Population-scale sequencing and the future of genomic medicine

Learning from past and present efforts

Talitha Dubow, Sonja Marjanovic

RAND EUROPE

For more information on this publication, visit www.rand.org/t/RR1520

Published by the RAND Corporation, Santa Monica, Calif., and Cambridge, UK

Preface

Genomic medicine, as a field, has the potential to change the way we prevent, manage and treat disease. However, the routine implementation of genomic medicine in clinical care remains a future prospect. This paper provides a reflection on the variety of population-scale genome-sequencing initiatives that have emerged over the past two decades and examines their social implications. We analyse the progress these initiatives have made, both in terms of their scientific, technological and biomedical contributions, and in terms of their influence on the institutions that govern science and innovation more widely. Based on our analysis, we identify five areas of action for future research and policy to consider.

RAND Europe is a not-for-profit organisation whose mission is to help improve policy and decisionmaking through research and analysis. RAND Europe's clients include European governments, institutions, NGOs and firms with a need for rigorous, independent, multidisciplinary analysis. This document has been peer-reviewed in accordance with RAND's quality assurance standards. For more information about this document please contact:

Dr Sonja Marjanovic
RAND Europe
Westbrook Centre, Milton Road,
Cambridge, CB4 1YG
Tel. +44 (1223) 353 329
smarjano@rand.org

Table of contents

Tables

Summary

Context and objectives

Genomic medicine, as a field, has the potential to change the way we prevent, manage and treat disease. However, the implementation of genomic medicine in routine clinical care – through 'personalised', 'precision' or 'stratified' medicine – remains a future prospect due to a range of scientific and social challenges. Frequently cited issues relate to test reliability and validity, cost-effectiveness, health system and workforce readiness, and regulatory and ethical concerns (McCarthy et al. 2013; Deloitte 2015; Milani et al. 2015; Manolio et al. 2015).

Paving the way for a genomic medicine era is in part being facilitated by the evolution of population-scale sequencing initiatives. A diversity of such efforts has emerged in response to scientific and technological advances in genomics over the past 20 years. They have varied aims and objectives, differ in scope and in scale of activities, and in management and governance arrangements. In this paper, we reflect on their variety and evolution, on the learning that they offer, and on implications for future research, policy and practice. We draw insights from national initiatives and international collaborations, both disease-centred and more general in their orientation, but all building on large-scale population-sequencing data. We highlight the social implications of scientific and technological progress in this transformative field, and particularly issues that influence how we might manage risk and reward in genomic medicine.

Methods

We conducted a scoping review, searching Google and Google Scholar for evidence on diverse population-scale sequencing initiatives and key wider literature on this topic (e.g. journal articles and reviews, websites of major initiatives, initiative reports, press releases and news articles). We complemented our search strategy with a snowballing approach. We do not claim to have profiled all population-scale sequencing initiatives that exist; through the initiatives we have reviewed, we have tried to represent the diversity that characterises the field. The 30 initiatives profiled in this study are listed in Table 1.

Below we highlight key insights gained. More detailed information and examples of initiatives associated with specific features and impacts are provided in the core report.

Key findings

Diversity in form and purpose

Population-based sequencing initiatives have diverse goals, but we have witnessed a general movement towards more clinically oriented efforts with time. Across the initiatives we profiled, objectives spanned advancing the knowledge base on genetic variation within and across populations; enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; the development of new tools and methods for genetic studies; capacity-building (human resource capacity, management and governance, infrastructure); catalysing translation and coordination across initiatives; and other aims (national security, surveillance, commercial services).

The plethora of opportunities that have been created by advances in genomic science are accompanied by a diversity of funders and partnerships, although the funding landscape is still dominated by public and third-sector players. This is in some ways reflective of the

state of the field and markets. The rationale for public investment or public-private partnership is partially based on the pre-competitive nature of the research conducted by many of the current initiatives, as well as on government efforts to place their countries at the forefront of genomic research and to help catalyse genomics-based life-science industries. Most projects take place within academic and research centres and in partnerships between them (nationally or internationally). Sequencing tasks are often outsourced to specialist firms (with BGI and Illumina dominating the provider market). Aside from this, only a small number of population-scale initiatives currently engage private-sector partners in the conduct of research.

Progress, achievements and impact

Over two decades of progress in this field have yielded numerous achievements and impacts – both direct and indirect. For example:

- Catalogues of population-specific genetic variation have enabled further research – namely, association studies that compare genetic and phenotypic traits in order to advance knowledge of the genetic basis of disease.
- Biobank-based projects are providing comprehensive, longitudinal datasets that could allow for a greater understanding of the interactions between genes, lifestyle and environmental exposures on phenotypic expression.
- Although still in its early days, progress in the clinical interpretation of genetic variation is starting to inform more effective disease management strategies as well as offering diagnoses to patients who suffer from previously undiagnosed conditions (e.g. rare developmental disorders).
- New sequencing and analysis techniques are opening up opportunities for further progress at pace and at scale.
- Diverse initiatives worldwide have made lasting contributions to the physical and human resource capacity that will allow for future research and innovation.
- Some studies have advanced our collective knowledge on patterns of human migration, divergence and evolution, as well as estimated rates of mutation in modern humans.

But these initiatives have also had a lasting influence on the institutions that govern science more widely. Examples include:

- *Transforming informed consent and research ethics practices and experimenting with new models of feeding data back to research participants*. Population-sequencing efforts, in particular those that are conducted as part of broader biobanking initiatives, have been at the centre of debates which have advanced group consent and community engagement processes in biomedical research, championed broad consent versus single-use consent principles, and introduced novel models for feeding back research findings. Some initiatives have opted for a full-feedback policy that would include incidental findings; others provide feedback on findings directly related to the core research aim only; and some have opted out of feedback provision (due to the absence of prospects for treatment, absence of informed consent on feedback provision issues, or lack of certainty on the finding implications).
- *Catalysing open access research practices and guidelines and driving debate over what is patentable matter*. Most initiatives have an open-access policy, for non-commercial, and in some cases commercial, uses of anonymised datasets. Some have ‘delayed data release provisions’ to enable researchers to publish their findings prior to making the data more widely available, and to enable a degree of competitive advantage. The principle of open data-sharing runs into more demanding challenges when the boundaries of what is considered pre-competitive and competitive research become more blurred, and when preserving data anonymity becomes higher risk. There is broad

consensus that primary sequence data should not be patentable, but some initiatives allow researchers to claim intellectual property rights on downstream discoveries, and initiatives with a more applied long-term drug discovery ambition place more emphasis on patenting.

- *Championing innovation in data management*. Within the wider social context in which these initiatives operate, fears of security and privacy breaches have spurred experimentation with data protection strategies bridging technological interventions (code-based systems including two-tiered or multi-level access systems to different types of data), regulatory policy and legal levers (e.g. vetting of researchers applying for access to pseudonymised but potentially identifiable information, access review bodies, requirements for physical presence at data storage premises), and community norms and behaviours (advisory groups, education on risk-minimising behaviours for research participants, emergency response plans). At the other end of the spectrum, one initiative considers it impossible (and unethical) to promise participants confidentiality and anonymity, considering the possibility of individual re-identification as high and at conflict with any consent given on a data confidentiality basis. In this model, participants engage with research only if they can demonstrate an in-depth understanding of the scientific and technological context and of the risks of engagement, and if they volunteer to disclose extensive genomic, phenotypic, clinical and lifestyle data with no confidentiality or privacy clause. Population-sequencing initiatives are at the forefront of debate on how to manage the inevitable tension between data security, privacy and ambitions for facilitating access to a wider range of actors who could help advance the translation of insights into genomic medicine practice.

Regarding the ultimate aim of many of these initiatives – that of translating research findings into new drugs and diagnostics and integrating them into genomic medicine services – larger-scale impacts are yet to accrue despite some promising examples of clinical change and transformation of practice (e.g. developments of drug candidates helped by Iceland's pioneering study; the potential for more comprehensive tools for the management of high-burden diseases in Qatar; integrating whole-exome sequencing for patients with rare, undiagnosed disorders in clinical practice in Estonia). Many of the most clinically oriented projects have also put enabling mechanisms in place in order to accelerate translational research (by building partnerships with industry and clinical services and by laying a conducive national landscape through training programmes for clinicians and investments in databases, registries and interoperable IT infrastructure).

A future research and policy agenda

It remains to be seen whether the various enabling mechanisms and achievements to date will deliver on their promise to bring genomics into routine clinical care. The regulatory, ethical, legal, scientific and socio-economic challenges to overcome remain substantial, but the implications of doing so are profound. Drawing on our analysis of the findings presented in this report, and our wider experience in science policy, we propose five key areas of action for researchers and policymakers to consider. These are likely to be important for building on current achievements and supporting future efforts:

1. *Scope for scaling-up international, interdisciplinary and cross-sector collaboration to enable clinically relevant sense-making from large amounts of distributed genotypic and phenotypic data.* This will require data-sharing and collaborative clinical interpretation that crosses disease, disciplinary and professional boundaries, and will call for new ways of designing studies and collaborating. Innovative means of study design will be needed to ensure sample representativeness and optimal collection for personalised medicine innovations.

2. *Engaging in new research on the implications of genomic interventions in a clinical setting.* This includes research on the health-economics of genomic medicine, capacity-building needs and implementation options, and implications for patient-clinician communications and liability management.
3. *Examining the likelihood of further changes in industry R&D models and in rationales for public and private investment in a genomic medicine era. In relation to this, there is a need to address public acceptability issues that accompany the joint pursuit of health and commercial interests.* This includes considering the implications of personalised medicine on blurred boundaries between pre-competitive and competitive research, new risk-reward calculations, and potentially changing market segments (e.g. will industry focus on targeted treatments for specific patient profiles across a disease life-cycle, or on specific disease states across multiple patient segments, as discussed in Chataway et al. 2012, 736).
4. *Consolidation of learning on the research ethics framework, based on the experiences of prior and current initiatives, and consideration of the legal framework needed for genomic medicine in practice.* Although research ethics challenges associated with data security and informed consent remain, and are subject to much debate, the legal framework for dealing with genomic medicine is far more nascent than the research ethics one. The recent landmark case of a woman who is suing a doctor for failing to disclose a family history of hereditary brain disease is illustrative of the legal framework challenges ahead.
5. *Wider evaluation and learning will also be central to accountable and effective progress in genomic medicine, for patient benefit.* This includes ex-post evaluation of completed efforts and evaluation in real-time of ongoing initiatives, to ensure both formative and summative learning and accountability for the investments made in this transformative field.

List of profiled initiatives

The 30 initiatives covered in this study are listed in Table 1 below, sorted according to the scope of their sample collection (either national or international):

Table 1. List of profiled initiatives

International
The Human Genome Diversity Project (HGDP)
The International HapMap Consortium
The Global Network of Personal Genome Projects (PGP)
The 1000 Genomes Project
The Human Heredity and Health in Africa (H3Africa) Initiative
The African Genome Variation Project
The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium
The International Cancer Genome Consortium (ICGC)
National
deCODE genetics (*Iceland*)
The Estonian Biobank / Estonian Genome Centre, University of Tartu (EGCUT)
The Singapore Genome Variation Project
Genome of the Netherlands (GoNL)
GenomeDenmark
The Faroe Genome Project (FarGen)
Cymru DNA Wales
The National Centre for Indigenous Genomics (NCIG) (*Australia*)
Kuwait legislation introducing mandatory DNA testing (no project name)
The Precision Medicine Initiative Cohort Program (*U.S.*)
SardiNIA
China Kadoorie Biobank (CKB)
UK Biobank
The Slim Initiative in Genomic Medicine for the Americas (SIGMA) (*Mexico*)
UK10K
The Deciphering Developmental Disorders (DDD) Study (*UK*)
Genomics England (The 100,000 Genomes Project)
A Weill Cornell Medical Study – Exome Sequencing Identifies Potential Risk Variants for Mendelian Disorders at High Prevalence in Qatar
The Saudi National Genome Program
The Belgium Medical Genomics Initiative (BeMGI)
The Initiative on Rare and Undiagnosed Diseases (*Japan*)
The National Centre for Excellence in Research in Parkinson's Disease (NCER-PD) (*Luxembourg*)

The impacts of population-scale sequencing initiatives

Direct impacts

Catalogues of population-specific genetic variation have enabled genotype-phenotype association studies to advance knowledge on the genetic basis of disease.

Biobank-based projects are providing comprehensive, longitudinal datasets that could allow for a greater understanding of the interactions between genes, lifestyle and environmental exposures on phenotypic expression.

Progress in the clinical interpretation of genetic variation is beginning to inform new diagnosis options and more effective disease management strategies (e.g. in the case of rare developmental disorders).

New sequencing and analysis techniques are opening up opportunities for further progress at pace and at scale.

Diverse population-sequencing initiatives are contributing to physical and human resource capacity building for future research and innovation.

Some studies have advanced our collective knowledge on patterns of human migration, divergence and evolution, as well as estimated rates of mutation in modern humans.

Indirect impacts

Population-scale sequencing initiatives have contributed to advances in informed consent and research ethics practices and introduced new models of feeding data back to research participants (e.g. they have informed group consent and community engagement processes, steered broad consent versus single-use consent principles, and spurred debate and experimentation in feedback practices).

Population-scale efforts are catalysing open access research practices and guidelines, and driving debate over what is patentable matter (e.g. open-access policy, delayed data release provisions, thresholds for patentable activity).

Diverse initiatives are championing innovation in data management to address the inevitable tensions between data security, privacy, and access for a wider range of actors who could help translate research findings into genomic medicine practice (e.g. technological code-based data protection solutions, regulatory policy and legal levers for privacy and data security, community norms-based regulation, and full disclosure volunteering agreements).

Acknowledgements

The authors would like to thank Professor Joanna Chataway and Dr Catriona Manville from RAND Europe for their quality assurance roles, as well as Celine Miani from RAND Europe for all her helpful suggestions throughout this project. This study was supported by RAND Europe internal resources and a three-month internship.

Abbreviations

CSF	Clinically Significant Finding
DAC	Data Access Committee
DNA	Deoxyribonucleic Acid
EGA	European Genome-phenome Archive
GWAS	Genome-Wide Association Study
IP	Intellectual Property
IPR	Intellectual Property Rights
NHGRI	National Human Genome Research Institute
NHS	National Health Service (UK)
NIH	National Institutes of Health (US)
R&D	Research and Development

Chapter One. Background and context

1.1. Introduction

Over the last two decades, we have witnessed rapid scientific and technological advances in genomics, which have underpinned the growth of population-scale sequencing initiatives globally. These initiatives are based on large-scale sequencing and analysis of DNA samples from particular human populations. They have diverse aims – from characterising genetic variation, to understanding the relationship between variants and their phenotypic expression, to harnessing disease-specific insights towards the development of new diagnostics and treatments for improved health outcomes. Although advances in genomics create new opportunities for improving population health, they also present new challenges for researchers, health services, firms and policymakers.

In this paper, we aim to capture key learning from a variety of population-scale sequencing efforts globally, in order to inform a future research, policy and practice agenda. We review some examples of the diversity of population-scale sequencing efforts that have emerged, and reflect on their aims and objectives, scope and scale, management and governance models, and impacts on the progress of genomics as a field, with a particular interest in informing genomic medicine futures. We focus on national initiatives and international collaborations, both disease-centred and more general in their orientation, but which build on large-scale population-sequencing data.

Such initiatives are as much social as they are technological, with social institutions and interactions shaping evolutionary pathways, progress and impacts, and co-evolving with the physical technologies themselves (Nelson & Sampat 2012, 31–54). They have emerged in response to scientific, technological and social drivers and have evolved to draw on a broader base of evidence relating to new scientific knowledge, patient-based data, healthcare systems priorities and operational capacities.

In the case of genomic medicine, advances in the science that guides the way research progresses require new arrangements between organisations and changes in the institutions that govern health innovation. As highlighted in Chataway et al. (2012, 736):

> …Institutional change is necessary because social and physical technologies emerge in the context of institutional structures, laws and norms. For instance, one of the institutions of fundamental importance to the pharmaceutical and biotechnology industries is IPR legislation and patent norms. Patent law has changed significantly over the past three decades, with a range of court decisions broadly expanding patentable subject matter.[1,2,3] This has influenced greatly the

1 Merges & Nelson (1994).
2 Jaffe & Lerner (2004).
3 Gaisser et al. (2009).

type of partnerships and collaborations that have been possible. Thus the three way dynamic between physical technologies, social technologies and general institutions sheds light on the particular configurations of technologies, organisations and institutions in particular areas of innovations.

Given the proliferation of population-scale sequencing efforts,[4] reflection on their variety, evolution, future prospects and associated systemic issues is timely and has substantial implications on how we might manage risk and reward for the future of genomic medicine. In this paper, we discuss these issues with the aim of highlighting the social implications of scientific and technological progress in this transformative field.

1.2. Over two decades of scientific and technological progress have created new opportunities

The sequencing of the first human genome (1990–2003) took 13 years of international collaboration and cost over £2 billion.[5] More than a decade of technological innovations have since driven down sequencing costs so that it now costs under £1,000 and can take as little as 24 hours to sequence a whole human genome (Middha et al. 2014, 1).[6] This sharp decline in time and cost demands has hinged primarily on the introduction of next-generation sequencing in 2008,[7] after which falling sequencing costs outpaced Moore's Law, the standard by which technological improvements are judged to be progressing well.[8] The growing feasibility of genome sequencing has resulted in a steady expansion of sequencing initiatives, both in terms of the number of initiatives and their scope of activities. Genome sequencing is no longer the exclusive domain of large international consortia of researchers like that which executed the Human Genome Project, but is now performed within some hospitals for certain patient profiles,[9] and is even offered as a service by consumer-orientated companies like 23andMe. In 2003, the Human Genome Project published the first human genome; Illumina now estimate that by 2017 1.6 million human genomes will have been sequenced worldwide.[10]

Growing attention to the potential of genome sequencing to enrich scientific knowledge on population diversity gained momentum soon after the launch of the Human Genome Project in 1990. For example, in 1991 the Human Genome Diversity Project (HGDP) was proposed as a study not of one composite genome (as was the case for the Human Genome Project) but of the range of human genetic diversity worldwide (Greely 2001, 222). Although the HGDP eventually failed to progress as expected (Greely 2001, 222–3), several other projects

4 Out of the many initiatives that have existed over the last 25 years, the present study focuses on a sample of 30, of which 1 was launched in the 1990–2000 period, 9 were launched in the 2000–2010 period, and 16 have been launched since 2010, or are still in the planning phases. The launch dates of the remaining four projects are unknown, but their study findings were first published in 2009, 2012, 2014 and 2015, respectively.

5 Genomics England: The 100,000 Genomes Project (2015).

6 Genomics England: The 100,000 Genomes Project (2015).

7 Next-generation sequencing (NGS), otherwise known as high-throughput or massively parallel sequencing, revolutionised sequencing methods by allowing millions of DNA fragments to be sequenced at the same time, whereas first-generation sequencing, known as Sanger sequencing, was only able to sequence one DNA fragment at a time. See Grada & Weinbrecht (2013).

8 National Human Genome Research Institute: DNA sequencing costs (2015).

9 See, for example, programmes offered at the Children's Hospital of Wisconsin (http://www.chw.org/medical-care/genetics-and-genomics-program/programs-and-services/rare-and-undiagnosed/whole-genome-sequencing/, last accessed 1 January 2016) and at Columbia University Medical Centre (http://newsroom.cumc.columbia.edu/blog/2015/10/22/precision-cancer-medicine-for-pediatric-patients/, last accessed 19 January 2016).

10 Regalado (2014).

both multinational (e.g. the International HapMap Consortium) and national (e.g. the Singapore Genome Variation Project) emerged to map genetic variation specific to different human populations across the globe. These early studies tended to have quite general research aims – namely, the investigation of genetic variation within and between human populations, in order to enable future studies with more targeted biomedical objectives.

