IBV" app for iPad or Android from your app store.

Related publications

Fraser, Heather, Chaturika Jayadewa, Jay Goodwyn, Peter Mooiweer, Dan Gordon and John Piccone. "Analytics across the ecosystem: A prescription for optimizing healthcare outcomes." IBM Institute for Business Value. September 2013. http://www-935.ibm.com/services/us/gbs/thoughtleadership/healthcare-ecosystem/

Mason, Barry, Gary Bacher, Harry Reynolds and Heather Fraser. "Collaborating beyond traditional boundaries: What convergence means for our health care systems." IBM Institute for Business Value. June 2013. http://www-935.ibm.com/services/us/gbs/thoughtleadership/healthcare-convergence/

Ferrucci, David, Anthony Levas, Sugato Bagchi, David Gondek and Erik T. Mueller. "Watson: Beyond Jeopardy!" Elsevier: Artificial Intelligence. ScienceDirect. May 2012. http://www.sciencedirect.com/science/article/pii/S0004370212000872

IBM Institute for Business Value

IBM Global Business Services, through the IBM Institute for Business Value, develops fact-based strategic insights for senior executives around critical public and private sector issues. This executive report is based on an in-depth study by the Institute's research team. It is part of an ongoing commitment by IBM Global Business Services to provide analysis and viewpoints that help companies realize business value. You may contact the authors or send an e-mail to iibv@us.ibm.com for more information.

About the authors

Aditya Pai is a healthcare and life sciences subject matter expert with the IBM Watson Group. He has 17 years of experience as a clinician and senior management consultant. His responsibilities include bringing deep healthcare, genomics and life sciences insights to clients. Trained in molecular genetics, genetic counseling and business administration, Aditya has a co-authored reports such as "Healthcare 2015: Win-Win or Lose-Lose, Healthcare in India—Caring for over a billion" and "Genomic Medicine—The tip of the iceberg." He also serves as a faculty instructor in the Health Industry Management Program at the Schulich School of Business, York University, Toronto teaching a course he co-designed on Commercialization in Biotechnology and Pharmaceuticals. Aditya can be reached at *aditya.pai@ca.ibm.com*.

Takahiko Koyama PhD, is the lead scientist in genomic medicine in IBM Research. Currently, he collaborates with New York Genome Center on the glioblastoma project. Originally, he trained as a physicist but changed his career through biomedical research at the University of Pennsylvania School of Medicine and the University of Oxford, Sir William Dunn School in Pathology. As a bioinformatician and a computational chemist at Takeda Pharmaceutical Company, he worked on ten research projects, primarily in oncology. Takahiko can be contacted at *koyama1@jp.ibm.com*.

Leonard Lee manages the worldwide IBM portfolio of software solutions for healthcare and life sciences. His responsibilities span strategy, mergers and acquisitions, development, sales execution and partnerships. He has worked on healthcare portals, predictive medical systems and drug discovery. In a "past life," he designed interactive toys, 3D visualization and other strange things at IBM Research. Leonard is a graduate of Carnegie Mellon University, where he researched pervasive computing systems, and virtual and augmented reality. He can be contacted at *thirteen@us.ibm.com*.

Contributors

This report was written with contributions made by a global team at IBM comprising genomics, healthcare and life sciences experts across Watson Group, Sales & Distribution, Software Group, Global Business Services and Research: Houtan Aghili, Ravishankar Balraj, Boaz Carmeli, Shahram Ebadollahi, Yaara Goldschmidt, Stephen J. Harvey, Richard Hennessy, Joe Jasinski, Prasad Kodali, Christine M Kretz, Janis Landry Lane, Frank Lee, Pietro Leo, Eleanor Lewin, Vanessa Michelini , Lauren O'Donnell, Arthur Panov, Laxmi Parida, Karen Parrish, Matthias Reumann, Michal Rosen-Zvi, Ajay K Royyuru, Amnon Shaabo, Charlie Schick, Raminderpal Singh, Tzy-hwa K Tzeng, Zeev Waks, Kelly Wyres and Jane Yu.

Acknowledgements

For their help in the development and production of this paper, thanks go to Heather Fraser and Eric Lesser both of the IBM Institute for Business Value; and Joni McDonald and Kristin Biron, both of ITSO Global Content Services, IBM Sales and Distribution.

Glossary

Adenocarcinoma—A malignant tumor formed from glandular structures in epithelial tissue.

