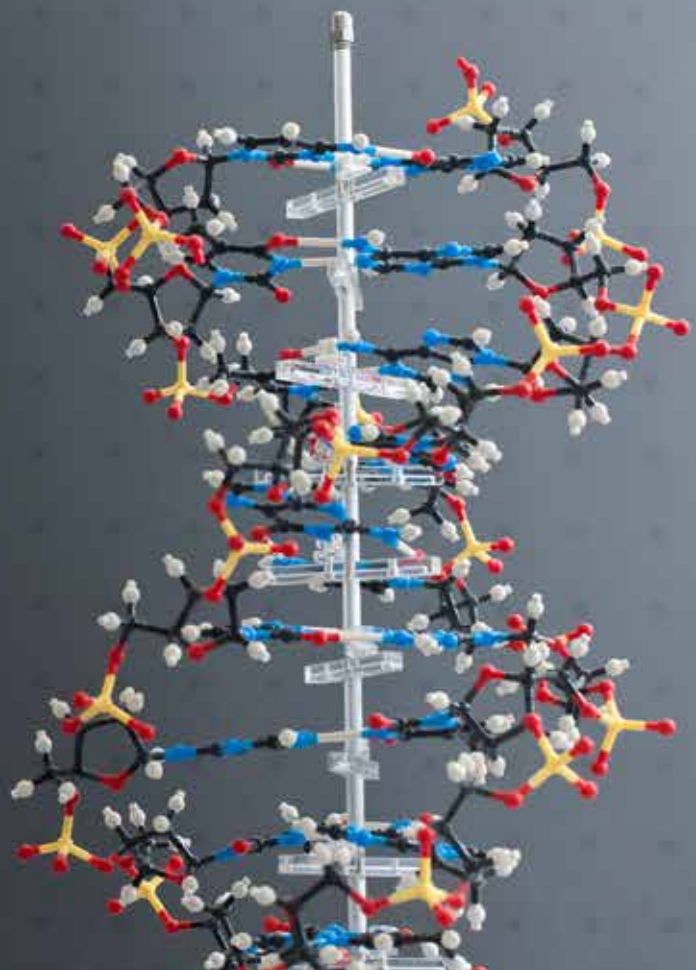
KPMG

Driving value from genomics in Life Sciences

Rising to the data challenge



KPMG International

kpmg.com

Genomics can bring significant benefits to healthcare systems, by accelerating clinical research and drug development, personalizing treatment regimens, improving patient outcomes, and reducing the cost of care.

Genomic data, the most personal of all human data, is part of the overall explosion in health data, enabled by the exponential growth in computing power and the commercialization of wearable technologies such as smartphones, fitness trackers and heart rate monitors. Managing this data has led to a change in the way such information is gathered, stored, analyzed and used; with privacy, reliability, and security posing prickly challenges.

Health and genomic data can now be shared with healthcare professionals and disparate partners. In addition to the many benefits, the ability to gather and share personal, sensitive information, has a paradoxical dark side. This can be found in the form of challenging privacy requirements and Cyber Security risks, which often delay research, lowering the financial return on investment and damaging the brand reputation.

The cornerstone for genomic data leadership is a robust data strategy, to support drug research and development, meet increasingly tough regulations and preserve ethical standards. This, along with the use of the most appropriate technologies — to process large data volumes and generate insights — should enable Life Sciences companies to become more fleet of foot in discovering and taking forward new drugs.

Five pillars of a genomic data strategy:

1. **Standardized sequencing and analysis** to assure data reliability

2. **Clear understanding of regulatory privacy requirements** and operationalization of privacy practices

3. **Security** to avoid breaches and theft of intellectual property

4. **Storage and data transfer** to ensure fast, safe access

5. **Harnessing analytical technology** to deliver useful insights for R&D and clinical use

Source: Driving value from genomics in Life Sciences. 2018.

Authors

Prof Hilary Thomas
Partner
KPMG in the UK

Hilary is KPMG's Chief Medical Adviser and a member of the global Centre of Excellence in Healthcare and Life Sciences. Prior to KPMG, she spent 23 years in the NHS — including Professor of Oncology at the University of Surrey and Medical Director of the Royal Surrey County Hospital. Hilary works at the interface of Healthcare and Life Sciences for public and private sector clients — redesigning care models and pathways and helping organizations to navigate the changing way that the public interact and the implications for organizational strategy and business models. She is a past winner of the First Women in Business Services and a 2015 luminary of the Healthcare Businesswomen's Association.

Caroline Rivett
Director, Life Sciences
KPMG in the UK

Caroline has over 20 years experience in technology. Caroline co-founded and sold a company which generated health scores for insurers based on people's health records. Caroline built a data analysis organization for a global professional services company as well as leading a change program for Aviva focused on costing technology and pricing technology services for shared services organizations. She served as a Chair of the Audit Committee and Board member in healthcare for eight years and is now responsible for leading the digital focus for the UK's Life Sciences team.

Dr Rebecca Pope
Lead Data Scientist
KPMG in the UK

Dr Rebecca Pope is a Clinical Neuroscientist and a Lead Data Scientist who specializes in the use of artificial intelligence within healthcare and preventative medicine. Rebecca works within KPMG's Healthcare & Life Sciences Practice. Rebecca has published a number of academic research papers and her post-doctoral research formed the basis of the 90-minute Channel 4 documentary 'Drugs Live: Cannabis on Trial'. She has written several scientific commentaries in the Guardian, and is the recipient of National awards for her research.

More recently, she delivered a TEDx talk entitled: How can AI help our NHS and should we be concerned? ; led the winning team in KPMG's Global Healthcare Hackathon and co-hosted a Microsoft Health Webinar Artificial Intelligence for Healthcare

Chantal Rademaker de Ridder

Partner Technology/Cyber and Head of Privacy & Data Protection
KPMG in the Netherlands

Chantal is a partner at KPMG Advisory in the Netherlands and one of the co-leaders of the Technology/Cyber Security group. She also heads the Privacy & Data Protection practice.

By combining her technical expertise with her business, marketing and legal knowledge, Chantal is uniquely positioned to support companies with digital transformation and to allow them to elevate cyber risks into business opportunities.

Her practice focusses on emerging technologies such as the Internet of Things, Artificial Intelligence, Machine Learning, Robotics, and Blockchain. She advises boards of directors about digital risk matters like cyber security, privacy & data protection, privacy enhancing technologies, data analytics, digital marketing, e-commerce, profiling & tracking, behavioral advertising, social engineering, digital identity, access management, digital risk monitoring & controls, GDPR, ePrivacy and regulatory compliance.

Martina Algeri

Privacy Manager
KPMG in the UK

Martina is a Privacy Manager in KPMG, specializing in providing support to global Pharmaceutical and Bio-technology clients and helping them address their complex, multi-jurisdictional Privacy challenges.

Martina has extensive experience working with global pharmaceutical companies on large Privacy Transformation Programs. For instance, Martina recently conducted a Privacy Maturity Assessment for a global pharmaceutical company. Upon conclusion of the analysis, she defined the organization's Privacy Risk Appetite and developed a detailed Privacy roadmap to help the client mitigate areas of Privacy risk exposure, whilst achieving its target Privacy Maturity posture. Martina also assisted the client in developing a business case and in gaining the required budget from the Board for implementing the desired Privacy improvement program.

Contents

The increasingly pervasive role of genomics in Life Sciences

Genomics has enormous potential to re-shape drug discovery and development — moving the Life Sciences industry from blockbusters to niche busters by leading the drive to personalized medicine.

The huge cost of taking a drug to market — estimated to be approximately 2.6 billion US dollars (US$)[1] — combined with increasingly tight healthcare budgets, has put intense pressure on research and development (R&D) teams to come up with new products faster, at a lower cost.