Following these initial exploratory efforts we observe a shift in focus towards harnessing the advances of genomics for improved human health outcomes. More recent studies such as the China Kadoorie Biobank or the Deciphering Developmental Disorders Study have incorporated medical records and more detailed phenotypic information in addition to sequencing data, in order to investigate the genetic basis of disease within either a specific ethnic population(s) or a certain disease community. Early studies such as the HapMap project built up a publically available catalogue of human variation which, beginning in 2005, was used to perform large-scale genome-wide association studies (GWAS)[11] to scan for markers that might be associated with specific diseases (Milani et al. 2015, 190). The utility of the GWAS approach has been the subject of much debate due to a number of criticisms that focus both on the limitations of the study design and also on the assumptions made about the underlying biology of complex disease (Visscher et al. 2012, 7–24). However, discussions have tended to concede that, despite their limitations, genome-wide association studies have had a transformative effect on the field of genomics (McCarthy et al. 2013, 1–2), yielding valuable insights into the genetic basis of disease, some of which have had direct clinical utility, such as the development of multiple sclerosis therapies and in the diagnosis of diabetes sub-types (Visscher et al. 2012, 16–17, 19).

Advances in genetic sequencing, coupled with proteomic profiling technologies[12] and innovations in metabolomics,[13] have also created new opportunities for more personalised (stratified) medical approaches. This new branch of medicine emphasises the potential of tailoring treatments to particular types of individuals through a better understanding of the interactions between drugs, other types of prevention or treatment-based interventions and particular patient profiles (Chataway et al. 2012, 732–40). It is hoped that the advantages of such an approach will be enhanced disease prevention, delayed disease progression, better patient outcomes and fewer predictable side effects for patients, and ultimately more cost-effective practice. As such, population-scale genetic sequencing has been a fundamental driver behind efforts to establish new models of drug discovery.

1.3. Science has evolved faster than society's readiness to harness it

At present, scientific progress has outpaced the institutional advances needed to make genomic medicine a reality. To an extent, increased certainty in scientific feasibility is being accompanied with increased uncertainty

11 Genome-wide association studies use chip-based microarray technology to rapidly scan large numbers of DNA sequences in order to detect the presence of specific variants. By comparing patterns of genetic variation between disease-affected individuals and an unaffected control group, scientists are able to identify genetic risk factors for particular diseases. See National Human Genome Research Institute: Genome-Wide Association Studies (2015).

12 Proteomic profiling provides information about protein expression in different tissues and across different samples. See NCI Dictionary of Cancer Terms: Proteomic Profile (2016).

13 Metabolomics is the analysis of metabolites (the intermediate products of metabolic reactions) in biological systems. See Mamas et al. (2011, 6); Harris (n.d.).

in systemic readiness – i.e. the readiness of institutions and actors to access, assimilate, translate, integrate and use new knowledge for further, targeted downstream R&D, effective health innovation and healthcare delivery models. This is not to say that progress in systemic readiness hasn't been made. In fact, and as we will discuss in Section 3.4, population-scale sequencing efforts have contributed to significant progress in the institutions governing informed consent and research ethics, data management, as well as influenced debate as to what is patentable matter. However, the implementation of genomic medicine in *routine* clinical care – through 'personalised', 'precision' or 'stratified' medicine – remains a future prospect. There is broad agreement on the reasons for this. Frequently cited challenges include: i) the lack of evidence of the cost-effectiveness of genomic interventions; ii) the lack of an evidentiary framework that would determine the validity and utility of genomic tests; iii) the need for standard-setting and regulation; iv) the high costs of investment and potentially unattractive rates of return; v) the probabilistic nature of interpreting genomic data; vi) concerns over consent, privacy and other legal, ethical and psychosocial issues; vii) and the unpreparedness of healthcare systems and workforces to integrate genomic technologies into the clinical workflow.[14]

More recently, various initiatives are trying to tackle some of these challenges and address a diversity of systemic issues, such as those associated with ethics, standards, regulation, intellectual property and innovative trial design for genomic medicine. For example, in the United Kingdom, Genomics England is delivering an ambitious effort to improve the use of genomics within the NHS through a coordinated approach to supporting scientific advances, NHS transformation and improved clinical services.[15] It is also aiming to catalyse a 'genomics industry' in the UK, to stimulate the translation of research into products for patient benefit.[16] Genomic Medicine Centres are being established by NHS England to help enroll participants onto the program, to accept NHS referrals and ensure appropriate sample and phenotypic information collection.[17] Genomes will be linked to phenotypic research data and clinical records to form a research data set, as well as to provide feedback to the individual patients via their clinicians.[18] The program is designed to leave a lasting legacy for patients, the NHS and the UK economy.

In Estonia, broad systemic changes undertaken following the successful creation of the Estonian Biobank provide further examples of efforts to tackle the challenges associated with implementing genomic medicine into routine care. The Biobank, set up in 1999, is maintained by the state, regulated by national legislation and managed by the Estonian Genome Center (EGCUT) (Milani et al. 2015, 188–90). Various infrastructural developments have subsequently laid the groundwork for the eventual implementation of genomic medicine nationwide: in 2002 Estonia introduced compulsory national electronic identification cards; in 2010 a digital prescription service was launched; and the progressive linking of health-related databases including those of the Estonian Biobank, various national healthcare providers, hospitals and other registries has built up the Estonian National Health Information System, creating individual electronic health records (EHRs) for every patient (Leitsalu, et al. 2015, 103). Between 2015

14 See McCarthy et al. (2013); Deloitte (2015); Milani et al. (2015); Manolio et al. (2015).

15 Genomics England News (2014).

16 Genomics England: The 100,000 Genomes Project (2015).

17 Genomics England: Frequently asked questions (2015).

18 Genomics England: The process (2015).

and 2018 a government-funded personalised medicine pilot project aims to produce and integrate individual-level genomic datasets with corresponding risk-prediction and drug-response analyses for every Biobank participant into their EHRs for use by clinicians (Leitsalu et al. 2015, 105; Milani et al. 2015, 196–7).

There have also been increased efforts to integrate research projects into large national or international consortia, and to respond to the need for greater sharing of evidence, tools and lessons learned for more efficient and effective translation into healthcare contexts (Manolio et al. 2015, 1–8). For example, between 2008 and 2013 the EU-funded ENGAGE project formed a consortium of 24 leading research organisations and biotechnology and pharmaceutical companies. In 2014 the US National Human Genome Research Institute (NHGRI) and the U.S. National Academy of Medicine brought together 90 leaders in genomic medicine from 26 countries for a Global Leaders in Genomic Medicine Symposium (Manolio et al. 2015, 1). And in the UK, Genomics England Clinical Interpretation Partnerships (GECIPs) represent pre-competitive research networks that bring together multidisciplinary groups of researchers, NHS professionals and trainees, and industry partners to help ensure translation and impact from the 100,000 Genomes Project through enhanced clinical interpretation of the 100,000 Genomes dataset. These are just a few examples of the many networks being established to facilitate communication between individual projects and diverse stakeholders that would otherwise be working in relative isolation.

Chapter Two. Methods

We conducted a scoping review to understand the diversity of population-scale sequencing initiatives and the progress of this field of research.[19] We searched Google and Google Scholar for evidence on diverse population-scale sequencing initiatives and key wider literature on this topic (e.g. journal articles and reviews, websites of major initiatives, press releases and news articles). We complemented our search strategy with a snowballing approach.

Our search comprised three strands. The first was a search of peer-reviewed academic literature such as articles, commentaries, editorials and reviews. The second strand was a grey literature review, which searched for press releases, initiative reports, policy documents and websites. A list of the search terms used is detailed in Table 6 in the Appendix. The third strand represents the snowballing part of our search strategy and was an iterative process; it was a search of both the academic and grey literature available, and was performed using the names of individual genome-sequencing initiatives as the search terms, each time that we came across the name of a new initiative. This third snowballing strand proved crucial to the gathering of data on individual initiatives, many of which were referenced only in the grey literature. In the event that our search generated a long list of hits, we sorted by relevance for strands one and two of our search, and by date if we were searching the name of a specific initiative as part of the third search strand. In both cases we looked at the first three pages of search hits.

Together, the combined keyword-based search and snowballing approach led to a core set of 105 relevant articles, reports and policy documents and 85 website pages. The search was conducted between November 2015 and January 2016. We aimed for a non-restrictive approach with key eligibility criteria including a population-type focus (e.g. disease specific or national/ethnic) and direct engagement with sequencing activities. Some initiatives also focused on supportive activities (capacity-building, ethics research, coordination, translation, etc.).

We focused on key initiatives for which there was sufficient publically available information. We do not claim to have provided a list of *all* population-scale sequencing initiatives that exist; through the initiatives we have reviewed, we have tried to represent the diversity that characterises this field. For example, some of the initiatives we came across are not included in this review because they were only launched very recently (which is illustrative of the pace of developments in the field),[20] or because we could not access information on them (i.e. they were established a long time ago but may have been terminated or haven't progressed very far).[21]

We have considered a sample of initiatives representing the *diversity* of projects and

19 The definition of a scoping review is one that 'aims to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available' (Arksey & O'Malley 2005, 19–32), with the basic approach following that of a systematic review: defining the research question, identifying relevant references and screening references for eligibility for inclusion (Centre for Reviews and Dissemination 2009).

20 See, for example, the very recently launched Genome Korea in Ulsan initiative (Burbidge 2015d). Another example is the Israeli plans for a population-scale genetic database (Taylor 2015).

21 For example, little information in English is available about a previously launched Korean Genome Project (see http://koreangenome.org/index.php/Main_Page, as of 12 January 2016).

reflecting the *global span* of research.[22] Those included are both national, country-based efforts and international collaborations, and both disease-oriented and more general. The key emphasis in this study has been on projects that are based on population-scale sequencing data and which actively engage in sequencing activities. This means that we have excluded projects with a purely translational or clinical focus (i.e. those which build on already completed genetic sequencing studies rather than carrying out sequencing work themselves, e.g. the Implementing Genomics in Practice (IGNITE) Network[23]) or which do carry out upstream genomic research but not with a focus on *population*-scale genetic variation data (e.g. the Structural Genomics Consortium,[24] the Roadmap Epigenomics Project[25]). Whereas such initiatives in themselves offer important learning potential, they raise related but distinct challenges and opportunities which are not the focus of this paper (for related studies, see Chataway et al. 2012 and Morgan Jones et al. 2014).

The 30 initiatives covered in this study are:[26]

- The Human Genome Diversity Project (HGDP)
- The International HapMap Consortium
- The Global Network of Personal Genome Projects (PGP)
- The 1000 Genomes Project
- The Human Heredity and Health in Africa Initiative (H3Africa)
- The African Genome Variation Project
- The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium
- The International Cancer Genome Consortium (ICGC)
- deCODE genetics
- The Estonian Biobank / Estonian Genome Centre, University of Tartu (EGCUT)
- UK Biobank
- The Singapore Genome Variation Project
- Genome of the Netherlands (GoNL)
- GenomeDenmark
- The Faroe Genome Project (FarGen)
- Cymru DNA Wales
- The National Centre for Indigenous Genomics (NCIG)
- Kuwait legislation introducing mandatory DNA testing (no project name)
- The Precision Medicine Initiative Cohort Program
- SardiNIA
- China Kadoorie Biobank (CKB)
- The Slim Initiative in Genomic Medicine for the Americas (SIGMA)
- UK10K
- The Deciphering Developmental Disorders (DDD) Study
- Genomics England (The 100,000 Genomes Project)
- A Weill Cornell Medical Study – Exome Sequencing Identifies Potential Risk Variants for Mendelian Disorders at High Prevalence in Qatar
- The Saudi National Genome Program
- The Belgium Medical Genomics Initiative (BeMGI)
- The Initiative on Rare and Undiagnosed Diseases (IRUD)
- The National Centre for Excellence in Research in Parkinson's Disease (NCER-PD)

22 We have tried to avoid including a disproportionate number of US- and UK-based projects, even though these might have the highest profile.

23 IGNITE (homepage) (2016).

24 For further information see Morgan Jones et al. (2014).

25 Roadmap Epigenomics Project (homepage) (2010).

26 Please see Appendix for more details on each.

Chapter Three. Results

In this chapter we reflect on the diversity of population-based sequencing initiatives that we have analysed, in light of their aims and objectives (3.1); scope and scale of activities (3.2); governance and partnership models (3.3); and policies on intellectual property (IP), data access and management and participant consent (3.4), based on publically available information. Each of these discussions is based on a cross-analysis of the 30 initiatives included in the study. More detail on each issue is presented in Tables 2–5 in the Appendix. In Chapter 4 we reflect on the achievements and evolving impacts of these initiatives, and consider future directions for a research, policy and service transformation agenda.

3.1. The diversity of aims and objectives across initiatives

Population-based sequencing initiatives vary in their aims and objectives (as summarized in Box 1), but we have witnessed a general movement towards more clinically oriented efforts with time. We discuss this further below, drawing on our analysis of the literature.

Early initiatives focused more broadly on identifying patterns of genetic variation – for example, the Human Genome Diversity Project was primarily interested in understanding the history of early human evolution and migration (Cavalli-Sforza 2005, 333). Cultural or anthropological interests have been of far lesser interest for subsequent studies, bar a few exceptions. One such exception is the recently established Australian National Centre for Indigenous Genomics (NCIG) which will 'support a broad range of anthropology, historical and genealogy studies' alongside research that it is hoped will have clinical implications for the indigenous Australian population.[27] As another example, Cymru DNA Wales, a commercial sequencing company which follows the model established by US company 23andMe, has been publicised as a research project exploring the history of the Welsh people.[28] However, it is principally a consumer-orientated, commercial venture that markets itself to people who wish to receive information on their ancestry as well as their carrier-status for red hair, blue eyes and

Box 1. A summary of the key aims and objectives across sequencing initiatives

1. Advancing the knowledge base on genetic variation within and across populations;
2. Enriching disease-specific, clinically relevant insights relating to diagnosis, risk prediction or treatment;
3. Development of new tools and methods for genetic studies;
4. Human resource capacity building for research;
5. Management and governance capacity building;
6. Physical capacity building;
7. Catalysing translation and coordination;
8. Furthering the cultural and anthropological knowledge base;
9. Other (e.g. national security, surveillance, commercial).

27 Indigenous Genomics (2015).

28 BBC News 2014.

baldness.[29] One example of an initiative which, although without cultural or anthropological aims, differs substantially from the rest by not aiming to further biomedical research, is the recent law enacted by the government of Kuwait that will introduce mandatory DNA testing of all citizens and foreign residents as a national security measure.[30]

In general, recent efforts have placed more emphasis on research of biomedical relevance with potential clinical applications, The vast majority of initiatives aim to advance the knowledge base on genetic mutations, either by identifying variants and/or by finding associations between genetic variants and phenotypic characteristics (see Table 2 in the Appendix for more detail). The two aims are closely related and interdependent. For example, the earlier cataloguing of population-based genetic variation by projects such as the HapMap and 1000 Genomes has enabled the development of microarrays for genome-wide association studies (Buchanan et al. 2012, 289–90).

Increased understanding of genetic variation at population levels, as well as the development of new tools for conducting genetic analyses, has also underpinned the proliferation of initiatives focused on specific disease areas. For example: SardiNIA studied age-related diseases in the Sardinian population (Pilia et al. 2006); GenomeDenmark has a demonstration project focused on cancer research;[31] the UK-based Deciphering Developmental Disorders Study looks at patients with undiagnosed developmental disorders;[32] similarly, Japan has launched an Initiative on Rare and Undiagnosed Diseases;[33] Luxembourg has recently established a National Centre for Excellence in Research in Parkinson's Disease;[34] and Genomics England is considering patients with cancer and a variety of rare diseases, infectious diseases and antimicrobial resistance areas (Caulfield, M. et al. 2015, 11–22, 33).[35]

Some initiatives have emerged to consider health conditions that are particularly pronounced in specific populations. Projects such as the Saudi National Genome Program and the Qatar-based Weill Cornell Medical Study have been launched with a view to tackling the two populations' high burden of severe inherited disorders, which are due to the high degree of consanguineous marriage within the two respective populations.[36] This will be discussed further in Section 3.2.

Other initiatives have tried to address global health disparities and research-capacity gaps. Projects such as The Slim Initiative in Genomic Medicine for the Americas (SIGMA),[37] the Human Heredity and Health in Africa Initiative (H3Africa)[38] and the African Genome Variation Project[39] have made it their explicit aim to address inequities in global health research. In line with their overarching mission, these initiatives share two core interrelated objectives: 1) to characterise genetic variation among particular populations (Latin American and Sub-Saharan respectively) in order to enable the

29 BBC News 2014.
30 Aljazeera (2015).
31 GenomeDenmark: About (2015).
32 Deciphering Developmental Disorders Study: What is the DDD Study? (2015).
33 Otake (2015).
34 Government of Luxembourg (2015).
35 Genomics England: The 100,000 Genomes Project (2015).
36 See Saudi Human Genome Program: Introduction (2015); Weill Cornell Medical College in Qatar (2014).
37 Carlos Slim Center for Health Research at the Broad Institute (2015).
38 H3Africa: Vision of the H3Africa Initiative (2013).
39 African Genome Variation Project (2015).

design of genome-wide association studies in the region; and 2) to build in-country and in-region research capacity in those countries in order to support the long-term improvement of local public health outcomes. While projects such as H3Africa, SIGMA and the African Genome Variation Project carry out capacity-building activities through the provision of administrative support, scientific consultation and advanced training to researchers, others focus on establishing the requisite physical infrastructure. For example, the Saudi National Genome Project aims to build a network of 15 genome sequencing laboratories across the country,[40] and GenomeDenmark has installed sequencing facilities and several super-computers for large-scale sequencing and analysis tasks.[41] Many other projects aim to build capacity in terms of research management and governance; for example, several initiatives including H3Africa (H3Africa Consortium 2014, 1348), UK10K (Kaye 2014), the Deciphering Developmental Disorders Study,[42] the Belgian Medical Genomics Initiative[43] and Genomics England[44] have made the development of ethical guidelines and protocol regarding genetic studies central to their research objectives.

Some studies have also produced new tools and methods for genetic research, most often in the form of population-specific reference panels for genome-wide association studies (as provided, for example, by the HapMap (International HapMap Consortium 2005), 1000 Genomes Project (1000 Genomes Consortium 2015), African Genome Variation Project (Gurdasani et al. 2015, 330), Singapore Genome Variation Project (Teo et al. 2009), Genome of the Netherlands (Genome of the Netherlands Consortium 2014), UK10K (UK10K Consortium 2015), and others). Others have explicitly aimed to contribute to methodological innovations in the field. One notable example is the CHARGE Consortium, which was formed in 2007 to facilitate genome-wide association study meta-analyses so that findings related to cardiovascular and ageing-related conditions could be reliably validated across multiple large-scale longitudinal cohort studies.[45] CHARGE is also an example of the many networks of researchers and projects that have recently been established with an explicit aim to catalyse translational research efforts through the collaboration and coordination of distributed activities – the Belgian Medical Genomics Initiative[46] and the International Cancer Genome Consortium are further examples.[47] As is the case with these last two examples, as well as with others such as the Estonian Genome Centre and Genomics England, diverse new sequencing projects continue to take place (either disease specific or population specific), but these are increasingly undertaken in combination with translation efforts under the same initiative, rather than in isolation.

3.2. Variety in scope and scale

The scope and scale of genome-sequencing initiatives varies considerably in terms of their geographical reach, population diversity and sample size, disease specific or general focus, types of data covered and scope of activities. In general, we have witnessed a move towards

40 Saudi Human Genome Program (homepage) (2015).
41 GenomeDenmark: Research Equipment (2015).
42 Deciphering Developmental Disorders Study: Ethics Research (2015).
43 BeMGI: Genomics and Society WG (2015).
44 Genomics England: Patient and Public Involvement (2015).
45 Charge Consortium (2015).
46 BeMGI: About BeMGI (2015).
47 International Cancer Genome Consortium: About (2015).

larger-scale efforts (partially driven by declines in sequencing costs), samples that combine more diverse types of data (genotypic and phenotypic), and, for some diseases, a growth in international collaborative efforts.

