CHAPTER 7

Harnessing in Silico Technologies to Develop and Augment Second-Generation Cell-Based Therapies

Crystal Ruff[1,2,a], Alain A. Vertès[3]
[1]Illumina Cambridge Ltd, Cambridge, United Kingdom; [2]DRI Biotechnologies, London, United Kingdom; [3]Sloan Fellow, London Business School, London, United Kingdom

CHAPTER OVERVIEW

The field of regenerative medicine is moving faster than ever. The sophistication of online databases is, likewise, on a sharp, upward trajectory. With Moore's Law producing, until recently, a near-doubling of technological advancement every 18 months, we have seen increasing integration of – and co-operation between – biological and in silico systems (Mack, 2011; Schaller, 1997; Waldrop, 2016).

As we move forward with cell therapies, it will be integral to characterise more fully the cellular response – both to injury and in transplant. Currently, cell replacement therapies exhibit limited engraftment, and very little is known about differentiation processes. Thorough studies into advanced tracking systems – as well as epigenetic and networked signalling control of differentiation and survival – will be the key to unlocking these complex mechanisms and optimising the performance of cell therapies. As science delves more into genomic, epigenomic, chemical and environmental signalling and regulation of the cellular microenvironment, scientists and clinicians are finding infinitely more physical, environmental, biochemical and microbiotic interactions that can influence how cells differentiate, respond to signals from and interact with their environment. Particularly with networked signalling cascades, Artificial Intelligence (AI) is currently the only way to accurately run in-parallel, multivariable, integrated assessments considering multiple effects. This, in turn, is enabling and will enable the use of cells for therapeutic and drug discovery purposes. Whether the cells can be used en masse with microfluidics to intuit function of large-scale living systems or whether data from cells can be fed back and forward to inform drug development, the use of in silico systems for regenerative medicine is at a crossroads.

Currently, the most common application for AI in regenerative medicine is creating smart manufacturing systems and processes (Adair et al., 2016; Williams and Thrasher, 2014; Wong, 2017); but beyond a more systematic use of 'virtual patients', new in silico approaches

[a] The opinions expressed in this manuscript are the author's own and do not reflect the views of Illumina Inc.

Second Generation Cell and Gene-based Therapies
ISBN 978-0-12-812034-7
https://doi.org/10.1016/B978-0-12-812034-7.00007-8

and big data crunching capabilities will enable scientists and clinicians to determine the pathobiology of diseases in finer detail than ever before. This will enable further delineation of previously uncharacterised or undercharacterised subforms of diseases to establish a 'phylogenic' tree of ailments – for example, based on the normal cytokine responses of the human body or microbiotic interactions. Combined with the breakthrough advances achieved in synthetic biology, these radical developments provide complete technical freedom to design the next generation of cell-based therapies attributes and their companion diagnostics to enable another step in the personalisation of the medicine of tomorrow (Barturen et al., 2018; Hofmann-Apitius et al., 2015; Schett et al., 2013; Zhou et al., 2018).

BIG DATA AND AI

Along with increased processing power, availability and storage capacity of computers, over the past several years, there has been a growing push to collect 'Big Data'. In virtually all areas – grocery shopping, transportation, online social behaviour – infinitesimal details can be collected and collated into servers for storage and analytics; these servers are becoming more powerful, faster and smaller as technology advances. This is perhaps most prominent in the medical field.

The first human genome took $2.7 billion and almost 15 years to complete in 2001 (Venter et al., 2001). Now, science is pushing to decrease that cost to $100 a genome with faster turnaround times than ever before (Illumina, 2017). For example, Illumina has developed the technology of Sequencing by Synthesis (SBS), which uses fluorescently-labelled nucleotides to sequence multiple clusters on a flow cell. This technology has been used to generate to this date more than 90% of the world's sequencing data. Remarkably, since the launch of the newest sequencing platforms, the number of sequenced human genomes has increased from 65,000 in 2015 to 500,000 in early 2017, a nearly eight times increase (Herper, 2017).

The meaningful extraction of, and access to, data has become faster and cheaper, and gradually, organisations are recognising the economic value of Big Data. As the capacity to organise and extract meaningful insights from data becomes increasingly powerful, having access to – and being in control of – large information resources is becoming a source of critical success factors; data are becoming commoditised.

What is Big Data?

The term 'Big Data' refers to large datasets; many of which are newly collected with our recent ability to store and encode more information quickly. Typically, these datasets are beyond the capability of common software tools to process in a reasonable period of time. Therefore advanced programmes must be run in parallel on any such large dataset to derive meaningful insights.

With increasing availability of and communication between technology, devices such as smart televisions, personal computers, phones, cameras, homes, wearable quantitative devices and sensors can collect large amounts of information easily, cheaply and

Metric	Value	Bytes
Byte (B)	1	1
Kilobyte (KB)	$1,024^{1}$	1,024
Megabyte (MB)	$1,024^{2}$	1,048,576
Gigabyte (GB)	$1,024^{3}$	1,073,741,824
Terabyte (TB)	$1,024^{4}$	1,099,511,627,776
Petabyte (PB)	$1,024^{5}$	1,125,899,906,842,624
Exabyte (EB)	$1,024^{6}$	1,152,921,504,606,846,976
Zettabyte (ZB)	$1,024^{7}$	1,180,591,620,717,411,303,424
Yottabyte (YB)	$1,024^{8}$	1,208,925,819,614,629,174,706,176

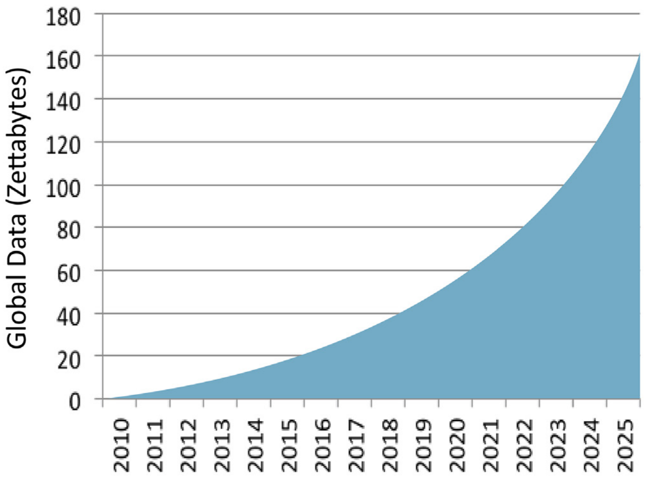

Figure 7.1 Progression of the global datasphere.

effectively. With so much information being recorded daily in society's typical digital footprint, Big Data is often a low-cost by-product of everyday digital interactions.

Most recent estimates show a predicted eightfold growth in the amount of global data creation between 2017 and 2025 (Reinsel et al., 2017), moving from 20 zettabytes generated in 2017 to 160 in 2025 (Fig. 7.1).

Furthermore, the amount of hypercritical data – that is, data with direct and immediate impact on health and well-being, such as medical, genomic, transcriptomic, proteomic and epigenomic data – is expected to grow with a compounded annual growth rate (CAGR) of 54% between 2015 and 2025 (Reinsel et al., 2017). Likewise, the percentage of those data stored on large, enterprise servers is expected to increase from ~35% to ~50% in the same period. There is a clear and present need for systems that can extract meaningful insights from compound and networked cloud datasets.

In 2001, the Gartner research group (then MetaGroup) analyst Doug Laney introduced what has become known as the three defining properties of Big Data (Laney, 2001). 'Big Data' consists of:

1. high-volume,
2. high-velocity and
3. high-variety

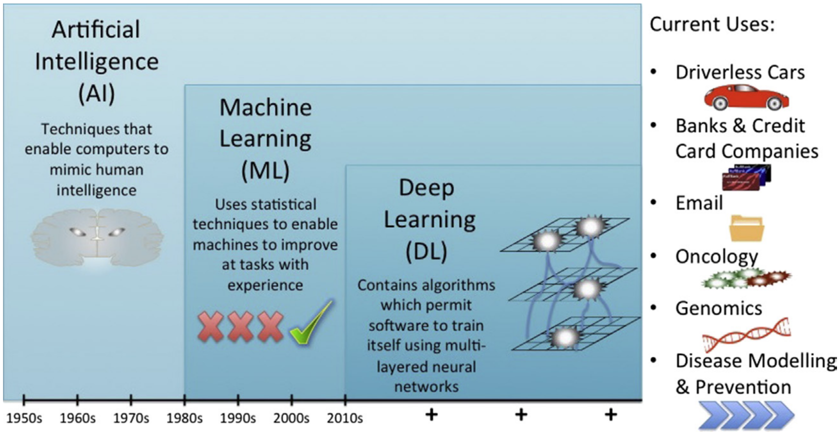
Artificial Intelligence (AI)
Techniques that enable computers to mimic human intelligence
Machine Learning (ML)
Uses statistical techniques to enable machines to improve at tasks with experience
Deep Learning (DL)
Contains algorithms which permit software to train itself using multi-layered neural networks
1950s
1960s
1970s
1980s
1990s
2000s
2010s
+
+
+
Current Uses:
Driverless Cars
Banks & Credit Card Companies
Email
Oncology
Genomics
Disease Modelling & Prevention

Figure 7.2 AI has evolved over time and is now commonplace.

information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. Thus big data often consists of raw, unsampled datasets, generated often in real-time from a variety of sources, in which data integrity is paramount.

What is AI?

First introduced in the early 1950s, AI has recently come into greater focus. AI is the branch of computer science that deals with the simulation or imitation of intelligent human behaviour using machines. It is a means by which one can intelligibly sort and extract data from large datasets. Increased focus on AI solutions has a direct correlation with increases in Big Data availability, rapid digitisation and technological advances that have made parallel processing of large amounts of data faster, cheaper and more powerful. Aside from cost and speed, AI has allowed for a new level of analysis that was previously not possible. From the push for 'Big Data' in the early 2010s arose the need for 'AI solutions' in the late 2010s (Jiang et al., 2017; Schadt et al., 2010; Tan et al., 2015; Turk-Browne, 2013).

The term 'AI' is fairly generic, and within the field, there exist many subdisciplines. AI can aid in numerous tasks, illustrated in (Fig. 7.2), however, of particular importance to Regenerative Medicine is that of Machine Learning – and within that, particularly Deep Learning.

Machine Learning

Machine Learning is the subdiscipline of AI that uses algorithms to parse data, learn from the data set and then solve a discrete problem drawn from that specific data set – all without human intercession. As a result, it enables machines to improve at tasks with repetition and experience. It is particularly useful in complex (or repetitive) system optimisation. Rather than manually inputting instructions each time, the machine is 'trained' to accomplish a specific task using large amounts of data and algorithms that enable it to learn to successfully perform the task in question. For example, instead of a researcher manually classifying or quantifying thousands of histological samples – which can take years – machine learning can quickly assess and quantify hundreds of thousands of images to immediately provide quantitative data and analyses that are equivalent to or even more accurate than those performed by humans (Awai et al., 2006; Ciompi et al., 2017; Li et al., 2005; Toney and Vesselle, 2014; Yassin et al., 2017). Often, it requires massively parallel software running on several thousands or hundreds-of-thousands of computers; but more recently, computing power has grown to allow the same power in fewer units – for example, NVIDIA has increased central processing unit (CPU) speed 10× and promises 96× faster training compared with previous standards with their DGX-1 system (NVIDIA, 2017). The relationship of machine learning within AI is thus Any written program that has human-like behaviour can be AI, but only with machine learning are the programme's parameters automatically learned from its datasets.