ALK (anaplastic lymphoma kinase)—an enzyme that in humans is encoded by the ALK gene

Bioinformatics—The science of collecting and analyzing complex biological data such as genetic codes

Biomarker—A measurable characteristic that reflects the severity or presence of some disease state

CCD (Continuity of Care Document)—An electronic document exchange standard for sharing patient summary information

CDR (Clinical Data Repository)—A storage interface between the data repositories of the Department of Defense and the U.S. Veterans Administration

CDS (Clinical Decision Support)—A process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health and healthcare delivery

CDSS (Clinical decision support system)—An application that analyzes data to help healthcare providers make clinical decisions

Cognitive computing—Systems that are trained using artificial intelligence and machine learning algorithms to sense, predict, infer, and in some ways, think

CPOE (Computerized physician order entry)—A process of electronic entry of medical practitioner instructions for the treatment of patients under his or her care

de novo—Without a reference genome

DNA (Deoxyribo Nucleic Acid)—A molecule that encodes an organism's genetic blueprint

ED—Emergency department

EGFR (Epidermal growth factor receptor)—The cell-surface receptor for members of the epidermal growth factor family of extracellular protein ligands

EMR (Electronic Medical Record)—A systematic collection of electronic medical information about an individual patient or population

Environmental Information—Includes information about air, water, soil, land, flora and fauna, energy, noise, waste and emissions, as well as information about decisions, policies and activities that affect the environment

Gene expression—The process by which information from a gene is used in the synthesis of a functional gene product

Genotype—The genetic constitution of an individual organism

Genomics—The branch of genetics that studies organisms in terms of their genomes (their full DNA sequences)

Gene Sequence—The process of determining the precise order of *nucleotides* within a DNA molecule

GWAS (Genome Wide Association Studies)—An examination of many common genetic variants in different individuals to see if any variant is associated with a trait

Glioblastoma—A malignant tumor of the central nervous system, usually occurring in the cerebrum of adults.

HIE (Health information exchange)—The mobilization of healthcare information electronically across organizations within a region, community or hospital system

HIPAA (Health Insurance Portability and Accountability Act)—U.S. healthcare law

HPC (High Performance Computing)—Any computational activity requiring more than a single computer to execute a task

Immunohistochemistry — The application of immunologic techniques to the chemical analysis of cells and tissues.

Metabolomics — The scientific study of the set of metabolites (a substance formed in, or necessary for, metabolism) present within an organism, cell or tissue

Metastasis — The development of secondary malignant growths at a distance from a primary site of cancer.

Me-too drugs — Compounds that are structurally very similar to already-known drugs with minor pharmacological differences

Microbiomics — The study of all the genomes of the microbes present in an ecosystem

Nucleotides — Organic molecules that serve as the monomers, or subunits, of nucleic acids like DNA and RNA

Oncology — The study and treatment of tumors

OP — Out-Patient

Pathogenic — A disease brought about by organisms in humans

PET-CT techniques (Positron emission tomography–computed tomography) — medical imaging tools that when combined allow physicians to pinpoint the location of cancer within the body before making treatment recommendations

Pharmacogenomics — The branch of genetics concerned with determining the likely response of an individual to therapeutic drugs

Phenotype — Trait caused by genes

Polymorphism — Variants of the same gene

Proteomics — The branch of genetics that studies the full set of proteins encoded by a genome

Rational drug design — The inventive process of finding new medications based on the knowledge of a biological target

Reference Genome — A digital nucleic acid sequence database, assembled by scientists as a representative example of a species' set of genes

RNA (Ribonucleic Acid) — A single-stranded polymer of nucleotides that contain the sugar ribose

RNA-Seq (RNA-Sequencing) — A technology that uses the capabilities of next-generation sequencing to reveal a snapshot of RNA presence and quantity from a genome at a given moment in time

R-PACS (Radiology picture archiving and communication system) — A computer network for digitized radiologic images and reports

SNP (Single Nucleotide Polymorphisms) — A variation among individuals at a single position in a DNA sequence

Software-as-a-service — Delivery of software remotely to customers via the Internet, with applications hosted by one or more service provider

Symbiotic — Having an interdependent relationship

Translational Medicine — Discipline within biomedical and public health research that aims to improve the health of individuals and the community by "translating" findings into diagnostic tools, medicines, procedures, policies and education

Transcriptomics — The examination of whole transcriptome (the sum total of all messenger RNA molecules expressed from an organism's genes) changes across a variety of biological conditions

TKIs (Tyrokine Kinase inhibitors) — Substances that block the action of enzymes called tyrosine kinases

Whole-genome sequencing — Laboratory process that determines the complete DNA sequence of an organism's genome at a single time

References

1 U.S. National Library of Medicine, National Institutes of Health. "What is a genome?" Genetics Home Reference: Your Guide to Understanding Genetic Conditions. April 21, 2014. http://ghr.nlm.nih.gov/handbook/hgp/genome (accessed April 28, 2014).

2 National Institutes of Health. "An Overview of the Human Genome Project: What was the Human Genome Project?" National Human Genome Research Institute. November 8, 2012. http://www.genome.gov/12011238 (accessed April 28, 2014).

3 McCarthy, J.J., H. L. McLeod and G. S. Ginsburg. "Genomic Medicine: A Decade of Successes, Challenges, and Opportunities." *Sci. Transl. Med.* 5, 189sr4 (2013).

4 Burstein, H.J. "The distinctive nature of HER2-positive breast cancers." *New England Journal of Medicine 353* (16): 1652–4. October 2005. http://www.nejm.org/doi/full/10.1056/NEJMp058197 (accessed April 23, 2014).