The size of initiatives will be driven by a range of concerns, including aims and objectives, as well as resource availability. Projects such as the Qatar-based Weill Cornell Medical Study (Rodriguez-Flores et al. 2013) and the Singapore Genome Variation Project (Teo et al. 2009) sequenced population samples as small as 100 and 268 individuals, respectively, whereas President Obama's Precision Medicine Initiative Cohort Program, announced in early 2015, aims to sequence the genomes of 1 million Americans.[48] In the UK, the 100,000 Genomes Project is focusing on 100,000 genomes of patients with rare diseases and their families, and patients with cancer,[49] and the Human Heredity and Health in Africa Initiative aims to sequence 50,000–75,000 genomes (H3Africa Consortium 2014, 1347) (see Table 3 in the Appendix). The sharp fall in sequencing costs in the last 15 years has made it more cost-effective to sequence larger numbers of genomes, enabling larger-scale efforts.

In terms of geographic reach, a number of initiatives have been driven by the genetic specificity of individual populations or ethnic groups and their links to particular diseases and phenotypic traits (Lu et al. 2014.). These (generally clinically orientated) population-based sequencing initiatives tend to focus their research on diseases at high prevalence within their populations. They do so with the aim of uncovering and understanding population-specific genetic variants for conditions that affect diverse populations globally (e.g. various types of cancers), as well as for conditions which are particularly prevalent within certain populations (e.g. recessive inherited disorders). For example, the Weill Cornell Medical Study analysed the genomes of 100 Qatari nationals in order to identify variants associated with Mendelian diseases such as cystic fibrosis, sickle-cell anaemia and muscular dystrophy, which are at high prevalence in Qatar because of the frequency of consanguineous marriages.[50] In China, where recent socio-economic changes have resulted in a significant shift in disease patterns and chronic diseases are now estimated to account for over 80 per cent of the population's deaths, the China Kadoorie Biobank was launched in order to enable studies of the aetiology of common chronic diseases such as stroke, ischaemic heart disease (IHD), diabetes, cancer and chronic obstructive pulmonary disease (COPD).[51] The Slim Initiative for Genomic Medicine in the Americas focuses research on cancer, diabetes and kidney disease in individuals of Latin American descent, and has uncovered a common genetic variant which predisposes Latin American populations to type 2 diabetes.[52]

Recently, there have also been increased efforts to pool resources, skills and knowledge and to enable greater efficiency and higher impact from genetic studies though international, networked approaches and larger population samples, specifically as related to disease-centred efforts. For example, the International Cancer Genome Consortium was formed in 2007 to maximise the efficiency of research into over 50 cancer types; so far it has brought together 78 projects from 16 countries, and over 27,000 DNA samples

48 PMI Cohort Program (2015).

49 Genomics England: The 100,000 Genomes Project (2015).

50 Weill Cornell Medical College in Qatar (2014).

51 China Kadoorie Biobank: About the study (2015).

52 Carlos Slim Center for Health Research at the Broad Institute (2015).

have been collected for sequencing.[53] Other key international efforts include: the CHARGE Consortium which, as discussed, brings together multiple prospective cohort studies in order to facilitate replication opportunities for the reliable identification of potential risk variants (Psaty et al. 2009); the Global Network of Personal Genome Projects, which has gained collaborators in Canada, Austria and the UK since launching a Harvard-based pilot study of ten participants;[54] and the Human Heredity and Health in Africa Initiative which aims to support research capacity for intracontinental collaborations between African scientists (H3Africa Consortium 2014).

In terms of data scope, more recent studies have incorporated increasingly extensive phenotypic data into their research. Earlier projects such as the Human Genome Diversity Project (Cavalli-Sforza 2005, 335), HapMap[55] and 1000 Genomes Project,[56] which aimed to characterise genetic variation globally, only linked very basic data such as sex and the geographic/population origin of samples to DNA data. However, as sequencing initiatives sought to understand the relationship between genetic variants and their phenotypic expression more closely, detailed data pertaining to lifestyle, environmental exposures, phenotypic profile, medical history, family history, and even different tissue samples have been more comprehensively collected. Biobanks provide a good example: the Estonian Biobank (Leitsalu et al. 2015, 99–100), China Kadoorie Biobank[57] and UK Biobank[58] collect extensive baseline phenotypic data, but also remain engaged with participants over time to ensure updates or supplements with additional types of data. They also seek consent from participants to link up their data with other databases such as death registries and hospital and insurance records in order to regularly update participants' profiles.[59] Comprehensive longitudinal data such as this is crucial for understanding the interactions between genes and environmental/lifestyle factors on the development of disease, and is expected to help enhance risk prediction in a clinical setting.

In terms of the scope of activities carried out, most initiatives focus:

- Primarily on research conduct (increasingly – though not exclusively – with clinical orientations, as discussed in Section 4.1); but some also integrate
- Capacity-building activities (e.g. PGP,[60] H3Africa (Adoga et al. 2014, 1), African Genome Variation Project,[61] SIGMA,[62] Saudi National Genome Program,[63] UK10K (Kaye 2014), UK Biobank,[64] the DDD Study,[65] BeMGI,[66] GenomeDenmark,[67] and Genomics England[68]);

53 International Cancer Genome Consortium: About (2015).
54 Personal Genome Project: Global Network (2015).
55 International HapMap Project: How are ethical issues being addressed? (2015).
56 1000 Genomes: About (2012).
57 See China Kadoorie Biobank: Study Design (2015); China Kadoorie Biobank: Long-term Follow-up (2015).
58 UK Biobank: About (2015).
59 See Leitsalu et al. (2015, 100–1); China Kadoorie Biobank: Long-term Follow-up (2015); and Sudlow et al. (2015, 3).
60 See, for example, Drmanac et al. (2010); Ball et al. (2012); Brock et al. (2012).
61 African Genome Variation Project (2015).
62 Carlos Slim Center for Health Research at the Broad Institute (2015).
63 Saudi Human Genome Program (homepage) (2015).
64 UK Biobank: Public consultation (2015).
65 Deciphering Developmental Disorders Study: Ethics Research (2015)
66 BeMGI: Genomics and Society WG (2015).
67 GenomeDenmark: Research Equipment (2015).
68 Genomics England: Patient and Public Involvement (2015).

- Coordination (BeMGI,[69] CHARGE (Psaty et al. 2009), ICGC,[70] H3Africa (Adoga et al. 2014), GenomeDenmark,[71] and the Luxembourg National Centre for Excellence in Research in Parkinson's Disease[72]);
- Translation work (e.g. the Estonian Genome Centre (Leitsalu et al. 2015), GenomeDenmark,[73] BeMGI,[74] the Precision Medicine Initiative Cohort Program,[75] FarGen[76] and the National Centre of Excellence in Research in Parkinson's Disease[77]).

More detail on the scope and scale of individual initiatives is available in Table 3 in the Appendix.

3.3. Funding and partnerships

The plethora of scientific and technological opportunities that have been created by advances in genomic science is accompanied by a diversity of funders and partnerships, although the funding landscape is still dominated by public and third-sector players.

The majority of initiatives are funded through government or charitable (philanthropic) funding or through public-charity partnership models. For example:

- Some initiatives are entirely government-funded. These are the Estonian Biobank (EGCUT),[78] Genome of the Netherlands,[79] SardiNIA,[80] the Saudi National Genome Program,[81] the 100,000 Genomes Project,[82] the Belgian Medical Genomics Initiative,[83] the Japanese Initiative on Rare and Undiagnosed Diseases,[84] and the Luxembourg-based National Centre for Excellence in Research in Parkinson's Disease[85];
- Joint charitable and government funding has supported the International HapMap Consortium,[86] 1000 Genomes Project,[87] H3Africa (Adoga et al. 2014, 1), UK Biobank,[88] China Kadoorie Biobank,[89] UK10K,[90] African Genome Variation Project[91] and the Deciphering Developmental Disorders Study[92];

69 BeMGI: About BeMGI (2015).
70 International Cancer Genome Consortium: About (2015).
71 GenomeDenmark: Research Equipment (2015).
72 Université de Luxembourg (2015).
73 GenomeDenmark: Vision (2015).
74 BeMGI: About BeMGI (2015).
75 PMI Cohort Program (2015).
76 Brice (2011).
77 Université de Luxembourg (2015).
78 Estonian Genome Center, University of Tartu (The Estonian Biobank) (2015).
79 GoNL (homepage) (2015).
80 National Institute on Aging (2008).
81 Saudi Human Genome Program: Introduction (2015, 1).
82 Genomics England: The 100,000 Genomes Project (2015).
83 BeMGI: About BeMGI (2015).
84 Otake (2015).
85 Université de Luxembourg (2015).
86 International HapMap Project: Groups Participating in the International HapMap Project (2015).
87 NIH News (2010).
88 UK Biobank: About (2015).
89 China Kadoorie Biobank: Funding agencies (2015).
90 UK10K: Funding (2011).
91 British Council (2015).
92 Deciphering Developmental Disorders Study: What is the DDD Study? (2015).

- The Slim Initiative in Genomic Medicine for the Americas[93] has been funded primarily by charities and philanthropic donors. Philanthropic funders of this and other initiatives include the Carlos Slim Foundation, the Bill and Melinda Gates Foundation, the W.M. Keck Foundation, the Delores Dore Eccles Foundation, the Wellcome Trust, Diabetes UK and the British Heart Foundation – please refer to Table 4 in the Appendix for more details.

The Wellcome Trust and US National Institutes of Health remain the largest funders of sequencing initiatives globally. A minority of current initiatives are funded through joint public-private investments (e.g. GenomeDenmark[94] and the National Centre for Indigenous Genomics at the Australian National University[95]).[96] In the case of some commercial ventures investment is fully private (e.g. deCODE genetics carries out R&D as a subsidiary of Amgen[97] and Cymru DNA Wales offers screening services to consumers[98]).

The rationale for public investment or public-private partnership is partially based on the pre-competitive nature of the research conducted by many of the current initiatives (Morgan Jones et al. 2014, 8–9), and on perceptions of population-based sequencing data as a 'public good' resource with potential further prospects for downstream innovation where the private sector would benefit from the research advances.[99] It is also, in some cases, about the efforts of governments to place their countries at the forefront of genomic research[100] and to help catalyse genomics-based life-science industries – with potential industry investment being attracted further down the line.[101] Some initiatives have active strategies for engaging the private sector in their activities or as 'spin-off efforts' to exploit results and develop new drugs, diagnostics and treatments, particularly in the personalised medicine states. As the Genomics England website states:

> Genomics England is working hard to realise the potential benefits for patients from the 100,000 Genomes Project as quickly as possible. This is why Genomics England is working with industry from the start and making the most of companies' expertise in developing new diagnostics and treatments.[102]

In terms of the locus of research and types of implementation partners, most projects take place within academic and research centres

93 Davis (2010).

94 GenomeDenmark: Partners (2015).

95 National Centre for Indigenous Genomics: Genomics and Bioinformatics (2015).

96 Although not the direct focus of this study, as it is an initiative with a more explicit precision-medicine focus rather than a population-based sequencing initiative per se, the Structural Genomics Consortium also received public-private funding (see Morgan Jones et al. 2014).

97 deCODE genetics: Company (2015).

98 Cymru DNA Wales (homepage).

99 See, for example, GenomeDenmark's explanation of their collaborative model (GenomeDenmark: Ethics FAQ 2015; Genome-Denmark: Commercialisation 2015).

100 The introduction to the Saudi Human Genome Project, for example, proclaims that 'This will be the largest disease gene discovery project ever undertaken, and will therefore also establish the Kingdom as a world leader in disease genetics research and Personalized Medicine' (Saudi Human Genome Program: Introduction 2015. 1). Similarly, the Genomics England website states: 'The supersonic age of genomics has begun. And just as the NHS has been at the forefront of scientific breakthroughs before, we want the NHS to be at the forefront again, with its patients benefiting from all that genomics offers, becoming the first mainstream health service in the world to offer genomic medicine as part of routine care for NHS patients.' (Genomics England: The 100,000 Genomes Project 2015).

101 See Genomics England: The 100,000 Genomes Project (2015).

102 Genomics England: How we are working with industry (2015).

and in partnerships between them (nationally or internationally). It is not uncommon for publically and/or philanthropically funded projects to partner with or subcontract sequencing work to specialist firms such as BGI and Illumina (see, for example, the HapMap Project,[103] 1000 Genomes Project,[104] 100,000 Genomes Project,[105] GenomeDenmark,[106] the Australian National Centre for Indigenous Genomics[107] and Genome of the Netherlands[108]), who offer specialised expertise and technologies. It is interesting to note that established companies like BGI and Illumina continue to be the leading market players in genome-sequencing technologies. Only a small number of initiatives currently engage private-sector companies as active research partners (aside from the sequencing technology providers). One example is GenomeDenmark, a collaboration between four Danish universities, two hospitals and two industry partners: the first, BGI, is responsible for sequencing and analysis and the second, Bavarian Nordic, will contribute its vaccine and development platform to the late stages of the Cancer and Pathogen demonstration project.[109] Clinical engagement is growing, particularly for initiatives with clear biomedical or clinical end goals (e.g. 100,000 Genomes and Genomics England in the UK,[110] the Global Network of Personal Genome Projects,[111] UK Biobank,[112] GenomeDenmark,[113] the Estonian Genome Centre (Milani et al. 2015, 196–7), and the Luxembourg National Centre of Excellence in Research in Parkinson's Disease[114]).

3.4. Data ownership, management and ethics

Population-based sequencing initiatives have underpinned advances in the fields of research ethics and data management. They raise new opportunities and challenges related to data access, informed consent and the management of research findings, including in terms of feedback of findings to research participants.[115]

They have also driven institutional innovations. Most – though not all – initiatives have an open-access policy, for non-commercial, and in some cases commercial, uses of anonymised datasets. Some have 'delayed data release provisions' to enable researchers to publish their findings prior to making the data more widely available, and to enable a degree of competitive advantage. Within the wider social context in which these initiatives operate, fears of security and privacy breaches have spurred experimentation with data protection strategies bridging technological interventions (code-based), regulatory policy and legal levers (e.g. vetting of researchers applying for access to data, review bodies) and community norms and behaviours (advisory groups, education on risk-minimising behaviours, emergency

103 International HapMap Project: Groups Participating in the International HapMap Project (2015).
104 NIH News (2012).
105 Genomics England News (2014).
106 GenomeDenmark: Partners (2015).
107 National Centre for Indigenous Genomics: Genomics and Bioinformatics (2015).
108 GoNL (homepage) (2015).
109 GenomeDenmark: Partner Info (2015).
110 Genomics England: Frequently asked questions (2015).
111 SickKids (2012).
112 UK Biobank: About (2015).
113 GenomeDenmark: Partner Info (2015).
114 Université de Luxembourg (2015).
115 For an overview of the debate on the feedback of incidental findings, see Wolf (2012).

response plans).[116] On the other extreme, the Personal Genome Project has chosen to invite volunteers to make all their data freely and publically available, with no guarantees of, or even efforts to ensure, the anonymity of participants' information, as discussed further below. Central to all these initiatives are issues of informed consent, spanning individual and group consent processes to ensure community-wide understanding of the research process, goals and intended outcomes; and single-use versus broad consent approaches. Closely related to this is a diversity of policies on the feedback of individual findings to participants. Some initiatives have opted for a full-feedback policy that would include incidental findings (e.g. the Personalised Medicine Initiative,[117] the Personal Genome Project (Ball et al. 2014, 4) and the Estonian Biobank (Leitsalu et al. 2015, 98); others provide feedback on findings directly related to the core research aim only (e.g. the Deciphering Developmental Disorders Study[118]), and some have opted out of feedback provision, especially in the absence of prospects for treatment for particular conditions, informed consent on feedback provisions, or certainty on the finding implications (e.g deCODE genetics,[119] UK Biobank[120]). We discuss each of these issues in more detail below, and each initiative's particular policies are outlined in Table 5 in the Appendix.

3.4.1. Open access policies or proprietary data

A cross-analysis of these initiatives suggests that – from the earliest to the more recently launched projects – there has been a strong trend to make non-identifiable data freely available for the public good. Of the 30 initiatives covered here (and which have published information on their data access policies), only four do not permit or are currently prohibited from sharing data outside of the project partnership for a combination of commercial interest, legal and regulatory reasons. These are deCODE genetics (Gudbjartsson et al. 2015b, 10), FarGen,[121] GenomeDenmark[122] and Genomics England.[123] Aside from these exceptions, most of these initiatives have explicit commitments to share their data with the wider research community, or even with the general public, insofar as they dedicate resources to the formation of data management mechanisms that are designed to guarantee anonymity and data probity, and appropriate data use. This includes internet portals and committees to review applications for data access. Various internet databases such as the European Genome-phenome Archive (used by H3Africa (H3Africa Consortium 2014, 1348), the African Genome Variation Project,[124] UK10K[125] and the Deciphering Developmental Disorders Study (Firth & Wright 2011, 703)) have been set up to manage the secure archiving and distribution of data from multiple genomic studies.

116 These examples relate specifically to the Precision Medicine Initiative, International Cancer Genome Consortium, Deciphering Developmental Disorders Study, and Saudi National Genome Program, which, alongside others, will be discussed in further detail in the paragraphs that follow.

117 PMI Cohort Program: Frequently Asked Questions (2015).

118 Deciphering Developmental Disorders Study: Frequently asked questions (2015).

119 Regalado (2015).

120 UK Biobank (2007, 6–8).

121 Lauerman (2013).

122 GenomeDenmark (homepage) (2015).

123 Genomics England: How we are working with industry – FAQs (2015).

124 African Genome Variation Project (2015).

125 UK10K: Data access (2013).

These efforts to develop robust open-access data policies and infrastructure have earned genomic science recognition as a leader in the promotion of data-sharing (Kaye et al. 2009, 331). Indeed, an ethos of data-sharing has been ingrained in genomic research since the Bermuda Principles were agreed at the First International Strategy Meeting on Human Genome Sequencing in 1996. These principles stated that primary genomic sequences should be rapidly released into the public domain upon completion.[126] Since then, all large funding bodies have required funding applications to consider data-sharing in their research proposals (Kaye et al. 2009, 331). Many sequencing initiatives also make it a condition of data access that secondary users of their sequencing data make their subsequent findings publically available, usually on the sequencing initiative's own data-sharing platform (see, for example, the Human Genome Diversity Project (Cavalli-Sforza. 2005, 334), the Estonian Biobank[127] and the UK Biobank[128]). Regarding the debate around gene patents and intellectual property rights, it is perhaps unsurprising that there seems to be broad consensus among these initiatives that primary sequence data should not be patentable (this is stated explicitly by the Human Genome Diversity Project (Cavalli-Sforza 2005, 334) and the International Cancer Genome Consortium (International Cancer Genome Consortium 2010, 995)). However, some initiatives do make clear that they are happy for researchers to claim intellectual property rights on downstream discoveries (for example, the HapMap Project (International HapMap Consortium 2003, 793) and the International Cancer Genome Consortium (International Cancer Genome Consortium 2010, 995)). Although not covered within the scope of the present study because they do not conduct population-scale sequencing, there are also a range of downstream initiatives focusing on more applied activities and early drug discovery R&D which have different intellectual property arrangements (for example, the Biomarkers Consortium and the stipulations set out by the UK model Industry Collaborative Research Agreement (Chataway et al. 2012, 734–5)).

In addition, open-access policies need to work within the wider social context of research. Some initiatives that have a strong capacity-building focus have found it challenging to reconcile the principle of data-sharing with their need to allow the researchers behind the data sufficient time to be the first to analyse and publish their findings. In response, the H3Africa Consortium is allowing the African scientists that lead H3Africa research exclusive access to the H3Africa data for a minimum of 11 months before it must be publically released. Upon release, H3Africa researchers have a publication lead of a further 12 months or until the first publication (De Vries et al. 2015). Similar principles underlie the China Kadoorie Biobank's data access policy,[129] which gives Chinese researchers a period of three months of exclusive access to the biobank datasets before they are made available to researchers worldwide.[130]

The principle of open data-sharing runs into more demanding challenges when the boundaries of what is considered pre-competitive and competitive research become more blurred, and when preserving data anonymity becomes higher risk. Our analysis shows that projects that incorporate little or no phenotypic data into their datasets tend to be the ones that allow

126 'Summary of Principles Agreed…' (1996).