One might know machine learning as the system that helps keep one's inbox (relatively) free of spam, monitors one's bank accounts for fraud or the 'bot' that sends one purchasing suggestions on social media, but it is much more ubiquitous than that. In any complex system, there comes a point where simple rules and operations that previously governed it become too complicated – and generate too many exceptions – to keep up with the pace of the data. The simple rules-based 'If word "x" is followed by word "y" within 15 spaces, then do operation "z"' no longer accounts for much of the variation within the data. False positive or false negatives – particularly in healthcare where one may be dealing with a cancer diagnosis or a prenatal genetic screen for a fatal disease – can lead to very serious consequences. To minimise this, machine learning generates a 'prediction score' of how close the data fit with the rules-based system (upon which it is based) in each scenario. The programmer can subsequently set a threshold whereby an action will be triggered – for example, manual analysis if the prediction score of a cancer image screen is under 95%. This can help to minimise errors in system (James et al., 2013; Kourou et al., 2015; Kruppa et al., 2012; Parmar et al., 2015; Yassin et al., 2017).

Machine Learning algorithms are typically robust – classification algorithms dating back to the 1950s and 1960s are still commonly used for pattern-based analysis, including Decision Trees, Cluster Analyses, Neural Networks and Genetic Algorithms (Hastie et al., 2002; Vapnik, 2013). More recently, deep learning – a type of neural net with multiple layers and nodes – has come to the forefront of data processing and engineering (Zanin et al., 2016).

Deep Learning

Deep learning (also known as deep structured learning or hierarchical learning) is a technique for implementing machine learning that utilises models with deeper processing layers. It collates blocks of machine learning algorithms into a final, structured, version called a 'function composition', the inputs of which can be adjusted to better predict final outcomes. Deep learning is based on the system in question 'learning' data representations – as opposed to task-specific algorithms – with multiple levels of abstraction. Learning can be supervised, partially supervised or unsupervised. The level of supervision relates to the amount of input or 'checking' the system has externally. For example, when one is asked to select all the 'signs' from pictures as a security measure on a Website, this acts in a dual mechanism. (1) it proves the user is a human and (2) it gives feedback to machine learning algorithms that have identified patterns in those images as 'potential' signs, thus training the model. Likewise, this deals with data inputs which can be structured or unstructured, examples of which are spreadsheets and PDFs, respectively.

In order to learn the complex functions of these high-level abstractions, systems must develop deep architectures. First, large amounts of data are encoded into neural networks. A neural network is a formation of data, the 'reasoning' of which is embedded in the behaviour of thousands of simulated neurons, arranged into dozens or even hundreds of intricately interconnected layers (Knight, 2017). First-layer neurons each receive an input signal, such as the intensity of a pixel in an image, perform a calculation and then output a new signal. Outputs are fed recursively and successively to the next layer and then the next layer in a process that is repeated until the final output is achieved. On top of this, deep learning algorithms can run multiple levels of nonlinear operations to extract meaningful insights from this large amount of encoded data. To do this, deep learning uses a back-propagation algorithm – which calculates the error contribution of each 'neuronal' unit – to influence how the machine should adjust its internal parameters. These weighted parameters are then applied to compute a representation in each layer from that of the previous layer (LeCun et al., 2015).

Recently, advances in CPUs and graphical processing units have allowed for increased computing power and parallel processing necessary to perform these advanced computations. Resultantly, machine learning has significantly improved outcomes in fields such as genomics, oncology and drug discovery, as well as more mainstream areas, such as voice/image/object recognition, translation and transport.

What is a Neural Network?

The concept of the neural network is based on the deep architecture of the mammalian brain. Neuronal units signal together in complex networks and have billions of interconnections, which can, in turn, propagate motor and sensory signals. An in silico neural network uses an interconnected network of virtual neuronal 'units'; however, unlike biological brains, where interneuronal connections are largely dependent on physical distance, artificial neural networks can have layers, connections and directions of data propagation, and if they do, those layers must be discrete (Rawat and Wang, 2017).

Successive layers in a deep neural network enable the system to recognise patterns and create outputs at different levels of abstraction. Take, for example, deep neural networks in the context of facial recognition. Lower layers detect simpler things like shapes, colours and brightness, whilst middle layers identify patterns within the shapes, such as eyes, ears, mouth and nose. Still further, top layers piece those together to recognise and identify a face. In the end, it produces a 'probability vector', based on the weighting, where, for example, the system might be 92% confident the image is a certain person, 70% confident it is a face and 2% confident it is a pizza – and the network architecture subsequently feeds back to tell the neural network whether it is correct or not (Rawat and Wang, 2017).

Currently, these processes are performed in a 'black box', with users unable to decipher the various adjustments the machine implements as it learns. However, there is a growing push for increased transparency – particularly with systems that output critical data, such as cancer diagnoses. In response, new systems are being created which can retrace their steps or produce examples, which are representative of their patterns, so paths within an algorithm can be traced (Yala et al., 2017).

Structured and Unstructured Data

As highlighted earlier, Big Data comes in two primary types: structured and unstructured. Structured data are typically what one would see in a massive excel spreadsheet. It possesses a clear and distinct schema – discrete classifications, rows and columns of data that correspond exactly to one another and that use identical terminology. In practice, most datasets are not that well prepared or co-ordinated. This is referred to as 'unstructured data'. Unstructured data are the type of data one would find in medical reports, PDFs or emails. Unstructured data use arbitrary abbreviations, do not always possess proper grammar, spelling and syntax, and are often written prose rather than succinct units of data. In order to extract usable information from unstructured data, Natural Language Processing (NLP) is often used.

Natural Language Processing

NLP is a tool for structuring data in a way that AI systems can process that deals with language. NLP uses AI to 'read' through a document and extract key information. For example, often the same abbreviations are used between related disciplines; MD can translate to Managing Director, Medical Doctor or even Mental Disease, depending on the context. Furthermore, 'Cold' can be a body temperature, an illness or a descriptor of climate. Even the 'NLP' acronym itself can stand for multiple terms: Natural Language Processing or Neuro-Linguistic Programming. NLP is a complicated subdiscipline of AI, which parses through and extracts data based on context. If a word or acronym is mentioned within a certain proximity to other key terms or in a certain paper, it can be classified and stored with a degree of certainty into a structured matrix, upon which machine learning can perform more complex calculations.

USING BIG DATASETS

In the context of regenerative medicine, Big Data insights are focused around three primary areas:

1. Drug Discovery and Development – including structural databases and process modelling
2. 'Omics – including genomics, proteomics, transcriptomics and epigenomics; and
3. Biosimulation for Disease Prevention and Prediction – including imaging and detection

Drug Discovery and Development

For the first time, NLP can extract and encode data from literature databases – such as PubMed and Google Scholar, structural databases, patent reference libraries, genomic data and any other big data stores of scientific information. These data can be organised, processed and kept up to date, allowing for the derivation of meaningful insights. With over a million articles published in PubMed alone each year (National Institutes of Health, 2018), it has become unfeasible for a human to be able to accurately interpret and encode all the information in a particular field in a timely manner. Therefore computers and AI are now being utilised to extract key information and relationships upon which insights can be based. This is typically organised around three categories:

1. Using Structural Databases to Design Molecules
2. Repositioning: Deriving Novel Pathway Insights and Drug/Disease Interactions and
3. Creating Network Linkages

Using Structural Databases to Design Molecules

Structural databases can be used to create predictive models of drug interactions or to design drugs with modified functional groups. Notably, Structure Activity Relationship data have been a cornerstone of small molecule design for many years. For example, if a drug failed clinical trials because of a certain adverse event, and that event has been linked to certain structural characteristics, AI can be used to explore the chemical space around those initial compounds and design drugs that contain the functional group of interest, whilst excluding those that might be linked to the adverse event in question (Elokely et al., 2016; Zürich, n.d.; Reker et al., 2014; Schneider, 2018; Schneider and Schneider, 2016; Sjögren et al., 2018). Likewise, based on their unique properties, drugs can be designed de novo to possess desired characteristics or interact with – or modify – certain cell types.

Repositioning: Deriving Novel Pathway Insights and Drug/Disease Interactions

On the other hand, AI can derive nth degree connections between drugs, pathways and diseases, which might not be feasible for humans. For example, if drug A affects pathway B, which influences pathway C, which sequentially activates pathways D, E, F and G, which was known to play a role in disease H, a computer could make the connection between drug A and disease H, whereas this would likely lie outside the realm of human capacity. In essence, this enables to bring to the next level the generation of experiment-derived learnings exemplified by the 'rule-of-five' (also known as 'Lipinski's rule of drug-likeness') that was empirically derived from pharmacochemistry research (Lipinski, 2004; Shultz, 2018). This is particularly powerful if A and J are involved in disparate biological systems, which may not typically be explored together in modern science. Here, put together, data from past clinical trials represent a bounty for drug designers.

Creating New Network Linkages

Besides biological systems, these functional links can also span geographies, languages and fields in ways which human insight cannot. For example, a Russian researcher working on a certain pathway for cancer might not necessarily be in contact with a US-based clinician investigating that same pathway but in ocular diseases. Furthermore, neither of them may know about a Korean trial designed to evaluate a drug that blocks that particular pathway, which could be useful to both of them in different ways. The use of AI is invaluable to link these otherwise disparate lines of investigation.

This necessarily changes the role of the human experimenter, from one of 'data processor' to one of 'insight deriver'. Indeed, Alibaba Group Holding Ltd. (Hangzhou, China) entrepreneur Jack Ma has publicly argued that future educational curricula should change to reflect the growing role of technology (Horowitz, 2017). There is no way that humans can keep up with the massive processing power of today's computers; therefore children must be taught to be proficient at those activities which computers cannot do – such as innovation and creativity.

Various Business Models

Within this area, there are two primary business models:

1. Those that provide the machinery for secondary companies to process separately, such as IBM Watson (Armonk, New York, USA)
2. Those that extract and process data in-house – either selling the finished product or developing it internally. Examples of which are BenevolentAI (London, Londonshire, UK), Merck (Kenilworth, New Jersey, USA) and 23andMe (Mountain View, California, USA)

Due to the technical difficulty in obtaining meaningful insights from high-quality data and designing usable user interfaces on existing systems, the most successful companies are most often those that iterate upon their own systems to derive value and either develop or sell the end product. Some companies currently investigating this are illustrated in Fig. 7.3.