When you add shortening patent periods, increased competition and greater use of generics and biosimilars into the mix, then all the conditions are in place for a perfect storm that threatens ROI and profitability.[2,3]

Payers, keen to make more efficient use of their budgets, have been driving a move towards value-based pricing: a form of 'pay-for-performance' where reimbursements are based upon treatment outcomes. Progress has been modest (only 25 or so branded drugs are in some form of value-based contracts in the US, for example[4]), in part due to lack of collected data to track and measure their effectiveness. Once these obstacles are overcome, expect to see a swift rise in value-based payment, based not just upon clinical outcomes, but also patient-related outcome measures (PROMs), as well as value delivered to payers and wider society.

In oncology, for example, outcomes-based schemes are evident across a large number of major players, enabling these companies to win market share (for more on this topic, see the KPMG papers *The future of oncology: A focused approach to winning in 2030*[5] *and Value-based pricing in pharmaceuticals: Hype or Hope?)*[6]

Reimagining drug development and therapy

Genomics offers several ways to address some of Life Sciences' biggest challenges:

Clinical trials
Using patients' genetic profiles, R&D teams should be able to narrow down patient groups into subsets with a higher chance of responding to a specific treatment. This should speed up trials and eliminate ineffective therapies at an earlier stage, getting products into the market sooner, at lower cost. With the age of the blockbuster coming to an end, genomics can help life sciences companies evolve into more nimble players, producing a faster pipeline of drugs aimed at smaller target populations.

Treatment
Improved targeting of therapy should increase the chance of successful outcomes, ushering in an era of personalized medicine, often involving

[1] *Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs*, Journal of Health Economics 47:20-33,2016. https://csdd.tufts.edu/recent-publications/

[2] *The future of oncology: A focused approach to winning in 2030*, KPMG International, 2017. https://home.kpmg.com/xx/en/home/insights/2017/07/the-future-of-oncology.html

[3] *Global Medicines Use in 2020; Outlook and Implications*, IMS Institute for Healthcare Informatics, November 2015. https://s3.amazonaws.com/assets.fiercemarkets.net/public/005-LifeSciences/imsglobalreport.pdf

[4] *Pricing for survival,* KPMG, 2018. https://advisory.kpmg.us/kpmg-strategy/thinking/healthcare-and-lifesciences/pricing-for-survival.html

[5] *The future of oncology: A focused approach to winning in 2030*, KPMG International, 2017. https://home.kpmg.com/xx/en/home/insights/2017/07/the-future-of-oncology.html

[6] *Value-based pricing in pharmaceuticals: Hype or Hope?*, KPMG International, 2016. https://home.kpmg.com/xx/en/home/insights/2017/07/the-future-of-oncology.html

combinations of drugs. The University of Florida Health, US, is just one example of an organization that has created a genotype testing process to help doctors tailor treatment based on patients' genetic information.[7]

Oncology is one area ripe for such an approach, for example, identifying patients with different breast cancer mutations, each of which may be responsive to specific treatments.

Gene editing may also catalyze prevention and cures, with much hope for revolutionizing treatment of conditions like neurological disorders and cancer. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is an RNA-guided genome editing tool that can delete harmful genes and insert 'helpful' genes.

From the laboratory to the laptop

Data is at the heart of genomics, with researchers using machine learning, a subfield of artificial intelligence (AI), to help oncologists choose the most effective, individualized cancer treatments, by swiftly sorting through research data. New and powerful algorithms may identify which genes are likely to be mutated and predict the most appropriate treatments.

Genetic sequencing company Adaptive Biotechnologies® is partnering with Microsoft to map the genetics of the human immune system, or immunome, in a bid to help detect early stages of cancers and other diseases. Immunosequencing is designed to help patients' immune systems fight diseases like cancer, by measuring the body's initial response and then using targeted drugs to stimulate a response. The initiative, which uses large-scale machine learning and cloud computing, is attempting to diagnose conditions based upon a simple blood test.[8] And advances such as Illumina's next-generation 'massively parallel' sequencing technology enable faster and larger-scale sequencing.[9]

These are exciting developments, but one should not get too carried away. Simply finding a problem in a genome does not automatically explain *how* this may affect the person in question. Currently any personal information needs to be integrated with other relevant clinical data — like family history — to improve its predictive power. Successful research increasingly integrates genetic data with phenotypic data (such as medical history).

In this paper we discuss how Life Sciences companies can gather, analyze and use genomic data effectively, to accelerate clinical trials, improve treatment regimens, and demonstrate drug efficacy. But as the amount of data held expands exponentially, it brings with it significant challenges in gathering, storage and analysis, as well as ongoing considerations relating to privacy, reliability and security.

GlaxoSmithKline (GSK) recently announced a 4-year collaboration with 23andMe. This will focus on R&D for potential new medicines using human genetics as the basis for discovery. The more than 5 million customers of 23andMe can choose to participate in the research and contribute their information to a database that is the world's largest genetic and phenotypic resource.

Dr Hal Barron, Chief Scientific Officer and President R&D, GSK said *"We are excited about this unique collaboration as we know that drug targets with genetic validation have a significantly higher chance of ultimately demonstrating benefit for patients and becoming medicines."*

Reference for information and quote in box above.[10]

[7] *Personalized medicine' drives better outcomes for certain heart patients*, University of Florida News, 9 November 2015. https://ufhealth.org/news/2015/personalized-medicine-drives-better-outcomes-certain-heart-patients

[8] *Adaptive Biotechnologies Announces Partnership with Microsoft to Decode the Human Immune System to Improve the Diagnosis of Disease*, Adaptive biotechnologies press release, 4 January 2018. https://www.adaptivebiotech.com/news/adaptive-biotechnologies-announces-partnership-microsoft-decode-human-immune-system-improve

[9] Illumina website, accessed 22 June 2018. https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html?langsel=/us/

[10] *GSK and 23andMe sign agreement to leverage genetic insights for the development of novel medicines*, GSK press release, 25 July 2018. https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/

Genomic data: The unstoppable growth

As the human genome becomes a more integral part of the Life Sciences landscape, the volume of genomic data continues to rise at an increasing rate.

It took more than 13 years and US$3 billion to sequence the first human genome in 2003[11], but today this can be done in a matter of hours for a tiny fraction of the cost. Consequently we are seeing numerous initiatives around the world in a bid to build up genomic data in so-called 'biobanks' to aid clinical development. Some of these are purely repositories of data for other parties to access, while others have teams of analysts to compute and analyze. In certain cases the sequencing data is combined with health and lifestyle information from other sources, including family health histories. In addition to the recent collaboration announced between 23andMe and GSK (on the previous page) there are a number of examples:

— Genomics England, a company set up by the UK's Department of Health & Social Care, is carrying out the *100,000 Genomes Project*, which aims to sequence 100,000 genomes from around 70,000 people with rare diseases, plus their families, as well as from patients with cancer.[12]

— Life Sciences giant AstraZeneca plans to go considerably further by sequencing 2 million genomes as part of its innovation efforts — an initiative likely to take a decade.[13]

— Roche recently acquired all of the shares in Flatiron Health, which includes oncology-specific electronic health record (EHR) software, as well as the curation and development of real-world evidence for cancer research[14], and a majority interest in Foundation Medicine, a molecular insights company that offers comprehensive genomic profiling and facilitates connecting physicians and their patients to the latest cancer treatment programs.[15]

— The UK Biobank follows the health and wellbeing of 500,000 volunteer participants. In partnership with GSK and the Regeneron Genetics Centre (RGC), it aims to generate genetic sequencing data from these people as part of the UK Government's Life Sciences Sector Deal, to apply genetic research to develop new medicines. Its ultimate goal is to prevent, diagnose and treat conditions like cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia.[16]

