1 Introduction

The Human Genome Project took approximately fourteen years to complete,[1] involved collaboration between twenty different centres based in six different countries,[2] and cost United States tax payers alone approximately $3 billion.[3] When James Watson's genome was sequenced in 2007 it cost approximately $1 million.[4] In 2009 a company called Complete Genomics announced that it would be able to read entire human genomes for $5,000.[5] Sequencing technology is now within the reach of many researchers and the availability of cheap sequencing is continuing to spread. At least one of the companies currently offering cancer genome analysis for research purposes is now reported to be planning to offer the service to patients and their doctors.[6]

As genetic testing enters primary healthcare there is the potential for large-scale, systematic collection of genetic data.[7] That data will be valuable for research purposes and questions about secondary uses of the data will have to be addressed. This development in the primary healthcare context will be taking place at the same time as there is an unprecedented growth in biobanks and research collections of genetic

[1] The project started in 1990 and the finished version of the euchromatic human genome was published in 2004. International Human Genome Sequencing Consortium, 'Finishing the euchromatic sequence of the human genome', *Nature* 431 (21 October 2004), 931–45.
[2] International Human Genome Sequencing Consortium, 'Finishing the euchromatic sequence of the human genome', 931–45.
[3] National Human Genome Research Institute, National Institute for Health, 'The Human Genome Project completion: frequently asked questions' (30 October 2010). www.genome.gov/11006943.
[4] P. Aldhous, 'Genome sequencing falls to $5000', *New Scientist* (6 February 2009).
[5] Aldhous, 'Genome Sequencing falls to $5000', although the question of whether routine genome sequencing can yet be done quite that cheaply may be debated. One laboratory that routinely sequences cancer genomes estimates the cost to be approximately $30,000. See E. Singer, 'Cancer genomics: Deciphering the genetics behind the disease', *Technology Review* (1 June 2011).
[6] Singer, 'Cancer genomics'.
[7] It is predicted that, for one type of cancer sufferer at least, genetic testing will become a routine part of informing patient care. See Singer 'Cancer genomics'.

data. It was estimated in 2009 that there were over 400 biobanks in Europe alone.[8] That was likely to be a conservative estimate even at the time.[9] As research biobanks gather strength, and the prospect of the regular and widespread collection of genetic data within the primary care context edges ever closer, individuals are themselves accessing direct-to-consumer genetic tests and they are beginning to publish the genetic information they discover online. Websites such as 23andme provide a personal genome service, offering a genetic testing service for over 100 different traits and diseases as well as information about ancestry.[10] Other sites, such as the Personal Genome Project, seek to recruit individuals willing to share genetic information openly through public profile pages.[11] All of this data is of potential significance within its original context, and with the possibilities of widespread access, relative permanence and increased future inter-operability between information platforms, the significance of the data shared now is likely only to grow over time.

There are international collaborations, such as P^3G,[12] that have as a core goal the facilitation of collaboration between biobanks. As genetic data research benefits from very large-scale datasets, it will undoubtedly be through such international efforts, and large-scale collaborations, that much of the promise of genetic research and future improvement in healthcare will be delivered. Alongside the hope and optimism there are, of course, concerns. Within the context of informational privacy, these might be divided crudely into two categories: those associated with unwanted access and those associated with unwanted uses of data. Security breach and unauthorised acquisition of genetic data are, of course, always a possibility. Unwanted access can, however, extend beyond the unauthorised.

Whether particular access, to particular data, for particular purposes is considered desirable is often a value judgment and is likely to be subject to a variety of, at times conflicting, views. These views might not even be stable over time, as people's attitudes are affected by personal experiences, the reported experiences of others and other kinds of education (and perhaps even misinformation). In a situation characterised by complex, conflicting and potentially unstable

[8] Editorial, 'Biobanks need pharma', *Nature* 461 (24 September 2009), 448.
[9] Over 400 biobanks are currently (June 2011) listed on the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) catalogue. www.bbmri.eu.
[10] 23andme.com. [11] www.personalgenomes.org.
[12] Public Population Project in Genomics. www.p3gobservatory.org.

preferences, how does one determine the 'proper' access to genetic data for the purposes of health research?

