From Blood to Data: An Ethnographic Account of the Construction of the Generation Scotland Population Genetic Database

Wendy Marsden

PhD
University of Edinburgh
2008

I declare that all the work within this thesis is my own

Wendy Marsden

Interpolated – that is, inserted – thus into the matrices of techno-scientific maps, we may or may not wish to take shape there. But, literate in the reading and writing practices proper to the technical-mythic territories of the laboratory, we have little choice. We inhabit these narratives, and they inhabit us.....We need stories for imagining how to be responsible within and for the zones in which we find ourselves.

Haraway1992:42

Abstract

This thesis is an examination of a population genetic database as both a social and scientific entity. Science and social science usually operate in a dichotomy this is a synergy of the two. The thesis examines practices and processes, and reveals how the formation of the Generation Scotland assemblage is the producer of multiple disconnections and connections layered in the science, technology, objects, people and places.

The story is based on a multi-sited ethnography that moves from the medical setting of blood sample and data collection, through the practices and processes of the laboratory, to end up in the much more diffuse settings of computer analysis. The blood sample is transformed into digital genetic data, and then connected to diverse other data for research. It traces the transformation and aggregation of heterogeneous elements which will become fixed in the population genetic database through scientific ordering and relationships which will be rendered immutable by the technology. In the processes described here, people's bodies, and information about them, are explicitly rendered as research 'resources'.

The thesis contributes to the growing knowledge of population genetic databases, and it is a response to calls from social science to understand better the science and technology that are currently changing the shape of the social world. Disconnections and connections are creating a framework of new referents between health and illness, identity and relationships in a way that rearticulates the body and the population.

Acknowledgements

My sincere thanks go to my supervisors Prof Janet Carsten, Prof Graeme Laurie and Prof David Porteous. Thank you to SCRIPT, the Arts and Humanities Research Council law and technology centre, and the School of Law, to friends and colleagues there especially the PhD students.

I am grateful to the whole Generation Scotland team, Genetic Health in the 21st Century and the Scottish Family Health Study, thanks to the colleagues and friends for help, support and encouragement including Gill Haddow, Rena Gertz, Sarah Cunningham-Burley, Shona Kerr and Blair Smith. The Chief Scientist Office of the Scottish Government and the Scottish Funding Council provide core support for Generation Scotland. This PhD studentship was specifically funded as part of Genetic Health in the 21st Century by the Scottish Funding Council (formerly Scottish Higher Education Funding Council), SRDG grant number HR03006.

Many thanks also for advisory support to Prof Sally Macintyre, Dr Kate Hunt and Mairi Langan from the Medical Research Council, Glasgow. Thanks to the Wellcome Trust Clinical Research Facility to the research nurses there, and to the research nurses on the projects in Dundee and Glasgow. My gratitude also goes to the WTCRF Genetics Core and to the lab technicians thank you for your patience and enthusiasm.

Thanks for reading early draft chapters to Robin Morton and Andy Macleod of Generation Scotland, and thanks for help with the information technology to Alan Blaikie. Lee Murphy in the WTCRF Genetics Core, thanks and appreciation

for your interest, pictures and comments. Thanks also to Alex Doney in Dundee, for your help. Thank you to Jean Sassoon for comments on an early version, and to Desmond Ryan for his words of wisdom.

My gratitude and love to Agnes Kruger for inspirational postcards, comments and proof reading, Catherine Middleton for patient listening, Carol Wiseman for indulgence therapy, Sandra and John Endicott and Margaret McCabe for encouragement and support.

Thank you and love to my family, Guy, Tess, George, Charlie, Linda and Maurice for encouragement, especially Hazel I wouldn't have finished without your help. Most of all, love and appreciation for unwavering faith and love of a student mother to Meredith, Jessamy and Rory.

Contents

List of Figures

Glossary

Allele	alternate forms of a gene
Biobank	a collection of DNA, blood tissue, organs or biological material, that can be linked to and/or include a phenotype database, environmental factors, physical measurements, mental test results, a personal details database, a genealogy database, and health record data. The term is sometimes used interchangeably with 'population genetic database' or 'databank'.
Gene	part of the DNA molecule that encodes for a protein
Genetic Database	A collection of DNA samples, and/ or the genetic data derived from DNA samples
Genotype	a single gene in a individual, the genetic constitution of an individual, or an aggregate of a single gene in a population
Phenotype	the observable traits or characteristics of an organism. It can refer to a single trait, a collection of traits in an individual, or the aggregate of a single trait in a population
Polymorphism	a difference in a DNA sequence among individuals. Genetic variations occurring in more than 1% of a population would be considered useful polymorphisms for genetic linkage analysis
Population Genetic Database	a large scale genetic database, that can be linked to and/or include, a biobank of samples (blood, tissue, biological material), phenotype database, environmental factors, physical measurements, mental test results, a personal details database, a genealogy database, and health record data. The term is sometimes used interchangeably with 'biobank' or 'databank'.

Acronyms

21CGH	Genetic Health in the 21st Century
AFM	Association Francaise contre les Myopathie
AHRC	Arts and Humanities Research Council
AHRC Centre	Centre for Studies in Intellectual Property and Technology Law
APoE	gene associated with Alzheimer's disease
BRCA	gene associated with breast cancer
BRIP	gene associated with breast cancer
CEPH	Centre d'Etude du Polymorphisme Humain
CHI	Community Health Index, a unique identifier
CLAM	Clean, anonymize and map
COE	Council of Europe
CSO	Chief Scientists Office
DoH	Department of Health
DNA	deoxyribonucleicacid
ECG	electrocardiogram (scan)
EDTA	ethylenediaminetetraacetic acid, anticoagulant and blood preservative
ELSA	Ethical, Legal and Social Aspects
ELSAGEN	Ethical, Legal and Social Aspects of Huuman Genetic Databases: A European Comparison Study
ELSI	Ethical, Legal and Social Implications
EMIS	Egton Medical Information System
EMSY	gene associated with breast cancer
ESRC	Economic and Social Research Council
FTA	Whatman FTA card, a chemically-treated fiber matrix for long-term blood sample storage
GHI	Genetic Health Initiative
GIG	Genetic Interest Group
GPASS	General Practice Administration System for Scotland
GS	Generation Scotland
GSAB	Generation Scotland Advisory Board
INNOGEN	Centre for Social and Economic Research on Innovation and Genomics
ISD	Information Services Division of NHS Scotland

IP	Intellectual Property
IT	Information Technology
MMC	Molecular Medicine Centre
MRC	Medical research Council
MREC	Multi-centre research ethics committee
NHS	National Health Service, or NHS Scotland
NHSiS	National Health Service in Scotland
NTRAC	National Translational Cancer Research Network
P3G	Public Population Project in Genomics
PBL	peripheral blood lymphocytes
PCR	Polymerase chain reaction
PIL	patient information leaflet
PXE	pseudoxanthoma elasticum, an inherited disorder
REVEAL	Risk evaluation and education for Alzheimer's disease study
SEHD	Scottish Executive Health Department
SFC	Scottish Funding Council, previously SHEFC
SFHS	Scottish Family Health Study
SHEFC	Scottish Higher Education Funding Council
SNBT	Scottish National Blood Transfusion Service
SNP	Single nucleotide polymorphism - a change in which a single base in the DNA differs from the usual base at that position.
SOP	standard operating procedure
SPSS	Statistical package for social scientists
SSPC	Scottish School of Primary Care
STATA	an interactive data management and statistical analysis program
VPN	Virtual private network, similar to a mobile phone network
WTCRF	Wellcome Trust Clinical Research Facility

Introduction

This thesis is an examination of Generation Scotland as a social and scientific entity. The interaction of science and information technology has given rise to complex alliances and relationships that make it possible to construct a population genetic database such as Generation Scotland. Based on a multisited ethnography that moves from the medical setting of blood sample and data collection, through the practices and processes of the laboratory, to end up in the more diffuse settings of computer analysis, the thesis follows the path of the blood sample as it is transformed into digital genetic information. Mapped onto this ethnography are key ethical, legal and social issues that have been raised within Generation Scotland, and by other studies of population genetic databases. In mapping the social science onto the science, I was aiming for a synthesis between the two, but it proved difficult to reconcile fields of inquiry that habitually operate as a dichotomy to create a complete picture of this multi-faceted entity; as a consequence, the thesis often grates, reflecting the complexity of the situation and of what I was trying to do. Importantly, both the ethnography and the crossing of disciplinary boundaries reveal disconnections and connections that are layered throughout the people, places and practices that make up Generation Scotland.

Generation Scotland

The population of Scotland suffers a high rate of mortality and morbidity from several diseases, including chronic heart disease, cancer and mental ill-health, in comparison to the rest of the UK and other western European countries (Popham 2006). The Scottish Office published a white paper in 1999 'Towards a healthier Scotland' in which *public health and preventive measures*

were identified as one of the three key elements in achieving this aim. The Annual Report of the Chief Medical Officer for Scotland supported 'the opportunities for Scottish health that could flow from the new genetics'(Chief Medical Officer 1999). The Science Strategy Review of 2000, commissioned by the First Minister, which included looking at science, particularly biotechnology, in Scotland in relation to the economy, and also a call for public engagement. In June 2003 the UK government published the White Paper, 'Our inheritance, our future - realising the power of genetics in the NHS' which 'set out a vision of how genetic techniques could benefit patients, and a £50million, three-year plan of implementation'. The convergence of these agenda set the background for the Generation Scotland proposal to emerge. The Royal Society of Edinburgh held a dinner to discuss Generation Scotland and there was a general consensus to support the proposal (The Royal Society of Edinburgh 2002).

A grant of £1.79million was awarded under a Strategic Research Development Grant initiative from the Scottish Higher Education Funding Council (SHEFC, now the Scottish Funding Council) to Genetic Health in the 21st Century (21CGH) for 'building the infrastructure, developing the science, enhancing the knowledge base, and engaging the public'(Project Proposal 2003). The intention was to create a platform for population genetic research in Scotland that included both Generation Scotland and the Scottish 'spoke' of UKBiobank. Generation Scotland is creating a Scottish family-based population genetic database resource. The Scottish 'spoke' of UKBiobank is one of the geographic areas combining to form a UK individual-based population genetic database. These are two separate projects, but they are complementary to each other as genetic research resources.

Generation Scotland (GS) is a consortium for a Scotland-wide project that includes medical schools, research institutions and other academic departments around the country in a multidisciplinary collaboration of health professionals, information technologists, geneticists, sociologists, health geographers, statisticians, lawyers and this anthropologist. The project is supported by the Scottish Executive, the Department of Health, the Chief Scientists' Office, the Scottish Higher Education Funding Council, and the NHS. The GS family-based population genetic database will serve as a powerful research resource. It is intended to enable researchers to hunt for and identify genes that predispose people to complex diseases such as cancer. The resource will also contain phenotypic data so that the relationship between genotypes and phenotypes can be analysed. It will, in addition, be used in analysis of the relationship between genotypes and drugs, and it is envisaged that it will be useful in future drug development by the pharmaceutical industry.

The 21CGH funding was awarded in August 2003. At that time the European Commission and the Council for Europe were conducting research into the ethical, legal and social implications of genetic testing, which incorporated issues pertaining to genetic databases and biological material (McNally and Cambon-Thomsen 2004; Council of Europe, 2005). The Economic and Social Research Council (ESRC) Centre for Social and Economic Research on Innovation and Genomics (INNOGEN), University of Edinburgh and the Open University, was conducting focus groups as the first step in public engagement on Generation Scotland (Haddow, Cunningham-Burley, Bruce and Parry 2004). Lawyers from the Arts and Humanities Research Council (AHRC) Research Centre for Studies in Intellectual Property and Technology Law, University of Edinburgh, had produced a report on the ethical and

legal aspects of Generation Scotland (Laurie and Gibson 2003). At the same time the principal investigators for Generation Scotland were working on a proposal for blood and data collection to submit to the Scottish funding agencies. The relationships between the Icelandic population, the proposed Health Sector Database, DeCode and the Icelandic government was controversial, and had become the focus of an ongoing bioethical debate on a range of issues. The bioethical issues arose, more generally, from 'the intersection of biomedical research and the market' (Palsson and Hardardottir 2002:271). These included ideas about: property and ownership (Gold 1996; De Witte and Ten Have 1997; Everett 2003; Parry 2004); informed, broad and presumed consent (Anderson and Arnasson 1999; Arnasson 2004; Kegley 2004); data protection, privacy and confidentiality (Morrow 2001; Laurie 2002; Lowerance 2002); anonymisation and withdrawal (Eriksson and Helgesson 2005); and, commodification of the body and commercialisation (Fortun 2001; Anscombe 2003; Wilson 2003; Merz, Glenn, McGee and Sankar 2004). Internationally, there were proposals for population genetic databases in other countries including Canada, Estonia, Sweden, Japan and Tonga. The science journals were regularly publishing papers on the discovery of 'the gene for' various conditions, and patenting of genes was already established, if controversial.