127 Estonian Genome Center, University of Tartu: Data access (2015).

128 UK Biobank (2007, 13).

129 China Kadoorie Biobank: Data Access Policy and Principles (2015).

130 China Kadoorie Biobank: CKB Data Access System Launched (2015).

unrestricted open access to their data (e.g. the Hap Map Project (International HapMap Consortium 2003, 792–3), 1000 Genomes Project[131] and the Singapore Variation Project[132]). As research conducted by population-based genome-sequencing initiatives has become more clinically orientated and relies on the integration of detailed phenotypic and clinical data, concerns have arisen over the competing need to ensure the privacy and confidentiality of sample donors, as well as to address potential commercial relevance. A paper published by Homer et al. in 2008, which demonstrated a method of determining whether a specific individual's DNA is present within an aggregate dataset, proved that it is impossible to guarantee sample donors complete anonymity even once their data has been anonymised within a larger set (Homer et al. 2008). In light of this evidence, both the NIH and the Wellcome Trust changed their data-release policies, deciding to remove single nucleotide polymorphisms (SNP) data from publically available databases (Kaye et al. 2009, 334).

Different strategies have been developed in order to mitigate the risk of privacy breaches. Two-tiered or multi-level access systems such as those used by the International Cancer Genome Consortium (International Cancer Genome Consortium 2010, 995) and Deciphering Developmental Disorders Study (Firth & Wright 2011, 703) and proposed by the Precision Medicine Initiative[133] and Saudi National Genome Program[134] have become common. In this system, data without any identifiers is released publically and without restrictions, whereas bona fide researchers must apply for access to data that includes pseudonymised but potentially identifiable information. Authorization is managed by a specific reviewing body and may only be given once the researcher's identity has been verified, the scientific and ethical basis of their proposed research judged sound, and they have agreed not to attempt to re-identify individual study participants. The Precision Medicine Initiative Cohort Program, which aims to sequence the genomes of 1 million Americans in the coming years, is furthermore committed to the ongoing development and testing of data security safeguards, will work to educate participants with regard to reducing their risk of re-identification, and will put in place emergency response plans in the event of a breach of privacy.[135] In the UK, Genomics England is considering additional security measures, such as a requirement for researcher presence at the physical premises where data will be stored. [136]

As mentioned above, the Personal Genome Project, which was first piloted as a US-based Harvard Study, has adopted a completely different approach. The founders of this project considered it unethical to promise participants confidentiality and anonymity, as the very real possibility of individual re-identification would render any consent given on this basis invalid (Lunshof et al. 2008, 406–8). Arguing, however, that the scientific community must have access to large numbers of genome sequences linked to extensive phenotypic datasets in order to advance clinical genomics, they decided to experiment with a new model of consent which is not based on the assurance of confidentiality and privacy (Lunshof et al. 2008, 408–9). Instead, only individuals who demonstrate a sound

131 1000 Genomes: About (2015).

132 Singapore Genome Variation Project (2015).

133 Precision Medicine Initiative (2015).

134 Saudi Human Genome Program: Data access (2015).

135 PMI Cohort Program: Frequently Asked Questions (2015).

136 This is an additional insight gained from a BMJ blog written by Mark Peplow, which was published after the core research presented in the present study was completed. This addition was added to the print version of this report on 15 April 2016.

understanding of genetics and, moreover, the risks of making their data publically available (tested through a written examination in which applicants must score 100 per cent) are invited to participate (Ball et al. 2014, 1–2). Participants then volunteer to disclose extensive genomic, phenotypic, clinical and lifestyle data, as well as analyses of their transcriptomes, microbiomes, epigenomes, *cis*-regulomes, VDJomes and microRNAs, which are linked to their project identifier and made freely and publically available (Angrist 2009, 3–4). Since starting the project, 185 participants in the Harvard PGP have shared their whole-genome or exome data, and close collaboration between the project researchers and participants has been ongoing in order to monitor the consequences of the public release of their data. According to their 2014 paper, no serious adverse effects have so far been reported (Ball et al. 2014, 6).

3.4.2. Informed consent

The 'open-consent' model trialled by the Personal Genome Project introduces two other key points of debate central to the research ethics of genome-sequencing initiatives: a) the issue of informed consent and b) the feedback of individual study findings (the latter is explored in Section 3.4.3). Since the US government introduced the 'Common Rule' in 1991 to protect the rights of human research subjects, the informed consent of research participants has been a legal requirement. These standards have set the norm for all the sequencing initiatives listed in the present study (with the exception of the compulsory DNA testing mandated by the government of Kuwait for security purposes[137]). However, the implementation of informed consent by initiatives spanning diverse cultural contexts can be challenging. The Human Genome Diversity Project (HGDP), which sought to collect DNA samples from indigenous populations worldwide, found itself at the centre of a controversy when it was accused by indigenous groups and activists of exploitation and 'biopiracy' (Greely 2001, 224–5). The HGDP tried to address these concerns by pointing out its non-commercial nature and by developing a Model Ethical Protocol. As part of these ethical guidelines, 'group consent' was introduced as a concept to ensure that research that might have consequences for a particular ethnic group obtained collective consent from the group as a whole (through its culturally appropriate authorities), as well as from individual sample donors (Greely 2001, 224–5).

Although the HGDP initiative never managed to recover its damaged reputation, the principles of 'group consent' have influenced later research initiatives. The HapMap Project, launched with very similar aims just over ten years after the HGDP, undertook community engagement processes and set up a Community Advisory Group for each participating sample donor community. These groups were created to act as a continuing liaison between the community and the HapMap researchers and to ensure that any future uses of the samples collected were consistent with the research uses that the community had consented to (International HapMap Consortium 2003, 792). More recently, the H3Africa Working Group on Ethics and Regulatory Issues has developed guidelines for engaging individuals and communities in an ongoing process in order to mitigate and address local suspicions of exploitation, cultural beliefs relating to the donation of body parts, and a lack of wider population knowledge about genetics research and DNA (H3Africa Consortium 2014, 1348).

These 'open-consent' and 'group consent' models, based on tested knowledge or community engagement processes, have

137 Human Rights Watch (2015).

been developed because informed consent must depend on study participants having a sound understanding of what exactly they are consenting to. Efforts to ensure this understanding from sample donors have also resulted in arguments for and against the relative advantages of 'single-use' and 'broad' consent models. Sometimes, 'single-use' consent is sought from donors for a specific use of their samples. However, increasingly, non-specific 'broad' consent is preferred, especially by biobanks, because samples may be collected for future research purposes, the exact nature of which may be as yet unknown.[138] The 'broad consent' approach relies on the building of trust between sample donors and the research organisation. Often a reviewing body, such as the Research Ethics Committee of the University of Tartu, which approves or prohibits proposed research on samples stored by the Estonian Biobank, is used as a safety check to ensure that future uses of donor samples are consistent with the broad medical research aims that donors originally consented to (Leitsalu et al. 2015, 98). In order to support the broad consent approach, consent processes are also becoming more dynamic. Plans for the Precision Medicine Initiative Cohort Program, for example, outline a particularly collaborative approach between participants and researchers, whereby participants will be given a voice not just to determine how their own data are used and shared, but to influence decisions on the governance of the cohort.[139]

3.4.3. Feedback of findings

The possible return of individually significant findings is a subject of intense debate[140] and experimentation across the variety of initiatives we profiled. deCODE genetics is currently at the centre of a complex bioethical debate because their study in Iceland has made it possible to identify about 2,000 people who carry a mutation the BRCA2 gene, associated with a sharply increased risk of developing breast or prostate cancer, and, as a result, a greatly reduced life expectancy.[141] Although informing these people of their increased risk would allow them to take preventative action, at the point of enrolment deCODE did not invite participants to consent to the feedback of clinically significant findings.[142] Moreover, many of these at-risk individuals never actually consented to participate in the study at all but have rather had their genomes inferred based on what is known about the DNA of closely related individuals who did donate samples for sequencing.[143] Informing people without their consent would therefore violate their right *not* to know about their own genetic risk factors.

In contrast, some initiatives such as the Precision Medicine Initiative Cohort Program,[144] Personal Genome Project (Ball et al. 2014, 4) and Estonian Biobank (Leitsalu et al. 2015, 98) allow participants full access to their individual results, along with the tools to interpret these data and, in the case of the Estonian Biobank/ EGCUT, genetic counselling. This approach, however, has been rejected by other initiatives, not only because it is resource-intensive, but

138 For a discussion of the use of broad consent, see Petrini (2010).

139 Precision Medicine Initiative (2015, 2).

140 For an example of this debate in the media, see Heger (2013). As previously mentioned, academic perspectives include Wolf (2012).

141 Regalado (2015).

142 Regalado (2015).

143 Regalado (2015).

144 PMI Cohort Program: Frequently Asked Questions (2015).

also because the inherently probabilistic nature of genetics, coupled with the fact that findings may not even be clinically actionable, may make the return of findings irresponsible from a bioethical point of view.[145]

As an intermediate approach, some initiatives such as the Deciphering Developmental Disorders Study allow the return only of findings which are relevant to the original research objectives (in this case a diagnosis for the patient's developmental disorder).[146] Others, such as GenomeDenmark, allow the feedback of both pertinent and incidental findings but only on the condition that they are clinically actionable, and subject to approval by a certified council.[147] Following the results of public engagement activities, Genomics England has recently been granted ethical approval to invite participants to choose whether they would like to be informed of *certain* incidental findings; they may consent to learning about their risk of developing ten 'serious but actionable' diseases identified by Genomics England, such as familial hypercholesterolaemia and certain types of cancer.[148] As these differing approaches indicate, many initiatives have, in the absence of established best practices, made research on these issues a central part of their research program (see, for example, UK10K (Kaye 2014), the Deciphering Developmental Disorders Study,[149] Genomics England,[150] and the Precision Medicine Initiative[151]).

145 See, for example, UK Biobank (2007, 6–8).
146 Deciphering Developmental Disorders Study: Frequently asked questions (2015).
147 GenomeDenmark: Ethics FAQ (2015).
148 Genomics England News (2015a).
149 Deciphering Developmental Disorders Study: Ethics Research (2015).
150 Genomics England: Patient and Public Involvement (2015).
151 National Institutes of Health (2015).

Chapter Four. In reflection

4.1. In reflection on achievements and evolving impacts

Based on our analysis of the 30 initiatives that were profiled in this study, over two decades of progress in the field has yielded numerous achievements and impacts from large-scale sequencing efforts (see Box 2).

The provision of population-specific datasets capturing common patterns of variation has been a fundamental step in enabling association studies. The HapMap was the first major catalogue of human genetic variation to be used by researchers worldwide, and the 1000 Genomes Project, launched six years later, extensively refined and expanded this earlier effort,[152] more than doubling the number of known variant sites in the human genome when the final phase of the project was published in 2015.[153] These two projects, in particular, have been considered 'vital' to the development of genome-wide association studies and genotyping arrays (Naidoo et al. 2011, 577, 589–90, 600–1), which have in turn contributed to the growing numbers

Box 2. Examples of achievement and impact

1. Catalogues of population-specific genetic variation have enabled further research – namely, association studies which compare genetic and phenotypic traits in order to advance knowledge of the genetic basis of disease;
2. Biobank-based projects are providing comprehensive, longitudinal datasets which could allow for a greater understanding of the interactions between genes, lifestyle and environmental exposures on phenotypic expression;
3. Progress in the clinical interpretation of genetic variation is starting to inform more effective disease management strategies as well as offering diagnoses to patients who suffer from previously undiagnosed conditions (e.g. rare developmental disorders);
4. New tools and research methods have been developed (e.g. sequencing and analysis techniques);
5. New physical infrastructure and human resource capacity has been established by diverse initiatives worldwide, creating new laboratories and information and communication capacity, a greater pool of highly trained researchers across academic and clinical research communities;
6. Population-scale genetic studies have benefited from but also made significant contributions to strengthening ICT infrastructure associated with cloud computing and data management innovation;
7. Genomic sequencing initiatives have driven innovations relating to ethical, legal and social concerns, establishing new protocols, practices and guidelines around open access, informed consent and privacy protection, and feedback of data to study participants;
8. Some studies have advanced our collective knowledge on patterns of human migration, divergence and evolution, as well as estimated rates of mutation in modern humans.

152 NIH News (2013).

153 NIH News (2015).

of genetic variants known to be associated with diverse clinical conditions. In addition, biobank-based projects are providing extensive longitudinal datasets which link genomic data with clinical records, phenotypic traits, lifestyle and environmental data, and which are enabling greater understanding of the complex interactions between genes, lifestyle and environment on phenotypic expression. These datasets will be particularly important for advancing knowledge of the aetiology of common complex diseases. Ten years on from the completion of the Human Genome Project, the US National Human Genome Research Institute reported that there have been a total of 1,542 GWAS published since their introduction in 2005; the number of replicated disease-associated genetic variants stood at roughly 2,900.[154]

As a result of progress made, some studies are providing patients with rare, previously undiagnosed diseases a diagnosis for the first time (especially related to single-gene mutation disorders). Informing patients and their families about the causes of their disorder is beginning to open up prospects for offering personalised therapeutic interventions. For example, a recent paper on the results of Deciphering Developmental Disorders reported that the study has diagnosed five children with genetic variants that are associated with inborn errors of metabolism, and which are therefore potentially treatable with existing therapeutic interventions, such as dietary restriction, supplementation or pharmacological intervention (Wright et al. 2015, 1311). Likewise, Genomics England announced in January 2016 that the 100,000 Genomes Project has provided its first diagnoses to two children with rare, undiagnosed genetic disorders, one of whom may also benefit from a special diet tailored to target the molecular basis of her condition.[155]

Characterising genetic variation across populations has also contributed to anthropological insights into the demographic history of human populations. Analyses of data produced by the Human Genome Diversity Project (Cavalli-Sforza 2005, 334–6), African Genome Variation Project (Gurdasani et al. 2015), deCODE genetics (Helgason et al. 2015, 455–6) and Singapore Genome Variation Project (Teo et al. 2009) have all contributed findings related to early human migration, divergence and evolution, as well as estimated rates of mutation in modern humans.

Underlying all these biomedical and anthropological findings has been the concomitant development of research capacity. Aside from contributions to the wider knowledge base, many of the studies listed here have published findings related to the development of sequencing techniques and methods for the design, analysis and interpretation of genomic studies (see, for example the Human Genome Diversity Project (Cavalli-Sforza 2005, 335, 339), Personal Genome Project,[156] 1000 Genomes Project (The 1000 Genomes Project Consortium 2015, 61–62), Genome of the Netherlands (Genome of the Netherlands Consortium 2014, 823), Deciphering Developmental Disorders Study (Wright et al. 2015, 1311), and Genomics England[157]). At a more fundamental level, many projects have built human and physical research capacity. For example, the collaboration between GenomeDenmark and the Beijing Genomics Institute (BGI) has resulted in the opening of BGI-Europe headquarters in Copenhagen, bringing state-of-the-art research

154 NHGRI (2013).

155 Genomics England News (2016).

156 See, for example, Drmanac et al. (2010); Ball et al. (2012); Brock et al. (2012).

157 Genomics England News (2015d).

facilities to the Copenhagen Bio Science Park and to researchers working at the University of Copenhagen.[158] Similarly, as a partner of Genomics England, Illumina will invest £162 million over four years into creating genome sequencing jobs and expertise in England, and, moreover, the Wellcome Trust has committed to build a £27 million world-class sequencing hub at its Genome Campus in Cambridge so that Genomics England researchers may work alongside experts at the Wellcome Trust's Sanger Institute.[159] In order to carry out the Saudi National Genome Project, a network of 15 genome-sequencing laboratories will be built across the country.[160] The African Genome Variation Project has held a two-week genetics workshop for analysts from all collaborating centres.[161] Under the Slim Initiative in Genomic Medicine for the Americas (SIGMA) collaboration, 36 Mexican researchers have been trained at the Broad Institute in the US and a total of 9,978 people have participated in the eight workshops, five symposiums and four conferences that have been conducted.[162]

It may be too early to assess the impact of projects that have a specific developing country focus, but the value of such initiatives is evidenced by examples such as SIGMA's identification of a common genetic variant predisposing Latin American populations to type 2 diabetes, and which had previously been overlooked because it is not present among European populations.[163] Furthermore, by providing the most comprehensive catalogue of common genetic variation in Africa, as well as an improved array design, the African Genome Variation Project has provided a critical mass of knowledge requisite for association studies in African populations such as those that H3Africa is carrying out (Ramesar 2015, 276–7; Gurdasani et al. 2015). The H3Africa Consortium, moreover, has already leveraged additional funding from the South African Department of Science and Technology for a project on cardiometabolic disease genomics (H3Africa Consortium 2014, 1348), suggesting high-level country buy-in.

A major development that has greatly impacted genomic research capacity worldwide is the harnessing of cloud computing to store, share and analyse the vast amounts of data produced by next-generation sequencing. For example, the 1000 Genomes Project was the first to use Amazon Web Services (AWS) to make large-scale datasets available to the wider research community in 2012 (1000 Genomes Project 2012, 3–4). At the same time, population-scale sequencing initiatives have contributed to strengthened ICT infrastructure and data management innovation. For example, The Human Genome Sequencing Centre at Baylor College of Medicine, responsible for sequencing samples collected by the CHARGE Consortium, partnered with AWS and DNAnexus to develop a cloud-based infrastructure that would allow them to conduct the large-scale analysis that the CHARGE project required. The innovative solution won the partnership the 2014 Bio-IT World Best Practices Award for IT Infrastructure and High Performance Computing.[164] The use of cloud computing has since become the norm for genome sequencing and analysis projects, and is essential for major international collaborations like the International Cancer Genome

158 University of Copenhagen (2012).
159 Genomics England News (2014).
160 Saudi Human Genome Program (homepage) (2015).
161 African Genome Variation Project (2015).
162 Carlos Slim Foundation: Our Programs (2015).
163 Carlos Slim Center for Health Research at the Broad Institute (2015).
164 Proffitt (2014).

Consortium because it offers a large increase in computing power without the need for additional capital investment (Stein et al. 2015).[165] Initial concerns that cloud-based data storage is less secure than local servers have been minimised; researchers are now calling for the major funding agencies to pay for the storage of important genomic data sets so that research is not further inhibited by limited institutional resources (Stein et al. 2015).

As we have discussed earlier in this paper, crucial also to the advancement of genomic research has been the development of ethical, legal and social principles and guidelines, driven, in part, by genomic sequencing studies. For example, the HGDP, although plagued by ethical concerns, introduced the concept of group consent (Greely 2001, 224–5), which left a lasting legacy and went on to influence the development of community engagement processes used by studies such as the International HapMap Consortium (International HapMap Consortium 2003, 792), 1000 Genomes[166] and H3Africa.[167] The participatory approach pioneered by the Personal Genome Project is subject to ongoing monitoring and evaluation (Ball et al. 2014), but seems to have had an important impact on plans for other research programs, such as the Personalised Medicine Initiative Cohort Program.[168]

The UK-based Deciphering Developmental Disorders Study has carried out various studies to investigate the preferences of would-be research participants. This includes a survey of 6,944 individuals from 75 different countries, which found that a large majority support the feedback of incidental findings although they do not expect researchers to actively search for results not relevant to their primary research aims (Middleton et al. 2015). Last year the consortium behind the UK10K project published a detailed management pathway that they have developed as a mechanism to support the feedback of clinically significant findings in an ethically sound way (Kaye 2014). These guidelines are by no means being adopted in a standardised way; nonetheless, the research and trialing of different methods by these and other initiatives is providing a more advanced and tailored evidence base for the genomic medicine field, and in particular for informed consent and decisionmaking on the design and implementation of large-scale genome-sequencing studies.