In recent history, the preferred answer has tended to be that, whatever the potential benefits for others or society more generally, an individual's own preferences on participation in health research should be respected. There are many good reasons for the dominance that the doctrine of informed consent has come to assume in this context. It needs to be expressly recognised, however, that – perhaps particularly in the case of health research using genetic data – there are limits to the ability of 'informed consent' to account adequately for *all* relevant preferences. There will, for example, be occasions when an individual's informed consent cannot be sought for practical reasons. If an individual cannot be asked to express a preference, then 'informed consent' is an inadequate tool by which her preferences might be judged. Alternatively, it might be that, at the time the data is collected, there is no more than the vaguest notion of what, precisely, it might be used for and to whom access might need to be granted. Any consent obtained in such circumstances can only be broad and general, and effectively incapable of expressing any detailed preferences on questions not yet considered. There may also be other occasions where the interests of others are so significantly implicated in the choices concerning access and use of data that to pretend the individual from whom consent is sought would be expressing their preference in a social vacuum would be to betray the interests, and potentially deny the preferences, of others.

These challenges to the adequacy of informed consent to provide a complete answer to the question of how one tackles the difficulties of conflicting preferences, and account for the interests of all implicated by decisions on research use of genetic data, are not posed *exclusively* by research use of genetic data. There is much about genetic data, however, as well as the kind of research that can make effective use of it, which makes these challenges particularly acute.

Genetic data

Genetic data is a difficult term to pin down. Throughout subsequent analysis, I will be relying upon a broad understanding of the idea. In fact, the conception of genetic data that I will be relying upon is broad enough to encompass many different forms of data, each capable of yielding many different kinds of personal information beyond that which might typically be described as 'genetic information'. Given subsequent use of this broad description of genetic data, one might wonder why I persist in

maintaining a focus upon *genetic* data at all: why not broaden the analysis to include all personal data, or at least, all personal health information? It will be seen that I do in fact tend to a particular position on the concept of 'personal data' as such. Analysis of the term 'personal data' is, however, best done in a context and genetic data provides a very suitable context in which to study the privacy protection currently provided to personal health data in a research context.

Genetic data's suitability as a case-study depends not only on the fact that decisions about research use can have implications that extend far beyond the individual research participant. The critique of privacy protection that I will advance relies upon recognising a particular relationship between the ideas of 'data' and 'information.' Specifically, I will describe information as a composite concept: the generation of information relies upon a particular interpretive framework being applied to data. Understanding the significance of fluid interpretive frameworks to the relationship between data and information is key to understanding the limitations associated with the law's current protection of privacy. The *distinction* between data and information is a distinction that genetic data is well placed to illustrate.

Genetic data provides an excellent example of data that might often, and plausibly, be placed within multiple, shifting, interpretive frameworks. The same genetic data might, in different contexts, over different periods of time, come to be understood to provide information about many different things, relating to many different persons. It is this *interpretive potential* of genetic data that helps to demonstrate the limitations of the current regulatory system as well as to understanding the multitude of different preferences that might be expressed regarding its access and use. In addition to its suitability as a vehicle to drive a more general critique, there are other reasons why genetic data in particular might provide a particularly suitable case-study.

Genetic science, particularly during the twentieth Century, came to be associated with some very dark moments in human history. The legacy of eugenic programmes, associated in the minds of many primarily with Nazi Germany but supported in less extreme form across Europe and North America, has cast a long shadow over science and the invocation of Science to support public policy agendas. What is more, through some more recent, and quite remarkable, achievements in genetic science – often associated with the hype that can accompany major public investment – research using genetic data has consolidated a contemporary perception that has heightened the significance associated with its access and use.