— A further joint initiative between GSK, the European Bioinformatics Institute, the Wellcome Trust Sanger Institute, Biogen and Takeda, should make available genetic and biological data for researchers to help develop future medicines.[17]

[11] *AstraZeneca to sequence 2m genomes in hunt for new drugs,* The Guardian, 22 April 2016. https://www.theguardian.com/business/2016/apr/22/astrazeneca-to-sequence-2m-genomes-in-hunt-for-new-drugs

[12] Genomics England website, accessed 22 May 2018. https://www.genomicsengland.co.uk/

[13] *AstraZeneca to sequence 2m genomes in hunt for new drugs*, The Guardian, 22 April 2016. https://www.theguardian.com/business/2016/apr/22/astrazeneca-to-sequence-2m-genomes-in-hunt-for-new-drugs

[14] *Roche to acquire Flatiron Health to accelerate industry-wide development and delivery of breakthrough medicines for patients with cancer,* Roche press release, 15 February 2018. https://www.roche.com/media/releases/med-cor-2018-02-15.htm

[15] *Roche and Foundation Medicine reach definitive merger agreement to accelerate broad availability of comprehensive genomic profiling in oncology,* Roche press release, 19 June 2018. https://www.roche.com/media/releases/med-cor-2018-06-19.htm

[16] *UK Biobank, GSK and Regeneron announce largest gene sequencing initiative on world's most detailed health database to improve drug discovery and disease diagnosis,* GSK press release, 23 March 2017. https://www.gsk.com/en-gb/media/press-releases/uk-biobank-gsk-and-regeneron-announce-largest-gene-sequencing-initiative-on-world-s-most-detailed-health-database-to-improve-drug-discovery-and-disease-diagnosis/

[17] *GSK welcomes launch of the UK Government's Life Sciences Sector Deal*, UK GSK website, 6 December 2017. https://www.gsk.com/en-gb/media/press-releases/gsk-welcomes-launch-of-the-uk-government-s-life-sciences-sector-deal/

> The Breast Cancer Now Generations Study, set up in 2004, follows more than 113,000 UK women for 40 years, to help understand why individual women are likely to develop/not develop the disease, looking at genetics, lifestyle and environmental factors.

Microsoft and BC Platforms — a genomic data management analysis company — have formed a partnership working with the Mexican health authorities to try to build Latin America's largest genotype biobank. Named *Codigo46*, the initiative initially focuses on diseases like diabetes, but should expand to include hereditary cancers and psychiatric disorders. The genotypes will be housed on the Microsoft Azure cloud and managed using software from BC Platforms. Data from electronic medical records and other health surveys will also be used to build a more complete picture of patients' health.

Reference for information in text box above.[21]

— One of the largest and long-established biobanks is the Danish National Biobank initiative, which has more than 9 million biosamples. As an example: between 1997-2004, blood samples from 100,000 pregnant women and their children were collected, and their health monitored over time, to see the cause and effect of lifestyle and diet, to detect links with diseases like cancer.[18, 19]

— The Breast Cancer Now Generations Study, set up in 2004, intends to follow more than 113,000 UK women for 40 years, to help understand why individual women are likely to develop/not develop the disease, looking at genetics, lifestyle and environmental factors.[20]

Meanwhile home DNA testing kits are proliferating. For a hundred dollars, and falling, consumers can find out more about their ancestry, and, potentially, help to predict the probability of developing conditions like multiple sclerosis, Alzheimer's, Parkinson's and cancer. The market for these types of direct-to-consumer services is expanding rapidly, again releasing massive amounts of data — although, as we discuss later, some of the promises need to be treated with a degree of skepticism. You can read more about this exciting sub-sector in KPMG's paper *Direct-to-consumer genetic testing: opportunities and risks in a rapidly evolving market*.[22]

When you add this explosion of genomic data to the broader clinical data from trials and treatment, as well as sources like wearables and Internet of Things (IoT) devices, then it is apparent that Life Sciences companies — as well as healthcare systems, consumer businesses and governmental agencies — have a huge task on their hands to manage and derive the most out of this valuable resource.

Collaboration increases the number of players

Data gathering and machine learning techniques are becoming increasingly important in defining treatment value. Algorithmic approaches can help predict those patients most likely to respond to therapies for trials and ongoing treatment regimes, as well as providing accurate evidence of drugs' effectiveness, to support R&D and reimbursement for value-based pricing. This is especially relevant, given that many of the treatments identified will be expensive and require longer term approaches to funding — in some cases across years. With continued advances in genomics, drug development is likely to move away from large, expensive randomized controlled trials in favor of alternative ways to measure a therapy's effectiveness.

But Life Sciences companies cannot go it alone and are sourcing data from an expanding range of parties, as well as outsourcing and collaborating to access specialist, clinical, laboratory and analytical resources. Amongst

[18] *Step Inside the Danish National Biobank,* accessed 19 June 2018. http://www.biobankdenmark.dk/assets/step-inside-the-danish-national-biobank.pdf

[19] Information about the Danish National Biobank, Accessed 19 June 2018. http://www.biobankdenmark.dk/assets/brochure-information-about-the-danish-national-biobank-ver-4.pdf

[20] The Breast Cancer Now Generations Study, accessed 19 June 2018. http://breastcancernow.org/breast-cancer-research/our-research-projects/the-breast-cancer-now-generations-study

[21] *Codigo, BC Platforms, Microsoft partner to build commercial repository of Latin American genotype data,* Microsoft website, January 17, 2017. https://enterprise.microsoft.com/en-us/customer-story/industries/health/genomics/codigo-bc-platforms-microsoft-partner-build-commercial-repository-latin-american-genotype-data/

[22] *Direct-to-consumer genetic testing: opportunities and risks in a rapidly evolving market,* KPMG International, 2018. https://home.kpmg.com/xx/en/home/insights/2018/07/direct-to-consumer-genetic-testing.html

the key players are biobanks, genetic interpretation services, laboratories, hospitals, genetic testing services, government projects and academia, all of whom may share data. Some examples are:

— St. Jude Children's Research Hospital, in Memphis, US, has built one of the world's largest public repositories of pediatric cancer genomics data. But the sheer size of the database made it virtually impossible to access information quickly. In response, the hospital has collaborated with Microsoft and DNAnexus, a cloud-based genome informatics and data management platform. DNAnexus can analyze large amounts of raw data, enabling companies to integrate sequencing with existing clinical records. Clients include biopharmaceutical companies, genome centers and diagnostic test providers. *St. Jude Cloud* aims to help researchers develop new treatments for pediatric diseases, by providing sequencing data and analysis. All the data is held on Microsoft Azure cloud, powered by rapid computing capabilities, which speeds up projects significantly. Such initiatives should give researchers the tools to better understand the specific genetic factors behind cancer and other diseases.[23]

— Nebula Genomics, a company run by scientists operating out of Harvard University, is using the power of blockchain to aid genomic-based R&D. It sequences genomes from individual consumers, puts these on a blockchain, and connects directly with research organizations. This should be a win-win for all parties, as the data itself remains anonymous and controlled by the owner, while life sciences scientists from both commercial and academic organizations can access valuable resources at an affordable price.[24]

These are just a few instances of parties contributing to a complex data ecosystem, in which dozens of organizations could be holding and/or processing genomic data for Life Sciences companies.

But, as we explain in the next section, this complexity exacerbates the challenge of keeping data secure, private and reliable, to ensure that it meets regulatory and ethical standards, and provides meaningful clinical analyses that can drive R&D and treatment.