Working with 'Omics Insights

Another interesting change over the past several years is the emergence and utilisation of large genomic datasets to derive therapeutic insight. Not long ago, in response to the dearth of human data, the biotechnology company Entelos Inc. (San Francisco, CA, USA)

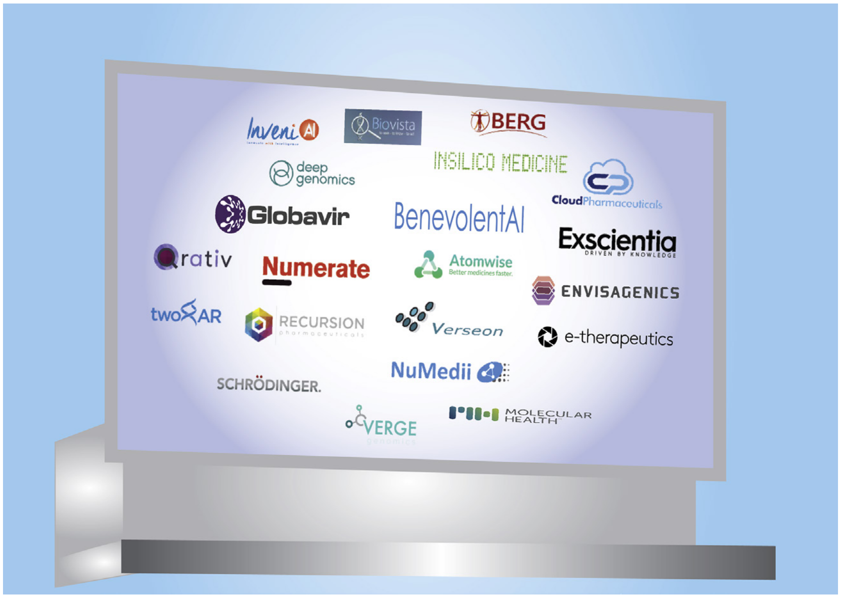

Figure 7.3 Companies using AI for drug discovery, development and repurposing.

characterised the concept of the virtual patient (Bangs, 2005; Ghosh et al., 2007; Michelson et al., 2006; Paterson and Bangs, 2006; Schmidt et al., 2013; Shoda et al., 2010). This 'virtual patient' concept uses AI and computer modelling to attempt to understand the impact of patient variability in predicting response to various therapies. Recently, with the rise and commoditisation of 'omics data, scientists and clinicians are now able to draw these insights on variability from observed genetic and phenotypic data, thus strengthening predictive power and ability to stratify patient subpopulations. Indeed, one of the largest players in consumer genomics – 23andMe – announced in 2017 that it was moving from passive data collection to active drug discovery based on its assembled database, which, in 2018, exceeded entries from 5 million patients, more than 80% of whom had consented to their data being used for research (23&me, n.d.).

Deriving 'Omics Insights

The human genome consists of 3 billion DNA base pairs, encoding ~20,000 genes. One to two percent of the genome is coding, while the remaining 98%–99% – noncoding regions – hold structural and functional information (International Human Genome Sequencing Consortium, 2004; Harrow et al., 2012; Venter et al., 2015). The Genome Reference Consortium, a successor of the Human Genome Project, is responsible for the maintenance and upkeep of the reference genome (Manzoni et al., 2016; Speir et al., 2016). Based on the reference genome and through the International HapMap Project – which maps common regions of genetic variability between populations (African, Asian and European ancestry) – scientists have determined that up to 99.8% of the human genome is

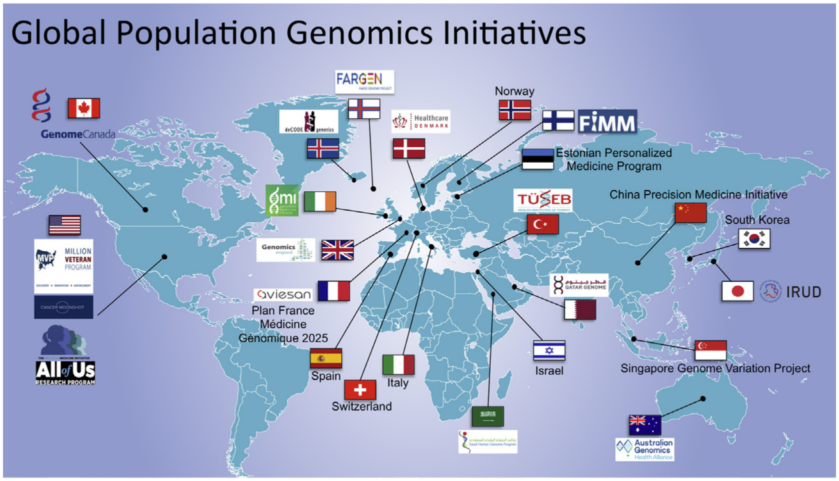
Global Population Genomics Initiatives
GenomeCanada
MILLION VETERAN PROGRAM
MVP
CANCER MOONSHOT
All of Us RESEARCH PROGRAM
FARGEN
deCODE genetics
Healthcare DENMARK
Norway
FIMM
Estonian Personalized Medicine Program
gmi
Genomics England
aviesan
Plan France Médicine Génomique 2025
Spain
Switzerland
Italy
TÜSEB
China Precision Medicine Initiative
South Korea
IRUD
QATAR GENOME
Israel
Singapore Genome Variation Project
Australian Genomics Health Alliance

Figure 7.4 Population-scale genomic initiatives are increasing alongside improvements in global health.

conserved across members of the population (Kidd et al., 2004). They have also been able to map tens of millions of single-nucleotide polymorphisms (SNPs) within the population. Applying this to disease states, scientists have developed the basis for genome-wide association studies (GWAS).

A GWAS is a study of a genome-wide set of genetic variants in different individuals to understand the genetic variations associated with different disease phenotypes. As of July 2018, the GWAS Catalogue contained 3471 publications and 65,793 unique SNP-trait associations (EBI, 2018).

Several key population genomics initiatives have been announced globally or instituted to help scientists and industry leaders understand more about genomics (from the genome), transcriptomics (involving transcribed RNA), proteomics (involving proteins translated from RNA), epigenomics (outside the genome) and metabolomics (involving metabolites), as well as more broad lifestyle and longitudinal factors (Aelion et al., 2016; Ashley, 2015; Australian Genomics, n.d.; Cyranoski, 2016; Dana, 2017; Dubow and Marjanovic, 2016; Gaziano et al., 2016; Geib, 2017; Genome Canada, 2018; Genome Canada, n.d.; Genomics Medicine, n.d.; Kaiser, 2016; Lévy, 2016; Oddmarsdóttir Gregersen, 2017; Petrone, 2017; Reardon, 2015; Sabah, 2018; Samad, 2018; TÜSEB, n.d.; Zayed, 2016); these are summarised in Fig. 7.4.

In line with this, the international Encyclopaedia of DNA Elements (ENCODE)'s remit is to map the entire genome and record each functional unit (ENCODE Project Consortium, 2012; Manzoni et al., 2016). Noncoding regions are mapped alongside coding regions, which are screened by GENCODE – an ENCODE subdirective. ENCODE can extricate nongenomic signatures as well as genetic ones, such as DNA methylation, histone modification and transcription factor, suppressor and polymerase binding patterns. It also collects data of RNA-binding proteins and the respective sequences that bind them. The ultimate product is an encyclopaedia, sorted by cell type, of DNA-based functional units, including interacting factors such as binding protein motifs, silencers, enhancers, promoters and structural or regulatory RNA (Kellis et al., 2014).

On top of this, disease-specific genomics databases are currently being compiled. These include the Cancer Genome Atlas (TCGA) – the National Institutes of Health (NIH) database of genomic information (https://cancergenome.nih.gov/) – which comprises genetic data from more than 500 human tumours representing a wide range of cancer types (Cancer Genome, n.d.). This could help with patient and tumour stratification in the future.

Applying 'omics Data to Cell Therapies

Typically, induced pluripotent stem cells (iPSCs) are used for monogenic and polygenic disease modelling or modelling of systems in which it is technically challenging or unfeasible to obtain samples. However, some of the most prevalent diseases are polygenic or do not have a readily identifiable genetic basis. With the democratisation of genetic data and unprecedented access to large datasets, increasingly, complex diseases are becoming diagnosable and polygenic or epigenetic signatures are being characterised using cell models. With greater understanding of major disease clusters, more complex diseases are increasingly being broken down into their more common component parts. Furthermore, previously undiscovered cancer neoantigens and biomarkers are now being elucidated as part of more complex systems.

Although GWAS has highlighted a large number of genetic variants with potential disease association, functional analysis remains a challenge.

Thus several groups have described approaches to functionally validate identified variants through the generation of large collections of iPSCs at population scale to form resources with which to study cellular function and pathophysiology. By assembling large iPSC cell banks, researchers hope to be able to utilise these cells more effectively to aid in understanding of phenotypic disease and hopefully move from more limited exome sequencing to whole genome analysis. Through this, scientists are now able to interrogate subtle differences between patients.

Some key players currently operating in this space include:

1. The 'Sumitomo Dainippon Manufacturing Plant for Regenerative Medicine and Cell Therapy' has opened in Japan and is the world's first commercial iPSC manufacturing facility dedicated to producing iPSCs for clinical trials (Daley, 2018; Goda and Shinden, 2018).
2. The HipSci Initiative in the United Kingdom has produced more than 3700 consistently derived iPSC lines – arguably the most extensive database of iPSC lines to date (Streeter et al., 2017). Many of these cells have accompanying Expression, Genotypic, Exome-seq or RNA-seq data (HipSci, n.d.).

3. The EU project, StemBANCC, aims to provide 1500 well-characterised patient-derived iPSC lines and associated biomaterials from 500 people in an accessible and sustainable biobank. Cells are characterised by genetic, proteomic and metabolomic profiles and are available for researchers for in vitro toxicology, disease modelling and drug discovery in hard-to-treat disorders (Morrison et al., 2015; STEMBANCC, 2012).
4. The European Bank for iPSCs is a centralised, Pan-European, not-for-profit iPSC bank. This public–private partnership project is coordinated by Pfizer Ltd. (New York, NY, USA) and managed by Roslin Cells Sciences Ltd. (Edinburgh, Scotland) and supported jointly by the Innovative Medicines Initiative and members of the European Federation of Pharmaceutical Industries and Associations. It contains 795 iPSC cell lines obtained from various disease and control individuals, with another 440 in various stages of quality control or derivation (De Sousa et al., 2017a; De Sousa et al., 2017b; EBiSC, 2018; Kurtz et al., 2018).
5. The California Institute for Regenerative Medicine (CIRM) bank is a publicly available iPSC repository for disease research and drug discovery. This initiative compiles iPSC resources derived from four leading California institutions into a central facility (CIRM, 2018). Their goal is to collect samples from various disease models (a total of 2451) and control samples (550) (Novak, 2015).
6. The NextGen Consortium in the United States has joined together nine central groups across the United States (plus their collaborators) to tackle target disease areas under the remit of the USA National Heart, Lung, and Blood Institute (Sweet, 2017).

Already, results from these initiatives show that they can represent over 95% of GWAS SNPs (Panopoulos et al., 2017) and indicate that a significant proportion (5%–46%) of variation in different iPSC phenotypes – including at epigenomic, transcriptomic and proteomic levels as well as differentiation capacity and cellular morphology – can arise from interindividual variation (Kilpinen et al., 2017). This has led to the development of the first map of common regulatory variants affecting the transcriptome of pluripotent cells in humans. They have shown new targets for nongenetic functional variations using data models and cell resources at scale that has never before been available (Carcamo-Orive et al., 2017; DeBoever et al., 2017; Pashos et al., 2017), highlighting how large population databases are necessary to obtain the experimental sensitivity that is necessary for high-level phenotypic classification across a genetically diverse population. For the first time, researchers have been able to extract key population-level insight required for determining epigenetic bases of diseases and to perform large-scale advanced disease modelling with affected and unaffected controls.