It is important that genetic data is seen to be subject to appropriate privacy protection. It is important that there is appropriate privacy protection not only for the sake of those individuals whose privacy might otherwise be infringed. It is important that genetic data is *seen to be* subject to appropriate privacy protection for the sake of those who rely upon participants' trust in the security and integrity of the research process. Research using genetic data is often reliant upon voluntary contribution and if participants lose confidence in the ability of the regulatory system to ensure their privacy is suitably assured, then this may have an impact upon their participation. Genetic data, thus, brings into particularly stark focus a number of things about privacy protection, many of which perhaps could be said about other personal data, but not always with the same implications for failure.

The legal protection of genetic privacy

Access to, and use of, genetic data for research purposes is regulated differently across the world. Also, genetic privacy protection *outside* the context of research may have implications for individual attitudes towards participation in research using genetic data. Concerns about the future uses that employers, insurers, government agencies, or one's friends and family might make of genetic data might discourage individuals from genetic testing in *any* context. A concern that individuals not be unduly discouraged from accessing information about their own genetic status, and not be subject to unfair stigmatisation or discrimination as a result of such access, has led a number of regulatory authorities to take steps to ensure that genetic privacy is protected and this protection is often directed at access and use outside of the research context.

The precise nature of the protection found necessary or desirable varies considerably around the world. This is not particularly surprising and might represent an appropriately tailored response to local needs. Countries reliant upon private healthcare insurance to meet citizens' healthcare needs might, for example, take a different position from those that have healthcare systems free at point of care, funded through public taxation. There is no need, nor space, here to undertake a comprehensive survey of different regulatory positions but a couple of relatively recent proposals, both in the United States, might serve to make the point. The state of Massachusetts has recently seen a 'Genetic Bill of Rights' proposed to State legislators. Similarly, the States of California and Vermont are considering

introducing legislation that would protect individuals' genetic information in circumstances that go beyond those recognised by the relevant Federal Law. None of these proposals is quite the same as another.[13]

The inconsistency in the approach towards regulating genetic privacy, with some legislators adopting (subtly different) bespoke legislation, and some relying upon more general privacy laws, is creating a situation of bewildering complexity. The complexity alone is creating enormous uncertainty, and no small amount of difficulty, for many researchers, and not least amongst these are those wishing to conduct international research.

When national inconsistencies challenge international co-operation, one might consider the instruments available to international lawyers to move the legal landscape, incrementally, towards a position of some harmony. There are a number of international legal standards that have application to medical research that involves people, or biological material taken from people, in the research process. When standards apply directly to *clinical* research[14] their concern is with research that involves people. They do not have, at the heart of their concern, research that only involves *data* that relates to people. The legal framework with perhaps the most practical significance for international research using 'only' genetic data is currently the law of data protection.

One important example of a relevant international legal standard in the context of data protection law is the EU Data Protection Directive 95/46/EC. The current framework of data protection adopted by a number of countries, across the EU but also beyond, implements the standards represented by this Directive. It represents a binding legal commitment by members of the European Union and is an important standard against which legal frameworks are judged for those wishing to receive data from the European Union.

The Directive relies upon the key concept of 'personal data' to establish its scope and application. This idea, and the associated idea of personal information, is one that is also found within other important international standards concerned with privacy protection. Unfortunately, the idea of 'personal data' is not compatible with appropriate privacy protection – at least, not in the area of research using genetic data. The idea is ill equipped to capture the full range of ways in which

[13] D. Vorhaus, 'Is the Genetic Rights Movement Picking Up Steam?', *Genomics Law Report* (16 March 2011). www.genomicslawreport.com/index.php/2011/03/16/is-the-genetic-rights-movement-picking-up-steam/.

[14] Such as that provided by the EU Clinical Trial Directive (2001/20/EC).

privacy might be affected by research using genetic data and the framework of privacy protection built around the idea is incapable of accounting for all relevant interests or preferences.