I joined Generation Scotland in the autumn of 2003 as a PhD student, and member of the Ethical, Legal and Social Implications (ELSI) team, which comprised two research fellows and me under the 21CGH project proposal. The research fellows joined GS in the spring of 2004, a lawyer (RG) in March and a sociologist (GH) in May. I was based in the AHRC Centre in the School of Law, with the dual roles of team member and individual researcher. I worked toward my own research agenda, but I was also making a

contribution to the teamwork of the GS project in support of, or in collaboration with, colleagues. In addition I was also working in collaboration with another PhD student at the University of Glasgow; funded by the Medical Research Council (MRC) she was taking a contemporary history approach to research on population genetic databases, in particular UKBiobank. Being located in an academic department but funded through a project is subtly different from being independently funded. It is a common practice to provide funding for PhD students through larger research grants, particularly in scientific and health research projects, less so in the social sciences. As scientific and health research projects become increasingly multidisciplinary, the number of opportunities for social science researchers to be funded has grown and is likely to continue to do so, at least for a while. Whether this integration is a good or bad thing is open to debate, as publicly funded social science research already comes under scientific research governance, there are those who suggest that this has created a 'slave social science' (Donovan 2005:598).

The people working on GS but located in the AHRC Centre had no immediate or regular contact with the science and technology people working on the project on the other side of the city. This did not cause me undue concern but I was conscious of feeling remote from much of the project and the people working on it, and of not being sure what was going on. For the first year, every time I attended a GS meeting that included the science and technology people the project seemed to have changed and I often felt that I was not sure what was being talked about. There was an ongoing process of development and planning of both GS as an organisation and the blood and data collections. It was in fact nothing like as 'concrete' as I had expected it to be. There was a mismatch between the representation of

the GS project as being something quite concrete, and the GS proposal which was changing and amorphous. The scientific knowledge and practices were in place, the genetics lab was already set up and running, but the practicalities of collecting blood samples and other data were not.

There appeared to be distinct groups of people working on the GS project, some met occasionally, others rarely, and indeed some not at all. When a multidisciplinary meeting took place, people used the same words to talk about aspects of the project. The scientists and non-scientists all seemed to understand what each other were saying, but I found when I sought explanations and clarification from different people of things that had been said about, for example, blood or DNA as tissue, biological material or information, that we were not all talking about the same thing. The talk was focused within, and by, particular epistemologies, whereas the scientific practices and processes were focused around particular objects, and I was not convinced that they were mapping onto each other.

The development of GS and the collection of blood and data to construct a population genetic database has led to the intersection of diverse knowledges, practices and interests that include: Scottish families; genetics and genes; ethical, legal and social implications, concerns and issues; medical research; biotechnology; public engagement and science; public understanding of science; phenotypes; disease and science (genetics); health and databases (information technology); genetic epidemiology and public health; academic science and commercial companies; therapeutic drugs and drug development, to name a few. All these converge within GS intersecting in a complex set of relationships. But these intersections are occurring across a gap between social science and science, which are constructed as separate

fields of enquiry and as having different specialist knowledge and disciplines. Different stakeholders view the others, often suspiciously, across this gap. For example, the very taken for grantedness of blood samples by clinicians and scientists struck me as interesting, since the whole enterprise is dependent on collecting enough of this substance. In medical science blood is so much a part of the everyday that it is unremarkable. As an anthropologist I was used to thinking about blood as a symbolic and ritual substance. In many cultures blood is invested with power and significance which makes it an exceptional substance. These two views do not sit easily together.

One of the intersections of particular interest was between the public and GS. GS was committed to a programme of public engagement research, the first part of which, as mentioned above, had been underway since the summer of 2003. Public engagement research on UKBiobank (Cragg Ross Dawson 2000, People Science and Policy 2002, Marsden, Sullivan, Duffy and McLaren 2002) had shown that there were a range of issues and concerns about participating in a population genetic database project.

The complexity of both GS as an organisation and the science and technology behind the creation of a population genetic database, with the aspirations for public health and drug development, led me increasingly to question what it was the public were being asked to engage with, and more importantly, to participate in. It seemed to be of growing importance to try and understand what a population genetic database is, as a local project and in a global context. I thought this would be a useful contribution to the ELSI work, to public understanding of science, and to the public engagement research.

I started my project with two very basic questions. First, what is it that Generation Scotland is asking the public to participate in? And second, how do you transform blood into digital data? I thought a multi-sited ethnography following the blood as it flows 'from arm to desktop' would be the best approach to understanding what participants become a part of through the scientific processing.

The blood flows from people in the social domain. It is collected in a medical setting. It then moves into the scientific domain, the laboratory, where it is processed and at the other end emerges as digital data. The digital genetic data flows into a database. The data in the genetic database can then link to various other data and transferred to a desktop for analysis. Finally, the results of the analysis will return to the social domain in various forms, including, risk calculations, public health policy and potentially personalised drug therapies. These elements strongly suggested to me that a multi-sited ethnography would contribute to the knowledge and understanding of population genetic databases, and answer the question of what it was the public were being asked to engage with and more importantly to participate in.

Outline of Chapters

Chapter One is a review of the literature that I drew on, it includes a range of authors and approaches to the study of science and genetics, the understanding of genetics as social and scientific endeavours, and research into genetic databases in particular. Chapter Two discusses the methods I used, describes the steps I took to organise the fieldwork and how I pieced together the multi-sited ethnography. It also explains the complicated

situation in which the work for the thesis was done, its effects on the ethnography and the constraints it placed on my research.

The third chapter is about Generation Scotland and genetic research design to examine the 'who' and the 'why' of this population genetic database. 'Generation Scotland' refers collectively to a brand name, an umbrella concept, an organisation and a genetic research resource. The public of Scotland are being asked to engage with and participate in something highly complex. GS characterises itself as 'a unique partnership'. The organisation has within it the knowledge and expertise to collect, process and control tens of thousands of DNA samples, and unlimited amounts of information that will come from the people of Scotland. I include a brief description of genetic research design because the databases as research resources have to fulfil the design requirements of genetic research projects. The background to GS and a brief analysis of the organisation reveal a complex organisation with an unlimited virtual capacity to absorb data from various sources and create connections that could be national or international. GS is comprised of not one, but several data collection projects and databases linked through an information technology platform.

The collection of blood samples is described in Chapter Four, and I show how a blood sample is a complex object. The chapter begins by asking – 'What is a blood sample?' In legislation, ethics and regulation, the answers are inconsistent. The blood samples are collected by research nurses who are skilled in the practice of venepuncture, which they regard as routine. The role of the research nurses is multifaceted for they also collect other information, administer tests and questionnaires and take physical measurements. They are responsible for checking that participants have read

and understood information sheets and ensuring that consent forms are signed. The role of the research nurse is nodal. They work at the interface between the project and the public. Their work is located at the point of connection between the participants and project, but it is simultaneously the point of multiple disconnections of blood and information from people. The collection of blood to construct a population genetic database means it must be disconnected physically from the body and conceptually from the social world. The language of the gift derived from *The Gift Relationship* (Titmuss 1970) is used to create an association between the blood for research samples and blood donation for transfusion. But, the blood sample for genetic research is not about saving a life; it is about collecting information. I would argue that the collection of information does not 'fit' the theory of the gift.

Chapter Five follows the blood to the laboratory. The blood arrives in the lab not as a gift but as 'nasty dirty stuff' that has to be handled wearing protective gloves. The story follows the blood, describing the practice of booking-in, after which it is stored in a freezer to await processing. The lab is a particular place where a particular type of work is done. The blood has moved from the social world across the divide and into the scientific domain. The lab seems disconnected, located at a physical and perceptual distance from the everyday of the social world and the other activities of GS. Here people and equipment enact a different version of the object that is the blood sample. Moreover, they take the blood, strip, purify and manipulate it to produce a new object, the DNA sample. There were fewer people in the lab and more equipment than at the collection point. They also tended to disappear, to become part of the processes, which were ordered and repetitive. There was a hierarchy of knowledge and each lab technicians had specific tasks to do. I describe one of these tasks, the extraction of DNA from

blood samples. The disconnection of the blood from the person is extended by the manipulation of the blood to produce a new object, the DNA sample.

Staying in the lab, Chapter Six describes how the DNA is put to work to produce digital data. The DNA samples are checked for quality, which needs to be consistent in concentrations and distribution in a buffer solution that forms the 'master stock' for a project. A 'working stock' is made up from the master stock for use in genotyping or sequencing. Before genotyping or sequencing, a small amount from each sample in a working stock is amplified using the process of polymerase chain reaction (PCR). The process of PCR includes 'denaturing' the DNA. The DNA is then inscribed in genotypes or sequences as digital data. The DNA is denatured and inscribed both literally and figuratively, disconnecting it from its substance to become digital data that can be ordered in a database. The disconnection between substance and data renders a new version of DNA - as information. In the lab these versions of DNA are enacted through specific practices. The social understandings of DNA, evident in metaphors of code, script or blueprint, reflect the lab versions of DNA but do not convey a sense of their *production*. There is a problem with DNA as information and there is an ongoing debate about genetic exceptionalism. Some argue that DNA is exceptional, others that it should be treated in the same way as other information held in medical records. I would agree with those that argue for genetic exceptionalism. The information held in medical records are not embodied in the person, although it is certainly about the person, nor can it be manipulated as is the substance DNA.

The value of the GS databases as a research resource lies in the ability to connect genetic data to other types of data. Chapter Seven examines what

these other types of data, phenotype, lifestyle and environmental are, and where they come from. In Chapter Four we see how the research nurses were also collecting a range of data from questionnaires, tests and physical measurements but these have been kept separately from the genetic data. I examine the notions of phenotype and environmental data, and analyse one of the GS project's questionnaires. The analysis demonstrates both the cultural construction and the contingency of this information. The questions bring families into focus, compared to the lab where they were not visible. The questions on family history and family illness record relationships between family members. This draws other people into the data of any participant whether they know it or not. Participation also has the potential to put known relationships at risk as GS routinely conducts paternity tests as a quality control measure. These phenotype, lifestyle and environmental data go through processes of disconnection to arrive in the phenotype database, and like the genetic data, their value lies in being able to (re)connect these data and in linking to other databases or datasets.

Chapter Eight starts with another source of data, health records. GS plans an ongoing collection of health record data, direct from patient records. There are two connections going on here, one is the direct linking of GS to patient records, the other, the data from health records will connect within the database to genetic data. I used my own health record in general practice as an example of what might be contained within personal health records that GS could access. Patient records are kept across the NHS under the ethics and regulations of confidentiality. The mechanisms in place to ensure people's confidentiality and privacy were investigated. They show that, although changes are underway, local mechanisms are currently inadequate to deal with the global forms of personal health and illness information in

digital data, information technology, knowledge transfer and data sharing. It cannot be assumed that the connection of large amounts of data, with known or unknown other data will remain confidential or private. The chapter goes on to examine why it is important to connect genetic and other data together for analysis, by looking at the genotype-phenotype relationship and the gene-environment relationship. This connection happens, or more accurately will happen, at 'the desktop', for there I found an emerging field of expertise.

Chapter Nine considers some of the connections that are being made beyond the GS organisation itself. Further and future connections beyond GS include a commitment to create and maintain a connection with the Scottish public in different ways that include participation, public engagement studies, governance and benefit-sharing mechanisms. The future uses of the database connect GS to a wider research community. The initial work of GS on gene identification, relationships between genes and phenotypes and genes and environment will be oriented to understanding risk of diseases in the population. The results of this research will (re)connect to the social world. We do not know how the public will respond to this sort of risk information but studies on understanding and interpreting risk from genetic testing raise a number of issues and problems particularly for families. As genetic information, risk factors and predispositions to disease become more widely discussed, the expansion of 'biosociality', another form of connection, is a possibility. However, there are several factors that mitigate against biosociality at least in an equitable way in Scotland. A further connection is envisaged between GS and the commercial sector through the aspiration to create wealth, as well as health, in Scotland. Finally, there are already suggestions being made about how population genetic databases in different countries might be linked or might share data. A comparison of other

population genetic database projects in other countries locates GS in an international network of projects that have informal connections through geneticists, IT specialists and ELSI researchers. The Public Population Project in Genomics (P3G) organisation is working to formalise connections in the interest of knowledge transfer, data sharing and support for research. However, local mechanisms of law and ethics are not necessarily equipped to deal with global forms, nor are technology and expertise evenly distributed globally. Much of the data, like these collected in Scotland, especially lifestyle and environment which is discussed in Chapter Seven, will be culturally constructed and possibly locally specific in a way that could make it either inadequate or inappropriate for use across international boundaries.