Deriving Genomic Insights in Oncology

Revealer

The machine learning algorithm REVEALER was developed to layer on top of the Cancer Genome Atlas with the global to better characterise the functional context of different types of genetic mutations leading to cancer (Kim et al., 2016). Notably, REVEALER could accurately identify many gene alterations known to be involved in tumour development and response to certain drugs. However, it also predicted novel gene mutations – for example, in the Beta-catenin and oxidative stress pathways – that are currently under investigation.

By actively and accurately tying together ‘omic datasets, scientists are able to make novel characterisations, which could potentially lead to life-saving new therapies and ways to stratify patient populations.

Measuring Single Cells

Greater technological sophistication has led to increased possibilities in the single-cell characterisation space. By attaching unique DNA barcodes to cells, scientists are not only able to ‘genome-sequence’ en masse, but they are also able to track individual cells in vivo to map lineage and differentiation to individual parent cells.

DNA-barcode-conjugated small molecules have been used successfully to develop chemical libraries for drug discovery (Castañón et al., 2018; Favalli et al., 2018; Neri and Lerner, 2018; Usanov et al., 2018; X-Chem, n.d.). This technology has opened up the possibility of exploring a huge chemical space, which is a tremendous advantage compared with conventional high-throughput screening. The concept of DNA-barcode-conjugation is now being extended into the cell therapy space.

Single-Cell Sequencing of Large Populations

By using a technique called single-cell genomic sequencing, scientists are now able to genome-sequence large cell populations at an ultra-high throughput (Lan et al., 2017). This method uses unique DNA barcode combinations, which are introduced on microfluidic droplets and a phenomenon called proximity ligation – where spatially close DNA pairs simultaneously form cross-links – to quickly label several tens of thousands of cells with unique DNA barcodes (Klein et al., 2015; Ramani et al., 2017; Vitak et al., 2017; Zhang et al., 2017). Pooled DNA can then be sequenced on a high-throughput DNA sequencer, and individual barcodes extracted in silico to map things like cell population composition, antibiotic resistance progression, viral mutations, microbial and bacterial mutagenesis in various ecosystems and tumour heterogeneity. Clear, cheap and routine sequencing of large single-cell populations will enable the deconvolution of complex components of genetic diversity in large, heterogeneous, rare or evolving cell populations.

Furthermore, as it allows for further in silico delineation of cell groups, scientists have compared this method to fluorescent-activated cell sorting (FACS) or used it in conjunction to identify new and unique cell populations from groups that had previously been identified as homogenous by FACS analysis (Bach et al., 2017; Cao et al., 2017; Chu et al., 2016; Hawkins et al., 2017; Kowalczyk et al., 2015; Macaulay et al., 2016; Molinaro and Pearson, 2016).

Furthermore, the ability to take this one step further and measure RNA expression (by reverse-transcribing into complementary DNA) at the single-cell level – using single-cell RNA sequencing or single-cell transcriptomics – has led to the discovery of new and unique cell populations, such as new types of brain (Darmanis et al., 2015; Lake et al., 2016; Pollen et al., 2014; Tasic et al., 2016; Zeisel et al., 2015), gut (Grün et al., 2015), retinal (Shekhar et al., 2016) and immune cells (Villani et al., 2017), as well as hybrid or intermediate developmental states not previously observed (Dong et al., 2018) and could begin to describe some of the heterogeneity observed in cell performance and lineage, even amongst populations which were previously thought to be uniform (Fig. 7.5).

Human Cell Atlas

In 2016, an international consortium of more than 130 biologists, clinicians, technologists, computational scientists, engineers, physicists and mathematicians assembled to attempt to create the first Human Cell Atlas (https://www.humancellatlas.org/). Primarily an RNA-sequencing project, this resource is only now possible with the advent of single-cell genomics and large-scale analysis (Rozenblatt-Rosen et al., 2017). Its purpose is to create a comprehensive, characterised map of the unique transcriptomes of all the cells in the typical human body,

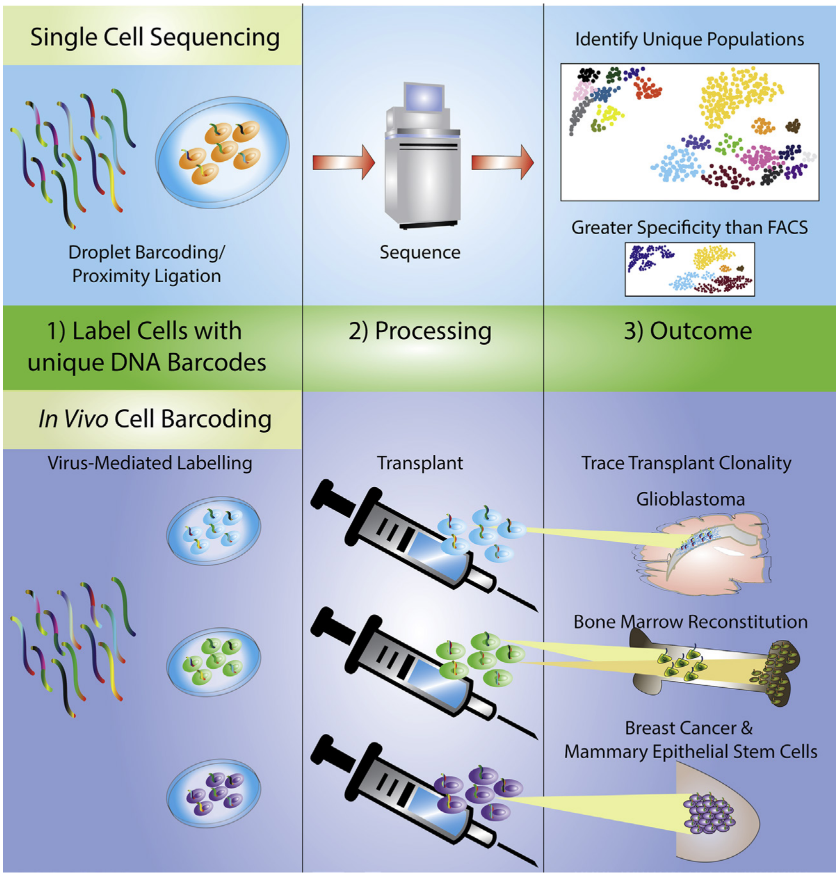
Single Cell Sequencing
Droplet Barcoding/
Proximity Ligation
Sequence
Identify Unique Populations
Greater Specificity than FACS
1) Label Cells with unique DNA Barcodes
2) Processing
3) Outcome
In Vivo Cell Barcoding
Virus-Mediated Labelling
Transplant
Trace Transplant Clonality
Glioblastoma
Bone Marrow Reconstitution
Breast Cancer &
Mammary Epithelial Stem Cells

Figure 7.5 Single-cell labelling allows for more detailed analysis than ever before.

which can then potentially be used to understand human health and disease states. This follows development and maturation and includes cell types, numbers, locations, relationships, three-dimensional interactions and molecular components. Understanding these changes and patterns, en masse, can potentially help to inform scientific understanding around subtle cell–cell and microenvironmental interactions. As of May 8 2018, 250,000 developmental and 1.08 million plus 530,000 immune cells from cord blood and bone marrow had been collected and sequenced in the Human Cell Atlas (Broad Institute, 2018; Human Cell Atlas, 2018; Human Cell Atlas, 2017; Karolinska Institute, 2018).

Cell Barcoding

Another interesting advancement is in vivo cell or DNA barcoding. Similar to the previous method, unique high-complexity DNA barcodes are attached to cell nuclei – typically here by viral vectors – and cells can be characterised both by pre- and post-implantation (Blundell and Levy, 2014; Dharampuriya et al., 2017; Ma et al., 2018; Spanjaard and Junker, 2017). This means that surviving cells can be cross-referenced against their preimplantation molecular signatures to help elucidate properties of transplant surviving or nonsurviving populations. Large numbers of cotransplanted stem cells can be studied simultaneously using this method, and researchers are able to track clonal expansion – for example, in tumours or following transplant – down to a single clonal parent cell. This can assist researchers by tracing tumour-initiating cells and determining which clones proliferate in vivo, thus contributing to biological outcomes.

This method has most extensively been used to characterise the heterogeneous and specialised nature, as well as mechanisms of carcinogenesis and kinetics, in breast cancer and mammary epithelial stem cell models (Nguyen et al., 2015; Nguyen et al., 2014b; Nguyen et al., 2014a; Wagenblast et al., 2015). It has also been employed to investigate potency and T cell development in clonal assays of hematopoietic precursors in the bone marrow niche (Aranyossy et al., 2017; Wu et al., 2018), as well as in models of glioblastoma (Lan et al., 2017), to delineate certain subpopulations that were either treatment-receptive or treatment-resistant. In terms of regenerative medicine, it could also be a useful tool to track transplant survival and the properties of cells that survive best in transplant paradigms.

Biosimulation for Disease Prevention and Prediction

On top of the wealth of these data, scientists are now able to derive insights from collections of images. These are currently being used to predict cell behaviour and in disease screening models for ailments ranging from diabetic retinopathy to cardiomyopathy.

Machine Learning to Predict the Phenotype of Stem Cells

Allen Cell Explorer

The Allen Cell Explorer (http://www.allencell.org/) is a flagship resource produced by the Allen Institute for Cell Science in Seattle, Washington (Allen Cell, n.d.). It consists of more than 6000 pictures of iPSCs, which have been altered using CRISPR/Cas9 gene editing to fluorescently label (and subsequently map) major cellular structures. It is a public online catalogue of three-dimensional stem cell images, the visuals of which were produced using deep learning analyses of these gene-edited cell lines. This resource allows researchers to track cell behaviour with various manipulations and eventually predict variations in cell layouts that are associated with cell development, differentiation and transplantation, as well as in cancer and other diseases. By using this tool, researchers are able to map and predict unique aspects of single cells at DNA, RNA and even protein levels.

The tool works by mapping the relationships between locations of cellular structures and examining how these relationships change over time. By applying machine learning to these data, systems can then use that information to extrapolate where the structures might be. For example, when the program is given only a few data points – such as nuclear positions – it is nonetheless able to generate predictions of relative positions of other intracellular structures, compare these to observed characteristics of real cells, 'learn' from it and subsequently to make further predictions about novel cell behaviours and structures with greater precision (Maxmen, 2017). Ultimately, such developments will enable better predictions of individual cell-specific differentiation and behaviour. Together with the aforementioned ability to track single cells, this could lead to large advances in the ability to predict, enrich for and even to *influence* high-performing cell populations.

Mechanobiology: Tracking the Role of Mechanical Forces on Cell Behaviour

During embryonic development, cell taxis and tumour growth, mechanical forces play a role in cell behaviour. Often on the scale of piconewtons, these forces can influence differentiation, growth trajectory, cell–cell interactions and movement; this has been extensively characterised in mesenchymal precursors (MacQueen et al., 2013; Wang and Chen, 2013). For the first time, scientists have developed the tools and the in silico capacity to map these physical forces in their minutiae (Eisenstein, 2017).