Sequencing the World Cup Final

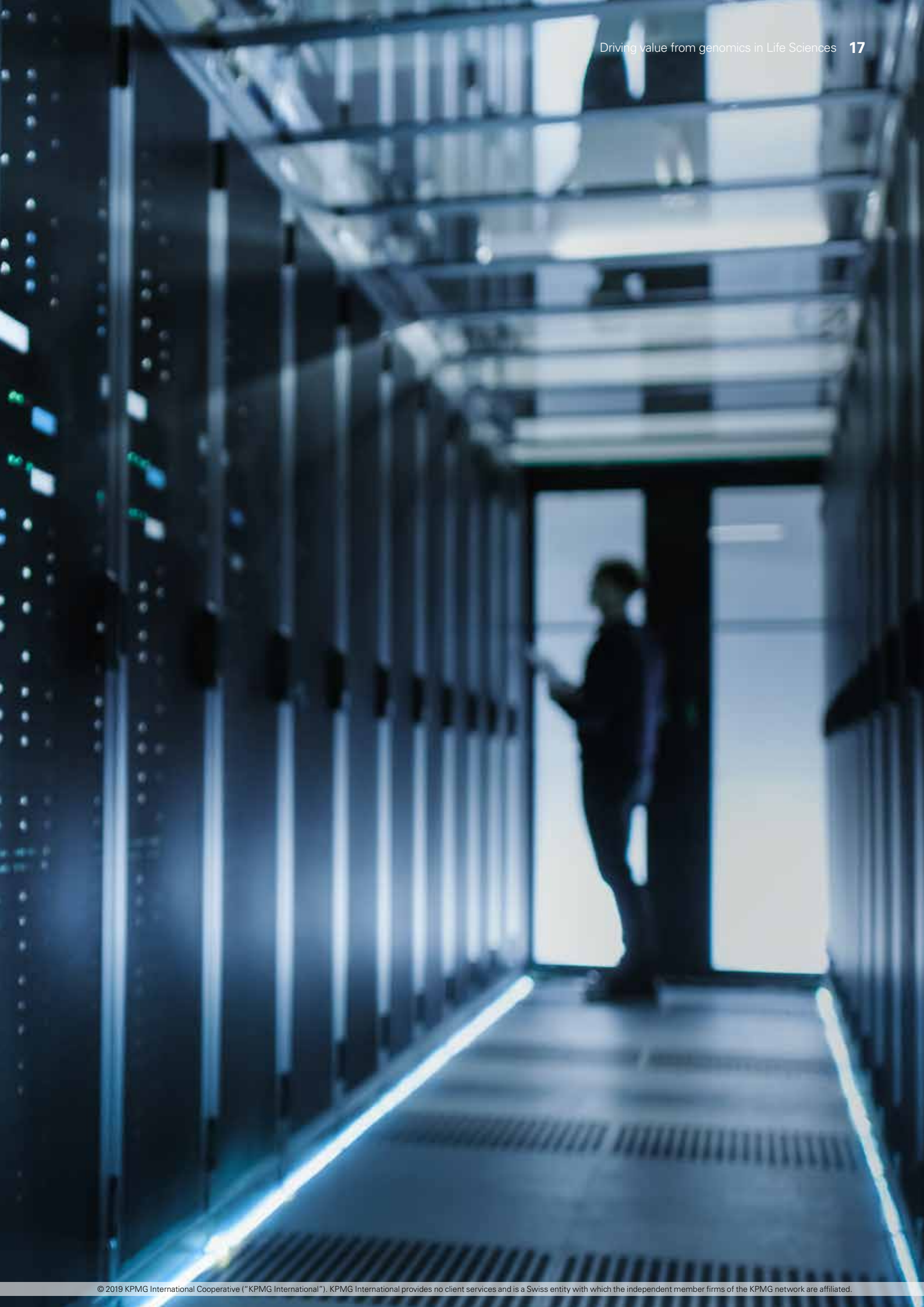
A single human genome takes 24 hours to sequence and contains 200 gigabytes of data. To sequence all 80,000 spectators at the 2018 World Cup Final at Moscow's Luzhniki Stadium, it would take 219 years involving 16 million gigabytes.

[23] *St. Jude, Microsoft and DNAnexus launch pediatric cancer genomics platform for researchers*, Fierce Biotech, 18 April 2018. https://www.fiercebiotech.com/cro/st-jude-microsoft-and-dnanexus-launch-pediatric-cancer-genomics-platform-for-researchers

[24] *This Company Wants to Sequence Your Genome, Put It On a Blockchain – And Pay You For It*, Fortune, 9 February 2018. http://fortune.com/2018/02/09/blockchain-genetic-testing-neb

Data dangers

Genomic data privacy, reliability and security is a rising concern, as Life Sciences companies strive to comply with strict regulations and verify trial and treatment results.



"The importance of data sharing for genomic medicine presents a challenge to the concept of restricting the usage of patients' data to their direct personal care."

UK House of Commons Science and Technology Committee report, 2018

Reference for text box above[25]

With ever-more genomic data emanating from an increasingly wide range of sources, and multiple parties collecting, storing, processing and analyzing the data, how can Life Sciences companies ensure that they keep control over information?

The key concerns center around genomic data privacy, reliability and security.

Privacy

Privacy laws and regulations governing the processing of personal data for research purposes are a complex area requiring careful navigation. Of all the new pieces of regulation on privacy across the world, the one with potentially the greatest impact is the EU's General Data Protection Regulation (GDPR). By covering the processing of personal data belonging to European citizens, regardless of whether the processing takes place in the EU or not, the GDPR applies to Life Sciences and technology companies developing IoT solutions and other third parties handling data on the data controllers' behalf.

The main considerations with regards to genomics relate to:

Transparency
Individuals using IoT devices through which personal data is collected and fed into clinical trials, should be aware of the organizations collecting their data, as well as the purposes for which the data is used. They also need to be informed of their privacy rights, such as the right of access, rectification and deletion of the data.

It is understandably difficult to provide such comprehensive information when the data is collected through wearable or ingestible devices — technologies which often leave very little scope for presenting transparent privacy notices.

Legal basis for processing
Another significant challenge faced by organizations collecting genomic data is finding a suitable legal basis to justify the data processing activity. Given the sensitive nature of such data, the justification could involve explicit consent from the individual, or — for entities falling under the scope of the GDPR — it could be argued that the data processing is in the interest of public health.

Whilst capturing informed consent shows respect for personal autonomy, and is an important ethical requirement in research, it may also present serious hurdles. Under the GDPR, consent must be specific, informed and unambiguous. Meeting these specificity requirements can be tough, due to the difficulty in identifying research purposes in advance. This is especially true for big data, where data mining techniques search for correlations within data sets without the baseline of a specific test hypothesis.

And whilst the GDPR allows individuals to give their consent only to certain areas of research, it is extremely difficult to manage a full audit trail of consent for each area of research and for each individual taking part.

[25] *Genomics and genome editing in the NHS,* House of Commons Science and Technology Committee, Third Report of Session 2017–19, 17 April 2018. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/349/349.pdf

By contrast, processing data for reasons of public interest — such as ensuring high standards of medicinal products or medical devices — may be less onerous.

Overall, Life Sciences companies need to balance ethical considerations against the practicalities of privacy management, to identify the most suitable basis for processing personal data relating to medical research.

Governance

Genomic data is just one example of how life sciences is becoming one of the most data-intensive industries, and consequently highly susceptible to privacy risks. To leverage personal information as an asset, Life Sciences organizations should develop a robust privacy and data governance framework, supported through clear ownership and accountability for privacy across the organization.

Cyber Security

Not surprisingly, Cyber Security looms large in the Life Sciences sector, especially for patient data generated from clinical trials. Genomic data, sensitive health data, or the formula for a complex molecule drug is worth far more on the black market than, for instance, credit card data.