My objective in following the blood and bringing the various aspects of the database together is to demonstrate that engagement and/or participation in a population genetic database is not a simple matter of 'helping other people'. There are complex scientific concepts, power relations, and social and political issues all at work here simultaneously. I conclude that the public are being asked to engage with and participate in an infinitely more complex process than could be imagined from the rhetoric used in the public domain. The rhetoric of the gift and the parallel drawn with blood donation does not fit. Finally, by following the practices and processes 'from blood to data' we can see how disconnections and connections are layered within the assemblage that is Generation Scotland. Also how GS has important implications for how we think about health and illness, the future uses of samples and data, and the management and governance of such databases. It also shows how a population genetic database is contingent and provisional, which is reflected both in the construction of Generation Scotland and this thesis.

Chapter 1

Literature Review

The literature on population genetic databases draws on a range of disciplines that includes science, medicine, bioethics, law and sociology. The headings under which population genetic databases are examined include, for example, epidemiology, consent, privacy and public engagement. Population genetic databases are constructed and critiqued within a multidisciplinary environment that is still evolving. It is a relatively small literature that is a side branch of a wider literature on the new genetics.

I sought a single authoritative framework in the literature in the belief that this would give me a solid base to work from, but I could not find one that offered this across the diverse aspects of my project. Instead, I found a range of authors who had taken different approaches to a field that is changing and evolving by the day. I have drawn on different authors at different times as I encountered and tried to understand various aspects of the project, such as blood, DNA, the laboratory, and information technology.

Prior to my fieldwork I had been looking at a range of literature which formed the background to my research. The history of the new genetics and the concept of the gene have not been without controversy and disagreement between biologists about the reductionist view of DNA as a 'code' or 'blueprint' for life (Condit 1999; Fox Keller 2000; Waldby 2001; Moss 2003). This deterministic 'dogma' (Lock 2005) was challenged both within the biological sciences and from without. The critique from social science challenged the 'geneticization' of

the social world (Lippman 1998). The hereditary nature of DNA and its implications for kinship have been examined by a number of authors (Lippman 1998; Finkler 2000; Nash 2004; Sachs 2004). Geneticisation, that is the reduction of differences between people to what lies in their DNA, pointed to a deterministic view of people as individuals, families and citizens (Petersen 2003; Rose and Novas 2003; Plows and Boddington 2006). This also has implications for ideas about race, ethnicity and disability (Shakespeare 1998, 1999; Atkin 2003; Tutton et al 2004).

The study of public understanding of science expanded to include genetics (Macintyre 1995; Kerr, Cunningham-Burley and Amos 1998; Hedgecoe 2000; Edwards 2002). Studies about public understanding of genetics are critical of the lack of engagement with the public about how the new knowledge generated by genetic research should be used, for example, in testing or surveillance programmes (Kerr and Cunningham-Burley 2000). The studies also show that the public has a complex informed understanding of genetics, albeit a non-technical one (Bates et al 2005). Furthermore these studies suggest that a greater democratisation of science through a closer engagement with the public is feasible (Kerr, Cunningham-Burley and Amos 1998).

The new medical technologies produce intersections of enquiry between various analytical approaches and the human actors, the tools, the entities and the bodies that are constitutive of the new medical technologies (Lock, Young and Cambrosio 2000:1). New objects and subjects call for new kinds of analysis. In this collection of essays the contributors variously approach 'working' and

'living' with the new medical technologies, which are taken as two sides of an equation. The contributors look at how these dichotomies are produced and struggle to bring these two sides together. The intersections are 'temporary convergences that can lead to advances on some particular problem with no pretence of providing a comprehensive world-view or theoretical manifesto' (Lock, Young and Cambrosio 2000:1). The studies created a dialogue across the disciplinary fields of medical anthropology and science studies. However neither medical anthropology nor science studies are characterised by unity but rather by diverse approaches that can draw on, for example, sociology or history or epistemology. The challenge to disciplines such as anthropology and sociology is that even when they are purportedly dealing with medical or scientific matters, they 'generally ignore[s] the production of clinical and laboratory objects and procedures, thus treating them as 'black boxes'' (Lock, Young and Cambrosio 2000:3). The editors argue that opening 'black boxes' has become a specific task of medical and scientific studies. The construction of the GS population genetic database has clearly led to many and various intersections, and I take it, at this point in time, as a 'black box' under construction.

Richard Tutton and Oonagh Corrigan edited the first book published to deal directly with genetic databases, it addresses the socio-ethical issues in the collection and use of DNA (Tutton and Corrigan 2004). The issues raised in this book include public participation, donation of blood, consent and commercialisation. These issues are raised within a multidisciplinary approach that has been particularly evident in the ethical, legal and social

implications/aspects (ELSI/ELSA) research attached to large scale projects such as the Human Genome Project and UKBiobank. The language of population genetic databases and in particular participation is problematic (Tutton and Corrigan 2004:7). For example, the various terms used, often interchangeably, include population genetic database, biobank, tissue bank, or databank, to describe collections that include any or all of the following - tissue, questionnaire data, physical measurements, health records, and genealogies. They go on to argue that there is no neutral language in which to talk about 'public participation'. The donation or giving of blood and personal data are 'informed by assumptions about the nature of people's involvement with research' (Tutton and Corrigan 2004:7). The idiom of the gift, influenced by Titmuss's (1970) account of blood donation, occurs in biomedical and ethical discourse as well as in regulations and guidelines in medical research. Tutton examines the use of the language of the gift in relation to person and property, he argues for a 'different approach to how human tissue provided for research should be conceived and regulated' (Tutton 2004:33). Helen Busby takes up the theme of blood donation for genetic research and shows how it is often dependent not on knowledge but the trust inhering in physical and practical relationships (Busby 2004:44). These relationships extend beyond persons such as doctors to include the NHS and universities as research hosts (Busby 2004:49).

After I had completed the fieldwork I had to extend the range of the literature in order to understand what I had found, in particular the disconnections and connections that were emerging from my fieldnotes, and how I could think

about them. Some of this literature had not been published prior to my fieldwork but became available to draw on as I was writing.

Paul Rabinow states that with regard to fieldwork 'The problem of where to look, how to proceed, and what to do once one gets there is persistently present' (Rabinow 1996:xiv), a comment that certainly reflected my own experience. I followed his advice that neither organisations nor technologies should be over-interpreted for their meaning, but that attention should be paid to sites and techniques because they are singular and temporally specific. He advocates restraint and a focus on the emergence of new objects, sites and forms (Rabinow 1996:43). The sites of molecular biology are where moments of triumph are short, their effects diverse, and a great deal of routinisation follows. There are cultural influences and 'something unquestionably modern in the desire and project to capture bits of living matter and manipulate the contexts in which they were found and the ways in which they worked and could be reworked' (Rabinow 2000:44). Rather than focusing on totalising categories such as society, culture or epochs, which he argues are in conceptual ruins, Rabinow prefers the 'event'. The formation of the GS and the construction of the database would be an 'event', which makes new research and ways of thinking about genes and populations possible

> from time to time new forms emerge that have something significant about them, something that catalyses previously present actors, things, institutions into a new mode of existence, a new assemblage, an assemblage that made things work in a different manner. A manner that made many other things more or less suddenly possible.
>
> Rabinow 2000: 44

However, the GS event is currently ongoing and if taken as an assemblage many things are happening simultaneously, as opposed to Rabinow's events which were in effect complete. While it is possible to focus on the emergence of objects, sites and forms, the effects will not be evident for some time to come.

Rabinow's ethnography of the invention of PCR is an account of one such event. The story of *Making PCR* not only tells the particularity of practices but frames them in the light of other forces at play - personalities, company politics and finances (Rabinow 1999:13). Moreover, he involves his informants in the telling of the story using their voices in collaboration with his own. Not only did the people shape the making of PCR, they also shaped the story of its making. He uses this approach as one possible way of dealing with the problem of who has the authority to represent experience and knowledge. But, people tend to disappear when the focus is on practices and objects, for example, in Annemarie Mol's work in *The Body Multiple* (2002) and in this thesis.

One of the concepts that Rabinow returns to repeatedly is Foucault's notion of 'biopower'. Biopower refers to the way in which life and its mechanisms become knowledge and power (Rabinow 1999:13). However, Rabinow sees the *bios* in biopower as problematic in the light of new genetic knowledges. The study of the GS database suggests that one of the problems with *bios* is that it has been transformed into information. The new knowledges are forming new assemblages with social and political networks but how these changes will interact with forms of power relations is 'open to question' (Rabinow 1999:15). I suggest in Chapter Nine that biosociality may not be an attractive option, nor

indeed available, to everyone. *French DNA* describes one such assemblage, which brought together France's leading genomics laboratory, the Centre d'Etude du Polymorphisme Humain (CEPH) and a patient group, the Association Francaise contre les Myopathies (AFM) (Rabinow 1999). It was a new type of alliance to combine genomics, public health and financing. The intention was to conduct world-class science with a different ethos. The newness lies not in the alliance between disease research and patient groups but what is distinctive – and contemporary - in this situation is not its radical newness but its assemblage of old and new elements (Rabinow 1999:25). The assemblage created a new space for public expression and agenda setting as a form of citizen participation. Likewise, the GS agenda includes the intention to engage with the public through governance and mechanisms of benefit-sharing.

Bruno Latour and Steve Woolgar see public engagement as problematic (1979). They show how the theory and practice of science are not easily accessible to public engagement. They made a significant contribution to understanding what science does and how it does it by challenging the assumption of scientists that technical culture was of little interest to 'outsiders' who lacked technical competence. Scientists routinely drew on a distinction between the social and the technical and systematically concealed the nature of the activity. Moreover, they suggested that 'the fact that scientists often change the manner and content when talking to outsiders causes problems both for outsiders' reconstruction of scientific events and for an appreciation of how science is done' (Latour and Woolgar 1979:28).

A good deal has changed since 1979 with the development of science and technology studies, public understanding of science and public engagement studies. Latour and Woolgar's study of *Laboratory Life* (1979) must, however, be essential reading for anyone who would venture into a laboratory. Latour and Woolgar suggested the observation of scientific practice would retrieve some of the 'craft character' of scientific activity. They describe the spaces, people and activities within the lab, the movement between spaces, the use of black boxes for inscription, and the juxtaposition of literatures. They were particularly concerned with the 'construction of order' and the way in which particular features were invoked to produce order out of a disordered array of observations. They suggested that science, a body of practices that was regarded by outsiders as well organised, logical, and coherent, was in fact a disordered array of observations that the scientists struggled with to produce order (Latour and Woolgar 1979:36). Construction of order is a defining feature of the GS database. The Genetics Core laboratory used by GS is of central importance in the construction and ordering of the database but, it is a service lab. The research and production of knowledge happens outside this lab. Latour and Woolgar (1979) are helpful in understanding what a lab is as a particular place of work, but their lab was not subject to the complex intersections characteristic of the GS projects.

Annemarie Mol suggests that rather than approach knowledge in an epistemological way, as do Latour and Woolgar (1979), we should attend to the ontologies because 'They inform and are informed by our bodies, the organisation of our health care systems, the rhythms and pains of our diseases,

and the shape of our technologies. All of these, all at once, all intertwined, all in tension.' (Mol 2002:7). She calls her inquiry 'praxiographic' and bases it on the notion of enactment. Practices and objects are enacted. 'This suggests that activities take place – but leaves the actors vague. It also suggests that in the act, and only then and there, something *is* – being enacted' (Mol 2002:33). Thus the object is never isolated from the practice, and the techniques that make things visible, tangible, knowable (Mol 2002:33). What this approach does is to allow that all kinds of objects and events be taken into consideration 'that no phenomenon can be ignored on the grounds that it belongs to another discipline' (Mol 2002:158).

This 'performative' approach to the practice of medicine shows how many versions of the same disease are enacted through the practices of people, places and equipment. Across multiple sites the same disease is enacted in diverse versions, these sites include medical practice and the lab. The lab version is different from the social versions and the medical versions. When I was struggling with my field notes and trying to make sense of what I had written I drew on Mol's ideas about enacting objects, and the possibility of different versions of an object. She also offered me a way of explaining why I was having such trouble recovering the people in these settings.

Rayna Rapp was also helpful in thinking about the lab as a place of work, but work that has implications that touch people's lives. She showed how there are different views of the genetic tests for Down's Syndrome and also how interpretation was important, the meaning in one setting could be quite

different in another place (Rapp 1999). Mol's focus on practices and sites reiterates Rabinow's suggestions, though their different methods and sites produce quite different results. Mol's objects generally maintain a sense of being embodied in a way that cannot be sustained by DNA and genes in a database.