The emerging field of mechanobiology investigates how mechanical signals can affect cell development and behaviour. For example, it is now possible to microscopically map the forces that skin cells exert as they crawl forward during wound healing or to program cells to blink on and off as proteins stretch and relax. Using machine learning, scientists can intuit mechanical cell signals that might instruct a cell to stop dividing and to differentiate into a certain, more terminal phenotype.

Moreover, researchers are currently investigating how mechanical forces might interact with cancer drugs to better inform chemotherapy treatments and enhance their efficacy (Coppola et al., 2017; Wu et al., 2017). As the software becomes more sophisticated, this could potentially be linked to intuit additional novel treatment modalities.

Automated Image Analysis Algorithms

Furthermore, sophisticated image analysis platforms are emerging, which are primarily being used for automated diagnosis of disease. By processing tens or hundreds of thousands of images, computers are able to spot patterns and can often predict outcomes with even greater reliability than their a human experimenter (examples in the following). Although the role of the human experimenter remains necessary for complex diagnosis, this technology could potentially be used to decrease clinical burden – for example, by eliminating 'high-confidence' positives and negatives so clinicians could focus on the most complex cases.

Already there are a number of reports that have been published describing investigations on the role of machine learning and image processing in diagnosis. For example, studies by Google Research (Gulshan et al., 2016) and Singapore Eye Research Institute/Duke-NUS Medical School (Ting et al., 2017) into ophthalmologic indications applied deep learning to diagnose diabetic retinopathy – and in the latter case glaucoma and age-related macular degeneration – in patients from photographs of their retina. From training sets of 128 and approximately 500,000 retinal images, respectively, the resulting algorithms were subsequently validated with a separate set of 12 and approximately 72,000 images that the software had not seen before. Not only the image analysis software recognised disease states just as well as human experts but it also did so much more consistently in both cases.

Similar results have been found in cardiomyopathy and constrictive pericarditis (Sengupta et al., 2016), where AI could accurately predict diagnosis 96% of the time in 2 months. This compares to human expert rates of between 50% and 75%.

Lastly, the disease area perhaps most used for clinical image analysis is oncology (Yassin et al., 2017). Particularly in the space of pulmonary oncology, computer-aided systems have been used for several years to increase productivity and output. Several studies show increases in the number of accurate predictions of pathology with computer-aided diagnosis over human diagnosis only (Awai et al., 2006; Ciompi et al., 2017; Li et al., 2005; Toney and Vesselle, 2014). This follows remarkable advances in breast cancer diagnosis, where machine learning has been able to link image processing to phenotypic differences in the tumour – for example, to delineate visually the Her-2 state of the tumour (Vandenberghe et al., 2017) or to characterise tumour microtexture (Singh et al., 2017). Similar early efficacy in machine learning solutions have also been observed in skin cancer and melanoma (Gautam et al., 2017; Møllersen et al., 2017; Mustafa et al., 2017).

Overall, the use of big datasets in the cell and molecular biology space and the ability to interrogate them using AI and machine learning have led to an unprecedented growth in clinical capability as well as in the scientific power of experiments. This technology will continue to advance to change the role of the human operator from that of a data processor to a data interpreter one.

PERSPECTIVES

For the first time, the scientific community is amassing huge cell and 'omics databases for regenerative medicine. However, only through advanced analytics can this resource by effectively put to use. Once appropriately processed into structured information, these signals can guide not only treatment decisions but also discovery decisions, as well as inform lifestyle management and assist in preventive healthcare. Furthermore, the successful use of personalised healthcare analytics is absolutely essential as expensive regenerative medicine technologies become more commonplace and as the global population grows and ages. Sound healthcare public policies aim at keeping the healthcare environment sustainable. To this end, large leaps in technology need to come of age. What is more, in a strained healthcare environment, there is a shift from a transaction-based reimbursement system to a more value-based, outcome-driven pay-per-results model. Increased quality of healthcare will lead to increased longevity, but not just that with increased years of life will need to come concomitant increase in *quality* of the later years of life. Advanced therapies will necessarily play an integral role in this.

Process modelling has taken the scientific community only so far with over 80% of the genome's biological function remaining unknown (ENCODE Project Consortium, 2012); it is interesting to think how this might be used in the future to interrogate cell–cell and cell–environment interactions. Taking into account current trends, one would also expect huge advances in efficiency in current drug discovery and development processes. The present generation may be the last one with such little knowledge and insight about one's own body. Would this be the last of the generations to die from cancer at a high incidence? In any case, it is likely that the present generation may be the last one to not inform healthcare decisions using personalised 'omics and digital signatures.

Phylogenetic trees of disease, or nosotaxonomies, will emerge, which will enable to better design armed living drugs, equipped for example with appropriate cytokines or drug payloads based on unique disease-level or individual-level elements, such as interleukin profiles, genetic or epigenetic factors and intracellular/extracellular microenvironments (Schett et al., 2013). Complex disease/treatment/cytokine/molecular taxonomies would inform medical decision-making, enabled solely because one would be able to utilise the whole of the up-to-date knowledge pool in a particular disease or population area using AI systems. Thus standards of care would evolve to employ personalised, situational treatments based on a specific pathway, pathology and person.

Ethics and regulation are also likely to follow suit. With mass uptake of 'omics-based screening and the ability to seamlessly germline-edit human embryos, ethical questions, will shift from '*can* we do it?' to '*should* we do it?' Shared moral authority and how that changes generationally and geographically will play a role in the future of healthcare and data-driven decision-making. New ethical considerations will form surrounding the use of data, the role of 'big pharma' and the moral obligation to use the latest advances in science and technology to help inform disease models and treat any patient with the best treatments that science and technology enable. Such considerations also imply a system that keeps healthcare affordable for most and for all in a sustainable manner.

The future is an expansive horizon; technologies are emerging which will, over the next decades, iterate on and even phase-shift from (1) those technologies that have already been deployed and (2) those that, for the time being mankind can only imagine. The innovation process typically proceeds in waves or S-curves of discovery; this has been the case for the emergence of mAbs that have revolutionised medicine and met heretofore unmet clinical needs (Vertès and Dowden, 2016); this is also the case for gene therapy and CAR-T cells, and likewise it will be the case for the next wave of genetically engineered living drugs.

Continuing on this 'wave' analogy, let us imagine this emergence and growth of living drugs as a physical wave on the ocean. Every surfer knows that in order to successfully catch and ride a wave, he or she needs to anticipate it, position in the right place and paddle hard when the wave finally comes. Likewise, as Society looks toward the huge wave of regenerative medicine advancing on the horizon, stakeholders must necessarily position themselves in the right place, anticipate the next moves and work hard to combine the right technologies in the best possible way to advance the field(s) and to help as many people as possible. Only then will mankind be able to reap the collective reward of a future where truly informed healthcare decisions are made to advance and enhance the well-being of patients.

REFERENCES

About Us - 23&me. [WWW Document]. 23andMe Media Cent; n.d. https://mediacenter.23andme.com/company/about-us/.

Adair JE, Waters T, Haworth KG, Kubek SP, Trobridge GD, Hocum JD, Heimfeld S, Kiem H-P. Semi-automated closed system manufacturing of lentivirus gene-modified haematopoietic stem cells for gene therapy. Nat Commun 2016;7:13173. https://doi.org/10.1038/ncomms13173.

Aelion CM, Airhihenbuwa CO, Alemagno S, Amler RW, Arnett DK, Balas A, Bertozzi S, Blakely CH, Boerwinkle E, Brandt-Rauf P, Buekens PM, Chandler GT, Chang RW, Clark JE, Cleary PD, Curran JW, Curry SJ, Diez Roux AV, Dittus R, Ellerbeck EF, El-Mohandes A, Eriksen MP, Erwin PC, Evans G, Finnegan JR, Fried LP, Frumkin H, Galea S, Goff DC, Goldman LR, Guilarte TR, Rivera-Gutiérrez R, Halverson PK, Hand GA, Harris CM, Healton CG, Hennig N, Heymann J, Hunter D, Hwang W, Jones RM, Klag MJ, Klesges LM, Lahey T, Lawlor EF, Maddock J, Martin WJ, Mazzaschi AJ, Michael M, Mohammed SD, Nasca PC, Nash D, Ogunseitan OA, Perez RA, Perri M, Petersen DJ, Peterson DV, Philbert M, Pinto-Martin J, Raczynski JM, Raskob GE, Rimer BK, Rohrbach LA, Rudkin LL, Siminoff L, Szapocznik J, Thombs D, Torabi MR, Weiler RM, Wetle TF, Williams PL, Wykoff R, Ying J. The US cancer moonshot initiative. Lancet Oncol 2016;17:e178–180. https://doi.org/10.1016/S1470-2045(16)30054-7.

Allen Cell Explorer. [WWW Document]. Allen Cell Explor; n.d. http://www.allencell.org/.

Aranyossy T, Thielecke L, Glauche I, Fehse B, Cornils K. Genetic barcodes facilitate competitive clonal analyses in vivo. Hum Gene Ther 2017;28:926–37. https://doi.org/10.1089/hum.2017.124.

Ashley EA. The precision medicine initiative: a new national effort. JAMA 2015;313:2119–20. https://doi.org/10.1001/jama.2015.3595.

Australian Genomics - Australian Genomics Health Alliance | Home. [WWW Document]; n.d. https://www.australiangenomics.org.au/.

Awai K, Murao K, Ozawa A, Nakayama Y, Nakaura T, Liu D, Kawanaka K, Funama Y, Morishita S, Yamashita Y. Pulmonary nodules: estimation of malignancy at thin-section helical CT—effect of computer-aided diagnosis on performance of radiologists. Radiology 2006;239:276–84. https://doi.org/10.1148/radiol.2383050167.

Bach K, Pensa S, Grzelak M, Hadfield J, Adams DJ, Marioni JC, Khaled WT. Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat Commun 2017;8:2128. https://doi.org/10.1038/s41467-017-02001-5.

Bangs A. Predictive biosimulation and virtual patients in pharmaceutical R and D. Stud Health Technol Inform 2005;111:37–42.

Barturen G, Beretta L, Cervera R, Van Vollenhoven R, Alarcón-Riquelme ME. Moving towards a molecular taxonomy of autoimmune rheumatic diseases. Nat Rev Rheumatol 2018;14:180. https://doi.org/10.1038/nrrheum.2018.23.

Blundell JR, Levy SF. Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. Genomics 2014;104:417–30. https://doi.org/10.1016/j.ygeno.2014.09.005.

Broad Institute. Researchers post genetic profiles of half a million human immune cells on human cell atlas online portal. [WWW Document] Broad Inst; 2018. https://www.broadinstitute.org/news/researchers-post-genetic-profiles-half-million-human-immune-cells-human-cell-atlas-online.

Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, Adey A, Waterston RH, Trapnell C, Shendure J. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 2017;357:661–7. https://doi.org/10.1126/science.aam8940.

Carcamo-Orive I, Hoffman GE, Cundiff P, Beckmann ND, D'Souza SL, Knowles JW, Patel A, Papatsenko D, Abbasi F, Reaven GM, Whalen S, Lee P, Shahbazi M, Henrion MYR, Zhu K, Wang S, Roussos P, Schadt EE, Pandey G, Chang R, Quertermous T, Lemischka I. Analysis of transcriptional variability in a large human iPSC library reveals genetic and non-genetic determinants of heterogeneity. Cell Stem Cell 2017;20:518–32.e9. https://doi.org/10.1016/j.stem.2016.11.005.