4.2. Looking forward to a future research, policy and service transformation agenda

As regards the ultimate aim of many of these initiatives – that of translating research findings into new drugs and diagnostics and integrating them into genomic medicine services – larger scale impacts are yet to accrue despite some promising examples of clinical change and transformation of practice. Evidence of progress includes an announcement by the head of R&D at Amgen at a conference in October 2014 that the data collected through deCODE's Iceland study have helped accelerate the company's development of drug candidates for heart disease and asthma.[169] In Qatar, the results of the Weill Cornell study could offer the Qatari population much more comprehensive tools for the management of their high-burden Mendelian

165 Cloud storage of genomic data greatly accelerates research as it eliminates the highly time-consuming download of data and allows analysis to be run over several servers at once. See 1000 Genomes Project (2012).

166 1000 Genomes Project (2015, 2).

167 H3Africa Working group on Ethics (2013).

168 Precision Medicine Initiative (PMI) Working Group Report (2015, 39–45).

169 Lauermann & Kitamura (2015).

diseases (Rodriguez-Flores 2013, 105). The study identified 37 genetic risk variants for monogenic disorders, only four of which are already tested for in premarital screening in Qatar; if further research confirms the high penetrance of these variants, they could be incorporated into existing screening programs (Rodriguez-Flores 2013, 105). And in Estonia, the Estonian Genome Centre has worked with medical geneticists at Tartu University Hospital in order to implement whole-exome sequencing into clinical practice; as of 2014 the Estonian Health Insurance Fund covers the costs of parent-offspring whole-exome sequencing for patients with rare, undiagnosed disorders (Milani et al. 2015, 193).

Many of the most clinically oriented projects have put enabling mechanisms in place in order to accelerate translational research. Some initiatives have only recently reached a stage of maturity where they are considering more downstream product development and service transformation activity (e.g. the International Cancer Genome Consortium,[170] the Slim Initiative in Genomic Medicine for the Americas[171]). Some are building in partnerships with industry and healthcare services from the onset, but have not yet reached product or technology implementation milestones nor yet impacted on healthcare delivery (e.g. Genomics England,[172] GenomeDenmark,[173] the Luxembourg National Centre for Excellence in Research in Parkinson's Disease[174]). And in some instances, substantial effort has already resulted in improvements in systems readiness to engage with a genomic medicine reality – for example, in Estonia, as further discussed below. More specifically:

- Public-private partnerships are designed to harness the expertise and resources of established industry players, generally for more downstream research but also to help optimise sample collection and research designs and to help in the interpretation of findings. For example, Genomics England has formed a Genomics Expert Network for Enterprises (GENE) Consortium in which ten biotechnology and pharmaceutical companies at present, as well as industry experts in big data, will conduct a year-long industry trial in order to see how industry partners can best collaborate with researchers and clinicians to develop diagnostics and treatments.[175]
- Some initiatives have recently entered into or made plans for a second phase of their existence, explicitly designed to translate research findings into clinical application. For example, in 2013 the Slim Initiative in Genomic Medicine for the Americas (SIGMA) launched SIGMA II in order to leverage findings from the original three-year collaboration towards the development of diagnostics and therapeutic 'roadmaps'.[176] Similarly, plans for the recent 10th International Cancer Genome Consortium (ICGC) Scientific Workshop focused on discussions of ICGC2 – an extension of the original scope of the ICGC program which would integrate detailed phenotypic and clinical data with genomic data and would require researchers to consider not only how to use genomic tests and genomics-based treatments in a clinical

170 Ontario Institute for Cancer Research (2015).

171 Carlos Slim Center for Health Research at the Broad Institute (2015).

172 See Genomics England: How we are working with industry (2015); Genomics England: The 100,000 Genomes Project (2015).

173 GenomeDenmark: Partners (2015).

174 Université de Luxembourg (2015).

175 See Genomics England: How we are working with industry (2015); Genomics England: How we are working with industry – FAQs (2015).

176 Carlos Slim Center for Health Research at the Broad Institute (2015).

setting, but also the cost-effectiveness of such interventions.[177]

- A third approach has been to lay the groundwork for personalised medicine by preparing a national infrastructure and healthcare system capable of integrating genomic medicine into routine healthcare. For example, parallel to the 100,000 Genomes Project, Health Education England will deliver a skills and training program for NHS workers[178] as well as a Master's program in Genomic Medicine.[179] In Estonia, significant steps have been taken by the Estonian Genome Centre (the body responsible for managing the Estonian Biobank) and the national government to develop system readiness for personalised medicine through the integration of health-related databases and registries, including biobank data where all samples will be genotyped, within a national IT infrastructure called X-road which securely links and communicates medical data in order to provide clinicians (and patients themselves) with access to complete individual electronic health records (Leitsalu et al. 2015, 103). The Estonian Genome Centre will analyse the data with automatic risk estimation and decision support software in order to deposit disease risk and drug response prediction reports for each biobank participant directly into the e-health system. Clinicians will be trained to use this data and, if this pilot phase is successful, genotyping and analysis will be offered to all adult residents of Estonia (Milani et al. 2015, 196–7).

It remains to be seen, however, whether these various enabling mechanisms will deliver on their promise to bring genomics into routine clinical care. The challenges to overcome are substantial, but the implications of doing so are profound and require coordinated action within countries and initiatives and across them. Drawing on our analysis of the findings presented in this report, we discuss some priority issues for a future research and policy agenda (see Box 3).

Box 3. Emerging issues for a policy and research agenda

- Scale up of coordinated international collaboration to enable clinically relevant sense-making from large amounts of distributed genotypic and phenotypic data;
- Innovative means of study design to ensure sample representativeness and optimal collection for personalised medicine innovations;
- A need for data-sharing and clinical interpretation that cross disease, disciplinary and professional boundaries will call for new ways of designing studies and of collaborating;
- New research on implications of genomic interventions in a clinical setting is needed (economic, capacity-building, and patient-clinician communications implications);
- A likelihood of further changes in industry R&D models and in rationales for public and private investment;
- Scope for further consolidation of learning on the research ethics framework, based on the experiences of prior and current initiatives and a need to consider legal frameworks for a genomic medicine era;
- Evaluation will be central to accountable and effective progress in genomic medicine, for patient benefit.

First, making sense of the data on diverse variants associated with a certain diseases will require international exchange of information. There will need to be cooperation not only between researchers worldwide, but also with a

177 Ontario Institute for Cancer Research (2015).

178 Genomics England: The 100,000 Genomes Project (2015).

179 Genomics England News (2015c).

wider pool of stakeholders, including clinicians, industry and regulatory bodies. The design of future studies remains a challenge in terms of optimising sample collection and ensuring the representativeness of data, which is crucial for robust clinical interpretation, and for informing personalised medicine innovation. Research will also have to further improve understandings of the implications of genomic interventions in a clinical setting, including the economic, capacity-building and patient-clinician communications processes required.

As discussed in Chataway et al. (2012, 736), the implications for the personalised medicine industry are also profound, with advances from genomic medicine studies already influencing change in traditional pharma business models and industry structures. For example, they raise questions about whether industry will focus on targeted treatments for specific patient profiles across a disease life-cycle, or on specific disease states across multiple patient segments (Chataway et al. 2012, 736). These advances are also blurring boundaries between pre-competitive and competitive research, and hence between public and private rationales for research funding and intellectual property ownership.

Second, our understanding of the boundaries of a single disease and disease classifications are changing as a result of genetic advances. For example, we now know that there are over 200 types of breast cancer and that, in fact, some types of breast cancer have more in common (genetically) with some types of prostate cancer, than with other breast cancers (Bezold & Peck 2005). As a result, data-sharing and clinical interpretation efforts will need to cross disease boundaries, and hence disciplinary and professional boundaries, which will require new ways of working and new ways of designing research studies and clinical trials. We are likely to see the emergence of deeper and broader public-private networks (Chataway et al. 2012).

Third, the challenges of ensuring informed consent and data security are widely accepted.[180] Current negotiations around EU regulation on data protection are subject to intense scrutiny and debate, and are being strongly resisted by some of the scientific community.[181] However, there is already scope for consolidation of learning from experimentation and progress in this space, and for cross-sectoral learning. By contrast, the legal framework for dealing with genomic medicine is far more nascent. The recent landmark case of a woman who is suing a doctor for failing to disclose a family history of hereditary brain disease, and who subsequently went on to have a child with a 50 per cent chance of developing Huntington's disease, is illustrative of the legal framework challenges ahead.[182]

It is clear that the success of genomic medicine will depend as much, if not more, on social determinants than on scientific and technological capacities. New organisational structures, spanning disease boundaries and geographies, will inevitably continue to emerge and mature before advances can be harnessed to their full potential. We have, in this paper, attempted to reflect on learning from key developments to date, and to highlight key issues that the wider scientific, healthcare and regulatory community will need to address (some of which have been identified in prior work). We have also attempted to go beyond the identification of issues, providing examples of how challenges are being grappled with and addressed to date, across the myriad initiatives that are likely to fundamentally change the way we think about health and health innovation, associated economic as well as political risks, and accountability and social

180 See: Lunshof et al. (2008); Ayuso et al. (2013); Kahn (2011).

181 Burbidge (2015c).

182 Gibb (2015).

responsibility for the results of scientific and technological progress. Building on our analyses, we argue that further investigations will need to evaluate the learning from past details in a summative and formative way. With the social and economic stakes at play, evaluation will be central to accountable and effective progress in genomic medicine, for patient benefit.

References

Bibliography sorted by initiative:

1000 Genomes Project

1000 Genomes: About. 2015. 1000genomes.org. As of 30 November 2015: http://www.1000genomes.org/about

1000 Genomes: Data. 2015. 1000genomes.org. As of 11 November 2015: http://www.1000genomes.org/data

1000 Genomes Project. 2012. '1000 Genomes Project Releases Data from Pilot Projects on Path to Providing Database for 2,500 Human Genomes.' 1000genomes.org, 21 June. As of 19 November 2015: http://www.1000genomes.org/sites/1000genomes.org/files/docs/100621.1000genomespressrelease.pdf

1000 Genomes Project. 2015. 'Ethical considerations for investigators proposing samples for inclusion in the 1000 Genomes Project.' As of 16 November 2015: http://www.1000genomes.org/sites/1000genomes.org/files/docs/Informed%20Consent%20Background%20Document.pdf

Burbidge, R. 2015a. 'International collaboration create global map of genetic diversity.' phgfoundation.org, 2 October. As of 5 November 2015: http://www.phgfoundation.org/news/16977

Clarke, L. et al. 2012. 'The 1000 Genomes Project: data management and community access.' *Nature Methods* 9(5): 459–62. doi:10.1038/nmeth.1974

NIH News. 2010. '1000 Genomes Project publishes analysis of completed pilot phase.' National Human Genome Research Institute, 27 October. As of 30 November 2015: https://www.genome.gov/27541917

NIH News. 2012. 'The 1000 Genomes Project more than doubles catalog of human genetic variation.' National Human Genome Research Institute, 31 October. As of 30 November 2015: http://www.genome.gov/27551417

NIH News. 2015. 'Scientists create world's largest catalog of human genomic variation.' National Institutes of Health, 30 September. As of 19 November 2015: http://www.nih.gov/news-events/news-releases/scientists-create-worlds-largest-catalog-human-genomic-variation

Peplow, M. 2016. 'The 100 000 Genomes Project.' *The BMJ* 353. doi: http://dx.doi.org/10.1136/bmj.i1757

Sudmant, P.H. et al. 2015. 'An integrated map of structural variation in 2,504 human genomes.' *Nature* 526 (October): 75–81. doi:10.1038/nature15394

The 1000 Genomes Project Consortium. 2015. 'A global reference for human genetic variation.' *Nature* 526 (October): 68–74. doi:10.1038/nature15393

The African Genome Variation Project

African Genome Variation Project. 2015. Wellcome Trust Sanger Institute. As of 1 December 2015: http://www.sanger.ac.uk/science/collaboration/african-genome-variation-project

British Council. 2015. ‘Mapping the genetic architecture of Africa.’ As of 1 December 2015: https://www.britishcouncil.org/cubed/science-breakthroughs/genetic-map-africa

Gurdasani. D. et al. 2015. ‘The African Genome Variation Project shapes medical genetics in Africa.’ *Nature* 517 (January): 327–32. doi:10.1038/nature13997

MacDougall, R. 2014. ‘NIH News: Researchers conduct comprehensive genomic study of sub-Saharan Africans.’ National Human Genome Research Institute, 3 December. As of 2 December 2015: https://www.genome.gov/27559699

Ramesar, R. 2015. ‘Genomics: African Dawn.’ *Nature* 517 (January): 276–7. doi:10.1038/nature14077

Belgian Medical Genomics Initiative (BeMGI)

BeMGI: About BeMGI. 2015. Belgian Medical Genomics Initiative. As of 8 November 2015: http://www.bemgi.be/about-bemgi/

BeMGI: Genomics and Society WG. 2015. Belgian Medical Genomics Initiative. As of 8 November 2015: http://www.bemgi.be/genomics-society-wg/

BeMGI: Research. 2015. Belgian Medical Genomics Initiative. As of 8 November 2015: http://www.bemgi.be/bemgi-research/

BeMGI: Work packages. 2015. Belgian Medical Genomics Initiative. As of 8 November 2015: http://www.bemgi.be/about-bemgi/bemgi-work-packages/

Frederickx, I. 2013. ‘Belgian universities establish nationwide genomics network.’ KU Leuven. As of 25 November 2015: http://www.kuleuven.be/english/news/2013/BeMGI

Charge Consortium

Charge Consortium: Aggregate Results Sharing Principles. 2012. washington.edu. As of 1 December 2015: http://depts.washington.edu/chargeco/wiki/DataSharingPrinciples

Charge Consortium: Charge Publications. 2015. chargeconsortium.com. As of 1 December 2015: http://web.chargeconsortium.com/main/publications

Charge Consortium. 2015. Baylor College of Medicine. As of 1 December 2015: https://www.hgsc.bcm.edu/charge-consortium

Charge Consortium (homepage). 2015. chargeconsortium.com. As of 1 December 2015: http://web.chargeconsortium.com/main

Proffitt, A. 2014. ‘Genome Analysis in the Cloud and for the Masses’. Bio.IT World, May 30. As of 1 December 2015: http://www.bio-itworld.com/2014/5/30/genome-analysis-cloud-for-masses.html

Psaty, B., et al. 2009. ‘Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies from 5 Cohorts.’ *Circulation: Cardiovascular Genetics* 2 (February): 73–80. doi: 10.1161/CIRCGENETICS.108.829747

China Kadoorie Biobank

Chen, Z. et al. 2011. ‘China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up.’ *International Journal of Epidemiology* 40(6): 1652–66. doi: 10.1093/ije/dyr120

China Kadoorie Biobank: About the Study. 2015. China Kadoorie Biobank. As of 6 December 2015: http://www.ckbiobank.org/site/About+the+Study

China Kadoorie Biobank: CKB Data Access System Launched. 2015. China Kadoorie Biobank, 30 September. As of 6 December 2015: http://www.ckbiobank.org/site/news/2015/data-access-launch.html

China Kadoorie Biobank: Collaborating Institutions. 2015. China Kadoorie Biobank. As of 6 December 2015: http://www.ckbiobank.org/site/About+the+Study/Collaborating+Institutions

China Kadoorie Biobank: Data Access Policy and Principles. 2015. China Kadoorie Biobank. As of December 2015: http://www.ckbiobank.org/site/Data+Access/Data+Access+Policy

China Kadoorie Biobank: Field Work. 2015. China Kadoorie Biobank. As of 6 December 2015: http://www.ckbiobank.org/site/About+the+Study/Field+Work

China Kadoorie Biobank: Funding Agencies. 2015. China Kadoorie Biobank. As of 6 December 2015: http://www.ckbiobank.org/site/About+the+Study/Funding+Agencies

China Kadoorie Biobank: Genotyping Data. 2015. China Kadoorie Biobank. As of 6 December 2015: http://www.ckbiobank.org/site/Study+Resources/Genotyping+Data

China Kadoorie Biobank: Long-term Follow-up. 2015. China Kadoorie Biobank. As of 6 December 2015: http://www.ckbiobank.org/site/About+the+Study/Long-TermFollow-up

China Kadoorie Biobank: Study Design. 2015. China Kadoorie Biobank. As of 6 December 2015: http://www.ckbiobank.org/site/About+the+Study/Study+Design

China Kadoorie Biobank (homepage). 2015. China Kadoorie Biobank. As of 6 December 2015: http://www.ckbiobank.org/site/

Cymru DNA Wales

BBC News. 2014. ‘New DNA project could give insight into history of Wales.’ BBC News, 24 September. As of 4 January 2016: http://www.bbc.com/news/uk-wales-29339804

Cymru DNA Wales (homepage). 2015. cymrudnawales.com. As of 9 December 2015: https://www.cymrudnawales.com/

Deciphering Developmental Disorders Study

Deciphering Developmental Disorders Study. 2013. Sanger Institute. As of 6 December 2015: http://www.sanger.ac.uk/research/areas/humangenetics/ddd/

Deciphering Developmental Disorders Study: Ethics Research. 2015. ddduk.org. As of 6 December 2015: http://www.ddduk.org/ethicsresearch.html

Deciphering Developmental Disorders Study: Frequently Asked Questions. 2015. ddduk.org. As of 6 December 2015: http://www.ddduk.org/faq.html

Deciphering Developmental Disorders Study: Publications. 2015. ddduk.org. As of 6 December 2015: http://www.ddduk.org/publications.html

Deciphering Developmental Disorders Study: What is the DDD Study? 2015. ddduk.org. As of 6 December 2015: http://www.ddduk.org/intro.html

Deciphering Developmental Disorders Study. 2015. ‘Large-scale discovery of novel genetic causes of developmental disorders.’ *Nature* 519 (March): 223–8. doi:10.1038/nature14135

Deciphering Developmental Disorders v.1. 2015. UK Clinical Research Network: Portfolio Database. As of 6 December 2015: http://public.ukcrn.org.uk/search/StudyDetail.aspx?StudyID=9955

Firth, H. & C.F. Wright. 2011. 'The Deciphering Developmental Disorders (DDD) study.' *Developmental Medicine and Child Neurology* 53(8): 702. doi: 10.1111/j.1469-8749.2011.04032.x

Wright, C. F. et al. 2015. 'Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data.' *The Lancet* 385(9975): 1305–14. doi: 10.1016/S0140-6736(14)61705-0

deCODE genetics

deCODE genetics: Company. 2015. decode.com. As of 23 November 2015: http://www.decode.com/company/

Gudbjartsson, D.F. et al. 2015a. 'Large-scale whole-genome sequencing of the Icelandic population.' *Nature Genetics* 47(5): 435–44. doi:10.1038/ng.3247

Gudbjartsson, D.F. et al. 2015b. 'Sequence variants from whole genome sequencing a large group of Icelanders.' *Scientific Data* 2. doi: 10.1038/sdata.2015.11

Helgason, A. et al. 2015. 'The Y-chromosome point mutation rate in humans.' *Nature Genetics* 47(5):453–7. doi:10.1038/ng.3171.