The database is a powerful tool for constructing order, and scientists work hard to create order (Latour and Woolgar 1979). But the use of databases in scientific work has raised questions about whether they may change the processes and the outcomes of scientific knowledge production (Hine 2006). This concern relates particularly to biomedicine and molecular biology which some have argued are becoming information science (Hine 2006:269). Christine Hine argues that databases embody and are embedded in already existing natural and social orderings. Computers do not impose logic but provide a mechanism for 'tying together' particular natural and social orderings (Hine 2006:269). The database does not act as an independent agent of change; what it does create is a new spatial organisation of science, with new communication regimes and new forms of collaboration (Hine 2006:270).

Hine reviews the relationships between databases and science and gives an ethnographic account of a mouse mapping resource. The mouse mapping resource was created through a collaboration of European laboratories. The aim of the collaboration was to enable faster and more accurate localisation of particular genes (Hine 2006:275). Although Hine's study was about a mouse mapping resource there were similarities with the GS project, and her ethnography drew my attention to and helped explain some aspects of working

with database technology. The mouse mapping resource 'can be seen to characterise a set of objects and to fix (or attempt to fix) a set of relationships between them' rendering the mouse genome into a manipulable object. (Hine 2006:277). The scientists had a dual role to pursue their own individual research goals and to provide a service to the laboratory and to outsiders (Hine 2006:280). While the whole resource was presented as a service to the scientific community, there was a tension between this and the claims of individual work.

The development of the database studied by Hine was not carried out in the lab but by computing specialists located elsewhere (Hine 2006:280). The database developers were providing a service to meet the users' needs. In order to do this, the computer specialists had to learn about and understand the ordering of the lab. Hine concludes that 'The 'digital ordering' represented by the database is highly contingent, representing the upshot of lengthy negotiations between the collaborators over the nature of the natural objects involved and of the scientific culture of the workplace' (Hine 2006:288). The database was enacted differently in different circumstances, sometimes as an object in its own right and sometimes only making sense as part of a complex of material artefacts and work practices (Hine 2006:290). The contingency of the ordering in the GS database is reflected in the collection of blood and data, and different versions of the GS database are enacted, like the mouse mapping database, at different times and in different places.

Attention to clinical and laboratory objects has been taken up by Margaret Lock in her work on Alzheimer's Disease and the ApOE gene (Lock 2005). In her

paper 'Eclipse of the Gene and the Return to Divination', Lock presents a review of the way in which the thinking has shifted both in molecular genetics to challenge the assumptions of genetic determinism, and in the social science literature that has considered the social effects of molecular genetics (Lock 2005:47). Lock goes on to show that the findings of her research into testing for Alzheimer's Disease 'illustrate the provisional nature of these bodies of knowledge and the complexity associated with susceptibility genes, which makes estimations of probabilities of individual risk unrealistic' (Lock 2005:47). Lock's paper was particularly helpful in thinking about the genetic risks that will be generated by GS in Chapter Nine. She discusses the social implications of testing for susceptibility genes and the way in which a particular group of individuals took the information they were given about their risk status and 'incorporated it into their already-well-established ideas about who in their families were likely to get Alzheimer's Disease in the future' (Lock 2005:59). In her conclusion, Lock makes the point that no systematic research has yet been carried out on the effect of recruiting many thousands of healthy people to population genetic research. She also asks as a matter of urgency

> what counts as well established knowledge in the world of genetics, what in effect is knowledge-in-the-making, what is essentially nonsense or bad science, and what impact this plethora of confused [risk] information is likely to have on publics as they are increasingly asked to contribute to research and submit to genetic testing
>
> Lock 2005:59

The apparent contradictions between the ideas of established genetic knowledge and the provisional nature of genetic knowledge are evident in the GS database as a project that is all about knowledge-in-the-making. Knowledge about

susceptibility genes and the predisposition to diseases generates information about 'risk'. The interpretation and understanding of risk from genetic testing, and its implications for individuals and families, identifies issues that could arise for participants in the GS database collection.

The second book published on population genetic databases comes from the Ethical, Legal and Social Aspects of Human Genetic Databases: A European Comparison (ELSAGEN) project and it continues with the themes of ethical questions, legal issues and social concerns raised within the ELSI discourse on human genetic databases (Hayry, Chadwick, Arnason and Arnason 2007). The contributors worked on the ELSAGEN project in Estonia, Iceland, Sweden and the UK and compared public perceptions, regulations and matters of harm and benefit between these countries. They find both global similarities and local differences between the countries and the projects that have been proposed or to some extent implemented in each case. Their work points to the possibility of framing population genetic databases as assemblages.

Stephen Collier and Aihwa Ong take up the idea of assemblages in *Global Assemblages*, where they frame technology, politics and ethics as anthropological problems (2005). They suggest 'the phenomena that concern social scientists assume spatial forms that are nonisomorphic with standard units of analysis' (Collier and Ong 2005:3). The forms of inquiry used by the authors in this book stay close to practices. Global forms are articulated in specific situations 'or territorialised in assemblages – they define new material, collective and discursive relationships' (Collier and Ong 2005:4). Assemblages create

connections and disconnections between what was there before and what is newly constructed. The forms and values of individual and collective existence are problematised in these sites because 'they are subject to technological, political and ethical reflection and intervention' (Collier and Ong 2005:4). The idea that GS is geographically 'local' but is responding to 'global' forces is a recurring theme in Chapters Three, Eight and Nine, but it is the idea of connection and disconnection that has become central to this thesis.

Gisli Palsson and Paul Rabinow also view population genetic databases as assemblages (2005:91). Palsson and Rabinow see an urgent need to reflect on the social implications of the production of biomedical knowledge, the social science engagement with this production and the transnational institutional conditions of existence (Palsson and Rabinow 2005:91). They consider the ethical issues of DeCode Genetics and the development of the Health Sector Database in Iceland. They also call for more comparative analysis of projects that are being developed in other countries, and the transnational development and practices of bioethics. They point to the particular relevance of the forces and interests at work in the 'market sphere'. Iceland has become 'the site' of biotech and bioethics while projects in other countries 'are almost completely ignored' (Palsson and Rabinow 2005:92).

Gisli Palsson and Kristin Hardardottir have previously addressed debates about biomedicine in the local context of Iceland (2002). They used the idea of moral landscapes to understand the debates that were informing public discussions. The debates focused on ownership and the marketing of biomedical material

and information, patient's rights, informed consent and the protection of privacy and personal autonomy (Palsson and Hardardottir 2002:271). They used a multi-method approach to chart the 'topography' of local debates showing that, unlike other countries, there has been extensive public debate in Iceland on the issues surrounding population genetic databases, which they attribute to 'earlier and ongoing debates on neoliberal politics' (Palsson and Hardardottir 2002:271), which drew on the reorganisation of fishing rights. They also show how 'folk' discourse on genetics and inheritance is complex and even contradictory. They conclude that the topography of the moral landscape is changing.

In *Anthropology of the New Genetics* (2007) Palsson considers biobanks, medical records and genetic databases. The combining of genetic, medical and genealogical information represents the extension and intensification of governance and the biomedical gaze (Palsson 2007:91). The use of human body components for extraction, storage, exchange and commodification are highly variable and context specific (Palsson 2007:92). Palsson goes on to examine assorted types of 'banks' and case studies the structure and histories of seven projects in different countries. All these projects have fundamental properties in common 'that distinguish them from other kinds of biomedical projects and assemblies' (Palsson 2007:111), although some are of a hybrid form so that they elude clear-cut categories. The projects also differ in some respects in the way they are assembled and used in 'the processes and criteria of sampling, property arrangements and management regimes, the nature and degree of commercial involvement, forms of consent as well as measures for the protection of personal

information' (Palsson 2007:112). Palsson also found that there was 'much formal and informal collaboration among them' (Palsson 2007:115). He concludes that anthropology can contribute to the study of genetic databases by providing a reasonable thick description of ongoing developments and a comparative perspective of existing plans and projects (Palsson 2007:122).

Many of the authors mentioned above call for social science research to pay attention to science and technology. They point to a need for a better understanding of the science and technology if we are to examine the relationships between the science and the social world. However, the technicalities of science and technology are not easy to come to grips with. The textbooks and papers on human genetics, which the lab technicians gave me to study, were way beyond my outdated school level biology. A hunt in the library discovered *Human Genetics for Social Scientists* (Carey 2003), and thinking this a hopeful title I borrowed it. I have relied heavily on it for explanations of scientific terms and processes. Roughly two years and several renewals later it occurs to me that there is not a big demand for this book, it has never been recalled.

The literature I drew on reflects two problems I was dealing with simultaneously. First, how to write about the complexity of transforming blood into digital data, and the various aspects congruent in the construction of a population genetic database. And second, of how to deal with the apparent divide between the science and the social world. There was no 'grand theory'

but concepts emerged that I could use to make sense of my fieldwork and guide the thesis.

Terms and Concepts

I struggled for words, terms to describe and relate to various ethnographic elements. I was not even sure that ethnographic work described exactly what I was doing. The features that most strongly emerged from the fieldwork were those of disconnection and connection. The disconnections and connections were remaking familiar relationships, for example kinship, giving them a new meaning and utility for a contemporary social world. The connections and disconnections of Generation Scotland and the collective enterprise of constructing a population genetic database was difficult to describe. I found the concept of assemblage useful here. The term is difficult to define precisely but its utility in this case is in its flexibility. Assemblages are about relationships, but something more than a collaboration, although that could be the starting point. Assemblage refers to a type of interaction that could be shared between either individuals or institutions, or both; it appears to have emerged strongly in studies of molecular, biology and technology interfaces. It has particular characteristics that include interdisciplinarity and/or multidisciplinarity, local and global applications, and an assemblage is contingent, unpredictable and complex. Assemblages are temporally specific and appear to either become more structured over time or to disaggregate (Collier and Ong 2005).

The use of the terms and concepts relies on an assumption that readers know what I mean, but the use of terms hang on unexplained and often questionable

assumptions which I make and invite the reader to make. For example, 'social world' is familiar but as Bruno Latour (2005) points out, problematic. Here I take 'social world' as referring to Scotland, but the unity of this geographic area includes people from diverse communities in rural and urban settings, cultural differences, variable economic situations from poor to wealthy, disparate backgrounds in education, assorted employment and unemployment, and groups such as Highland dancers or prison inmates. Collectively they are Scottish and can be presented as a particular social world, yet this includes a great deal of diversity, and represents simultaneously individuals, families, groups and the population. This 'social world' is explicitly 'local' but has the potential to become 'global'.

I also had problems with using science and social as terms of separation, because I think that science is integral to the social world. However, I use the term science in the conventional way to differentiate a field of study, a way of thinking about the world with a particular set of methods and practices that are perceived to be different from social enquiry. 'Science', however, embraces a broad range of subjects, people with specialist knowledge, places, spaces and equipment. Here, I use science to refer to genetics in the field of medical research. Likewise, 'technology' is a blanket term for a plethora of practices, people and equipment. I use it to refer to the information technology of the database and the scientific technologies of the genetic lab used in transforming blood into digital data. The scientific terms and concepts are also layered, ambiguous or based on assumptions. I examine key concepts such as gene,

genotype and phenotype in more detail in Chapters Six and Seven to show how they have multiple and complex meanings.

The utility of the concept of biosociality in explaining the formation of social groups around ideas of genes and risk (Rabinow, 1992), also has a broader use for thinking about the separation and linking together of science and the social world. In Chapter Nine I use biosociality in its narrower sense to consider the connections and disconnections of genetic risk and group formation in Scotland. However, in the broader sense the thesis shows the population genetic database as a version of biosociality. Here, scientists are collecting 'facts' about social bodies, relationships and experiences and transforming them into digital scientific data called phenotypes and environment. If the social world is searching for ways to understand and relate the new genetics to lived experience, then, the new geneticists are finding ways to relate social data to scientific objects such as genotypes, in order to understand them better. This broader notion of biosociality is not developed in the thesis as it would add another layer to an already complicated study, but it does suggest interesting possibilities for further work.

This thesis does not locate itself easily in the literature, it is part science and technology study, and part social study, but orientated toward public understanding of science and public engagement. It comes closest to the work done on assemblages, as a particular type of anthropological problem.

Chapter 2

Methodology

I was throughout this project both an individual researcher and a team member. I could not rigorously separate these roles they intersected and often happened simultaneously. I used different research methods including comparative analysis of literature and verbal discourse, much of which was taking place across disciplines, observation of the collection of blood samples, the practices and processes in the lab, and the data analysis at the desktop. I was participant in the meetings as the 21CGH project was developing, and participant-observer at seminars and conferences in the dual roles of postgraduate student and representative of Generation Scotland. I also drew on a diverse range of literature. This 'methodological eclecticism' has been characterised as 'reactive' and 'edifying'.