One of the problems associated with the concept of personal data is that it tends to assume that there will be a single identifiable individual to whom personal data will 'relate' *and* that this individual's privacy is only at risk for as long as they are identifiably associated with that data. These are both assumptions that can be readily challenged within the context of research use of genetic data. They are, however, also assumptions that sit comfortably with the widely adopted mechanisms of 'consent' and 'anonymisation' as ways to protect fundamental rights and freedoms, including the right to privacy. The Data Protection Directive 95/46/EC has effectively consolidated the mechanisms of 'ask' or 'anonymise' as key routes towards lawful processing of personal data. The emphasis placed upon these two mechanisms is unfortunate for privacy protection.

In order to challenge the adequacy of consent and anonymisation as regulatory expectations, while continuing to recognise their importance for the protection of individual privacy in some cases, one might demonstrate the inability of the concept of 'personal data' either to recognise or to protect the full range of privacy preferences on questions of access and use of genetic data in a number of broadly applicable circumstances. Data protection should supplement these traditional mechanisms of protection in those cases where they are inadequate to the task. It should also seek to do so in a way that might bring *increased* certainty, consistency and transparency to the regulatory arena.

Assessing privacy protection

Not only is privacy itself a contested concept but the appropriateness of privacy protection can only be judged when it has been placed alongside other affected interests in a particular scenario. This introduces a number of variables into what is already a complex judgment. The *relative* significance of any privacy infringement can only be assessed according to a particular scheme of values when the practical implications of the infringement are understood: privacy infringement 'X' in situation 'A' might ordinarily be considered impermissible but, if in situation 'B' infringement 'X' is necessary in order to protect more significant interest 'Y', then the prevention of privacy infringement 'X' may be inappropriate. Without constructing a coherent world-view, acceptable by all as representative of the correct values in the correct measure, how might one assess the appropriateness of a legal framework designed to protect certain preferences regarding access to, and use of,

genetic data? What is more, how can one embark on this impossible task without first obtaining agreement on the, not undisputed, question of what is meant by privacy?

If one is to maintain confidence in a regulatory framework, then the individuals subject to regulation must be reasonably content that the things that matter to them are *at least* taken into account by the regulatory process. For this reason, I adopt a definition of privacy that places the patterns of behaviour and *preferences* of individuals, relative to access and use of information, at its core. I suggest that the first step towards a realistic assessment of the appropriateness of any legal framework, intended to protect privacy, must be to consider whether it is capable of even accounting for particular patterns and preferences. These patterns and preferences regarding access to genetic data I call 'norms of exclusivity'.

If a legal framework is incapable of bringing particular norms into view, then it is necessarily incapable of appropriately assessing the relative significance of the preferences that they represent. *If* the full range of relevant interests *were* accounted for by a regulatory system, then the next step would be to determine whether the assessment of relative significance was defensible according to a particular normative framework. Next, one might consider whether, if protection was to be *effective*, the protection identified as appropriate could be delivered in practice. In this book, I intend to demonstrate how, in many cases, the existing legal framework can be critiqued according to its inability to take even the first step towards adequate privacy protection: the concept of 'personal data' is incapable of accounting for the norms of exclusivity regarding research use of genetic data.

Public cf. private interest

If the argument is successful, then it has implications beyond demonstrating the inability of the current legal framework to protect the full range of individual privacy preferences. It has implications also for the ability of the existing framework satisfactorily to protect the public interest in research using genetic data. This is not only because a failure to protect privacy may undermine participation in research projects, although that is a distinct possibility. It is because there is a public interest in appropriate privacy protection itself: the public interest is served by the proper protection of privacy.