Lauermann, J. & M. Kitamura. 2015. 'Deals for Genetic Data Raise Issues of Privacy, Sharing.' Bloomberg Business, 14 January. As of 25 November 2015: http://www.bloomberg.com/news/articles/2015-01-13/deals-for-genetic-data-raise-issues-of-privacy-sharing

Regalado, A. 2015. 'Genome Study Predicts DNA of the Whole of Iceland.' MIT Technology Review, 25 March. As of 6 December 2015: http://www.technologyreview.com/news/536096/genome-study-predicts-dna-of-the-whole-of-iceland/

Vaydylevich, Y. 2015. 'Iceland study provides insights into disease, paves way for large-scale genomic studies.' National Human Genome Research Institute, 1 May. As of 27 November 2015: http://www.genome.gov/27561444

Warren-Gash, C. 2015. 'Sequencing the nation: Iceland's genomic profile.' 9 April. PHG Foundation. As of 6 December 2015: http://www.phgfoundation.org/blog/16642/

The Estonian Biobank and the Estonian Genome Center (EGCUT)

Estonian Genome Center, University of Tartu (The Estonian Biobank). 2015. University of Tartu. As of 11 November 2015: http://www.geenivaramu.ee/en/about-us

Estonian Genome Center, University of Tartu: Data access. 2015. University of Tartu. As of 23 November 2015: http://www.geenivaramu.ee/en/access-biopank/data-access

Leitsalu, L., H. Alavere, M. Tammesoo, E. Leego & A. Metspalu. 2015. 'Linking a Population Biobank with National Health Registries – The Estonian Experience.' *Journal of Personalized Medicine* 5(2): 96–106. doi:10.3390/jpm5020096

Milani, L., L. Leitsalu & A. Metspalu. 2015. 'An epidemiological perspective of personalized medicine: the Estonian experience.' *Journal of Internal Medicine* 277(2): 188–200. doi:10.1111/joim.12320

Schwab, Katherine. 2015. 'Estonia wants to collect the DNA of all its citizens.' The Atlantic, 8 October. As of 27 November 2015: http://www.theatlantic.com/health/archive/2015/10/is-a-biobank-system-the-future-of-personalized-medicine/409558/

The Faroe Genome Project (FarGen)

Brice, P. 2011. ‘World first as genome sequencing offered to entire Faroe Islands.’ PHG Foundation, 30 September. As of 18 November 2015: http://www.phgfoundation.org/news/10106

Brice, P. 2013. ‘Pace slips for Faroes genomics project.’ PHG Foundation, 29 August. As of 18 November: http://www.phgfoundation.org/news/14552

FarGen (homepage). 2015. FarGen. As of 9 November 2015: http://www.fargen.fo/en

Lauerman, J. 2013. ‘Faroes’ 50,000 Residents Leap Into DNA Testing Quagmire.’ Bloomberg Business, 25 February. As of 9 November 2015: http://www.bloomberg.com/news/articles/2013-02-25/faroes-50-000-residents-leap-into-dna-testing-quagmire

Genome of the Netherlands (GoNL)

Boomsma, D.I. et al. 2014. ‘The Genome of the Netherlands: design, and project goals.’ *European Journal of Human Genetics* 22: 221–7. doi:10.1038/ejhg.2013.118

GoNL (homepage). 2015. Genome of the Netherlands. As of 7 December 2015: http://www.nlgenome.nl/

Genome of the Netherlands Consortium. 2014. ‘Whole-genome sequencing variation, population structure and demographic history of the Dutch population.’ *Nature Genetics* 46(8): 818–35. doi:10.1038/ng.3021

Van Leeuwen, E.M. et al. 2015. ‘Genome of the Netherlands population-specific imputations identify ABCA6 variant associated with cholesterol levels.’ *Nature Communications* 6: 6065. doi:10.1038/ncomms7065

GenomeDenmark

Brix, L. 2015. ‘Scientists are mapping the entire Danish genome.’ ScienceNordic, 25 January. As of 9 December 2015: http://sciencenordic.com/scientists-are-mapping-entire-danish-genome

GenomeDenmark: About. 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.com/english/about/

GenomeDenmark: Commercialisation. 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.com/english/partners/commercialisation/

GenomeDenmark: Ethics FAQ. 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.com/english/faq/

GenomeDenmark: Research Equipment. 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.dk/english/equipment/

GenomeDenmark: Partners. 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.com/english/partners/

GenomeDenmark: Partner Info. 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.dk/english/partners/partnerinfo/

GenomeDenmark: The Cancer and Pathogen Project. 2015. genomedenmark.com As of 9 December 2015: http://www.genomedenmark.com/english/about/cancer-pathogen/

GenomeDenmark: The Danish Reference Genome Project. 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.com/english/about/referencegenome/

GenomeDenmark: Vision. 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.com/english/vision/

GenomeDenmark (homepage). 2015. genomedenmark.com. As of 9 December 2015: http://www.genomedenmark.com/english/

University of Copenhagen. 2012. 'New international genome research centre to open in Denmark.' University of Copenhagen, 10 February. As of 9 December 2015: http://news.ku.dk/all_news/2012/2012.2/new-international-genome-research-centre-to-open-in-denmark/

Genomics England and the 100,000 Genomes Project

Caulfield, M. et al. 2015. 'The 100,000 Genomes Project Protocol.' Genomics England, February. As of 3 December 2015: http://www.genomicsengland.co.uk/wp-content/uploads/2015/03/GenomicEnglandProtocol_030315_v8.pdf

Genomics England News. 2014. 'UK to become world number one in DNA testing with plan to revolutionise fight against cancer and rare diseases.' Genomics England, 1 August. As of 1 December 2015: http://www.genomicsengland.co.uk/uk-to-become-world-number-one-in-dna-testing-with-plan-to-revolutionise-fight-against-cancer-and-rare-diseases/

Genomics England News. 2015a. '100,000 Genomes Project gains ethical approval to offer NHS patients further information about their genomic results.' Genomics England, 11 March. As of 12 November 2015: http://www.genomicsengland.co.uk/100k-genomes-projectp-gains-ethical-approval-to-offer-nhs-patients-further-information-about-their-genomic-results/

Genomics England News. 2015b. 'First patients diagnosed through the 100,000 Genomes Project.' Genomics England, 11 March. As of 12 December 2015: http://www.genomicsengland.co.uk/first-patients-diagnosed-through-the-100000-genomes-project/

Genomics England News. 2015c. 'Masters in Genomic Medicine.' Genomics England, 1 June. As of 12 December 2015: http://www.genomicsengland.co.uk/masters-in-genomic-medicine/

Genomics England News. 2015d. 'New rare disease gene tool launched – Panel App.' Genomics England, 9 September. As of 12 December 2015: http://www.genomicsengland.co.uk/rare-disease-gene-panelapp-launched/

Genomics England News. 2016. 'First children receive diagnoses through 100,000 Genomes Project.' Genomics England, 11 January. As of 22 January 2016: http://www.genomicsengland.co.uk/first-children-recieve-diagnoses-through-100000-genomes-project/

Genomics England: Data access and use. 2015. Genomics England. As of 12 December 2015: http://www.genomicsengland.co.uk/the-100000-genomes-project/faqs/data-faqs/

Genomics England: Frequently asked questions. 2015. Genomics England. As of 12 December 2015: http://www.genomicsengland.co.uk/the-100000-genomes-project/faqs/

Genomics England: How we are working with industry. 2015. Genomics England. As of 12 December 2015: http://www.genomicsengland.co.uk/working-with-industry/

Genomics England: How we are working with industry – FAQs. 2015. Genomics England. As of 12 December 2015: http://www.genomicsengland.co.uk/working-with-industry/working-with-industry-faqs/

Genomics England: Patient and Public Involvement. 2015. Genomics England. As of 12 December 2015: http://www.genomicsengland.co.uk/about-genomics-england/how-we-work/patient-and-public-involvement/

Genomics England: The 100,000 Genomes Project. 2015. Genomics England. As of 12 December 2015: http://www.genomicsengland.co.uk/the-100000-genomes-project/

Genomics England: The 100,000 Genomes Project by numbers. 2015. Genomics England. As of 12 December 2015: http://www.genomicsengland.co.uk/the-100000-genomes-project-by-numbers/

Genomics England: The process. 2015. Genomics England. As of 12 December 2015: http://www.genomicsengland.co.uk/the-100000-genomes-project/taking-part/the-process/

Human Heredity and Health in Africa Initiative (H3Africa)

Adoga et al. 2014. ‘H3Africa: a tipping point for a revolution in bioinformatics, genomics and health research in Africa.’ *Source Code for Biology and Medicine* 9(10). doi:10.1186/1751-0473-9-10

De Vries, J. et al. 2015. ‘The H3Africa policy framework: negotiating fairness in genomics.’ *Trends in Genetics* 31(3): 117–9. doi:10.1016/j.tig.2014.11.004

H3Africa Consortium. 2014. ‘Enabling the genomic revolution in Africa.’ *Science* 344(6190): 1346–8. doi:10.1126/science.1251546

H3Africa: Ethics and Governance. 2013. H3Africa. As of 2 December 2015: http://www.h3africa.org/about/ethics-and-governance

H3Africa: Frequently Asked Questions. 2013. H3Africa. As of 2 December 2015: http://www.h3africa.org/about/faq

H3Africa: Vision of the H3Africa Initiative. 2013. H3Africa. As of 2 December 2015: http://www.h3africa.org/about/vision

H3Africa Working group on Ethics and Regulatory Issues for the Human Heredity and Health (H3Africa) Consortium. 2013. ‘H3Africa Guidelines for Informed Consent.’ h3africa.org. As of 16 November 2015: http://www.h3africa.org/images/FinalPoliciesDocuments/H3A%20WG%20Guidelines%20Informed%20Consent_FINAL_01082013.pdf

Human Genome Diversity Project

Cavalli-Sforza, L. 2005. ‘The Human Genome Diversity Project: past, present and future.’ *Nature Reviews Genetics* 6(4): 333–40. doi:10.1038/nrg1596

Greely, H. T. 2001. ‘Human genome diversity: What about the other human genome project?’ *Nature Reviews Genetics* 2(3): 222–7. doi:10.1038/35056071

Human Genome Diversity (HGD) Committee of HUGO. 1993. ‘Alghero Summary Report.’ Stanford University. As of 29 November 2015: http://hsblogs.stanford.edu/morrison/files/2011/03/Alghero.pdf

Human Genome Diversity Project. 2007. Stanford University. As of 29 November 2015: http://www.hagsc.org/hgdp/

Li., J. et al. 2008. ‘Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation’. *Science* 319(5866): 1100–4. doi:10.1126/science.1153717

Morrison Institute for Population and Resource Studies. 2011. 'Human Genome Diversity Project Committee Members.' Stanford University, 10 March. As of 11 November 2015: http://hsblogs.stanford.edu/morrison/2011/03/10/human-genome-diversity-project-committee-members/

Morrison Institute for Population and Resource Studies 2015. 'Human Genome Diversity Project.' Stanford University. As of 11 November 2015: http://hsblogs.stanford.edu/morrison/human-genome-diversity-project/

The Initiative on Rare and Undiagnosed Diseases (IRUD)

Otake, T. 2015. 'Genome project aims to diagnose patients with rare diseases.' *Japan Times*, 23 July. As of 8 December 2015: http://www.japantimes.co.jp/news/2015/07/23/national/science-health/genome-project-aims-diagnose-patients-rare-diseases/#.VhaBt_NwaUl

The International Cancer Genome Consortium (ICGC)

International Cancer Genome Consortium: About. 2015. International Cancer Genome Consortium. As of 5 December 2015: https://icgc.org/icgc#about

International Cancer Genome Consortium. 2008. 'International Cancer Genome Consortium Goals, Structure, Policies & Guidelines.' icgc.org. As of 5 December 2015: https://icgc.org/icgc/goals-structure-policies-guidelines

International Cancer Genome Consortium. 2010. 'International network of cancer genome projects.' *Nature Perspectives* 464 (April): 993–8. doi:10.1038/nature08987

Ontario Institute for Cancer Research. 2015. 'ICGC launches six new projects from Australia, France, Germany, Japan and Singapore.' Canada/ICGC, 4 February. As of 25 November 2015: http://oicr.on.ca//files/public/ICGC%20launches%20six%20new%20projects.pdf

The International HapMap Consortium

International HapMap Project: Groups Participating in the International HapMap Project. 2015. nih.gov. As of 23 November 2015: http://hapmap.ncbi.nlm.nih.gov/groups.html

International HapMap Project: How are ethical issues being addressed?. 2015. nih.gov. As of 23 November 2015: http://hapmap.ncbi.nlm.nih.gov/ethicalconcerns.html

International HapMap Consortium. 2003. 'The International HapMap project.' *Nature* 426 (December): 789–96. As of 30 November 2015: http://www.nature.com/nature/journal/v426/n6968/abs/nature02168.html

International HapMap Consortium. 2005. 'A haplotype map of the human genome.' *Nature* 437 (October): 1299–320. doi:10.1038/nature04226

International HapMap3 Consortium. 2010a. 'Integrating common and rare genetic variation in diverse human populations.' *Nature* 467 (September): 52–8. doi:10.1038/nature09298

International HapMap3 Consortium. 2010b. 'Integrating common and rare genetic variation in diverse human populations: Editor's summary.' *Nature* 467 (September). As of 30 November 2015: http://www.nature.com/nature/journal/v467/n7311/abs/nature09298.html

Kuwait legislation (no project name)

Aljazeera. 2015. 'Can Kuwait justify mandatory DNA testing?' Aljazeera, 13 July. As of 16 November 2015: http://www.aljazeera.com/news/2015/07/kuwait-justify-mandatory-dna-testing-150713200708228.html

Field, D. 2015. 'Kuwait's War on ISIS and DNA.' Center for genetics and Society, 3 September. As of 9 December 2015: http://www.geneticsandsociety.org/article.php?id=8830

Human Rights Watch. 2015. 'Kuwait: New Counterterror Law Sets Mandatory DNA Testing.' hwr.org, 20 July. As of 24 November 2015: https://www.hrw.org/news/2015/07/20/kuwait-new-counterterror-law-sets-mandatory-dna-testing

National Centre for Indigenous Genomics

Indigenous Genomics. 2015. bioplatforms.com. As of 7 December 2015: http://www.bioplatforms.com/indigenous-genomics/

Indigenous Health News. 2014. 'Launch of the National centre for Indigenous Genomics.' Australian Indigenous HealthInfoNet, 19 February. As of 12 November 2015: http://www.healthinfonet.ecu.edu.au/about/news/1896

National Centre for Indigenous Genomics. 2014. Media Release: New centre to boost indigenous health research. Australian National University, 19 February. As of 12 November 2015: http://ncig.anu.edu.au/news-events/media-release

National Centre for Indigenous Genomics: Genomics and Bioinformatics. 2015. Australian National University. As of 7 December 2015: http://ncig.anu.edu.au/ncig-collection/current-projects/genomics-and-bioinformatics

National Centre for Indigenous Genomics: The NCIG Collection. 2015. Australian National University. As of 7 December 2015: http://ncig.anu.edu.au/ncig-collection

National Centre of Excellence in Research on Parkinson's Disease

Université de Luxembourg. 2015. 'A National Centre of Excellence in Research on Parkinson's disease.' Université de Luxembourg, 2 April. As of 8 December 2015: http://wwwen.uni.lu/university/news/latest_news/a_national_centre_of_excellence_in_research_on_parkinson_s_disease

Government of Luxembourg. 2015. 'Launch of National Centre of Excellence in Research in Parkinson's Disease.' innovation.public.lu, 3 April. As of 8 December 2015: http://www.innovation.public.lu/en/actualites/2015/04/centre-parkinson/index.html

Personal Genome Project

Abraham, C. 2012. 'Would you make your DNA and health data public if it may help cure disease?' *The Globe and Mail*, 7 December. As of 2 December 2015: http://www.theglobeandmail.com/news/national/time-to-lead/would-you-make-your-dna-and-health-data-public-if-it-may-help-cure-disease/article6090038/?page=1

Angrist, M. 2009. 'Eyes wide open: the personal genome project, citizen science and veracity in informed consent.' *Personalized Medicine* 6(6): 691–9. doi:10.2217/pme.09.48.

Ball, M.P. et al. 2012. 'A Public Resource Facilitating Clinical Use of Genomes.' *PNAS* 109(30): 11920–7. doi:10.1073/pnas.1201904109

Ball, M.P. et al. 2014. 'Harvard Personal Genome Project: lessons from participatory public research.' *Genome Medicine* 6(10). doi:10.1186/gm527

Brock, P.A. et al. 2012. 'Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells.' *Nature* 487 (July): 190–5. doi:10.1038/nature11236

Drmanac, R. et al. 2010. 'Human Genome Sequencing Using Unchained Base Reads on Self-Assembly DNA Nanoarrays.' *Science* 327(5961): 78–81. doi:10.1126/science.1181498

Lunshof, J.E., R. Chadwick, D.B. Vorhaus & G.M Church. 2008. 'From genetic privacy to open consent.' *Nature Reviews Genetics* 9(5). doi:10.1038/nrg2360

Lunshof, J.E. et al. 2010. 'Personal genomes in progress: from the Human Genome Project to the Personal Genome Project.' *Dialogues in Clinical Neuroscience* 12(1): 47–60. As of 18 November 2015: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181947/

Personal Genome Project: Global Network. 2015. personalgenomes.org. As of 2 December 2015: http://www.personalgenomes.org/organization/network

Personal Genome Project: Mission. 2015. PersonalGenomes.org. As of 2 December 2015: http://www.personalgenomes.org/organization/mission

Personal Genome Project: Publications. 2015. PersonalGenomes.org. As of 2 December 2015: http://www.personalgenomes.org/organization/publications

SickKids. 2012. 'Personal Genome Project Canada Launches.' SickKids, 10 December. As of 3 December 2015: http://www.sickkids.ca/AboutSickKids/Newsroom/Past-News/2012/Personal-Genome-Project-Canada-Launches.html

Precision Medicine Initiative Cohort Program (PMI)

National Institutes of Health. 2015. 'NIH forms a team of experts to chart course for the President's Precision Medicine Initiative research network.' NIH, 30 March. As of 24 November 2015: http://www.nih.gov/news-events/news-releases/nih-forms-team-experts-chart-course-presidents-precision-medicine-initiative-research-network

PMI Cohort Program. 2015. National Institutes of Health. As of 9 December 2015: http://www.nih.gov/precision-medicine-initiative-cohort-program

PMI Cohort Program: Frequently Asked Questions. 2015. National Institutes of Health. As of 9 December 2015: http://www.nih.gov/precisionmedicine/faq-cohort.htm

PMI Cohort Program: Scale and Scope. 2015. National Institutes of Health. As of 11 November 2015: http://www.nih.gov/precision-medicine-initiative-cohort-program/scale-scope

Precision Medicine Initiative. 2015. 'Precision Medicine Initiative: Proposed Privacy and Trust Principles.' The White House, 8 July. As of 12 November 2015: https://www.whitehouse.gov/sites/default/files/docs/pmi_privacy_and_trust_principles_july_2015.pdf

Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director, NIH. 2015. 'The Precision Medicine Initiative Cohort Program – Building a Research Foundation for 21st Century Medicine.' National Institutes of Health, 17 September. As of 12 November 2015: http://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf

White House. 2015. ‘Fact Sheet: President Obama’s Precision Medicine Initiative.’ The White House, 30 January. As of 22 January 2016:
https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative

The SardiNIA Study

National Institute on Aging. 2008. ‘ SardiNIA Project steps up global research in genetics, aging.’ National Institute on Aging, 1 March. As of 8 December 2015:
https://www.nia.nih.gov/newsroom/features/sardinia-project-steps-global-research-genetics-aging

Pilia, G. et al. 2006. ‘Heritability of Cardiovascular and Personality Traits in 6,148 Sardinians.’ *PLoS Genetics* 2(8): 1207–23. doi:10.1371/journal.pgen.0020132

Scuteri, A. et al. 2010. ‘The central arterial burden of the metabolic syndrome is similar in men and women: the SardiNIA Study.’ *European Heart Journal* 31: 602–13. doi:10.1093/eurheartj/ehp491

Saudi Human Genome Project

Abu-Elmagd, M. et al. 2015. ‘Individualized medicine enabled by genomics in Saudi Arabia.’ *BMC Medical Genomics* 8 (Supplement 1). doi:10.1186/1755-8794-8-S1-S3

GenomeWeb Daily News. 2013. ‘Life Tech Becomes Partner in Saudi Human Genome Program, International Barcode of Life Project.’ genomeweb.com, 9 December. As of 12 November 2015:
https://www.genomeweb.com/sequencing/life-tech-becomes-partner-saudi-human-genome-program-international-barcode-life

Saudi Human Genome Program: Data access. 2015. Saudi Human Genome Program. As of 8 December 2015:
http://shgp.kacst.edu.sa/site/Data_Access.html