> An edifying discipline, by contrast [to a systematic one that seeks objectivity], distrusts the notion of essences and is dubious about claims that reality can be accurately, holistically, singularly or disinterestedly explained and described; for not only is there contingency and diversity of existing epistemological regimes, there is also the poetry of the new. In the face of an essentialist enquiry, therefore, the edifying account seeks to maintain a conversation between different ways of being in the world and eschews any singular, authoritative framing.
>
> Rapport and Overing 2000:248

I wanted to use the thesis to initiate such a 'conversation' by combining descriptions of the practices and processes being used to construct a population genetic database, with explanations of the genetic science, and incorporating the

ethical, legal and social (ELSI) issues that have been raised by specialists and studies of public engagement. By taking this approach the aim was to construct a synergy of perspectives across the 'divide' between science and social science; to go 'beyond the two-culture divide' of anthropology and science (Goodman, Heath, Lindee 2003). It is a somewhat experimental approach and less than poetic.

The story follows the blood from arm to desktop. It includes, as far as I was able, the scientific explanations because these are the reasons and knowledge that shape the database and inform the processes. But, writing the science into the text was difficult because it sits uneasily, loaded as it is with technical terms and unfamiliar concepts. Equally, it was important to include the ELSI issues that arise in a social world that condones, even demands, genetic research to find cures for diseases on the one hand but is often appalled by the possible consequences on the other.

Science and technology studies tend to use two approaches, one to examine the culture of science and technology, the other the practice. I chose to focus on the practices because I wanted to foreground the scientific objects, to come as close as possible to understanding the objects that were being produced and manipulated through the work of constructing a population genetic database (Rabinow 1996; Mol 2002).

The fieldwork followed the blood, mapping the topography across which the blood was flowing, but it also made a pathway through GS as the database was

being constructed. This gave me a unique view; moving around the local sites for a time, I saw connections and disconnections between people, places and objects that others did not. My activity not only connected me to different sites and people but, moving between them, I created connections between people that rarely or never met. I found myself carrying stories and information about other people and their work between sites, and answering questions about what people were doing in other places. Indeed, through the multi-sited research I became a 'circumstantial activist', renegotiating identities in different sites, and with cross-cutting and contradictory personal commitments (Marcus 1995:113).

Ethical Approval

I did not need medical ethical approval to observe the collection of blood samples. I contacted the Multiple-sited Research Ethics Committee (MREC) to enquire about ethical approval because observing the collection of blood samples would of necessity mean that participants were also involved, and it was probable that I would hear and possibly see personal information. I wanted to observe research nurses collecting blood; I did not intend to ask any questions of the participants. I was asked to send an outline of what I was intending to the MREC secretary who said he would put it before the chairman. I received a reply stating that I did not need ethical approval to observe research nurses at work. I felt that because participants were involved this was a grey area and contacted the MREC a second time just to be absolutely sure. I was told that what I wanted to do was 'not research' and that I definitely did not require ethical approval. I also contacted a member of the School of Law for advice about ethical approval from the school. The reply I received said that research

ethics were under review and they would get back to me. I never heard any more.

It appeared that as my project was in effect social research into medical science research it fell into a strange grey area that was clearly not clinical, yet could potentially involve access to personal information about people participating in clinical research projects. It seemed to be taken for granted that I would treat any personal information of research participants that I came across as confidential.

It was important to me to be as ethical as possible in my work. I made it clear to everyone that I approached that it was likely that I would write about them in the thesis. I offered to show what I had written to everyone that I talked to or observed before I showed it to anyone else. I undertook to give consideration to any comments they might have, though I did not commit myself to making changes.

Being in Generation Scotland

My studentship specified that my research should contribute to GS public engagement and that I would be a member of the Ethical, Legal and Social Implications (ELSI) team. I return to this below to explain why this thesis is not about public engagement. The remit for being a member of the ELSI team was not specific, but from discussions at my interview, I would be required to contribute any useful experience or knowledge I might have to whatever matters arose, collaborate with and support my colleagues. I did attempt to

construct a mental and physical separation between my work on GS and my PhD. This was achieved partly through differentiating my role, either as member of the ELSI team and therefore working on GS, and as PhD student, therefore studying GS. These I enacted in different ways, firstly through the use of space. I used my time in the office and with other members of the ELSI team as GS work, and moved into an isolated space when doing my own work. I kept separate notebooks, one for GS and one for my PhD work and thesis. I also chose to locate the research for my thesis in what appeared to be a space between the ELSI and the scientific work. I thought this would create a separation between the project work and my work which would avoid clashes and overlaps with what other people were doing, and in theory reduce confusion and possible conflicts of objectives and timeframes.

However, the knowledge that I had about GS became a problem for me when I came to write up, there were things that I knew that seeped into the writing but I could not, or would not, offer an explanation of how I knew these things. This came about for two reasons. First, right at the beginning at my first supervisory meeting I had asked for access to documents, memos, letters pertaining to the setting up of GS, I also asked for permission to sit in on meetings between Principal Investigators and others who were planning and organising the GS bid. These requests were denied. The reason given was that they were still in negotiation and it was too sensitive. It was made quite clear that I should not think about investigating this aspect of GS.

I resolved not to do this but at the same time I could not ignore what was happening in front of me. I have a stack of notebooks that contain quantities of notes and reflections on what was happening that I kept firmly shut when I came to write up. Much of the story within them is fragmented, partial and probably regarded by at least some senior members of GS as confidential. Scientists in particular are not in the habit of exposing or discussing in public the more messy aspects of projects that go on behind the scenes, uncertainties, changes of mind, new directions, negotiations and so on, as Bruno Latour showed (1987).

The second reason I felt I could not use the data I had collected and my inside knowledge of GS in the thesis comes down to a matter of ethics. As Klaus Hoeyer, Lisa Dahlager and Neils Lynoe write there is a mismatch between medical ethics and social science ethics (Hoeyer, Dahlager and Lynoe 2005). Because I had not intended to write about GS I never sought permission or made it clear to any of the team that I was likely to use what they said either at meetings or in conversation. To do so now would be a betrayal of trust that is invested in the personal and professional relationships I have established with many people. Would I do it differently another time? I don't know, I think it would depend on the circumstances. I think the suggestion of negotiation throughout a project, where as an ethnographer your perception of a colleague or collaborator can change to that of informant, could be useful (Hoeyer et al 2005). But I wonder how that would change the nature and dynamics of relationships of trust. Would I risk my personal relationships? I can say in this situation that there are some that I would be willing to risk and others that I

would not. The problem is one of 'taking it personally'. If you write about people you work with there is a danger that they could see a critique of (or even a question about) their role, position or actions as a personal criticism (Hoeyer et al 2005). I found on several occasions that comments I made were taken as personal criticism rather than critical questioning of assumptions by people unused to this approach. They tended to view comments as fixed opinion, or as creating complications, rather than an opportunity to open up a debate or discussion. It seemed to be one of the fundamental differences in a multi-disciplinary setting, scientific methods work to close down options into immutable facts, while the social science seeks to open up debate and discussion.

I had a difficult time when I came to write up because although I had tried to keep the knowledge and data separate it was in fact all interconnected, meshed together, and interdependent. Every time I started to write a new section I struggled with what I knew and what I could write, and how to handle my field notes. I hunted through the literature for help. It became evident that different authors had taken different approaches, some more useful to my agenda than others. For example, I appreciated Paul Rabinow's work on French DNA and Making PCR, but he had used interviews which I had not, and clearly with full knowledge and consent of his informants, to the extent that they often became co-producers of the text. I had brief and fragmented bits of conversations with the research nurses, lab technicians and others, but I had not used interviews. Because my attention was primarily on the scientific objects and not the social aspects of the project I did not think at the time that this would be a problem. I

had also decided not to use interviews because they would have introduced a formality and a different type of data to the informal conversations of participant-observation. Nor in many cases were my informants keen to be 'seen' or indeed 'heard'. They were just doing their jobs and I did not encounter any situations where any of them wanted to involve themselves in the 'politics' of the projects they were working on. I did not ask them to reflect on their roles in genetic research. This might have been different if my stated intention had not been to focus on the blood and digital data. But as Hoeyer et al (2005) state, the ethnographer cannot always explain or be specific about their interests so that other people can understand their intentions.

Does this make me a squeamish researcher? I have heard it argued that all data is 'fair game', and that anyone who lets an anthropologist through the door is open to anything they write. I can see their point, but I think those who are not familiar with anthropology, coupled often with our inability to be explicit about what we are doing, leaves people open to being shocked and feeling betrayed (Hoeyer et al 2005). I think the fair game approach while probably producing very interesting work risks creating problems for individuals and more generally for anthropology, particularly in multidisciplinary projects. I personally do not wish to have a reputation for lacking integrity or being unworthy of trust, even at the expense of my own work.

The power relations of this type of work are not always clear, and I found that I moved through different positions of power within GS. At times I was a seemingly powerless PhD student attached but not part of the project, then a

member of the ELSI team which changed my position to one of having a contributing voice and the possibility of influencing pieces of work at times, to one of author with the power to write whatever I wanted. Yet, as author I was conflicted because I was aware that i) I had to produce a piece of work that was acceptable for examination as a scholarly piece of work, ii) that there were people within GS who could object and even block what I wrote (all papers for publication have to be approved) iii) that there were people that could view what I wrote as personal criticism and feel that their trust had been betrayed, and iv) that I held my own personal views of what would and would not be ethical.

Writing a background chapter about GS gave me the most problems in terms of how to approach it. Having previously resolved not to write about GS and then finding it necessary to attempt an explanation of the context and mechanisms through which the population genetic database was being constructed, I searched for literature on collaborations but found none in either social science, management, or business studies. It is possible that I looked in the wrong places but even so it strikes me as strange that in a world where there seem to be a rapidly growing number of collaborative enterprises, for example funders active promotion of these type of projects, that I could only find a couple of papers on how to run a collaborative project and none on how to analyse such a venture. I turned to literature on organisations, again not helpful on the subject of collaboration. In the end I returned to the Gellner and Hirsch *Inside Organisations* (2001) (in which I had previously read a chapter on lab work) and found the Mascarenhas-Keyes chapter on doing a brief organisational analysis.

This suggested not just a useful way of structuring an analysis but that a degree of distance could be achieved, as I was anxious to remove myself and my experiences from a central position, and the conflicting problems of truth and trust.

Teamwork and Communication

I observed as I moved around GS locations that people were keen to talk about what they were doing, their feelings about GS, successes and frustrations that were not shared on a day to day basis with others. At one point I realised that I was spending a lot of my time calling in on other people to see how they were and although I could not necessarily help them solve their problems I provided a listening ear and supportive encouragement. There were two reasons for this, first I had worked on previous projects where 'team' members were isolated from each other, and I was aware that in a multidisciplinary group people can feel quite cut off; second, as a trainer in teamwork elsewhere I knew that communication and contact need to be maintained if people are to have a sense of being part of a team and working collectively toward a single shared objective. People also talked to me about their personal relationships, homes, children, holidays - their lives outside GS.

As an undergraduate and contract researcher I was aware of a certain element of rivalry between the Scottish higher education institutions, their medical schools and research agendas. I would speculate that the formation of both GS and UKBiobank Scottish Spoke were not without their elements of rivalry, political negotiations and strategic alliances between individuals, departments and

institutions. Not that I would suggest anything acrimonious, rather the existence or creation of tensions between loyalty and expediency. I have no evidence to support this, but occasional comments and the fact that I knew people in different places led me to suspect that on occasions this might have been the case. As previously stated I had initially asked for permission to see early documents pertaining to the collaboration and attended meetings, however, I was denied permission on the grounds that this information was too sensitive and was not an appropriate area for my research. A couple of years later another anthropologist did join the GS team for a while to attend meetings to see if or how findings from the public engagement work being done did in fact have any effect on project decision making. I met her briefly once and I have not seen the results of her work. From my point of view as this aspect of GS was being covered by someone else and was an element of a different agenda this was not an area that I should contemplate addressing.

Only one of the research fellows (GH) from the ELSI team was involved with the Scientific Management Committee. This made a difference to her level of integration in the project, level of involvement, knowledge of individuals who were making decisions and opportunities to develop relationships within the group that were in effect denied to the other research fellow (RG). In terms of collaborative work it is such events that, whether intentional or by chance, can or might, shape the dynamics of an assemblage in terms of people feeling connected or disconnected. Teams only hold together if they are connected through shared experience, objectives or work. However, in a multi-disciplinary team it is not easy to predict at the outset who will end up working together and

who will not. Those that work together have more to share and so create connections that are not available to others. It also meant as GS developed, because I was involved in some areas and not involved in others, that I saw some parts but not all.