The suggestion that the public interest might lie in proper privacy protection might strike some as odd. After all, the public interest is often presented as a foil to private interests. While the relationship they share is undoubtedly complex, and at times it may well be fractious, it is

perfectly proper to suggest that the public interest is served by the proper protection of privacy. 'Proper' privacy protection is assessed by a value judgment in a way similar to how public interest is assessed: a consistent world-view will reconcile their demands. This is not to suggest that each is simply or inevitably qualified by the other. This may be part of their relationship, but there is more to it than that. *From an individual's perspective*, if there are particular private interests recognised to be worthy of protection, then, all other things being equal, she must prefer a society in which she is protected. From a societal perspective, the public interest will lie in preserving and promoting reasonable accounts of preferred societies. Put crudely, one might respond to the question 'What is in the public interest?' by asking 'What kind of society do you prefer?' *If* agreement could be reached on the latter question, then that agreement would provide an answer to the former.

In the absence of agreement to the latter question, the public interest lies in finding a way to adjudicate legitimately between competing preferences. *Legitimacy* requires, minimally, that these competing preferences be *accounted for* within the regulatory process. The existence and nature of different preferences is a matter of fact to be investigated and their relative significance a matter of value to be debated. The point is, however, that if preferences are to be overridden or trumped in any circumstance, then the *acknowledgment* of those interests is an important precursor to a legitimate decision. If an individual or group is negatively affected by a decision or circumstance, then acceptance of the hardship will be influenced by the perception of whether the process has taken the interests of that individual or group into reasonable account.

Admittedly, legitimacy requires more than simple transparency. Ultimately, people have no reason to accept authority unless they perceive that acceptance to be instrumental to protection of things that they value. If their preferences were to be systematically overridden on a regular basis, then the fact that they were transparently acknowledged by a system before being overridden would provide little comfort or reason for future confidence. Being able to account *reasonably* for the broadest ranges of preferences does, however, first require that regulatory decisions are able to account for interests so that they might satisfy them where they can. Accounting for their interests and preferences also provides people with a way into an argument if they think that their position is not properly being understood. Where it is impossible to reconcile competing demands to everyone's satisfaction, then reasonable account also implies that it is possible to justify the decision to override particular interests in terms that would be acceptable to a reasonable person. For these reasons, there is a *public* interest in the *proper*

protection of privacy in research using genetic data. This public interest cannot be fully served by the idea of 'personal data'.

Structure

This book is divided into three sections. Part I, 'The context', attempts to set the scene for subsequent critique. It describes, in Chapter 2, the contested nature of privacy, in Chapter 3, the alternative ways in which the term genetic data might be understood, and, in Chapter 4, the existing legal framework available to protect privacy in research using genetic data. Part II, 'The critique', offers a critical perspective. It focuses upon the legal framework established by the European Data Protection Directive 95/46/EC and challenges its reliance upon the concept of 'personal data'. The limitations of the concept are considered in Chapters 5 and 6. Chapter 5 underlines the limitations that follow from the fact that genetic data is held in common between people. Chapter 6 seeks to underline the significance that genetic data retains, and the preferences regarding access and use that it might attract, even if it is not 'identifiable' (as the term is currently understood by law). Throughout this section, while the European Data Protection Directive is taken as a prime exemplar of a regulatory approach that orbits around the term 'personal data', comments made have significance for privacy protecting frameworks that are based upon similar ideas of 'personal data' or 'personal information'. Part II also includes, in Chapter 7, a critical consideration of the information/sample distinction and, in Chapter 8, a critical consideration of the suggestion that any of the limitations identified to this point should be addressed through bespoke genetic privacy protection. Part III, 'The consequences', seeks in a single chapter to consider the implications for analysis for the future of data protection legislation and for appropriate privacy protection in the context of research using genetic data. The proposals made throughout the book are brought together and certain suggestions made for reforms. These reforms include additional responsibilities to account for a wider range of preferences (including those relating to certain third parties and de-identified data) but within a regulatory environment that ensures that *all* responsibilities are not only much clearer, but also more explicitly *proportionate* to the research in question. The uncertainty surrounding the current framework is unhelpful to research and it is important that research using genetic data is facilitated where it is in the public interest to do so. Protection of the public interest in research is entirely consistent with appropriate privacy protection but not with the current fixation upon the idea of 'personal data'.