5 Bollag, G., J. Tsai, J. Zhang, C. Zhang, P. Ibrahim, K. Nolop and P. Hirth. "Vemurafenib: the first drug approved for BRAF-mutant cancer." *Nature Reviews Drug Discovery* 11 (11): 873–886. November 2012. http://www.nature.com/nrd/journal/v11/n11/full/nrd3847.html?WT.ec_id=NRD-201211 (accessed April 23, 2014).

6 Takeuchi, F. et al. "A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose." *PLoS Genetics* 5(3):e1000433. doi: 10.1371/journal.pgen.1000433. March 2009. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2652833/ (accessed April 23, 2014).

7 Life Technologies "Ion AmpliSeq™ Comprehensive Cancer Panel." http://www.lifetechnologies.com/order/catalog/product/4477685 (accessed April 23, 2014).

8 Nagahiro, S. "Targeted therapies: Tyrosine-kinase inhibitors—new standard for NSCLC therapy." Nature Reviews Clinical Oncology 7, 618-619. November 2010. http://www.nature.com/nrclinonc/journal/v7/n11/full/nrclinonc.2010.168.html (accessed April 23, 2014).

9 "The future of the personal genome." The value of openness blog by patientslikeme. May 21, 2012. http://blog. patientslikeme.com/2012/05/21/the-future-of-thepersonal-genome/ (accessed May 19, 2014).

10 Polli, F. "Why 23andme Deserves A Second Chance" *Forbes*. January 2014. http://www.forbes.com/sites/fridapolli/2014/01/14/why-23andme-deserves-a-second-chance/ (accessed April 23, 2014).

11 "Genetic Information Nondiscrimination Act (GINA) of 2008." National Human Genome Institute. http://www.genome.gov/24519851 (accessed April 23, 2014).

12 Perkel, J.M."Gene Patents Decision: Everybody Wins." *The Scientist*. June 2013. http://www.the-scientist.com/?articles.view/articleNo/36076/title/Gene-Patents-Decision--Everybody-Wins/ (accessed April 23, 2014).

13 Fridlyand, J. et al "Considerations for the successful co-development of targeted cancer therapies and companion diagnostics." *Nature Reviews Drug Discovery*. September 2013. 12, 743–755 2013. http://www.nature.com/nrd/journal/v12/n10/abs/nrd4101.html?WT.ec_id=NRD-201310 (accessed April 23, 2014).

14 "DNA Sequencing Costs." National Human Genome Institute. www.genome.gov/sequencingcosts (accessed April 23, 2014).

15 For example, see Genetests. www.genetests.org (accessed May 19, 2014).

16 "EMR Adoption Model." HIMSS Analytics. http://www.himssanalytics.org/hc_providers/emr_adoption.asp (accessed April 23, 2014).

17 Ibid.

18 Ibid.

19 "Cognitive computing." IBM Research. http://www.research.ibm.com/cognitive-computing/index.shtml (accessed April 23, 2014).

20 Rizzo, J., Buck, M."Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing." *Cancer Prev Res* 2012;5:887-900. July 2012. http://cancerpreventionresearch.aacrjournals.org/content/5/7/887.long (accessed April 23, 2014).

21 "Rizzoli Orthopedic Institute: Applying advanced analytics to better treat hereditary diseases." IBM Smarter Planet Leadership Series. 2011. http://www.ibm.com/smarterplanet/us/en/leadership/rizzoli/ (accessed April 23, 2014).

22 "Computer Wins on 'Jeopardy!' Trivial, It's Not." *The New York Times*. February 2011. http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?_r=0 (accessed April 23, 2014).

23 Burstein, H.J. "The distinctive nature of HER2-positive breast cancers." *New England Journal of Medicine*. 353 (16): 1652–4. October 2005. http://www.nejm.org/doi/full/10.1056/NEJMp058197 (accessed April 23, 2014).

24 Pai, Aditya. "Genomic Medicine—The Tip of the Iceberg." *Electronic Healthcare.* Volume 8; No 1 2009. www-03.ibm.com/industries/ca/en/healthcare/.../eh_vol8_no1_pai.pdf (accessed April 23, 2014).

25 "The New York Genome Center And IBM Watson Group Announce Collaboration To Advance Genomic Medicine."Reuters. March 2014. http://www.reuters.com/article/2014/03/19/ibm-ny-genome-center-idUSnPnbNdN4x+94+PRN20140319 (accessed April 23, 2014).

26 Garattini, S. "Are me-too drugs justified?"*Journal of Nephrology*. 1997 Nov-Dec;10(6):283-94 (accessed May 19,2014). http://www.ncbi.nlm.nih.gov/pubmed/9442441

27 Friends of Cancer Research. "Friends of Cancer Research Moves Revolutionary New Clinical Trial Design Forward: Update—Lung Cancer Master Protocol Activation Announcement." November 11, 2013. http://www.focr.org/master-protocol (accessed April 29, 2014).

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
May 2014

Please Recycle

GBE03605-USEN-01