GS split into the Genetic Health in the 21st Century (21CGH) project and the Scottish Family Health Study (SFHS) project, which will be explained in Chapter Three. RG and I had thought we would be part of the SFHS project as well as 21CGH since we were part of GS, but in the event neither of us was involved with the SFHS. The public engagement aspect of GS shifted to focus on the SFHS, and took only one member of the ELSI team with it. This in effect created a cut off point for both RG and myself. Without access to what was going on or an idea of what we might contribute to the SFHS project we shifted our attention away from questions about families, recruitment and participation. The spilt in the GS projects had a definitive effect on my previously proposed involvement in the public engagement and I shifted my full attention to another aspect of the project that had been interesting me, that of blood and data.

Being on the Genetic Health in the 21st Century studentship

Supervision of the studentship was already decided and came from within 21CGH, a geneticist and Principal Investigator (DP) and a lawyer (GL). What I did not initially comprehend was that this set up a tension between my desire to produce a competent anthropology thesis and the interests of GS. I had started the studentship under the impression that I would be supervised by a sociologist (SM) from the Medical Research Council (MRC) unit in Glasgow.

What agreements were made between the various supervisors I do not know, but from what ensued I surmise that they were informal and they were not sustained throughout the studentship. Nor do I know how it was decided that I should be based in the School of Law. However, although not an ideal situation, in the interest of collaboration, which I knew to be central to GS, I thought it would be workable. There was another PhD student (ML), a medical historian, at the MRC being supervised by SM and KH, another member of the unit. It was proposed that the four supervisors would form an advisory group for both of us and that we PhD students would work in collaboration, with both of us studying both GS and UKBiobank. Meetings were a problem because the supervisors were busy and it was difficult to get a date when all could be present, and it required that three people also had to travel across the country adding time onto the meetings. For the main part I saw GL with occasional visits to Glasgow to see SM when I felt I had a problem that I wanted to talk about. The enthusiasm and support of GL was unquestionable but I quite often felt he did not understand what I was thinking about, or why. I also frequently felt that he saw things in a particular way that I could not quite grasp. The occasional meetings in Glasgow where I felt what I was doing was better understood were not frequent enough but were reassuring. I saw little of DP. My co-PhD student in Glasgow ML chose a top down approach to her work on UKBiobank through interviewing stakeholders. It seemed to me that if I took a bottom-up approach it would be complementary to this, and it also appealed to my anthropological sensibilities.

It was initially envisioned that both ML and I would carry out research relating to both GS and UKBiobank, but as our initial work progressed it became clear that circumstantially I was closer to GS and she was closer to UKBiobank. This was not a deliberate differentiation. It also became clear that although there were strong connections between the two projects, particularly in Scotland through the multiple involvements of many individuals, both the projects were progressing in quite different ways and to work on both simultaneously would undoubtedly lead to an even more complex field of research than we were already facing. Attempting a differentiation between the two projects at that point seemed quite an important objective since they were so often confused and blurred together in the Scottish context. Circumstantially GS was strongest in Edinburgh, i.e. had the most people working on it there, and UKBiobank Scottish Spoke was strongest in Glasgow where its co-ordination centre was located. Key people for the different projects were located on each side of the country. The geographical split might have been due to political influence or might have been arbitrary based on the already existing location of key people. It was not clear if this was a deliberate strategy to generate a distribution of resources or coincidental.

In my research the geneticist (DP) wanted me to propose a hypothesis and then prove it through my research, whereas the lawyer (GL) wanted me to construct an argument. I could see what they wanted me to do and why, but I felt capable of neither. A lack of knowledge and theoretical underpinning left me in a position where I found it difficult to articulate more than vague questions and without a framework around which to construct my work. In the midst of a

multidisciplinary project, where everyone else seems to have very clear ideas about what they think you should be doing, it is a distinct disadvantage to have doubts, and certainly not a position from which you can argue with any conviction. What I did have was an idea about 'following the blood', a multi-sited ethnography around which to structure fiiieldwork, analysis and writing. This was my one clear thought and I clung to it hoping that it would hold together.

I have a tendency to want to take a broad sweep approach rather than an in-depth focus when addressing a new topic. I think from an anthropological perspective you want to question everything. The problem then becomes one of sheer volume. There are so many questions – how do you find a focus? How do you reduce what is in front of you to the manageable proportions required of a single piece of work within a limited time scale? Now a more disciplined mind than mine would cut straight to the chase, or at least so I imagine. I had some vague recollection of very basic genetics from my school biology, and was familiar with the idea of inherited traits running in the family. There were of course the media announcements on the discovery of the 'gene for….' various human diseases, conditions such as baldness, and interest in criminal and social behaviours. I did not have an in depth knowledge of the new genetics or the practices of science, which I return to later in this chapter.

Being in the School of Law

I had come to the AHRC Centre in the School of Law from a Department of General Practice where I had become accustomed to being located within a

different discipline. I have not spent time in an anthropology department since I was an undergraduate and I wonder now, as an aspiring social anthropologist, if that accounts for at least some of my lack of confidence and courage in my own convictions. I have perhaps not learned to present and argue my position from my anthropology peers but rather have become accustomed to feeling that most of what I have to say is not understood by my fellow researchers, nor indeed that I can always understand what they are saying. Instead of becoming competent in concepts, theories and methods of anthropology I have been continuously learning to come to grips with those of other disciplines. I have come to think of myself as something of a 'mongrel' or perhaps now I would claim a 'marginal' (Cohen 1985) anthropologist. Marginal in anthropological terms means that I can cross-boundaries, move between people, places, ideas and concepts, but it also means I don't really belong anywhere. If I now claim to be an experienced multi-disciplinarian my worry is not that I personally cannot fit in but that my work does not. That said, this marginal anthropologist aspires to use this situation to create and describe a space that opens up the possibility of a particular interaction between science and social science in the creation of a population genetic database. To give this notion an analogy – I imagine this thesis as my Pompidou Centre.

The Pompidou Centre in Paris is a distinctive building surrounded by more conventional architecture. The first time I saw it I thought it ugly and out of place but up close it made me see things in a different way. First of all the 'guts', the working parts of the building, plumbing, electrics and so on, are on the outside where you can see them. Second, you can see inside from the outside,

and outside from the inside. But most impressive are the spaces, big open spaces inside and out so that you can see across, and, of particular importance, see what people are doing all around. As a structure it may not be aesthetically pleasing, and removing habitual barriers leaves everyone exposed, but the spaces are in fact all about interaction. The designers of the building, Richard Rogers and Renzo Piano, wanted to create spaces, not compartments, big open spaces so that people could see each other and what was going on, to open up the possibility of new interactions, new ways of seeing each other and communicating.

Deploying the anthropological gaze in such a setting of course creates a further problem – where to look and how to look. A narrowly focused gaze in a confined space will give depth and detail. What I ended up with looking across the fascinating space of GS were a whole range of activities and actors. Many actors were largely unaware of the others inhabiting the same conceptual space that Generation Scotland created, with no sight of the others or the shifting assemblage to which they belong. But I had to focus my work and in the end I looked at what I thought of as the guts of GS, at the construction of the database itself. If it works well, or even if it does not, I would hope that it could suggest ways for other researchers to think about how to approach this type of work.

In the first year I co-authored a chapter for a book with my lawyer colleague (Marsden and Gertz, with publisher). We were working on a presentation (with an intended paper to follow) on families to be made at an international conference. We were initially given a count of 3000 words for the chapter but

when we submitted we were asked for a further 2000. The chapter was accepted for publication and describes how we had worked together on the presentation and paper. It was undoubtedly naïve, and reflected in some ways our dependence on each other and isolation from the rest of the GS team in two ways. First, that we were separate from the scientific, technical and practical aspects of the project and second, from the discussion and decision-making layer of the project. We finished the chapter but we never finished the paper we were going to write. The paper fell into the abyss of unfinished work because firstly we didn't have any data to add to what in effect was a review of recruitment practices, and it coincided with the GS split into 21CGH and SFHS.

The chapter and the presentation were written from the ELSI position which was peripheral to the scientific and technological work. While it was clear that there was much (particularly for the lawyer) in the practicalities that we could not contribute to, there was also a sense of being excluded at times. I was also somewhat frustrated because I had practical experience from previous work that I felt I could contribute had I been given the opportunity. For me there was, at least in the first couple of years, a continuous tension between being part of the GS project and getting on with the PhD which I was advised to think of as my own work, and therefore somehow separate.

GL went on sabbatical for twelve months at the end of my first year, I was not perturbed by this as the period would cover my planned fieldwork. However, the first day I walked into the lab I felt I was in trouble. I felt completely unprepared for what I encountered and panicked. Now, I am aware that this is

not an unusual fieldwork experience but then I thought I had made a dreadful mistake. After a flurry of e-mails to GL and a meeting with the DP later it was agreed that I could seek an anthropologist as a third supervisor. Fortunately for me this was successful, JC agreed, and having poured out my concerns she told me in the calmest and most down to earth way to go and do my fieldwork. She recommended I took Bruno Latour to bed with me - for a little night-time reading.

How much Science?

One of the more controversial questions with regard to ethical, legal and social aspects/implications research is - how much of the science or technology do you need to understand in order to address the issues that arise? Scientists, health professionals, clinical researchers and politicians have been arguing for some time now that the public need to understand science better but what exactly they mean by that is not clear. Do the public, and ELSI researchers, need to understand, the theories, concepts, processes, practices, results, scientists, or the culture of science? The study of science raises divergent views. I asked fellow PhD students and people that I met at ELSI meetings or conferences if they thought they needed to understand the science. The majority thought not, with many saying that it was too difficult and a few that there was no need. In these informal conversations I heard it argued, on the one hand that social scientists should study science subjects, take courses or even degrees in their subject of interest, but on the other, social scientists don't need to understand the science to comment and critique its practices and implications for the social world. A few thought that social scientists should study the science if they were to write

about it. In the literature there are calls by some authors for social scientists to study the science, which poses the question – how much science do non-scientists need to understand?

Studying science as a scientist requires the individual to learn from a conceptually specific perspective. The crossing of the culturally established disciplinary divide between 'arts' and 'science' demands that one commonsense view of the world is suspended or even displaced in favour of another. I think in Scottish (and other western) society the idea of a non-scientist studying science is difficult because we have all been streamed through the educational system at secondary and tertiary levels to be either one or the other. My solution to the apparent contradiction was to approach the science as an anthropologist trying to understand a subject of investigation, just like any other. I learned the science in a particular way, through observation. If I had learned it as a science student I would have 'learned' it differently, as someone who would 'become' a scientist.

When I showed early drafts of chapters to the lab technicians and statistical geneticist they all said they would like to see more of the science included. When I showed the same drafts to non-scientists they mainly commented that there was too much science. I wanted a balance between the social and scientific aspects of the database. I did not want to loose sight of the science to, what to many of my colleagues was, the more interesting social aspects of the database project. Conversely, I found that by putting a lot of effort into the science I did at times loose sight of the social factors and significances, and struggled, often unsuccessfully, to reconcile the two different views of blood, DNA, data and technology. This is a key difficulty when writing in the conceptualised gap

between the two fields of research. Different disciplinary views, methods and language habitually work to mark themselves as separate and resist attempts to merge them together, added to which I was also having to overcome my own conditioning. The science did at times 'suck me in'. When I felt it was getting too much I would look for something anthropological to read or find a social scientist to talk to or go to a seminar to get my head back in the 'right' place and remember how to think like an anthropologist. When I was writing I had to switch between the two modes of thinking, which meant there were many attempts at each chapter before I could find a balance.

There are however limitations to learning the science through observation. I did not have the knowledge, skills and training, required to participate. Constrained to the role of observer places limits, not so much on what can be said but rather how it can be said and how it can be 'known', for example, the kinesic knowledge (Rapp 1999) is absent from the descriptions of practices.

I have written the chapters, particularly chapters four (venepuncture), five and six (lab work), to reflect the science and the social aspects of the project as they were 'on the ground' at the time. The science is therefore often presented, as it was in practical terms, separated from the other aspects of the project. This was not intended to reify the science but rather to show that there are different views and arenas in which people enact the construction of a population genetic database.

Historical Accuracy

The thesis was not written with the intention of constructing an historical account. Retrospectively it was a mistake to overlook this aspect of the data collection and writing. This pertains in particular in chapter three which is intended to provide a background to the database construction. The development of GS, the projects collecting the data and the construction of the database were ongoing, and indeed still continue, throughout the four years from the time I joined GS to the completion of writing the thesis. Many of the details of the project changed during that time. Some of what is presented here is out of date, whilst much remains current. I started out aiming to capture a 'moment' in time in the construction of a population genetic database. However, by choosing to follow a 'pathway', which of itself created a momentum, I did not clearly define the moment. Rather, the story covers a 'period' during which many things were happening, sometimes simultaneously, but in which none actually came to a conclusion. Consequently the aim shifted to capture a sense of these changes and shifts without explicitly focussing on the exact timing. Dates are as accurate as I could make them, but the changing nature and multiple sites meant that sometimes I did not hear of things until after the event, or something that was decided at a particular point was subsequently altered.