Saudi Human Genome Program: Introduction. 2015. Saudi Human Genome Program. As of 8 December 2015:
http://shgp.kacst.edu.sa/site/Introduction.html

Saudi Human Genome Program (homepage). 2015. Saudi Human Genome Program. As of 8 December 2015:
http://shgp.kacst.edu.sa/site/index.html

Singapore Genome Variation Program

Singapore Genome Variation Project. 2015. National University of Singapore. As of 8 December 2015:
http://www.statgen.nus.edu.sg/~SGVP/

Teo, Y. et al. 2009. ‘Singapore Genome Variation Project: A haplotype map of three Southeast Asian populations.’ *Genome Research* 19(11): 2154–62. doi:10.1101/gr.095000

The Slim Initiative in Genomic Medicine for the Americas (SIGMA)

Carlos Slim Center for Health Research at the Broad Institute. 2015. Broadinstitute.org. As of 7 December 2015:
http://www.broadinstitute.org/scientific-community/science/programs/carlos-slim-center-health-research/carlos-slim-center-health

Carlos Slim Foundation: Our Programs. 2015. carloslslim.org. As of 7 December 2015:
http://www.salud.carlosslim.org/english2/en-cifras/

Davis, N. 2010. ‘Mexico-US collaboration launched.’ Broad Communications, 19 January. As of 25 November 2015:
http://www.broadinstitute.org/news/1405

Type 2 Diabetes Knowledge Portal: About the data. type2diabetesgenetics.org. As of 3 December 2015:
http://www.type2diabetesgenetics.org/informational/aboutthedata

Type 2 Diabetes Knowledge Portal: Policies. 2015. type2diabetesgenetics.org. As of 3 December 2015: http://www.type2diabetesgenetics.org/informational/policies

Type 2 Diabetes Knowledge Portal: SIGMAT2D. 2015. type2diabetesgenetics.org. As of 7 December 2015: http://www.type2diabetesgenetics.org/sigma/index?section=data

UK Biobank

Elliott, P. & T.C Peakman (on behalf of UK Biobank). 2008. 'The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine.' *International Journal of Epidemiology* 37(2): 234–44. doi:10.1093/ije/dym276

Sudlow, C. et al. 2015. 'UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age.' *PLoS Medicine* 12(3): e1001779. doi:10.1371/journal.pmed.1001779

UK Biobank. 2007. 'Ethics and Governance Framework Version 3.0.' ukbiobank.ac.uk. As of 5 December 2015: http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf?phpMyAdmin=trmKQlYdjjnQIgJ%2CfAzikMhEnx6

UK Biobank. 2015. 'Genotype imputation and genetic association studies of UK Biobank: Interim Data Release, May 2015.' ukbiobank.ac.uk. As of 6 December 2015: http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf

UK Biobank: About. 2015. ukbiobank.ac.uk. As of 5 December 2015: http://www.ukbiobank.ac.uk/about-biobank-uk/

UK Biobank: Public consultation. 2015. ukbiobank.ac.uk As of 5 December 2015: http://www.ukbiobank.ac.uk/public-consultation/

UK Biobank: Published papers. 2015. ukbiobank.ac.uk As of 5 December 2015: http://www.ukbiobank.ac.uk/published-papers/

UK10K

Brice, P. 2010. 'UK 10,000 human genomes project launched.' phgfoundation.org, 25 June. As of 5 December 2015: http://www.phgfoundation.org/news/5553

Burbidge, R. 2015b. 'UK 10K genomes project shares its data.' phgfoundation.org, 15 September. As of 5 December 2015: http://www.phgfoundation.org/news/16948

Ethical Advisory Group of the UK10K project. 2010. 'Ethical Governance Framework: Version 21.' uk10k.org. As of 12 November 2015: http://www.uk10k.org/assets/ef_uk10k_v21.pdf

Kaye, J. 2014. 'Managing clinically significant findings in research: the UK10K example.' *European Journal of Human Genetics*, 22: 1100–4. doi:10.1038/ejhg.2013.290

UK10K: Funding. 2011. uk10k.org. As of 5 December 2015: http://www.uk10k.org/funding.html

UK10K: Data access. 2013. uk10k.org. As of 12 December 2015: http://www.uk10k.org/data_access.html

UK10K (homepage). 2014. uk10k.org. As of 5 December 2015: http://www.uk10k.org/

UK10K Consortium. 2015. The UK10K project identifies rare variants in health and disease.' *Nature* 526 (October): 82–90. doi:10.1038/nature14962

A Weill Cornell Medical study – Exome Sequencing Identifies Potential Risks Variants for Mendelian Disorders at High Prevalence in Qatar

Rodriguez-Flores, J. et al. 2013. 'Exome Sequencing Identifies Potential Risk Variants for Mendelian Disorders at High prevalence in Qatar.' *Human Mutation* 35(1): 105–16. doi:10.1371/journal.pone.0047614

Weill Cornell Medical College in Qatar. 2014. 'Mapping the Qatari genome to prevent inherited diseases.' Weill Cornell Medical College in Qatar – Office of Public Affairs. As of 12 November 2015: https://qatar-weill.cornell.edu/media/reports/2013/qatariGenome.html

Bibliography of other works cited:

Arksey H. & L. O'Malley. 2005. 'Scoping studies: towards a methodological framework.' *International Journal of Social Research Methodology* 8(1): 19–32. doi:10.1080/1364557032000119616

Bezold, C. & J. Peck. 2005. 'Drug regulation 2056'. *Food and Drug Law Journal* 60: 127–36. As of 2 December 2015: http://heinonline.org/HOL/Page?handle=hein.journals/foodlj60&div=23&g_sent=1&collection=journals

Buchanan, C.C., E.S. Torstenson, W.S. Bush & M.D. Ritchie. 2012. 'A comparison of catalogued variation between International HapMap Consortium and 1000 Genomes Project data.' *Journal of the American Medical Informatics Association* 19(2). doi:10.1136/amiajnl-2011-000652.

Burbidge, R. 2015c. 'PHG Foundation backs data research plea.' phg foundation, 19 November. As of 2 December 2015: http://www.phgfoundation.org/news/17059

Burbidge, R. 2015d. 'Korea's largest personal genome project launched.' phg foundation, 2 December. As of 4 December 2015: http://www.phgfoundation.org/news/17074/

Centre for Reviews and Dissemination (CRD). 2009. *CRD's Guidance for Undertaking Reviews in Healthcare*. York: CRD.

Chataway, J., C. Fry, S. Marjanovic & O. Yaqub. 2012. 'Public-private collaborations and partnerships in stratified medicine: making sense of new interactions.' *New Biotechnology* 29(6): 732–40. doi:10.1016/j.nbt.2012.03.006.

Deloitte. 2015. 'Genomics in the UK: An industry study for the Office of Life Sciences.' London: Office for Life Sciences. As of 2 December 2015: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/464088/BIS-15-543-genomics-in-the-UK.pdf

Gaisser, S., et al. 2009. 'The phantom menace of gene patents.' *Nature* 458: 407–8. doi:10.1038/458407a

Gibb, F. 2015. 'Mother sues doctors who hid family brain disease.' *The Times,* 27 November. As of 2 December 2015: http://www.thetimes.co.uk/tto/law/article4625157.ece

Grada, A. & K. Weinbrecht. 2013. 'Next-Generation Sequencing: Methodology and Application.' *Journal of Investigative Dermatology* 133(e11): 1–4. doi:10.1038/jid.2013.248

Harris, E.D. n.d. 'Biochemical Facts behind the Definition and Properties of Metabolites.' Food and Drug Administration. As of 17 January 2016: http://www.fda.gov/ohrms/dockets/ac/03/briefing/3942b1_08_Harris%20Paper.pdf

Heger, M. 2013. 'Debate Heats up on ACMG's Incidental Findings Recommendations.' genomeweb, 22 March. As of 21 January 2015: https://www.genomeweb.com/sequencing/debate-heats-acmgs-incidental-findings-recommendations

Homer, N. et al. 2008. 'Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays.' *PLoS Genetics* 4(8): e1000167. doi:10.1371/journal.pgen.1000167

IGNITE (homepage). 2016. As of 12 January 2016: http://www.ignite-genomics.org/

Jaffe, A. & J. Lerner. 2004. *Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It*. Princeton: Princeton University Press

Kahn, S.D. 2011. 'On the Future of Genomic Data.' *Science* 331: 728–9. doi:10.1126/science.1197891

Kaye, J., C. Heeney, N. Hawkins, J. de Vries & P. Boddington. 2009. 'Data sharing in genomics – re-shaping scientific practice.' *Nature Reviews Genetics* 10 (May): 331–5. doi: 10.1038/nrg2573

Lu, Y., D.B. Goldstein, M. Angrist & G. Cavalleri. 2014. 'Personalized Medicine and Human Genetic Diversity.' *Cold Spring Harbour Perspectives in Medicine* 4(9): a008581. doi:10.1101/cshperspect.a008581.

Mamas, M., W. B. Dunn, L. Neyses & R. Goodacre. 2011. 'The role of metabolites and metabolomics in clinically applicable biomarkers of disease.' *Archives of Toxicology* 85(1): 5–17. doi: 10.1007/s00204-010-0609-6.

Mancheno, M. & R. Dal-Ré. 2013. 'Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process.' *European Journal of Human Genetics* 21: 1054–9. doi:10.1038/ejhg.2012.297

Manolio, T.A., M. Abramowicz, F. Al-Mulla et al. 2015. 'Global implementation of genomic medicine: We are not alone.' *Science Translational Medicine* 7(290): 1–8. doi:10.1126/scitranslmed.aab0194.

McCarthy, J.J., H.L. McLeod & G.S. Ginsburg. 2013. 'Genomic Medicine: A Decade of Successes, Challenges, and Opportunities.' *Science Translational Medicine* 5(189): 189sr4. doi:10.1126/scitranslmed.3005785.

Merges, R.P. & R.R. Nelson. 1994. 'On limiting or encouraging rivalry in technical progress: the effect of patent scope decisions.' *Journal of Economic Behaviour and Organization* 25(1): 1–24, doi:10.1016/0167-2681(94)90083-3

Middha, S., S. Baheti, S.N. Hart & J.A. Kocher. 2014. 'From Days to Hours: Reporting Clinically Actionable Variants from Whole Genome Sequencing.' *PLoS One* 9(2): 1–4. doi:10.1371/journal.pone.0086803

Morgan Jones, M., S. Castle-Clarke, D. Brooker, E. Nason, F. Huzair & J. Chataway. 2014. *The Structural Genomics Consortium: a knowledge platform for drug discovery.* Santa Monica, CA: RAND Corporation. As of 3 December 2015: http://www.rand.org/pubs/research_reports/RR512.html

Naidoo, N., Y. Pawitan, R. Soong, D. Cooper & C. Ku. 2011. 'Human genetics and genomics a decade after the release of the draft sequence of the human genome.' *Human Genomics* 5(6): 577–622. doi:10.1186/1479-7364-5-6-577

National Human Genome Research Institute. 2013. 'Quantitative Advances Since the Human Genome Project (HGP).' As of 19 November 2015: http://www.genome.gov/27553526

National Human Genome Research Institute: DNA sequencing costs. 2015. National Human Genome Research Institute, 2 October. As of 14 December 2015: http://www.genome.gov/sequencingcosts/

National Human Genome Research Institute: Genome-Wide Association Studies. 2015. National Human Genome Research Institute. As of 27 November 2015: http://www.genome.gov/20019523

National Institutes of Health. 2012. '1000 Genomes Project data available on Amazon Cloud.' NIH, 29 March. As of 19 November 2015: http://www.nih.gov/news-events/news-releases/1000-genomes-project-data-available-amazon-cloud

NCI Dictionary of Cancer Terms: Proteomic Profile. 2016. National Cancer Institute. As of 17 January 2016: http://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=367447

Nelson, R.R. & B.N. Sampat. 2001. 'Making sense of institutions as a factor shaping economic performance.' *Journal of Economic Behaviour and Organization* 44(1): 31–54.

NIH News. 2013. 'NHGRI celebrates 10th anniversary of the Human Genome Project.' National Human Genome Research Institute, 12 April. As of 19 November 2015: http://www.genome.gov/27553526

Petrini, C. 2010. '"Broad" consent, exceptions to consent and the question of using biological samples for research purposes different from the initial collection purpose.' *Social Science and Medicine* 70(2): 217–20. doi:10.1016/j.socscimed.2009.10.004

Regalado, A. 2014. 'EmTech: Illumina Says 228,000 Human Genomes Will Be Sequenced This Year.' MIT Technology Review, 24 September. As of 20 December 2015: http://www.technologyreview.com/news/531091/emtech-illumina-says-228000-human-genomes-will-be-sequenced-this-year/

Roadmap Epigenomics Project (homepage). 2010. As of 12 January 2016: http://www.roadmapepigenomics.org/

Stein, L.D., B.M. Knoppers, P. Campbell, G. Getz & J.O. Korbel. 2015. 'Data analysis: Create a cloud commons.' *Nature* 523: 149–51. doi:10.1038/523149a

'Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing.' 1996. 25–26 February, Bermuda, as reported by HUGO. As of 16 November 2015: http://www.casimir.org.uk/storyfiles/64.0.summary_of_bermuda_principles.pdf

Taylor, N.P. 2015. 'Israel mulls setting up a population-scale genetic database.' Fierce BioTech IT, 19 May. As of 2 December 2015: http://www.fiercebiotechit.com/story/israel-mulls-setting-population-scale-genetic-database/2015-05-19

Visscher, P.M., M.A. Brown, M.I. McCarthy & J. Yang. 2012. 'Five Years of GWAS Discovery.' *American Journal of Human Genetics* 90: 7–24. As of 2 December 2015: http://www.sciencedirect.com/science/article/pii/S0002929711005337

Wolf, S.M. 2012. 'The past, present, and future of the debate over return of research results and incidental findings.' *Genetics in Medicine* 14(4): 355–7. doi:10.1038/gim.2012.26

Appendix

Table 2. Aims and objectives across profiled initiatives

Name of initiative	Aims and objectives
Human Genome Diversity Project (1991–date unknown)	• Advancing the knowledge base on genetic variation within and across populations; • Development of new tools and methods for genetic studies; • Furthering the cultural and anthropological knowledge base.
The International HapMap Consortium (2002–2009)	• Advancing the knowledge base on genetic variation within and across populations; • Development of new tools and methods for genetic studies; • Furthering the cultural and anthropological knowledge base.
Global Network of Personal Genome Projects (2005–ongoing)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Management and governance capacity building; • Development of new tools and methods for genetic studies.
1000 Genomes Project (2008–2015)	• Advancing the knowledge base on genetic variation within and across populations; • Development of new tools and methods for genetic studies.
H3Africa (2010–ongoing)	• Advancing the knowledge base on genetic variation within and across populations; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Human resource capacity building for research; • Management and governance capacity building; • Furthering the cultural and anthropological knowledge base.
The African Genome Variation Project (2011–2014)	• Advancing the knowledge base on genetic variation within and across populations; • Development of new tools and methods for genetic studies; • Human resource capacity building for research; • Furthering the cultural and anthropological knowledge base.
CHARGE Consortium (2007–ongoing)	• Development of new tools and methods for genetic studies; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination.
ICGC (2007–ongoing)	• Catalysing translation and coordination; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment.
deCODE genetics (dates unknown, but findings published 2015)	• Advancing the knowledge base on genetic variation within and across populations; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment.

Name of initiative	Aims and objectives
The Estonian Biobank/ (EGCUT) (2000–ongoing)	• Advancing the knowledge base on genetic variation within and across populations; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination.
Singapore Genome Variation Program (dates unknown, but findings published 2009)	• Advancing the knowledge base on genetic variation within and across populations; • Development of new tools and methods for genetic studies.
GoNL (dates unknown, but findings published 2014)	• Advancing the knowledge base on genetic variation within and across populations; • Development of new tools and methods for genetic studies.
GenomeDenmark (2012–ongoing)	• Advancing the knowledge base on genetic variation within and across populations; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination; • Physical capacity building.
FarGen (2013–ongoing)	• Advancing the knowledge base on genetic variation within and across populations; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination.
Cymru DNA Wales project (2014–ongoing)	• Furthering the cultural and anthropological knowledge base; • Other (commercial).
National Centre for Indigenous Genomics (2014–ongoing)	• Advancing the knowledge base on genetic variation within and across populations; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Furthering the cultural and anthropological knowledge base.
Kuwait legislation (planned 2015–2016)	• Other (national security, surveillance).
PMI Cohort Program (2015–ongoing)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination; • Management and governance capacity building.
SardiNIA (2001– date unknown)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment.
China Kadoorie Biobank (2004–ongoing)	• Advancing the knowledge base on genetic variation within and across populations; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment.
UK Biobank (2006–ongoing)	• Advancing the knowledge base on genetic variation within and across populations; • Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Management and governance capacity building.

Name of initiative	Aims and objectives
SIGMA (2010–ongoing)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Human resource capacity building for research.
UK10K Project (2010–2015)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Management and governance capacity building.
DDD Study (2011–2016)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Management and governance capacity building.
Genomics England (2012–2017)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination; • Management and governance capacity building.
A Weill Cornell Medical Study (dates unknown, but findings first published 2012)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment.
Saudi National Genome Program (2013–ongoing)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination; • Physical capacity building.
BeMGI (2013–ongoing)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination; • Management and governance capacity building.
The Initiative on Rare and Undiagnosed Diseases (2015–ongoing)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment.
National Centre for Excellence in Research in Parkinson's Disease (2015–ongoing)	• Enriching disease specific, clinically relevant insights relating to diagnosis, risk prediction or treatment; • Catalysing translation and coordination.