Castañón J, Román JP, Jessop TC, de Blas J, Haro R. Design and development of a technology platform for DNA-encoded library production and affinity selection. SLAS Discov Adv Life Sci R D 2018. https://doi.org/10.1177/2472555217752091.

Chu L-F, Leng N, Zhang J, Hou Z, Mamott D, Vereide DT, Choi J, Kendziorski C, Stewart R, Thomson JA. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol 2016;17:173. https://doi.org/10.1186/s13059-016-1033-x.

Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, Scholten ET, Schaefer-Prokop C, Wille MMW, Marchianò A, Pastorino U, Prokop M, van Ginneken B. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 2017;7:46479. https://doi.org/10.1038/srep46479.

[WWW Document]. CIRM - Induced pluripotent stem cell repository website. 2018. https://www.cirm.ca.gov/researchers/ipsc-repository/about.

Coppola S, Carnevale I, Danen EHJ, Peters GJ, Schmidt T, Assaraf YG, Giovannetti E. A mechanopharmacology approach to overcome chemoresistance in pancreatic cancer. Drug Resist Updat 2017;31:43–51. https://doi.org/10.1016/j.drup.2017.07.001.

Cyranoski D. China embraces precision medicine on a massive scale. Nat News 2016;529:9. https://doi.org/10.1038/529009a.

Daley J. World's first commercial iPSC cell plant opens in Japan. Scientist 2018. https://www.the-scientist.com/the-nutshell/worlds-first-commercial-ipsc-cell-plant-opens-in-japan-29915.

Dana G. 3 ways China is leading the way in precision medicine. World Economic Forum; 2017.

Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, Quake SR. A survey of human brain transcriptome diversity at the single cell level. Proc Natl Acad Sci USA 2015;112:7285–90. https://doi.org/10.1073/pnas.1507125112.

De Sousa PA, Steeg R, Kreisel B, Allsopp TE. Hot start to European pluripotent stem cell banking. Trends Biotechnol 2017;35:573–6. https://doi.org/10.1016/j.tibtech.2017.04.006.

De Sousa PA, Steeg R, Wachter E, Bruce K, King J, Hoeve M, Khadun S, McConnachie G, Holder J, Kurtz A, Seltmann S, Dewender J, Reimann S, Stacey G, O'Shea O, Chapman C, Healy L, Zimmermann H, Bolton B, Rawat T, Atkin I, Veiga A, Kuebler B, Serano BM, Saric T, Hescheler J, Brüstle O, Peitz M, Thiele C, Geijsen N, Holst B, Clausen C, Lako M, Armstrong L, Gupta SK, Kvist AJ, Hicks R, Jonebring A, Brolén G, Ebneth A, Cabrera-Socorro A, Foerch P, Geraerts M, Stummann TC, Harmon S, George C, Streeter I, Clarke L, Parkinson H, Harrison PW, Faulconbridge A, Cherubin L, Burdett T, Trigueros C, Patel MJ, Lucas C, Hardy B, Predan R, Dokler J, Brajnik M, Keminer O, Pless O, Gribbon P, Claussen C, Ringwald A, Kreisel B, Courtney A, Allsopp TE. Rapid establishment of the European Bank for induced pluripotent stem cells (EBiSC) - the hot start experience. Stem Cell Res 2017;20:105–14. https://doi.org/10.1016/j.scr.2017.03.002.

DeBoever C, Li H, Jakubosky D, Benaglio P, Reyna J, Olson KM, Huang H, Biggs W, Sandoval E, D'Antonio M, Jepsen K, Matsui H, Arias A, Ren B, Nariai N, Smith EN, D'Antonio-Chronowska A, Farley EK, Frazer KA. Large-scale profiling reveals the influence of genetic variation on gene expression in human induced pluripotent stem cells. Cell Stem Cell 2017;20:533–46.e7. https://doi.org/10.1016/j.stem.2017.03.009.

Dharampuriya PR, Scapin G, Wong C, John Wagner K, Cillis JL, Shah DI. Tracking the origin, development, and differentiation of hematopoietic stem cells. Curr Opin Cell Biol 2017;49:108–15. https://doi.org/10.1016/j.ceb.2018.01.002.

Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, Guo H, Wen L, Tang F. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. Genome Biol 2018;19:31. https://doi.org/10.1186/s13059-018-1416-2.

Dubow T, Marjanovic S. Population-scale sequencing and the future of genomic medicine. [WWW Document] 2016. https://www.rand.org/pubs/research_reports/RR1520.html.

GWAS Catalog. [WWW Document]; 2018. https://www.ebi.ac.uk/gwas/.

[WWW Document]. EBiSC – European bank for induced pluripotent stem cells website. 2018. https://www.ebisc.org/.

Eisenstein M. Mechanobiology: a measure of molecular muscle. Nature 2017;544:544255a. https://doi.org/10.1038/544255a.

Elokely K, Velisetty P, Delemotte L, Palovcak E, Klein ML, Rohacs T, Carnevale V. Understanding TRPV1 activation by ligands: insights from the binding modes of capsaicin and resiniferatoxin. Proc Natl Acad Sci USA 2016;113:E137–45. https://doi.org/10.1073/pnas.1517288113.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74. https://doi.org/10.1038/nature11247.

Favalli N, Bassi G, Scheuermann J, Neri D. DNA-encoded chemical libraries: achievements and remaining challenges. FEBS Lett 2018. https://doi.org/10.1002/1873-3468.13068.

Gautam D, Ahmed M, Meena YK, Haq AU. Machine learning based diagnosis of melanoma using macro images. Int J Numer Methods Biomed Eng 2018;34(5):e2953. https://doi.org/10.1002/cnm.2953.

Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, Guarino P, Aslan M, Anderson D, LaFleur R, Hammond T, Schaa K, Moser J, Huang G, Muralidhar S, Przygodzki R, O'Leary TJ. Million veteran program: a mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 2016;70:214–23. https://doi.org/10.1016/j.jclinepi.2015.09.016.

Geib C. A Chinese province is sequencing 1 million of its residents' genomes. Futurism.com; 2017.

Genome Canada. Canadian patients to benefit from major investment in genomics and precision health research. Genome Canada; 2018.

Genome Canada | [WWW Document]; n.d. https://www.genomecanada.ca/en.

Genomics Medicine Ireland | Scientific Research & Discovery. [WWW Document]; n.d. http://genomicsmed.ie/.

Ghosh S, Young DL, Gadkar KG, Wennerberg L, Basu K. Towards optimal virtual patients: an online adaptive control approach. Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2007;2007:3292–5. https://doi.org/10.1109/IEMBS.2007.4353033.

Goda R, Shinden S. World's first commercial iPS cell-making plant opens: the Asahi Shimbun. Asahi Shimbun; 2018.

Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, van Oudenaarden A. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 2015;525:251–5. https://doi.org/10.1038/nature14966.

Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402–10. https://doi.org/10.1001/jama.2016.17216.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis

M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ. GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res 2012;22:1760–74. https://doi.org/10.1101/gr.135350.111.

Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Biometrics 2002. https://web.stanford.edu/~hastie/ElemStatLearn/.

Hawkins F, Kramer P, Jacob A, Driver I, Thomas DC, McCauley KB, Skvir N, Crane AM, Kurmann AA, Hollenberg AN, Nguyen S, Wong BG, Khalil AS, Huang SXL, Guttentag S, Rock JR, Shannon JM, Davis BR, Kotton DN. Prospective isolation of NKX2-1–expressing human lung progenitors derived from pluripotent stem cells. J Clin Investig 2017;127:2277–94. https://doi.org/10.1172/JCI89950.

Herper M. Illumina promises to sequence human genome for $100 -- but not quite yet. Forbes 2017;2.

HipSci Catalogue. [WWW Document]; n.d. http://www.hipsci.org/lines/#/lines.

Hofmann-Apitius M, Alarcón-Riquelme ME, Chamberlain C, McHale D. Towards the taxonomy of human disease. Nat Rev Drug Discov 2015;14:75–6. https://doi.org/10.1038/nrd4537.

Horowitz J. Jack Ma: we need to stop training our kids for manufacturing jobs. CNNMoney; 2017.

Human Cell Atlas. The international human cell atlas publishes strategic blueprint; announces data from first one million cells. [WWW Document] 2017. https://www.humancellatlas.org/news/14.

Human Cell Atlas. Human cell atlas sequences first 250K developmental cells. [WWW Document] 2018https://www.humancellatlas.org/news/15.

Illumina. Press release: high throughput sequencing instruments capable of transforming genomics to improve human health at an unprecedented scale. [WWW Document] 2017. https://www.illumina.com/company/news-center/press-releases/press-release-details.html?newsid=2236383.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931–45. https://doi.org/10.1038/nature03001.

James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Springer; 2013.

Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2:230–43. https://doi.org/10.1136/svn-2017-000101.

Kaiser J. Qatar's genome effort slowly gears up. Science 2016;354:1220. https://doi.org/10.1126/science.354.6317.1220.

Karolinska Institute. 250,000 developmental cells sequenced. [WWW Document] 2018. https://ki.se/en/news/250000-developmental-cells-sequenced.

Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, Dunham I, Elnitski LL, Farnham PJ, Feingold EA, Gerstein M, Giddings MC, Gilbert DM, Gingeras TR, Green ED, Guigo R, Hubbard T, Kent J, Lieb JD, Myers RM, Pazin MJ, Ren B, Stamatoyannopoulos JA, Weng Z, White KP, Hardison RC. Defining functional DNA elements in the human genome. Proc Natl Acad Sci USA 2014;111:6131–8. https://doi.org/10.1073/pnas.1318948111.

Kidd KK, Pakstis AJ, Speed WC, Kidd JR. Understanding human DNA sequence variation. J Hered 2004;95:406–20. https://doi.org/10.1093/jhered/esh060.

Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, Bala S, Bensaddek D, Casale FP, Culley OJ, Danecek P, Faulconbridge A, Harrison PW, Kathuria A, McCarthy D, McCarthy SA, Meleckyte R, Memari Y, Moens N, Soares F, Mann A, Streeter I, Agu CA, Alderton A, Nelson R, Harper S, Patel M, White A, Patel SR, Clarke L, Halai R, Kirton CM, Kolb-Kokocinski A, Beales P, Birney E, Danovi D, Lamond AI, Ouwehand WH, Vallier L, Watt FM, Durbin R, Stegle O, Gaffney DJ. Common genetic variation drives molecular heterogeneity in human iPSCs. Nature 2017;546:370–5. https://doi.org/10.1038/nature22403.

Kim JW, Botvinnik OB, Abudayyeh O, Birger C, Rosenbluh J, Shrestha Y, Abazeed ME, Hammerman PS, DiCara D, Konieczkowski DJ, Johannessen CM, Liberzon A, Alizad-Rahvar AR, Alexe G, Aguirre A, Ghandi M, Greulich H, Vazquez F, Weir BA, Van Allen EM, Tsherniak A, Shao DD, Zack TI, Noble M, Getz G, Beroukhim R, Garraway LA, Ardakani M, Romualdi C, Sales G, Barbie DA, Boehm JS, Hahn WC, Mesirov JP, Tamayo P. Characterizing genomic alterations in cancer by complementary functional associations. Nat Biotechnol 2016;34:539–46. https://doi.org/10.1038/nbt.3527.

Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 2015;161:1187–201. https://doi.org/10.1016/j.cell.2015.04.044.

Knight W. There's a big problem with AI: even its creators can't explain how it works. [WWW Document] MIT Technol Rev 2017. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.

Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8–17.

Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, Schneider RK, Wagers AJ, Ebert BL, Regev A. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res 2015;25:1860–72. https://doi.org/10.1101/gr.192237.115.

Kruppa J, Ziegler A, König IR. Risk estimation and risk prediction using machine-learning methods. Hum Genet 2012;131:1639–54. https://doi.org/10.1007/s00439-012-1194-y.

Kurtz A, Seltmann S, Bairoch A, Bittner M-S, Bruce K, Capes-Davis A, Clarke L, Crook JM, Daheron L, Dewender J, Faulconbridge A, Fujibuchi W, Gutteridge A, Hei DJ, Kim Y-O, Kim J-H, Kokocinski AK, Lekschas F, Lomax GP, Loring JF, Ludwig T, Mah N, Matsui T, Müller R, Parkinson H, Sheldon M, Smith K, Stachelscheid H, Stacey G, Streeter I, Veiga A, Xu R-H. A standard nomenclature for referencing and authentication of pluripotent stem cells. Stem Cell Rep 2018;10:1–6. https://doi.org/10.1016/j.stemcr.2017.12.002.

Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, Wildberg A, Gao D, Fung H-L, Chen S, Vijayaraghavan R, Wong J, Chen A, Sheng X, Kaper F, Shen R, Ronaghi M, Fan J-B, Wang W, Chun J, Zhang K. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 2016;352:1586–90. https://doi.org/10.1126/science.aaf1204.

Lan F, Demaree B, Ahmed N, Abate A. SiC-Seq: single-cell genome sequencing at ultra high-throughput with microfluidic droplet barcoding. Nat Biotechnol 2017;35:640–6. https://doi.org/10.1038/nbt.3880.

Lan X, Jörg DJ, Cavalli FMG, Richards LM, Nguyen LV, Vanner RJ, Guilhamon P, Lee L, Kushida MM, Pellacani D, Park NI, Coutinho FJ, Whetstone H, Selvadurai HJ, Che C, Luu B, Carles A, Moksa M, Rastegar N, Head R, Dolma S, Prinos P, Cusimano MD, Das S, Bernstein M, Arrowsmith CH, Mungall AJ, Moore RA, Ma Y, Gallo M, Lupien M, Pugh TJ, Taylor MD, Hirst M, Eaves CJ, Simons BD, Dirks PB. Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. Nature 2017;549:227–32. https://doi.org/10.1038/nature23666.

Laney D. 3D data management: controlling data volume, velocity and variety. META Group Res Note 2001;6:70.

LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44. https://doi.org/10.1038/nature14539.

Lévy Y. Genomic medicine 2025: France in the race for precision medicine. Lancet (Lond Engl) 2016;388:2872. https://doi.org/10.1016/S0140-6736(16)32467-9.

Li F, Arimura H, Suzuki K, Shiraishi J, Li Q, Abe H, Engelmann R, Sone S, MacMahon H, Doi K. Computer-aided detection of peripheral lung cancers missed at CT: ROC analyses without and with localization. Radiology 2005;237:684–90. https://doi.org/10.1148/radiol.2372041555.

Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov. Today Technol 2004;1(4):337–41.

Ma J, Shen Z, Yu Y-C, Shi S-H. Neural lineage tracing in the mammalian brain. Curr Opin Neurobiol 2018;50:7–16. https://doi.org/10.1016/j.conb.2017.10.013.

Macaulay IC, Svensson V, Labalette C, Ferreira L, Hamey F, Voet T, Teichmann SA, Cvejic A. Single-cell RNA-sequencing reveals a continuous spectrum of differentiation in hematopoietic cells. Cell Rep 2016;14:966–77. https://doi.org/10.1016/j.celrep.2015.12.082.

Mack CA. Fifty years of Moore's law. IEEE Trans Semicond Manuf 2011;24:202–7. https://doi.org/10.1109/TSM.2010.2096437.

MacQueen L, Sun Y, Simmons CA. Mesenchymal stem cell mechanobiology and emerging experimental platforms. J R Soc Interface 2013;10. https://doi.org/10.1098/rsif.2013.0179.

Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, Ferrari R. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform 2016. https://doi.org/10.1093/bib/bbw114.

Maxmen A. Machine learning predicts the look of stem cells. Nat News 2017. https://doi.org/10.1038/nature.2017.21769.

Michelson S, Sehgal A, Friedrich C. In silico prediction of clinical efficacy. Curr Opin Biotechnol 2006;17:666–70. https://doi.org/10.1016/j.copbio.2006.09.004.

Molinaro AM, Pearson BJ. In silico lineage tracing through single cell transcriptomics identifies a neural stem cell population in planarians. Genome Biol 2016;17:87. https://doi.org/10.1186/s13059-016-0937-9.

Møllersen K, Zortea M, Schopf TR, Kirchesch H, Godtliebsen F. Comparison of computer systems and ranking criteria for automatic melanoma detection in dermoscopic images. PLoS One 2017;12:e0190112.

Morrison M, Klein C, Clemann N, Collier DA, Hardy J, Heisserer B, Cader MZ, Graf M, Kaye J. StemBANCC: governing access to material and data in a large stem cell research consortium. Stem Cell Rev 2015;11:681–7. https://doi.org/10.1007/s12015-015-9599-3.

Mustafa S, Dauda AB, Dauda M. Image processing and SVM classification for melanoma detection. In: Computing networking and informatics (ICCNI), 2017 international conference on IEEE. 2017. p. 1–5.

National Institutes of Health. Yearly citation totals from 2017 MEDLINE/PubMed baseline: 26,759,399 citations found. [WWW Document] 2018https://www.nlm.nih.gov/bsd/licensee/2017_stats/2017_Totals.html.

Neri D, Lerner RA. DNA-encoded chemical libraries: a selection system based on endowing organic compounds with amplifiable information. Annu Rev Biochem 2018. https://doi.org/10.1146/annurev-biochem-062917-012550.

Nguyen LV, Cox CL, Eirew P, Knapp DJHF, Pellacani D, Kannan N, Carles A, Moksa M, Balani S, Shah S, Hirst M, Aparicio S, Eaves CJ. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. Nat Commun 2014;5:5871. https://doi.org/10.1038/ncomms6871.

Nguyen LV, Makarem M, Carles A, Moksa M, Kannan N, Pandoh P, Eirew P, Osako T, Kardel M, Cheung AMS, Kennedy W, Tse K, Zeng T, Zhao Y, Humphries RK, Aparicio S, Eaves CJ, Hirst M. Clonal analysis via barcoding reveals diverse growth and differentiation of transplanted mouse and human mammary stem cells. Cell Stem Cell 2014;14:253–63. https://doi.org/10.1016/j.stem.2013.12.011.

Nguyen LV, Pellacani D, Lefort S, Kannan N, Osako T, Makarem M, Cox CL, Kennedy W, Beer P, Carles A, Moksa M, Bilenky M, Balani S, Babovic S, Sun I, Rosin M, Aparicio S, Hirst M, Eaves CJ. Barcoding reveals complex clonal dynamics of de novo transformed human mammary cells. Nature 2015;528:267–71. https://doi.org/10.1038/nature15742.

Novak J. The California Institute for regenerative medicine's human iPSC initiative. Drug Discov 2015;47.

[WWW Document]. NVIDIA DGX-1: essential instrument of AI research. NVIDIA; 2017. https://www.nvidia.com/en-us/data-center/dgx-1/.

Oddmarsdóttir Gregersen N. Population scale genome study of the Faroese population using linked-reads. Nature.com Webcasts. 2017.

Panopoulos AD, D'Antonio M, Benaglio P, Williams R, Hashem SI, Schuldt BM, DeBoever C, Arias AD, Garcia M, Nelson BC, Harismendy O, Jakubosky DA, Donovan MKR, Greenwald WW, Farnam K, Cook M, Borja V, Miller CA, Grinstein JD, Drees F, Okubo J, Diffenderfer KE, Hishida Y, Modesto V, Dargitz CT, Feiring R, Zhao C, Aguirre A, McGarry TJ, Matsui H, Li H, Reyna J, Rao F, O'Connor DT, Yeo GW, Evans SM, Chi NC, Jepsen K, Nariai N, Müller F-J, Goldstein LSB, Izpisua Belmonte JC, Adler E, Loring JF, Berggren WT, D'Antonio-Chronowska A, Smith EN, Frazer KA. iPSCORE: a resource of 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. Stem Cell Rep 2017;8:1086–100. https://doi.org/10.1016/j.stemcr.2017.03.012.

Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine learning methods for quantitative radiomic biomarkers. Sci Rep 2015;5:13087.

Pashos EE, Park Y, Wang X, Raghavan A, Yang W, Abbey D, Peters DT, Arbelaez J, Hernandez M, Kuperwasser N, Li W, Lian Z, Liu Y, Lv W, Lytle-Gabbin SL, Marchadier DH, Rogov P, Shi J, Slovik KJ, Stylianou IM, Wang L, Yan R, Zhang X, Kathiresan S, Duncan SA, Mikkelsen TS, Morrisey EE, Rader DJ, Brown CD, Musunuru K. Large, diverse population cohorts of hiPSCs and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. Cell Stem Cell 2017;20:558–70.e10. https://doi.org/10.1016/j.stem.2017.03.017.

Paterson TS, Bangs AL. Method and apparatus for conducting linked simulation operations utilizing a computer-based system model. 2006.

Petrone J. Estonia invests €5M to genotype 100K people in 2018 as part of personalized medicine project | GenomeWeb. GenomeWeb; 2017.

Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, Ramalingam N, Sun G, Thu M, Norris M, Lebofsky R, Toppani D, Kemp DW, Wong M, Clerkson B, Jones BN, Wu S, Knutsson L, Alvarado B, Wang J, Weaver LS, May AP, Jones RC, Unger MA, Kriegstein AR, West JAA. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat Biotechnol 2014;32:1053–8. https://doi.org/10.1038/nbt.2967.

Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J. Massively multiplex single-cell Hi-C. Nat Methods 2017;14:263–6. https://doi.org/10.1038/nmeth.4155.

Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput 2017;29:2352–449. https://doi.org/10.1162/NECO_a_00990.

Reardon S. Giant study poses DNA data-sharing dilemma. Nat News 2015;525:16. https://doi.org/10.1038/525016a.

Reinsel D, Gantz J, Rydning J. Data age 2025: the evolution of data to life-critical don't focus on big data; focus on the data that's big. 2017.

Reker D, Rodrigues T, Schneider P, Schneider G. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proc Natl Acad Sci USA 2014;111:4067–72. https://doi.org/10.1073/pnas.1320001111.

Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. The human cell atlas: from vision to reality. Nat News 2017;550:451. https://doi.org/10.1038/550451a.

Sabah D. Scanning genomes, Turkish scientists look for cures, answers for longevity - Daily Sabah. Dly. Sabah Turk; 2018.