The historical unfixity of this approach also gives rise to a problem of tenses, to write all of the analysis in the past tense would be inaccurate because much still persists, but to write all in the present would be inaccurate because some things have changed, and in the later chapters eight and nine much is still conditional. I found I was switching between tenses, which has the potential to confuse the reader further in an already complicated description, and one that persists

throughout the thesis. It does, however, capture a sense of the temporal contingency and shifting field in which the thesis was written.

Fieldwork

Blood collection.

I approached the Wellcome Trust Clinical Research Facility (WTCRF) at the Western General Hospital, Edinburgh to seek permission to: i) identify projects that were collecting blood samples for genotyping; ii) approach these projects for permission to observe the collection of their blood samples; and iii) talk to and observe research nurses, if they were agreeable. The director of the research nurses gave me permission to talk to and observe the research nurses, and approach the projects which were collecting blood samples for genotyping. The lead research nurse helped me identify which projects would be suitable to approach for permission, and introduced me to the research nurses working in the clinical research facility. I was asked to sign a confidentiality agreement, which I did.

I approached the principal investigators of three projects that were running and one that was about to start for permission to observe the collection of blood samples. Two gave permission for me to observe the blood collection. Then it was a matter of appointments. The numbers of participants coming in to the clinical research facility for these studies were low, averaging one or two per week. I checked with the lead research nurse every few days for participant appointments. I attended the facility for each appointment and waited while the research nurse asked the participant if they would permit me to observe the

collection of their blood sample. Most participants permitted me to observe the whole appointment and often chatted to me when the research nurse was occupied with equipment, paperwork or as a distraction while the needle was being inserted into their vein.

I had also contacted a research nurse I had worked with previously. I knew she was working on a project that was collecting blood samples for genotyping at a different university hospital. She approached the principal investigator of the project on my behalf with a letter requesting permission to observe the blood collection on that project. I was quite hopeful of a positive response because the principal investigator was also involved in the GS collaboration. Permission was given. The research nurses working on this project told me they were also doing the recruitment, sending out information to patients and arranging appointments. They were working in clinics and general practice. I was permitted to observe in the general practices but not in the clinics.

The appointments in general practice had to be organised around the availability of rooms in medical centres. The research nurses would take several weeks to organise a room and then fill the time with appointments. The research nurses then travelled to the medical centre with all their equipment and saw as many participants as they could get to attend that session. I checked in with the research nurses every few days to find out about sessions. Most of their work was being done in the clinics so there was a wait until I could finally go to a session with them. As previously I waited outside the room until the research nurse had asked the participant for permission for me to observe. All the

participants agreed and as before some of them chose to chat to me when the research nurse was busy or inserting the needle.

A third possibility opened up around the time I was starting my fieldwork with the research nurses. Two research nurses from another research project were due to join the GS project and I was able to contact them and receive permission to go and observe them working on the final appointments of their current project. I had estimated ten to twelve observations of collecting blood samples. The addition of this site brought the total number of observations up to fifteen.

The total number of observations of venepuncture was probably more than was necessary, but when I began to write up the fieldwork notes it became obvious that the role of the research nurses was more important and complex than I think they are usually given credit for. Because I saw the whole appointment of most of the participants in different settings and for several projects it became evident that the role of the research nurses was multifaceted. They were interacting with the participants in diverse ways whilst collecting various types of data including the blood sample, and all against the clock.

Once I had written up the first draft of the chapter on the collection of blood samples I sent copies to the research nurses at each of the sites for comment. I received none. I made one follow up contact to each site to ask if anyone wished to make a comment but received no responses.

The Laboratory

I sought permission from the acting lab manager to observe the processing of blood samples into digital data carried out in the Genetics Core laboratory at the WTCRF. This was the lab where all the blood samples from GS projects would come for DNA extraction, genotyping and sequencing. The acting lab manager had been closely involved with the development of the GS proposal for several years. I was known to her through GS meetings and discussions. I was given permission to spend a couple of weeks in the lab observing the lab technicians and learning about the processes that were carried out there. I felt that this would not be long enough but accepted that this was an opportunity not to be passed up and that I would have to make the best of it. At the end of the two weeks I asked for two more and was granted these on the understanding that I would finish before the new lab manager came into post. In the event the new lab manager arrived while I was still there. This proved fortuitous for me because he was interested in the ELSI issues surrounding their work. He was subsequently very helpful when I had questions or became stuck writing up the science and he read drafts of the chapters I wrote about the lab, checking them for accuracy, for which I am indebted.

Once I had written the first draft chapters on the lab I sent them to the lab manager and lab technicians for comment. I had used their first names in the text. The lab manager contacted me to say that they had discussed the chapters during one of the weekly meetings and that the lab technicians were uncomfortable seeing their names. They asked that I remove their names. This left me with a dilemma. I had wanted to make them more present and had used their names as a way of doing this, but they did not want to be 'seen'. I complied

with their request, removed their names and substituted the term 'lab tech', with the numbers from their lab coats that hung in a row of hooks outside the office door, to distinguish them. It made them disappear into the text in the same way that they tended to disappear into the machinery when in the lab.

Desktop

I continued the fieldwork by seeking permission to follow the digital data from the lab to the desktop. This presented an unexpected problem. The studies collecting blood samples were not yet doing analytical work because the collection was incomplete. There were complete collections of DNA samples in the freezers from other projects, some were waiting for funding to carry out the genotyping or sequencing, others had not yet applied. There were no projects where I could follow the data I saw being generated in the lab to the desktop. I used every contact I had to find someone, anyone, working on genotypic datasets. It appeared that many studies were collecting blood samples for DNA extraction with some possible future research project in mind. In the end, I found one clinical geneticist working on a genotypic dataset. I was looking for the construction of knowledge, what I found was a new field of analysis in an emergent state.

The clinical geneticist was working intermittently on the genotyped dataset. I was on standby to drop everything and chase over to his office 40 minutes away. He rarely planned when he would work in the office but fitted it in opportunistically around his clinical commitments. I did explore the possibility of desk-space in the building so that I could be at hand if he came in.

Unfortunately there were only occasional hot-desking possibilities and, as it was not possible to identify the days on which he was likely to come in, I could not be specific about my requirements. We did make several appointments most of which he was unable to keep. In the end we managed two short sessions over a period of six weeks, and on both occasions I was aware that he was in a rush to be elsewhere. This was clearly a problem for both of us, so I gleaned what I could and left the man in peace. It was a less than ideal situation, and I did consider switching to an interview. There were two problems with that, first I did not think that an interview would give a sense of the practice and second without observation I was not sure what I would want to ask. It seemed more important to try and capture the practicalities of his work on the dataset.

I also used the timely (for me) arrival of the newly appointed GS statistical geneticist. This is a very small though rapidly expanding field of expertise in Scotland. The technology to create the data is in place but the knowledge and skills to work with it are still scarce. It became apparent that many of the studies collecting DNA would either have to employ someone to do the analytical work or train someone on the project before the data could be used. Many clinical studies routinely include the collection of blood samples for DNA extraction but do not have the expertise to use it, yet. The expertise needed to analyse these data may well come from within, and be driven by GS. Not only is GS in a strong position to do this, but indeed it may well be essential that GS provides this type of training if the databases are to be used to anything like their potential.

Data Manager and Project Manager (21CGH)

The GS data manager was temporarily using desk-space in the office in the lab at the time I was there. This meant that at intervals during my time in the lab I also spent some time with him discussing the early stages in the development of the data collection and management and some of the problems of the technology and programming. We also discussed the separation of the GS data into different databases. Later I was able to check the details of the data collection and the databases with him. The GS project manager was located in a nearby university building within the hospital complex and I also saw him occasionally either at his then temporary office or when he came to the lab. At the time he was working on the MREC application for GS, and developing a network of communications across the GS sites. This placed him in a nodal position within the GS organisation. He was also responsible for the co-ordination of the different aspects of the project. I had discussions with him about GS as an organisation and later he read and commented on my analysis of GS.

The fieldwork was planned and framed by my choices and decisions, but in effect it often worked out as response to opportunity, events and personal contacts. The advantage was a flexibility to respond to the changing circumstances and sites. The disadvantage was that it was difficult to anchor the research in a particular theory. The loose-ended incomplete data from the fieldwork left me worried about how to handle an analysis and writing up. I

was faced with a choice, either work with what I had or try and gather more data. The hit-and-miss experiences with regard to the timing of suitable projects doing the same type of work suggested it could take many more months for a small increase in the data. It was also evident that, as this is a developing field, there were going to be changes. Finally, GS had been changing and there were more developments to come. The data collection had not started. The pragmatic decision was to work with the data I had, to capture a moment in an emerging field of research. The data was messy, but reflected the messiness of developing a large scale population genetic database project. The state of my incomplete and piecemeal data in this situation reflects to some extent the problems that face an anthropologist undertaking this type of research.

Finding the thesis

The face of GS looked concrete, coherent and fixed, it belied the frenetic activity behind the scenes. Being located within the organisation, I had a unique opportunity to look at the inner workings. These inner workings are a complex interaction of science, medicine, technology and their satellite ELSI activities. The inner workings were simultaneously connected (within the GS project) and disconnected (by people, places and practices). In the everyday work of the project many of the connections and disconnections are not evident or questioned by those involved. The disconnections and connections emerged strongly from my fieldwork notes when I started the writing process.

Connections and disconnections are layered throughout the whole GS assemblage. Connections are made through the organization and institutional

collaboration, multidisciplinarity, people and their movement, and by diverse technologies. For example, GS connects scientific expertise, funding, and technology into an organisational infrastructure. Disconnections are made through the organisation, people and their skills, different places, knowledge and concepts, and the technologies. For example, GS disconnects participants from their blood and information through physical removal, anonymisation and information technology.

The disconnections and connections within the GS assemblage can take different forms e.g. physical or conceptual. The physical disconnections and connections occur between people, places, objects and equipment. The conceptual disconnections and connections occur as subjects change, new subjects and objects appear, and as discourses compete both within a multidisciplinary field, and with the public. Information technologies serve at the interface between disparate elements of the assemblage, sometimes used as a barrier to disconnect, at others as a facilitator to connect.

Ultimately, disconnections must be maintained to ensure the privacy of participants. The blood sample and information from other sources are disconnected within GS through practices, processes and the technology of the database. The disconnections produced through technology and anonymisation are not absolute, they work to separate DNA and information, which exists simultaneously in the social world and the scientific domain. The disconnections are constructed by and in this setting, but they are not fixed. Furthermore, these

data can be linked. Indeed, the greatest value of the data lies in the ability to connect it to other data.

Conclusion

It seemed to me that a multisited ethnography that followed a particular object – blood – offered a mechanism that would draw the diverse aspects of the project together into a relationship in a structured way. That blood more than anything was a way of linking the social world and the science, as an entity that has both social and scientific uses and significance.

I had thought that I could write about the blood being transformed into digital data or at least constructed as being cut off from the rest of the project by sheer dint of writing a bounded and separate piece of work. However, it became clear that this was not going to work well as a strategy, for the significance and utility of the digital data obtained from the blood, lay in its connection to other data. Moreover the blood and DNA and other data were connected to and disconnected from individual participants who in turn were connected to their families and a wider Scottish social world. These connections occurred at the start of the process through participation (and engagement) and will emerge again at the end of analysis when results return to the Scottish public through policy decisions, public health actions, and therapeutics. The thesis was never intended to be about the social world of GS, it is about the construction of a population genetic database, about trying to understand the science and technology and the relationship between the data being collected and processed,

and the 'outside' social world it is assumed to represent rather than its own micro social world.

Being inside the project shaped my work and informed the ethnography. Working with people from different disciplines pushed me to consider questions which I might otherwise have missed or ignored. Events caused changes in direction and focus. The combined effects of people and events required me to be flexible in response and forced me to think and rethink what I was doing. These factors also pushed me to attempt a complex thesis that would reflect the complexity of the project, to show how the relationships between diverse interests from both scientific and social perspectives emerged as disconnections and connections that were layered through and interwoven across the construction of the GS population genetic database.

Chapter 3

'A Unique Partnership': Creating Generation Scotland

This chapter is intended to provide some of the background to the construction of the Generation Scotland database, both the people and institutions who want to create this resource, and the scientific reasons, based on genetic research design, why it will be useful. As such it is not intended as an historical account and there are inconsistencies in the time frame due to the approach that I took to the analysis. Most of what I have written below is based on the period from late 2003 to the beginning of 2006, but there are elements and issues that are ongoing, right up to the present.