Table 3. Scope and scale of activities across profiled initiatives

Name of initiative	Geographic data scope	Number of individuals' genomes sequenced	Research focus	Linked data
Human Genome Diversity Project (1991– date unknown)	International: *51 populations from around the world*	1,403 individuals	General	Only basic information on sex and population origin
International HapMap Consortium (2002–2009)	International: *11 populations from around the world*	1,184 individuals	General	Only basic information on sex and population origin, as well as on family structure if coming from a parent-child trio
Global Network of Personal Genome Projects (2005–ongoing)	International: *U.S., Canada, UK and Austria*	Sequencing is ongoing. U.S and Canada aim to sequence genomes from 100,000 individuals each	General	Yes: participants volunteer to share extensive phenotypic, clinical and lifestyle data, as well as omics-related analyses
1000 Genomes Project (2008–2015)	International: *26 populations from around the world*	2,504 individuals	General	Only basic information on sex and population origin
H3Africa (2010–ongoing)	International: *African continent*	Aims to sequence 50,000 to 75,000 genomes	General	*Information not available*
The African Genome Variation Project (2011–2014)	International: *Sub-Saharan Africa*	1,481 individuals representing 18 ethno-linguistic groups	General	Only basic information on sex, population origin and language group
CHARGE Consortium (2007–ongoing)	International: *Data come from a voluntary federation of large, complex cohort studies*	In 2009 the consortium included GWAS data on about 38,000 individuals	Cardiovascular and ageing-related diseases	Yes: the individual cohort studies have collected health-related phenotypic data measured in similar ways
ICGC (2007–ongoing)	International: *Data come from studies worldwide*	Sequencing is ongoing and will be performed on at least 27,000 samples that have already been collected. So far, samples from over 12,807 donors have been sequenced	Cancer	Only limited phenotypic/ clinical data is linked to samples. A potential ICGC2 project would link more extensive patient data
deCODE genetics (dates unknown, but findings published in 2015)	National: *Iceland*	Whole-genome sequencing data have been obtained from 2,636 individuals. The DNA sequences of over 100,000 individuals were then imputed	General	Yes: sequencing data is combined with very detailed clinical, phenotypic and drug-reaction data

Name of initiative	Geographic data scope	Number of individuals' genomes sequenced	Research focus	Linked data
The Estonian Biobank/ EGCUT (2000–ongoing)	National: *Estonia*	2,000 whole genome and 3,000 exome sequences so far (as well as additional RNA and associated sample data)	General	Yes: extensive baseline data collected; data are continuously updated through re-contacting projects and through links with other electronic databases
Singapore Genome Variation Program (dates unknown, but findings published 2009)	National: *Singapore*	268 individuals from three ethnic groups in Singapore of Chinese, Malay and Indian descent, respectively	General	*Information not available*
GoNL (dates unknown, but findings published 2014)	National: *Netherlands*	750 genomes	General	Yes: all participants are part of active biobanks, so there is extensive phenotypic and lifestyle factor information available
Genome Denmark (2012–ongoing)	National: *Denmark*	The Danish Reference Genome Project will require the sequencing of 150 genomes; so far 30 have been published.	One pilot study has general research aims; the other is focused on cancer research	*Information not available*
FarGen (2013–ongoing)	National: *Faroe Islands*	Sequencing work has not yet begun. Aims to eventually sequence the genomes of all Faroese citizens	General	Yes: data will be linked to clinical records
Cymru DNA Wales (2014–ongoing)	National: *Welsh people, living both in and outside of Wales*	*Information not known*	General	No
National Centre for Indigenous Genomics (2014–ongoing)	National: *Australia*	Around 7,000 pre-existing DNA samples from indigenous Australians will be sequenced	General	*Information not available*
Kuwait legislation (planned 2015–2016)	National: *Kuwait*	All 1.3 million citizens and 2.9 foreign residents	N/A	*Information not available*
PMI Cohort Program (2015–ongoing)	National: *U.S*	Aims to sequence the genomes of 1 million Americans	General	Yes: clinical and phenotypic, as well as lifestyle and environmental exposures, possibly

Name of initiative	Geographic data scope	Number of individuals' genomes sequenced	Research focus	Linked data
SardiNIA (2001–date unknown)	National: *Sardinia*	Participants totalled 6,148 individuals. It is, however, not clear how many of these had their genomes sequenced	Ageing-related diseases	Yes: baseline data is collected on 98 traits
China Kadoorie Biobank (2004–ongoing)	National: *China*	32,000 individuals have been genotyped so far. Second phase will extend genotyping to 100,000 individuals by mid-2016.	Common chronic diseases	Yes: very detailed information collected, through baseline surveys, periodic re-surveys and through linkage with death, disease and insurance registries
UK Biobank (2006–ongoing)	National: *UK*	Samples from all 500,000 donors should be genotyped by the end of 2015.	Middle- and old-age related diseases	Yes: very detailed data. Repeat measures are also taken, and donors are invited to contribute further datasets
SIGMA (2010–ongoing)	International: *Individuals of Mexican/Latin American descent*	Genotyping data has been obtained from 8,214 individuals; whole exome data from 3,756 individuals	Cancer, diabetes and kidney disease	*Information not available*
UK10K Project (2010–2015)	National: *UK*	Whole genomes sequences have been obtained from 4,000 individuals; whole-exome sequences have been obtained from 6,000.	Age-related diseases and other severe medical conditions	Yes: detailed phenotypic data
DDD Study (2011–2016)	National: *UK*	Over 12,000 undiagnosed children and adults in the UK with developmental disorders and their parents have had their genomes sequenced	Developmental disorders	Yes: clinical and phenotypic information as well as antenatal and growth data, developmental milestones and family history
Genomics England (2012–2017)	National: *UK*	Aims to sequence 100,000 genomes	Rare diseases and cancer	Yes: clinical and phenotypic
A Weill Cornell Medical Study (dates unknown, but findings published 2012)	National: *Qatar*	100 Qatari individuals representing the three major subgroups: the Bedouin; those of Persian-South Asian descent; and those of African descent	Mendelian diseases	*Information not available*
Saudi National Genome Program (2013–ongoing)	National: *Saudi Arabia*	Aims to sequence 100,000 genomes	Multiple diseases	Yes
BeMGI (2013–ongoing)	National: *Information not available*	*Information not available*	Multiple diseases	*Information not available*

Name of initiative	Geographic data scope	Number of individuals' genomes sequenced	Research focus	Linked data
The Initiative on Rare and Undiagnosed Diseases (2015–ongoing)	National: *Japan*	Pilot will study 1,000 genomes	Undiagnosed, rare, early-onset diseases	Yes
National Centre for Excellence in Research in Parkinson's Disease (2015–ongoing)	National (primarily): *Luxembourg and neighbouring countries*	*Information not available*	Parkinson's disease	*Information not available*

Table 4. Funding and partnership arrangements

Name of initiative	Type of partnership	Partners	Funders
Human Genome Diversity Project (1991–date unknown)	International: *Academic and research institutes*	Primarily Stanford University and the Centre d'étude du polymorphisme humain (CEPH); as well as other international researchers	Public-charity: *NIH, National Research Council, MacArthur Foundation, and others*
International HapMap Consortium (2002–2009)	International: *Academic and research institutes; industry*	Various universities; Wellcome Trust Sanger Institute; Cold Spring Harbour Laboratory; Broad Institute; Baylor College of Medicine; BGI; Illumina	Public-charity: *Various government agencies, the Wellcome Trust, W.M. Keck Foundation, Delores Dore Eccles Foundation, the SNP Consortium*
Global Network of Personal Genome Projects (2005–ongoing)	International: *Academic and research institutes; hospital; private clinic*	Harvard University; University College London; the University of Toronto's McLaughlin Centre; The Hospital for Sick Children; The MedCan Clinic, Toronto; CeMM Research Center of Molecular Medicine of the Austrian Academy of Sciences	Private-charity: *Foundations, companies and private donors (further details not known)*
1000 Genomes Project (2008–2015)	International: *Academic and research institutes; industry*	NHGRI's large-scale sequencing centres; the Wellcome Trust Sanger Institute; BGI Shenzhen; the Max Planck Institute for Molecular Genetics in Berlin; Illumina; and others	Public-charity: *Various governmental agencies and foundations*
H3Africa (2010–ongoing)	International: *Academic and research institutes*	The Wellcome Trust; NIH; African Society of Human Genetics; various African research centres	Public-charity: *Funded primarily by the Wellcome Trust and NIH*
The African Genome Variation Project (2011–2014)	International: Academic and research institutes	African Partnership for Chronic Disease Research; the Wellcome Trust Sanger Institute; Centre for Research on Genomics and Global Health; 1000 Genomes Project; and others	Public-charity: *Wellcome Trust; Bill and Melinda Gates Foundation; NIH; UK Medical Research Council*

Name of initiative	Type of partnership	Partners	Funders
CHARGE Consortium (2007–ongoing)	International*: Formed of pre-existing cohort studies*	Original members include: the Age, Gene/Environment Susceptibility-Reykjavik Study; the Atherosclerosis Risk in Communities Study; the Cardiovascular Health Study; the Framingham Heart Study; and the Rotterdam Study	Public: *Each cohort study is financed separately, by various sources of public funding*
ICGC (2007–ongoing)	International: *Collaborating studies span 16 countries worldwide*	There are currently 78 projects included within in the consortium	Public-private: *Research members are responsible for securing their own funds*
deCODE genetics (dates unknown but findings published 2015)	International: *Industry*	deCODE genetics; with support from Illumina	Private: *deCODE genetics (subsidiary of Amgen)*
The Estonian Biobank/ (EGCUT) (2000–ongoing)	National: *Academic and research institutes*	The Estonian Genome Centre at the University of Tartu (EGCUT); the Estonian Biobank; (sequencing is outsourced to the Broad Institute)	Public: *Estonian Ministry of Social Affairs and Ministry of Education and Research*
Singapore Genome Variation (dates unknown, but findings published 2009)	National: *Academic institute*	National University of Singapore	Public: *University institutes and departments*
GoNL (dates unknown, but findings published 2014)	National: *Academic institutes*	Various Dutch universities (UMCG, LUMC, Erasmus MC; VU University, and AMCU); (sequencing work sub-contracted to BGI Hong Kong)	Public: *Netherlands Organization for Scientific Research*
Genome Denmark (2012–ongoing)	International: *Academic and research institutes; industry*	Four Danish universities (KU, AU, DTU and AAU); two hospitals (Herlev and Vendsyssel); and two private firms (Bavarian Nordic and BGI-Europe)	Public-private: *Funded primarily by the Innovation Fund Denmark, but with contributions from the private partners too*
FarGen (2013–ongoing)	*Information not available*	*Information not available*	Public: *Some government funding already promised*
Cymru DNA Wales (2014–ongoing)	National: *Industry*	S4C, the Western Mail, the Daily Post, Green Bay Media and research company ScotlandsDNA	Private: *Private sponsorship and consumer fees*
National Centre for Indigenous Genomics (2014–ongoing)	National: *Academic and research institutes*	National Centre for Indigenous Genomics (NCIG) at ANU; the Kinghorn Centre for Clinical Genomics at the Garvan Institute; and the National Computational Infrastructure (NCI)	Public-private: *Australian National University; Bioplatforms Australia*
Kuwait legislation (planned 2015–2016)	*Information not available*	*Information not available*	*Information not available*

Name of initiative	Type of partnership	Partners	Funders
PMI Cohort Program (2015–ongoing)	Information not available: *Academic and research institutes; industry*	Will partner with existing cohorts, patient groups, and the private sector	Public: *NIH*
SardiNIA (2001–date unknown)	International: *Academic and research institutes*	National Institute of Aging (NIA); the Institute of Neurogenetics and Neuropharmacology of the Italian Research Council, University of Michigan	Public: *US National Institute of Aging*
China Kadoorie Biobank (2004–ongoing)	International: *Academic and research institutes; industry*	University of Oxford's Clinical Trial Service Unit & Epidemiological Studies Unit; the Chinese Academy of Medical Sciences; BGI; Fudan University; The George Institute for Global Health; International Agency for Research on Cancer; Lund University; Oulu University; Peking University	Public-charity: *Various private foundations and research charities as well as Chinese government agencies*
UK Biobank (2006–ongoing)	National: *Academic and research institutes; NHS*	Hosted by the University of Manchester; with collaborators at other UK universities; supported by the NHS	Public-charity: *Various UK government agencies; the Wellcome Trust; British Heart Foundation; Diabetes UK*
SIGMA (2010–ongoing)	International: *Academic and research institutes*	Carlos Slim Health Institute; Mexican National Human Genome Research Institute; the Broad Institute	Charity: *Carlos Slim Foundation*
UK10K Project (2010–2015)	National: *Academic and research institutes*	Bristol University; King's College London; Wellcome Trust Sanger Institute; University of Cambridge; and others	Public-charity: *Wellcome Trust, Medical Research Council and UK Department of Health*
DDD Study (2011–2021)	National: *Academic and research institutes*	NHS Regional Genetics Services and the Wellcome Trust Sanger Institute	Public-charity: *Health Innovation Challenge Fund (Wellcome Trust and the UK Department of Health); Wellcome Trust Sanger Institute; NHS National Institute for Health Research*
Genomics England (2012–2017)	National: *Academic and research institutes; industry*	NHS England; Health Education England; Public Health England; and 73 NHS Trusts and hospitals across England. Industry partners include Cognizant, AstraZeneca, Biogen, GSK and others.	Public: *Genomics England is a registered company entirely owned by the UK Department of Health*
A Weill Cornell Medical Study (dates unknown, but findings published in 2013)	International: *Academic and research institutes; industry*	Weill Cornell Medical College in Qatar (WCMC-Q); Weill Cornell Medical College New York (WCMC-NY); Cornell University in Ithaca and Hamad Medical Corporation	*Information unavailable*

Name of initiative	Type of partnership	Partners	Funders
Saudi National Genome Program (2013–ongoing)	National: *Research institutes; industry*	Ten national research centres (with an additional five to be built in the future); Life Technologies	Public: *King Abdulaziz City for Science and Technology*
BeMGI (2013–ongoing)	International: *Academic and research institutes*	Scientists from several Belgian universities, as well as two professors from universities in the Netherlands and Australia	Public: *Phase VII Interuniversity Attraction Poles (IAP) programme of the Belgian Federal Science Policy Office*
The Initiative on Rare and Undiagnosed Diseases (2015–ongoing)	National: *Academic and research institutes*	Japan Agency for Medical Research and Development; the National Center for Child Health and Development; and three Japanese universities	Public: *Japan Agency for Medical Research and Development (AMED)*
National Centre for Excellence in Research in Parkinson's Disease (2015–ongoing)	International: *Academic and research institutes; hospital*	Five national hospitals, academic and research institutes as well as the Oxford Parkinson's Disease Centre, the Hertie-Institut für klinische Hirnforschung in Tübingen, the Paracelsus-Elena-Klinik in Kassel and NIH in the USA	Public: *Luxembourg National Research Fund*

Table 5. Data ownership, management and ethics

Name of initiative	Data access and management	Participant consent	Feedback of results to donors
Human Genome Diversity Project (1991– date unknown)	Managed-access: HGDP opposes the patenting of DNA, therefore DNA from samples is made available to non-profit researchers at cost, on the condition that they submit their results to an open-access CEPH database for other researchers to use	Group informed consent	N/A
International HapMap Consortium (2002–2009)	Open-access: All data rapidly released into the public domain through an online portal. Only condition for access is that users must agree not to restrict use of the HapMap data by others. Any patents sought on downstream discoveries must not block others' access to the HapMap data	Individual informed consent accompanied by community engagement processes	N/A
Global Network of Personal Genome Projects (2005–ongoing)	Open-access: Datasets and tissues are made publically and freely available with little or no restrictions on access	Informed 'open' consent based on exam-tested knowledge	Full feedback of results
1000 Genomes Project (2008–2015)	Open-access: All data made freely and publically available upon completion; cell lines are also made available to researchers from the Coriell Cell Repository	Informed consent based on context-specific ethical guidelines	N/A

Name of initiative	Data access and management	Participant consent	Feedback of results to donors
H3Africa (2010–ongoing)	Managed-access: Data will be made available through the EGA, where public access is controlled by a Data Access Committee. Current policy states that access to biospecimens will be controlled by a Data and Biospecimen Access Committee. However, delayed release provisions' will allow African researchers to be the first to use and analyse data and samples	Informed consent based on context-specific ethical guidelines	Community-level CSFs will be communicated to the relevant community. Policy on individual incidental findings has not yet been determined
The African Genome Variation Project (2011–2014)	Open-access: Data will be made available to researchers through the EGA and the H3Africa Bionet	Informed consent	N/A
CHARGE Consortium (2007–ongoing)	Open-access: The consortium follows NIH GWAS policy on intellectual property, i.e. genotype-phenotype associations must be made publically available to all researchers	Written informed consent	*Information not available*
ICGC (2007–ongoing)	Open and controlled access: Researchers may not claim for intellectual property rights on primary data. A two-tiered data access system makes non-identifiable data publically available while access to germline genomic and detailed clinical data associated with unique individuals is managed by a Data Access Compliance Office	Informed consent from donors or their family	*Information not available*
deCODE genetics (dates unknown but findings published 2015)	Not publically available: Icelandic law prohibits the release of individual level and personally identifying data. However, data is being shared with Iceland's healthcare system, and collaborators performing meta-analyses have access to summary-level statistics or are able to travel to deCODE's facilities for local data access	Informed consent	No
The Estonian Biobank/ EGCUT (2000–ongoing)	Managed-access: Researchers may apply for access to data and DNA samples, which will be provided at cost and only for approved research projects. Any results obtained from the data must be submitted to the EGCUT database within a specified period of time	Broad consent	Full feedback of results
Singapore Genome Variation (dates unknown, but findings published 2009)	Open-access: Data publically available online	Ethical approval to extend original study on drug response was granted by two independent Institutional Review Boards	N/A
GoNL (dates unknown, but findings published 2014)	Open-access: Data has been made publically available online	Managed through the individual biobanks that contributed samples	Managed through the individual biobanks that contributed samples

Name of initiative	Data access and management	Participant consent	Feedback of results to donors
Genome Denmark (2012–ongoing)	Not publically available: Data will only be accessible by researchers on the project and certified partners. IP will be shared between the public research facilities and private partners. Sequencing data will not be patentable but downstream inventions will be	Informed consent	Only findings related to clinically significant, actionable diseases may be reported
FarGen (2013–ongoing)	Access restricted by national laws: Faroese law requires DNA data to be kept in a secure database, only accessible by doctors who have reason to access it for clinical reasons	*Information not available*	Not yet determined
Cymru DNA Wales (2014–ongoing)	*Information not available*	Donors buy their own 'spit kits' and send in their samples for analysis	Donors may receive information on their ancestry and carrier status for various non-clinical traits
National Centre for Indigenous Genomics (2014–ongoing)	Managed-access: Sequence data will be made available to researchers for approved research projects	Informed consent based on community engagement	*Information not available*
Kuwait legislation (planned 2015–2016)	*Information not available*	Testing will be mandatory; non-compliance will incur a fine and/or prison sentence	*Information not available*
PMI Cohort Program (2015–ongoing)	Managed access: PMI Working Group recommends that various levels of access are designed for data of different levels of sensitivity. Researchers using PMI data will be required to make their results publically accessible	Dynamic, ongoing, informed consent processes	Full feedback of results
SardiNIA (2001–date unknown)	Managed-access: Bona fide researchers may apply for access to the data, which is available through an online portal	Informed consent	*Information not available*
China Kadoorie Biobank (2004–ongoing)	Managed-access: Secure online portal (CKB Data Access System) allows researchers access to the data for approved research purposes. Data is currently available to Chinese researchers and will be made available to the rest of the world on 1 January 2016	Informed broad written consent	*Information not available*
UK Biobank (2006–ongoing)	Managed-access: Access to data and/or samples is granted as a single-use license for scientifically and ethically approved research projects, and for a fee. Users must subsequently publish all results in the UK biobank database	Informed broad consent	Only baseline measurements are communicated (e.g. no genomic information)
SIGMA (2010–ongoing)	Managed-access: Some data on diabetes patients has been made available through an online portal managed by the Broad Institute's Data Coordinating Centre	*Information not available*	*Information not available*

Name of initiative	Data access and management	Participant consent	Feedback of results to donors
UK10K Project (2010–2015)	Managed-access: Data is submitted to the EGA where access to sequence and phenotype data is managed through a Data Access Committee	Informed consent	Clinically significant findings may be communicated under certain conditions
DDD Study (2011–2016)	Managed-access: Summary linked-anonymized data will be displayed on the DECIPHER database. More detailed data will be submitted to the EGA where it will be made available to bona fide researchers through a managed access mechanism	Informed consent is given by participants or their parent/ legal guardian	Only findings related to the original research aims are reported
Genomics England (2012–2017)	Managed-access: Data will be kept within Genomics England's data structures. Access will be given to bona fide researchers, following a vetting process. NHS clinicians will also have access to data on participants in their care. Industry will need to pay a fee for data access. As an additional security measure, access is likely to be restricted to on-site presence of the researcher at the premise where the data is stored.	Informed consent	Relevant, and certain clinically significant incidental findings, may be communicated to individuals
A Weill Cornell Medical Study (dates unknown, but findings published 2012)	*Information not available*	Written informed consent	*Information not available*
Saudi National Genome Program (2013–ongoing)	Managed-access: Different levels of access will apply. Researchers will be able to apply for access to controlled-access data and will have to agree to several legally binding conditions	*Information not available*	*Information not available*
BeMGI (2013–ongoing)	*Information not available*	*Information not available*	*Information not available*
The Initiative on Rare and Undiagnosed Diseases (2015–ongoing)	*Information not available*	*Information not available*	*Information not available*
National Centre for Excellence in Research in Parkinson's Disease (2015–ongoing)	*Information not available*	*Information not available*	*Information not available*

Table 6. List of search terms

Search term category	List of search terms used
1	(national OR international OR country-based)
2	(population OR population based OR population-based OR population scale OR population-scale)
3	(genomics OR genome OR genome sequencing OR genome-sequencing OR sequencing OR genomic medicine OR personalised medicine OR personalized medicine OR precision medicine OR stratified medicine)
4	(initiative OR project OR program OR programme OR study OR collaboration OR research)

The logic links between the categories should be the following:

1 AND 2 AND 3 AND 4