To mitigate privacy and Cyber Security risks, more and more organizations are trying to de-identify personal data. However, as a recent article in *Wired* magazine puts it: *"To completely eliminate the risk of outing an individual based on their DNA records, you'd have to strip it of the same identifying details that make it scientifically useful."*[26]

The theft of trade secrets by hackers and company insiders is deemed to be an even larger problem.

The UK Government identified pharmaceutical companies as the primary target for cyber-theft of intellectual property (IP). Since the beginning of this decennium there has been an estimated US$12 billion of damages in the UK alone, of which US$2.4 billion was attributed to theft of pharmaceutical, biotechnology, and healthcare trade secrets.[27] Overall the cost of cyber-attacks continues to rise exponentially each year.

In the US, the Life Sciences sector is a primary target for cyber-theft of IP. Prominent pharmaceutical companies like Abbott Laboratories, Boston Scientific and Pfizer have already experienced major attacks. The hack of the Food & Drug Administration's computer center in Maryland exposed sensitive data including drug trial information, chemical formulas and other data for almost every important drug sold in the US.[30]

In past years, Life Sciences organizations have been on an acquisition spree, expanding their footprints, defining and realizing synergies, moving into emerging markets, and trying to personalize products and services to get as close to the patient as possible. M&A of this scale, along with the tidal wave of new technology and data capabilities, means unprecedented exposure to even more nefarious cyber threats and privacy risks. These risks originate most often from competitors and nation-states scheming to capitalize on organizations' R&D and IP data assets — and even from disgruntled or displaced workers.[31]

Companies naturally want to protect their IP and trade secrets and could suffer serious commercial and reputational damage if such information were to be hacked or go public. With a growing number

Keeping data safe and useful

Scientists are investigating ways to use vital genomic data whilst reducing the risks of identifying patients. One method being pioneered is cryptographic 'genome cloaking', which claims to hide 97 percent of each participant's unique genetic information. 'Homomorphic encryption' is another route (favored by Microsoft's cryptography team), where scientists are able to decode the final results without ever seeing the source. With genetic data increasingly moving off local servers and onto the cloud, expect more such initiatives to emerge.

References for text in box above [28, 29]

[26] *To protect genetic privacy, encrypt your DNA,* Wired, 23 August 2017. https://www.wired.com/story/to-protect-genetic-privacy-encrypt-your-dna/

[27] *The Cost of Cyber Crime,* A Detica report in partnership with the Office of Cyber Security and Information Assurance in the Cabinet Office, 2011. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60943/the-cost-of-cyber-crime-full-report.pdf

[28] IBID.

[29] *Deriving genomic diagnoses without revealing patient genomes,* Jagadeesh et al, Science, 18 August 2017. https://www.ncbi.nlm.nih.gov/pubmed/28818945

[30] *China-Based Hacking of 760 Companies Shows Cyber Cold War,* Bloomberg, 14 December 2011. http://www.bloomberg.com/news/2011-12-13/china-based-hacking-of-760-companies-reflects-undeclared-global-cyber-war.html

[31] *Life sciences innovation and cyber security: Inseparable,* KPMG International, 2017. http://www.kpmg-institutes.com/content/dam/kpmg/healthcarelifesciencesinstitute/pdf/2017/cyber-report-life-sciences.pdf

Clinical trials' reliability under question

A major study of reproducibility of medical scientific publications estimated that 80 percent of studies were non-reproducible, due to errors, misconduct or fraud. Another study of data from more than 5000 clinical trials from recent years claimed that *"Dozens …contain suspicious statistical patterns that could indicate incorrect or falsified data."* The study went on to suggest that many of these flaws could lead to corrections and even retraction of the trials.

References for text in comment box above [32, 33, 34, 35, 36, 37]

of players holding genomic data, it becomes harder to enforce strict data protection policies, which increases the chance of breaches arising from insufficient security measures.

Life Science executives need to help their organizations maintain a keen Cyber Security focus, instilling cyber-risk awareness and instigating action throughout the organization. In short: they should check the risks and not just check the box.

Reliability

Genomic data may promise a speedier time-to-market and a higher hit rate, but it could also raise the expectations of shareholders and put pressure on R&D teams to generate new, competitive breakthrough drugs.

With more parties involved in the clinical trial process, the chance of data being falsified or tampered with increases enormously. On top of this, the small target populations for therapies may force companies to look further afield for real-world evidence, including data from wearables and ingestible devices. If this data flow is not fully secure, we may see potentially corrupted data enter the trial.

To make matters even more complicated, the growing reliance on AI, machine learning and algorithms could make analysis more vulnerable to error — especially if the initial underlying assumptions are flawed, or, indeed, if the data itself is untrustworthy. It could prove catastrophic where clinical trial results found to be inaccurate or, worse still, tampered with. Such an incident would render the trial invalid, holding up availability of the drug and damaging the company's reputation. It is a similar story when analyzing treatment results for use in value-based pricing. Risk specialists, such as internal auditors, have started to explore ways to audit and give assurance over AI considering aspects such as:

— ownership and accountability for AI

— corporate values, culture and ethics

— hypothesis management

— completeness and accuracy

— controls over identity and access including the overall Cyber Security controls

— logic validation

KPMG's paper, *Trust in Artificial Intelligence*,[38] explores this issue further.

Genomic home testing kits have also come under fire. A 2018 analysis of direct-to-consumer genetic tests revealed a 40 percent false-positive rate and concluded that *"While having access to raw genotyping data can be informative and empowering for patients, this type of information can also be inaccurate and misinterpreted."* [39]

Some health insurers are offering lower premiums or rewards related to demonstrably healthier lifestyles, often based upon data from medical devices and wearables and, possibly, genomic profiles. The potential for consumer fraud is very real, with people providing false data in order to pay less.

[32] *Why most published research findings are false,* Ioannidis, 2:e124, PLoS Medicine, 2005. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124

[33] *Genetic associations: False or true?* Ioannidis, Trends Molecular Medicine, 9:135–8, 2003. https://www.ncbi.nlm.nih.gov/pubmed/12727138

[34] *Problems of reporting genetic associations with complex outcomes,* Colhoun et al,;361:865–72, The Lancet, 2003. https://www.ncbi.nlm.nih.gov/pubmed/12642066

[35] *Blockchain technology for improving clinical research quality,* Benchoufi and Ravaud, 18:335, Trials, 2017. https://trialsjournal.biomedcentral.com/track/pdf/10.1186/s13063-017-2035-z

[36] *Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals*, Anaesthesia, 4 June 2017. https://onlinelibrary.wiley.com/doi/full/10.1111/anae.13938

[37] *Dozens of recent clinical trials may contain wrong or falsified data, claims study,* The Guardian, 5 June 2017. https://www.theguardian.com/science/2017/jun/05/dozens-of-recent-clinical-trials-contain-wrong-or-falsified-data-claims-study

[38] *Trust in Artificial Intelligence,* KPMG, 2018. https://home.kpmg.com/uk/en/home/insights/2018/06/trust-in-artificial-intelligence.html

[39] *False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care*, Genetics in Medicine, 22 March 2018. https://www.nature.com/articles/gim201838

In an age where trust in institutions — particularly ones that hold personal data — is low, Life Sciences companies should think carefully about how they manage genomic data, to ensure they meet the highest standards and deliver credible clinical insights.

Shift to cloud technology

Any or all of these data dangers has the potential to compromise clinical trials and treatment and hold back clinical and commercial development. Over the past 2 years we have seen a significant shift from on-premises technology into cloud technology, in part to give flexibility of capacity, operational capability and operating costs, and also to provide greater control over genomic data privacy, reliability, and security.

In the next section we discuss how Life Sciences companies can take control of data to build a competitive edge.