Samad S. Unlocking the power of the genome. In: William Blair growth stock conference, June 13, 2018. 2018.

Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP. Computational solutions to large-scale data management and analysis. Nat Rev Genet 2010;11:647–57. https://doi.org/10.1038/nrg2857.

Schaller RR. Moore's law: past, present and future. IEEE Spectr 1997;34:52–9. https://doi.org/10.1109/6.591665.

Schett G, Elewaut D, McInnes IB, Dayer J-M, Neurath MF. How cytokine networks fuel inflammation: toward a cytokine-based disease taxonomy. Nat Med 2013;19:822–4. https://doi.org/10.1038/nm.3260.

Schmidt BJ, Casey FP, Paterson T, Chan JR. Alternate virtual populations elucidate the type I interferon signature predictive of the response to rituximab in rheumatoid arthritis. BMC Bioinform 2013;14:221. https://doi.org/10.1186/1471-2105-14-221.

Schneider P, Schneider G. De novo design at the edge of chaos. J Med Chem 2016;59:4077–86. https://doi.org/10.1021/acs.jmedchem.5b01849.

Schneider G. Automating drug discovery. Nat Rev Drug Discov 2018;17:97–113. https://doi.org/10.1038/nrd.2017.232.

Sengupta PP, Huang Y-M, Bansal M, Ashrafi A, Fisher M, Shameer K, Gall W, Dudley JT. Cognitive machine-learning algorithm for cardiac imaging: a pilot study for differentiating constrictive pericarditis from restrictive cardiomyopathy. Circ Cardiovasc Imaging 2016;9. https://doi.org/10.1161/CIRCIMAGING.115.004330.

Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin JZ, Nemesh J, Goldman M, McCarroll SA, Cepko CL, Regev A, Sanes JR. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. Cell 2016;166:1308–23.e30. https://doi.org/10.1016/j.cell.2016.07.054.

Shoda L, Kreuwel H, Gadkar K, Zheng Y, Whiting C, Atkinson M, Bluestone J, Mathis D, Young D, Ramanujan S. The Type 1 Diabetes PhysioLab platform: a validated physiologically based mathematical model of pathogenesis in the non-obese diabetic mouse. Clin Exp Immunol 2010;161:250–67. https://doi.org/10.1111/j.1365-2249.2010.04166.x.

Shultz MD. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs: Miniperspective. J. Med. Chem. 2018;62(4):1701–14.

Singh VK, Romani S, Torrents-Barrena J, Akram F, Pandey N, Sarker MMK. Classification of breast cancer molecular subtypes from their micro-texture in mammograms using a VGGNet-based convolutional neural network. In: Recent advances in artificial intelligence research and development: proceedings of the 20th international conference of the Catalan Association for artificial intelligence, Deltebre, Terres de L'Ebre, Spain, October 25–27, 2017. IOS Press; 2017. p. 76.

Sjögren E, Halldin MM, Stålberg O, Sundgren-Andersson AK. Preclinical characterization of three transient receptor potential vanilloid receptor 1 antagonists for early use in human intradermal microdose analgesic studies. Eur J Pain 2018;22:889–903. https://doi.org/10.1002/ejp.1175.

Spanjaard B, Junker JP. Methods for lineage tracing on the organism-wide level. Curr Opin Cell Biol 2017;49:16–21. https://doi.org/10.1016/j.ceb.2017.11.004.

Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS, Heitner S, Harte RA, Haeussler M, Guruvadoo L, Fujita PA, Eisenhart C, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ. The UCSC Genome Browser database: 2016 update. Nucleic Acids Res 2016;44:D717–25. https://doi.org/10.1093/nar/gkv1275.

[WWW Document]. STEMBANCC website. STEMBANCC; 2012. http://stembancc.org/index.php?option=com_content&view=featured&Itemid=101.

Streeter I, Harrison PW, Faulconbridge A, The HipSci Consortium, Flicek P, Parkinson H, Clarke L. The human-induced pluripotent stem cell initiative-data resources for cellular genetics. Nucleic Acids Res 2017;45:D691–7. https://doi.org/10.1093/nar/gkw928.

Sweet DJ. iPSCs meet GWAS: the NextGen consortium. Cell Stem Cell 2017;20:417–8. https://doi.org/10.1016/j.stem.2017.03.020.

Tan SS-L, Gao G, Koch S. Big data and analytics in healthcare. Methods Inf Med 2015;54:546–7. https://doi.org/10.3414/ME15-06-1001.

Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, Koch C, Zeng H. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci 2016;19:335–46. https://doi.org/10.1038/nn.4216.

The Cancer Genome Atlas Home Page. [WWW Document]. Cancer Genome Atlas - Natl. Cancer Inst.; n.d. https://cancergenome.nih.gov/.

Ting DSW, Cheung CY-L, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, Yeo IYS, Lee SY, Wong EYM, Sabanayagam C, Baskaran M, Ibrahim F, Tan NC, Finkelstein EA, Lamoureux EL, Wong IY, Bressler NM, Sivaprasad S, Varma R, Jonas JB, He MG, Cheng C-Y, Cheung GCM, Aung T, Hsu W, Lee ML, Wong TY. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017;318:2211–23. https://doi.org/10.1001/jama.2017.18152.

Toney LK, Vesselle HJ. Neural networks for nodal staging of non–small cell lung cancer with FDG PET and CT: importance of combining uptake values and sizes of nodes and primary tumor. Radiology 2014;270:91–8. https://doi.org/10.1148/radiol.13122427.

Turk-Browne NB. Functional interactions as big data in the human brain. Science 2013;342:580–4. https://doi.org/10.1126/science.1238409.

TÜSEB Projects. [WWW Document]; n.d. http://www.tuseb.gov.tr/tuseb-projeler.

Usanov DL, Chan AI, Maianti JP, Liu DR. Second-generation DNA-templated macrocycle libraries for the discovery of bioactive small molecules. Nat Chem 2018. https://doi.org/10.1038/s41557-018-0033-8.

Vandenberghe ME, Scott MLJ, Scorer PW, Söderberg M, Balcerzak D, Barker C. Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. Sci Rep 2017;7:45938. https://doi.org/10.1038/srep45938.

Vapnik V. The nature of statistical learning theory. Springer Science & Business Media; 2013.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli

S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. Science 2001;291:1304–51. https://doi.org/10.1126/science.1058040.

Venter JC, Smith HO, Adams MD. The sequence of the human genome. Clin Chem 2015;61:1207–8. https://doi.org/10.1373/clinchem.2014.237016.

Vertès AA, Dowden NJ. History of monoclonal antibodies and lessons for the development of stem cell therapeutics. In: Stem cells in regenerative medicine. Vertès AA, Qureshi N, Caplan AI, Babiss L, (Eds.), Wiley-Blackwell; 2016. p. 665–92. https://doi.org/10.1002/9781118846193.ch33.

Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S, Jardine L, Dixon D, Stephenson E, Nilsson E, Grundberg I, McDonald D, Filby A, Li W, De Jager PL, Rozenblatt-Rosen O, Lane AA, Haniffa M, Regev A, Hacohen N. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 2017;356. https://doi.org/10.1126/science.aah4573.

Vitak SA, Torkenczy KA, Rosenkrantz JL, Fields AJ, Christiansen L, Wong MH, Carbone L, Steemers FJ, Adey A. Sequencing thousands of single-cell genomes with combinatorial indexing. Nat Methods 2017;14:302–8. https://doi.org/10.1038/nmeth.4154.

Wagenblast E, Soto M, Gutiérrez-Ángel S, Hartl CA, Gable AL, Maceli AR, Erard N, Williams AM, Kim SY, Dickopf S, Harrell JC, Smith AD, Perou CM, Wilkinson JE, Hannon GJ, Knott SRV. A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. Nature 2015;520:358–62. https://doi.org/10.1038/nature14403.

Waldrop MM. The chips are down for Moore's law. Nat News 2016;530:144. https://doi.org/10.1038/530144a.

Wang Y-K, Chen CS. Cell adhesion and mechanical stimulation in the regulation of mesenchymal stem cell differentiation. J Cell Mol Med 2013;17:823–32. https://doi.org/10.1111/jcmm.12061.

Williams DA, Thrasher AJ. Concise review: lessons learned from clinical trials of gene therapy in monogenic immunodeficiency diseases. Stem Cells Transl Med 2014;3:636–42. https://doi.org/10.5966/sctm.2013-0206.

Wong N. Stem cells meet AI in quest to mass-produce key disease fighters. Bloomberg; 2017.

Wu Y-L, Engl W, Hu B, Cai P, Leow WR, Tan NS, Lim CT, Chen X. Nanomechanically visualizing drug–cell interaction at the early stage of chemotherapy. ACS Nano 2017;11:6996–7005. https://doi.org/10.1021/acsnano.7b02376.

Wu C, Espinoza DA, Koelle SJ, Potter EL, Lu R, Li B, Yang D, Fan X, Donahue RE, Roederer M, Dunbar CE. Geographic clonal tracking in macaques provides insights into HSPC migration and differentiation. J Exp Med 2018;215:217–32. https://doi.org/10.1084/jem.20171341.

X-Chem. [WWW Document]; n.d. http://www.x-chemrx.com/.

Yala A, Barzilay R, Salama L, Griffin M, Sollender G, Bardia A, Lehman C, Buckley JM, Coopey SB, Polubriaginof F, Garber JE, Smith BL, Gadd MA, Specht MC, Gudewicz TM, Guidi AJ, Taghian A, Hughes KS. Using machine learning to parse breast pathology reports. Breast Cancer Res Treat 2017;161:203–11. https://doi.org/10.1007/s10549-016-4035-1.

Yassin NI, Omran S, El Houby EM, Allam H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. Comput Methods Programs Biomed 2018;156:25–45. https://doi.org/10.1016/j.cmpb.2017.12.012.

Zanin M, Papo D, Sousa PA, Menasalvas E, Nicchi A, Kubik E, Boccaletti S. Combining complex networks and data mining: why and how. Phys Rep 2016;635:1–44. https://doi.org/10.1016/j.physrep.2016.04.005.

Zayed H. The Qatar genome project: translation of whole-genome sequencing into clinical practice. Int J Clin Pract 2016;70:832–4. https://doi.org/10.1111/ijcp.12871.

Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 2015;347:1138–42. https://doi.org/10.1126/science.aaa1934.

Zhang F, Christiansen L, Thomas J, Pokholok D, Jackson R, Morrell N, Zhao Y, Wiley M, Welch E, Jaeger E, Granat A, Norberg SJ, Halpern A, C Rogert M, Ronaghi M, Shendure J, Gormley N, Gunderson KL, Steemers FJ. Haplotype phasing of whole human genomes using bead-based barcode partitioning in a single tube. Nat Biotechnol 2017. https://doi.org/10.1038/nbt.3897.

Zhou X, Lei L, Liu J, Halu A, Zhang Y, Li B, Guo Z, Liu G, Sun C, Loscalzo J, Sharma A, Wang Z. A systems approach to refine disease taxonomy by integrating phenotypic and molecular networks. EBioMedicine 2018;31:79–91. https://doi.org/10.1016/j.ebiom.2018.04.002.

Zürich ETH. SPiDER. [WWW Document]; n.d. http://www.cadd.ethz.ch/software/spider.html.