The Theranos fraud scandal

Theranos, a Silicon Valley startup once valued at US$9 billion, offered quick, simple and pain-free fingerprick diagnosis for consumers, and claimed to be able to assess 70 different markers from a single drop of blood from a pinprick. The company subsequently signed contracts with major US retailers. However, an investigative Wall Street Journal reporter alleged that many of the company's claims were misleading, producing flawed testing and false results. To put this into context: a false positive on a blood test could lead to unnecessary procedures, while a false negative might mean a patient dying from an undiagnosed but serious condition.

Subsequently both the founder Elizabeth Holmes and the former president Ramesh Balwani have been charged by the SEC (US Securities and Exchange Commission) with "massive fraud", exaggerating claims about the company's technology, business and financial performance in order to attract investors. Holmes was fined US$500,000 and banned from being an officer or director of any public company for 10 years. In addition, Theranos had to settle with private investors who lost an estimated US$600 million.

References for Theranos copy in the box above [40, 41, 42, 43]

[40] *Everything you need to know about the Theranos saga so far,* Wired, 4 May 2016. https://www.wired.com/2016/05/everything-need-know-theranos-saga-far/

[41] *Theranos founder Elizabeth Holmes charged with massive fraud,* CNNtech, 14 March 2018. http://money.cnn.com/2018/03/14/technology/theranos-fraud-scandal/index.html

[42] *How Theranos used the media to create the emperor's new startup,* The Guardian, 3 June 2018. https://www.theguardian.com/commentisfree/2018/jun/03/theranos-elizabeth-holmes-media-emperors-new-startup

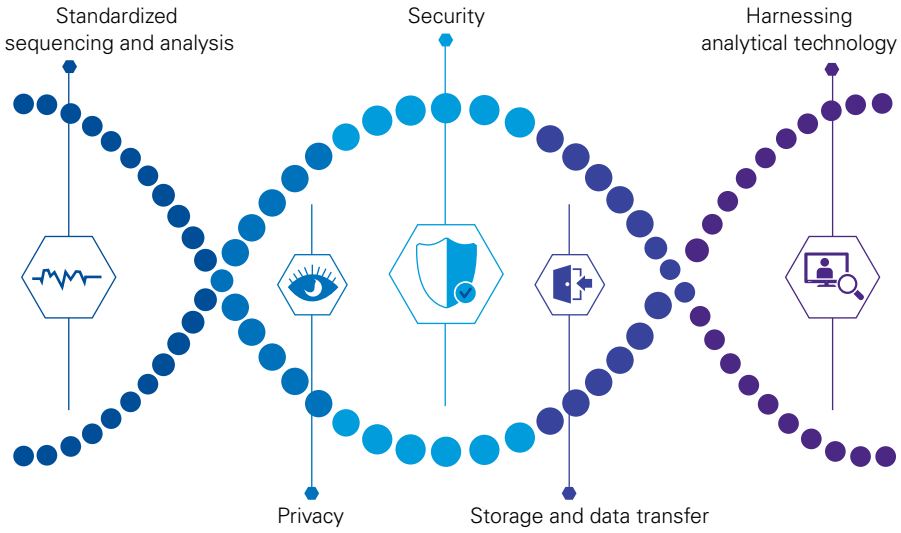
[43] *The Theranos Scandal: What Happens When You Misunderstand Steve Jobs,* Forbes, 18 June 2018 https://www.forbes.com/sites/dereklidow/2018/06/18/the-theranos-scandal-what-happens-when-you-misunderstand-steve-jobs/

Strategies to take control of genomic data: five pillars

Establishing a genomic data strategy that drives more effective R&D and meets regulatory requirements.

As regulators bed down rules on privacy and consent, R&D heads are thinking about how to speed up the pace of innovation. The future research agenda, therefore, should address the commercial, regulatory, technological and ethical challenges of getting new products to market quicker.

All Life Sciences companies need robust internal policies for keeping data safe, secure and reliable; policies that extend to all customers/patients, suppliers and partners, ensuring anonymity and pseudonymity, and avoiding any data tampering. With fake news, data leaks and breaches hitting the headlines, trust in data has become a hugely important issue. In this section we outline the five pillars of a robust genomic data strategy.



Source: Driving value from genomics in Life Sciences. 2018.

1. Standardized sequencing and analysis

The science behind genomics should be reproducible, to ensure that results are reliable and can be compared. Reproducibility is the 'litmus test' of robust science. Regrettably, R&D units and sequencing labs often sequence on different platforms, using a variety of software and analysis techniques. Without a common standard, something as simple as the amount of time a blood sample spends in a centrifuge, or the temperature during testing, can distort findings and prevent meaningful comparisons.

With more and more community platforms being used to access biobanks and other repositories, there are literally dozens of strategies available to sequence genomes. According to one study on technical analysis: *"Each approach makes trade-offs between the cost of sequencing, time to results, and type and frequency of errors. This means that different approaches may produce different results and these differences may have important clinical implications. To move toward precise genomic medicine, we must be able to reliably sequence and decipher the difficult regions of the genome."*[44]

> "The future research agenda, therefore, should address the commercial, regulatory, technological and ethical challenges of getting new products to market quicker.

[44] *Medical implications of technical accuracy in genome sequencing,* Goldfeder et al, Genomic Medicine, 2 March 2016. https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-016-0269-0

There is some hope that from a data perspective, use of blockchain technology can help bring greater consistency to sequencing during clinical trials. With its immutable nature, this technology helps to ensure traceability and data integrity, so trials can be automated and standardized.[45]

2. Privacy

To maintain and foster individuals' trust, personal data should be used for limited purposes and in line with the expectations set with the individual at the point of data collection. Privacy notices should be clear and accessible to individuals whose genomic, health and lifestyle data is being collected. Due consideration should also be given to the best format for notice requirements: for instance, 'Just in Time' notices can go a long way towards meeting the transparency requirements for individuals using wearable devices.

Organizations using genomic data must also consider the most suitable legal basis for processing the data (with attention given to the GDPR requirements and other privacy laws applying to the organization), and establish their ethical position on using individuals' data for medical research.

Governance
Companies should appoint a Data Protection Officer (DPO) to oversee IT, Legal and other functions involved in processing genomic information, to vouch for reliability, trustworthiness and completeness of data, and confirm legal consent for its use. One of the key governance goals is to balance the desire for new products with a commitment to accurate and credible trial reporting, to avoid any inaccuracies that could backfire in future.

How data is processed is not just a regulatory matter — security should also be considered.

3. Cyber Security

Life Sciences companies and the industry together will need to commit more resources to managing cyber risk as the threat continues to evolve — with high stakes. European governments have regular meetings with major life sciences companies to participate in cross-industry working groups.

Cyber Security is everyone's responsibility — and it starts at the top. Leadership and all members of the executive management team should be committed, and that commitment should radiate throughout every level of every department.

Best practice for Cyber Security involves raising awareness, performing training and simulation exercises, monitoring threats, assessing and detecting vulnerabilities, establishing processes to address weaknesses, adopting disclosure policies, and building systems that mitigate cyber risks.

Cloud technology
In safeguarding data, the guiding principle is to constantly assume the possibility of a breach. Cloud technology providers are increasingly expected to provide a high baseline of security. Consequently, Life Sciences companies and their cloud security providers should be carefully assessed to ensure compliance with HIPAA (Health Insurance Portability and Accountability Act of 1996), the GDPR, and other regulations to minimize the impact of data breaches.

In assessing the security of cloud technology providers, Life Sciences companies should also check for encryption techniques for transmitting and stationary data, assurance over third party contracts and security vetting of employees and other insiders.

The recently updated Cyber Security Framework of the US National Institute of Standards & Technology (NIST) provides a good overview of the various policies and measures for Life Sciences companies to implement to protect itself against cyber risks:

Identify
Develop an organizational understanding to manage Cyber Security risk to systems, people, assets, data, and capabilities.

Protect
Develop and implement appropriate safeguards to ensure delivery of critical services.

Detect
Develop and implement appropriate activities to identify the occurrence of a Cyber Security event.

Respond
Develop and implement appropriate activities to take action regarding a detected Cyber Security incident.

Recover
Develop and implement appropriate activities to maintain plans for resilience and to restore any capabilities or services that were impaired due to a Cyber Security incident.

Source for the NIST Cyber Security Framework[46]

[45] *Blockchain technology for improving clinical research quality,* Benchoufi and Ravaud, 18:335, Trials, 2017. https://trialsjournal.biomedcentral.com/track/pdf/10.1186/s13063-017-2035-z

[46] *Framework for Improving Critical Infrastructure Cybersecurity,* version 1.1, National Institute of Standards and Technology, 16 April 2018. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf

Data beneath the sea

Microsoft has placed data centers in the sea as part of a research project to develop greener ways to store information. The latest data center is situated off the coast of Scotland and uses seawater to cool the server racks. Another benefit of storing data in the sea is the short distance information needs to travel to reach the shore.

Reference for text in box above[47]

Managing access to genomics data

Identity and access management technology is increasingly important for managing general and privileged access to assets for employees, customers and other third parties. This technology is important not only for security, but also to streamline digital transactions across the Life Sciences ecosystem.

Think ahead and make practice a priority

No system can ever be completely secure, so continuous threat monitoring and regular testing of organizational Cyber Security practices is a must, to stay on one's digital toes and avoid any security or privacy lapses. A crisis response plan is another important safeguard, to build the capability to respond swiftly and appropriately should a breach occur, to stop any further leaks, and communicate openly to stakeholders and the media.

4. Storage and data transfer

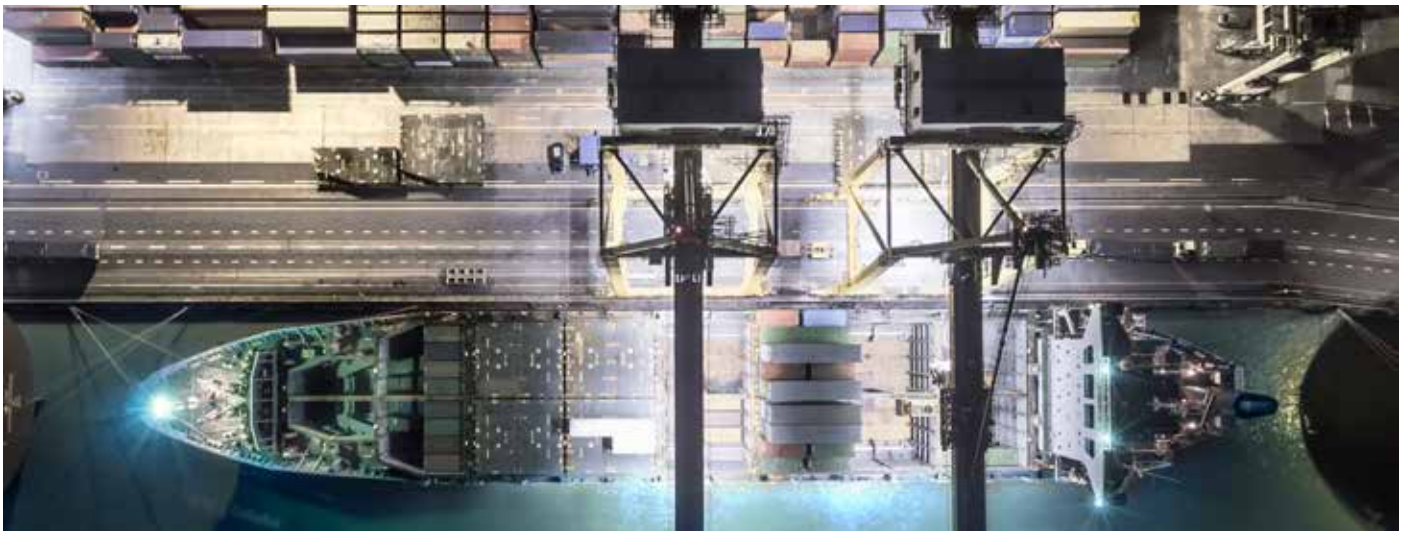
Research is typically conducted in multiple locations, so a central team of data managers should continuously monitor how clinical data is captured, to help verify the accuracy and reliability of source data, with alerts triggered when inconsistent data patterns are spotted.[48]

The global nature of research means that data is often collected, processed, and transmitted in multiple countries, and then integrated into a global clinical trial database. With the GDPR's restrictions on international data transfers, organizations need to ensure that they have adequate safeguards for the protection of personal data that is sent outside the EEA (European Economic Area), paying particular attention to model clauses, ad hoc contractual clauses, etc.

And as genomic data volumes expand, storage becomes a bigger issue. With trials typically using petabytes of data, the bandwidth required to move such huge amounts from remote cloud-based data centers — often in lower-cost countries — is impractical. In essence, it has become harder to move sensitive personal genomic data across borders, for both regulatory and cost reasons, so Life Sciences companies have rethought the footprint of their data centers and located them closer to the R&D facilities. Realistically, most, if not all will be shifted to global cloud providers, who can offer secure storage closer to the point of use, enabling safe transfer of massive volumes of data.

[47] *Under the sea, Microsoft tests a datacenter that's quick to deploy, could provide internet connectivity for years,* Microsoft website, 5 June, 2018. https://news.microsoft.com/features/under-the-sea-microsoft-tests-a-datacenter-thats-quick-to-deploy-could-provide-internet-connectivity-for-years/

[48] *The Cost of a Failed Clinical Data Strategy and How to Avoid It,* Pharmpro.com, 4 August 2017. https://www.pharmpro.com/article/2017/08/cost-failed-clinical-data-strategy-and-how-avoid-it



5. Harnessing analytical technology

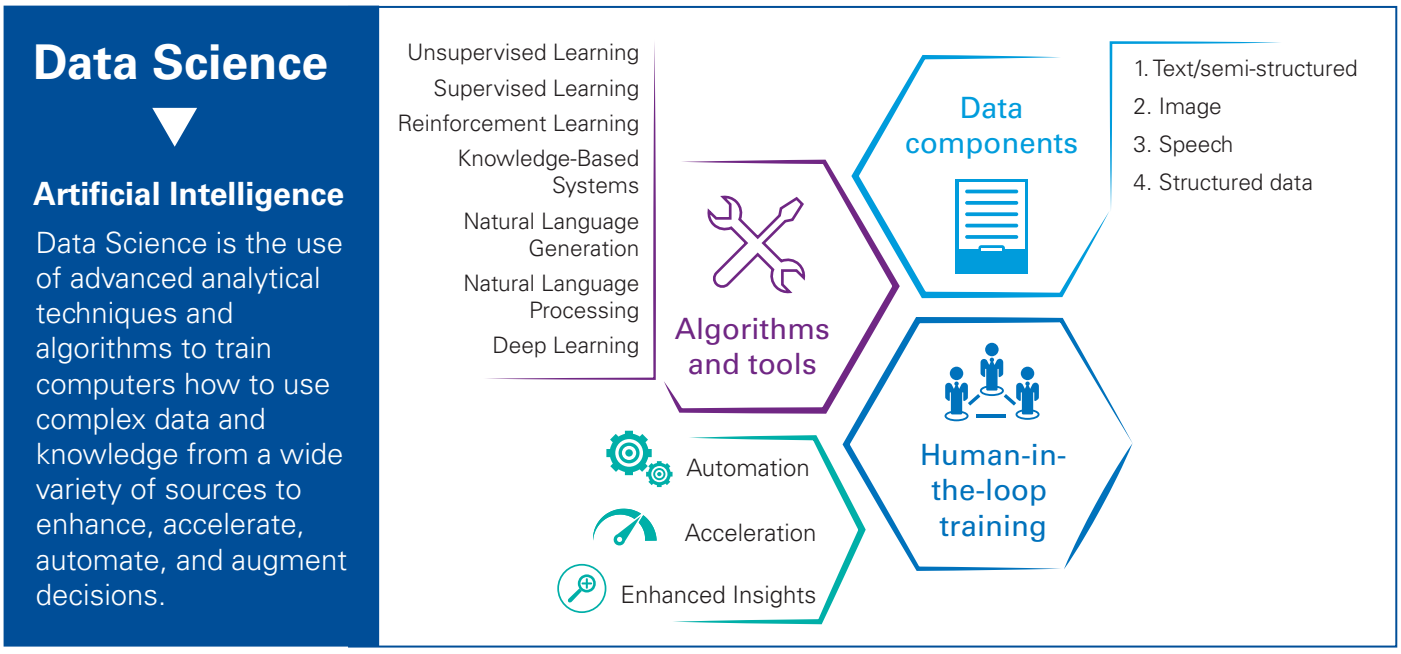
With rare diseases and oncology set to benefit greatly from genomics, the role of technology cannot be underestimated, with vast and increasing volumes of genomic data involved in clinical trials, decision support during treatment, and payments. R&D is likely to require fast, powerful and incisive algorithmic capabilities involving specialized tools.

For example, deep learning algorithms have already demonstrated revolutionary achievements in the field of AI (e.g. image recognition, object detection, audio recognition and natural language processing). The intersection of deep learning and genomic research offers huge promise in understanding human disease, particularly with the introduction of high-throughput sequencing.[49] Life Sciences companies need to acquire and/or gain access to such skills — as well as huge computing power — which is arguably only realistically available via the cloud.

A second example includes Microsoft's Project InnerEye, a research project that uses machine learning technology to analyze 3D radiological images, including those containing cancerous tumors. The process is designed to help radiologists save time and cost.[50]

In a further example of AI, SOPHiA GENETICS has developed a universal technology for genomic data analysis. Its aim is to share knowledge and enable genomic testing worldwide through its software-as-a-service platform, enabling processing and analyzing of raw genomic data to help diagnoses in oncology, metabolism, pediatrics, cardiology and hereditary cancer. This so-called "democratization of data" encourages collaboration from clinicians to improve diagnosis and treatment.[51]

Finally, we have the emerging 'digital twin' technology, which attempts to simulate a whole human through genomics, physiology, lifestyle and environment, using genetic information overlaid with other data from IoT. Potentially, we can then simulate and predict healthcare outcomes and evaluate which drugs may be most appropriate for particular conditions. Not surprisingly, this is highly complex and currently very expensive.



Source: Driving value from genomics in Life Sciences. 2018.

[49] *Deep Learning for Genomics: A Concise Overview,* Tianwei Yue, Haohan Wang, Cornell University Library, 8 May 2018, https://arxiv.org/abs/1802.00810

[50] *Machine Learning and the InnerEye for Cancer Treatment with Dr. Antonio Criminisi,* Microsoft website, 21 February 2018. https://www.microsoft.com/en-us/research/blog/machine-learning-and-the-innereye-for-cancer-treatment-with-dr-antonio-criminisi/

[51] SOPHiA Genetics website, accessed 18 June 2018. https://www.sophiagenetics.com/home.html

Closing thoughts and takeaways

Key steps for Life Sciences companies

This is a very exciting time for R&D in Life Sciences. Genomic data can take the sector into an era of highly personalized medicine, where patients get treatment tailored to their genetic make-up, with a greater chance of success, delivering value to the healthcare system.

While R&D heads ponder how to make the most of genomic data while remaining compliant, they could also enter into an open dialogue with patients on how genomic data is used, and the extent to which patients want their genomic and medical data to remain private.

One challenge the industry needs to resolve is patient numbers for trials. As genomic targeting gets more precise, the volume of potential trial participants continues to fall, to the point where it can be very difficult to reach a significant sample size. More innovative trial design and ways of comparing groups of patients are required. Data sharing between companies that are traditionally competitors may be a step change in mindset, but could open the door to faster trials and subsequent product approval.

And, as we have mentioned, genomics is a great step forward, but should be supplemented by wider clinical and lifestyle data, supported by data analytics to achieve pattern recognition, as well as diagnostic technologies like radiology.

Genomic data strategy checklist

Have you established a consistent approach to testing and reporting?

Have you considered technologies like blockchain?

Have you placed customers at the center of your genomic data privacy strategy?

Can anyone easily find out how data is submitted, stored, accessed and used?

Have you identified the suitable legal basis for processing data for research purposes?

Do you monitor privacy and security practices and carry out regular checks on data reliability and accuracy?

Have you re-appraised contracts with third parties to ensure privacy and security?

Are all employees and third parties fully aware of all key regulations and practices?

Security starts at the top: Have you put a senior executive at the helm?

Are you taking action and practicing before you experience security breaches?

Do you know your crown jewels and are you focusing on securing the heart of the business?

Do you have best practice Cyber Security policies including access restrictions, firewalls and alerts?

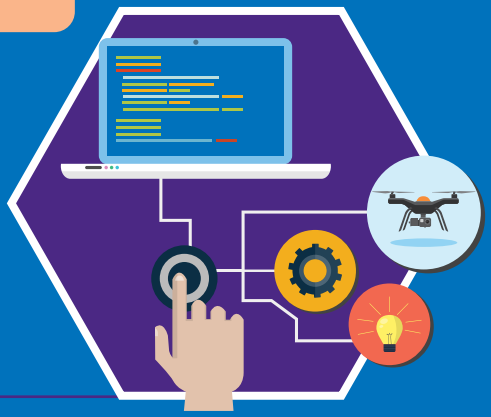
Do you have a crisis plan in case of a breach?

Are you continuously monitoring and translating the cyber threat landscape?

Security reaches beyond your organization: Do you include partners, suppliers, and vendors?

Are you confident in your data handling capabilities, in terms of volume and accessibility?

Can you scale genomic data activity up at optimal cost and effectiveness?



Are you making the most of machine learning, AI and algorithms?

Have you accessed key capabilities by buying directly, acquiring companies or outsourcing?

Source: Driving value from genomics in Life Sciences. 2018.

Contacts

Larry Raff
Global Sector Chair for Life Sciences
KPMG in the US
T: +1 212 872 3387
E: iraff@kpmg.com

Caroline Rivett
Director, Life Sciences
KPMG in the UK
T: +44 20 73112366
E: caroline.rivett@kpmg.co.uk

Prof Hilary Thomas
Partner
KPMG in the UK
T: +44 20 73114154
E: hilary.thomas@kpmg.co.uk

Dr Rebecca Pope
Lead Data Scientist
KPMG in the UK
T: +44 20 76941890
E: rebecca.pope@kpmg.co.uk

Chantal Rademaker de Ridder
Partner Technology/Cyber and Head of Privacy & Data Protection
KPMG in the Netherlands
T: +31206 567317
E: rademakerderidder.chantal@kpmg.nl

Martina Algeri
Privacy Manager
KPMG in the UK
T: +44 20 73118405
E: martina.algeri@kpmg.co.uk

Thought Leadership

R&D 2030

Digitalization in Life Science

Genetic testing

Oncology

kpmg.com

Designed by Evalueserve.

Publication name: Driving value from genomics in Life Sciences

Publication number: 135904-G

Publication date: January 2019