Generation Scotland is one of a new type of research project, indeed it characterises itself as 'a unique partnership', that is emerging from the possibilities presented by the development of technology, both information technology and scientific technology. It is also, like many health research projects, a response to policy initiatives, in this case from the Scottish Executive, the Department of Health, NHS Scotland, as well as the priorities of funding organisations. Generation Scotland (GS) started off as an umbrella concept for a consortium of researchers from a range of disciplines, including health professionals, across Scotland; it then became a project proposal, but was unsuccessful in its initial bid for funding and now constitutes a corporate image for a multi-institutional collaboration, with a brand name owned by the University of Edinburgh and licensed to other universities across Scotland.

One of the primary aims of Generation Scotland is the identification of genes that contribute to the genetic predisposition for the main diseases that afflict the population of Scotland, including chronic heart disease, cancer and mental ill-health. One of the expectations of this project is that identifying the genes that predispose people to particular diseases, and then linking this data to other sources of phenotypic data and patient records will help in understanding disease aetiology. Research carried out using the GS database will assist policy makers and NHS Scotland in making effective health policy and planning for the future. In addition, it is intended that the genotypic data collection will be useful for research into the effectiveness of drug treatments and the development of new therapeutic drugs. Generation Scotland has three key features: it is a family based project, aiming at gene identification of complex diseases (initially, chronic heart disease, cancer, and mental health), and the development of targeted drug treatments (http://www.generationscotland.org 15.11.06).

Generation Scotland is however something more than the usual medical science research project, it has the appearance of an organisation. For example, GS has a logo, a newsletter and a website at www.generationscotland.org. This chapter examines Generation Scotland as an organisation, i.e. the people and institutions that are creating a population genetic database as a research resource, and who will control its uses in the future.

When looking at Generation Scotland in terms of its organisation, I was unable to find any existing models dealing with this type of collaboration on which I

could base my analysis. My search led me to *Inside Organizations: Anthropologists at Work* (Gellner and Hirsch 2001), where Stella Mascarenhas-Keyes (2001) suggests a check list for a rapid organisational analysis 'in order to understand the institutional and political complexity' (Mascarenhas-Keyes 2001:208) of a working environment. The check list identifies key areas that should be taken into consideration when analysing an organisation and includes: type of organisation; financial matters i.e. funding; management structure; key actors; human resources; communication systems; power and influence in the organisation; intra-organisational dynamics; inter-organisational politics; time frame; and research outcomes. According to Mascarenhas-Keyes, the shape and structure of the GS organisation should have been revealed by a rapid organisational analysis.

What this rapid organisational analysis of GS showed was a degree of flux between levels and topics that made it difficult to locate and fix the shape and structure of GS as an organisation; instead, it revealed unfixity, movement, networks, lack of clear coherence, new connections and disconnections. These are characteristics not of an organisation, but rather of an assemblage (Rabinow 2003, Collier and Ong 2005). While an assemblage may have features of 'an organisation', it lacks the continuous perceptible structure of the latter. Moreover, my analysis showed that Generation Scotland was an assemblage that has created a virtual space, into which millions of 'bits' of data from different sources can be poured, and which has the potential to access and link data from other databases. Networks and systems were being developed through a combination of people, places, hardware and software that will

contain data and be used to manage it. I started my organisational analysis by looking at the 'type of organisation'. This is followed by a description of genetic research design which contributes to an understanding of why it is deemed necessary to collect large numbers of blood samples from the population and from families.

Type of Organisation

Generation Scotland is a particular type of organisation: centrally, it is an academic collaboration, funded by government grants, with present involvement from NHS Scotland and Scottish Enterprise, and conceived with an intention of engaging with the private sector through commercialisation. It is comprised of a complex network of relationships constituted through institutions, people, technology, and equipment. There are diverse objectives, some shared, others divergent, regarding genetic research and public health. Describing Generation Scotland is not straightforward because it is founded on a multifaceted approach to the creation of a population genetic database for research, so that concepts, people, and places can often be conflated or linked together, sometimes in contradictory or unexpected ways. The Generation Scotland corporate image, as presented on the website, creates an appearance of coherence that belies the complexity and changes at its core. To talk about Generation Scotland is to talk about genetics, medicine, science, technology, health, disease, blood samples, health records, phenotypes, lifestyles, DNA, genes, participants, patients, families, public health, ethics, law, social issues, and drug therapies, as well as medical schools, research institutes, university departments, technology transfer, commercialisation and drug companies, some

of it in the present, much of it in the future - all in one breath. This gave rise to lists as a way of trying to encapsulate the enterprise such as the one published on the Generation Scotland website:

> Generation Scotland is a multi-institution, cross-disciplinary collaboration between all 5 Scottish University Medical Schools (Edinburgh, Glasgow, Aberdeen, Dundee and St. Andrews); the MRC Human Genetics Unit, Edinburgh; the MRC Social and Public Health Science Unit, Glasgow; the Wellcome Trust Clinical Research Facility, Edinburgh; the National e-Science Centre, Edinburgh and Glasgow; the Scottish School of Primary Care, representing all of the academic Departments of Primary Care in Scotland; and the Information Services Division within the NHS National Services in Scotland, which has both a public health planning and academic research function.
>
> Generation Scotland website 14.04.06

However, lists were an unsatisfactory way of trying to characterise the organisation since they never contained the same elements twice, nor did they show the connections between different elements. The lists changed as people, departments and organisations joined or contributed to the collaboration. The lists were also a result of who drew them up, from which perspective, and on the basis of their knowledge of other members of the collaboration. An alternative representation was also used on the website and on presentation slides, comprised of the logos of participating institutions.

There were often inconsistencies between the information on presentation slides and the website, for example, the absence of the University of St Andrews from the slide on the next page.

Figure 3.1. Generation Scotland Partnership: PIs Presentation Slide October 2005

The slide came from a set for a presentation given by two of the Principal Investigators. Lists and collections of logos reflected the speed with which particulars of GS changed, the fragmented information that individuals within GS had at a given time, and how GS was presented could even be determined by the topic and target audience. Both lists and logos were authoritative in the way they emphasised the extent of the collaboration. Indeed, there were 17 logos displayed collectively on the links page of the website as partner organisations in the Generation Scotland collaboration.

There are different types of institutions and organisations within the collaboration that comprises Generation Scotland, many of which have

similarities, particularly across the academic institutions, but there are a number of other organisations which are distinctively different in their principal function. I do not propose to venture an analysis of each organisation, but acknowledge that there may be distinctive similarities and differences which will influence the way these organisations engage in the collaboration and the relationships between them. GS is more closely integrated with health and medical research than other academic research or departments, and the extent to which agenda and aims may be connected varies. For example, the guiding principles of the Scottish School of Primary Care (SSPC), which is closely involved in the SFHS, coincide with much of the GS agenda in terms of developing health research. The SSPC is funded by the Scottish Executive Health Department through NHS Education for Scotland.

The main aims of SSPC are to:

> 1. Provide high quality research evidence needed to inform decisions made by patients, practitioners, managers and policy makers.
> 2. Increase research capacity and capability with Scotland.
> 3. Increase the relevance and use of research-based evidence in health policy and practice.
>
> These translate into three main areas of activity:
> - the development of *programmes of research*;
> - the development and implementation of a strategy to *increase capacity and capability*; and
> - a programme to develop a more synergistic relationship between *research and service or policy development.*
>
> Scottish School of Primary Care website 15.11.06

By contrast, the aims of, for example the AHRC Centre which is part of the University of Edinburgh's School of Law, are wider ranging than health research:

> The Centre's research themes examine the synergies between intellectual property law and information technology law together with work on medical law and ethics.
> Our research is about the synergetic relationship between law, technology, commerce and society in the widest possible sense. As well as IT and IP, we and our associates are concerned with the adjunct areas of biotechnology, genetics and medical jurisprudence and ethics; law and artificial intelligence, including the distribution of legal knowledge via the Web; regulation of electronic commerce, the Internet and the information society; and law as it affects information management and cultural production and archiving.
>
> AHRC Centre website 15.11.06

These examples of just two of the collaborating institutes show how there would not normally be any connection between them, but that they have become connected through GS.

Funding

The GS project is supported by the Scottish Executive, the Department of Health, the Chief Scientist's Office, the Scottish Higher Education Funding Council, and the NHS. Generation Scotland has been shaped by two streams of funding, Scottish Higher Education Funding Council (SHEFC) and Genetics Health Initiative (GHI) which in effect produced a split in how the collaboration is organised.

Genetic Health in the 21st Century (21CGH)

The proposals for both Generation Scotland and the UKBiobank Scottish Spoke called for investment in resources and personnel across a wide range of disciplines and expertise. The 21CGH funding by the Scottish Higher Education Funding Council in October 2003 was the first step in making the proposals

concrete through developing an infrastructure for record linkage, data management and knowledge transfer platforms, public engagement, work on legal and ethical issues, and the collection of samples and data. Whilst the research design and agenda for Generation Scotland had previously been worked out, the SHEFC funding ensured the proposal would be further implemented by putting people into posts.

The objectives, for example, of the Ethical, Legal and Social Implications (ELSI) team were not to look at questions of 'if' but 'how' the Generation Scotland proposal could be implemented in a socially acceptable, legally and ethically robust way. There was no ethicist on the ELSI team, ethical issues are addressed by both the lawyers and the social scientists involved in GS. The legal issues that were being researched with regard to genetic databases were concerned with the production and protection of knowledge. The social perspective was focused on public engagement, the issues raised by public(s), ways of addressing these issues, and the issues surrounding family-based genetic research. These concerns impact on the practicalities of recruitment of participants, consent forms, information leaflets, data collection, handling and storage, confidentiality, privacy, data protection, governance, and benefit sharing. It is important to the success of Generation Scotland that information about the project is accessible so that 1) participants understand what will be done with their samples and data; 2) the rights of participants and their families are protected; 3) the rights of the researchers are protected; 4) data are anonymous and stored in a secure way; 5) data can be shared between other research projects; and 6) discoveries can be exploited for the public good and for profit.

At the same time the technical infrastructure was being developed with scientific, medical and information technology members of GS working on strategies for data collection, sample storage, sample management, IT design was being developed, equipment tried out and evaluated, and the design of the database and data management were under discussion.

Genetic Health Initiative (GHI)

The Genetics Health Initiative was announced in 2003 as part of the research development strategy for genetic research of the Scottish Executive and the Department of Health through the Chief Scientists Office (CSO), which allocates funding resources. The Generation Scotland proposal had been discussed and broad agreement had been reached across the institutions and funding bodies, support for a funding application to GHI appeared to be present. A joint application was written by the principal investigators of the consortium. The proposal identified ten fields of research, in which a genetic component would add to the value of current research. The ten fields included those that are considered to have priority in Scotland (chronic heart disease, cancer and mental health) and also encompassed the areas of specialisation of the consortium institutions. The CSO convened a special committee to review the funding application. The application was rejected. At the next GS (Edinburgh) meeting it was clear that this decision had caused some dismay and surprise as previous informal discussions between funders and principal investigators had suggested that the application would meet with approval. Following the rejection of the GS application, the CSO issued a set of guidelines for

applications for GHI funding which emphasised a focus on project specific applications. This was a shift away from the broader concept of GS and the control of a budget across the consortium. In effect, this meant that the CSO retained control over research agenda and priorities through the funding, rather than allowing GS to determine those within the consortium, despite the fact that the GS application closely reflected the health priorities stated by the CSO at that time. During the weeks following the rejection of the GS application there was some reflection and regrouping within the consortium. This led to a second application for funding of a project called the Scottish Family Health Study (SFHS), led by Dundee and Glasgow in conjunction with the Scottish School of Primary Care (SSPC), which was successful. Subsequent discussions led to the licensing of the Generation Scotland brand to the other universities in the consortium in order to maintain the corporate image. The two projects that had emerged from the separate streams of funding were brought together under the GS umbrella as GS: 21CGH and GS: SFHS. However, the funding continues to be managed separately, through participating academic institutions.

Management Structure

The GS umbrella functions through a committee, called by some the Scientific Committee (SC) and others the Scientific Management Committee (SMC), both referring to the same body. Comprised of people not necessarily known to everyone across the different projects and institutions, its function is to consolidate disparate activities into a cohesive corporate type organisation that maintains and co-ordinates the overview. Both management and governance are

the responsibility of the SC/SMC, which is officially comprised of approximately 50 members with voting rights.

> The GS SC oversees the governance of GS and consists of all PIs [Principal Investigators] on SFHS and 21CGH and other co-applicants and experts, we are also looking to extend this to include public interest group(s) and lay representatives. 21CGH and SFHS are projects overseen by this. 21CGH has a separate Management Group composed of some of the PIs from the GS SC, management of 21CGH is discussed at GS SC meetings, with other communication/telecom/meetings as necessary, increasingly this is regarded as part of the GS portfolio of projects rather than a separate entity, with the 21CGH infrastructure funding supporting GS as a whole. SFHS is co-ordinated by the GS Implementation Group
>
> E-mail from the 21CGH Project Manager 10.04.06

There is also a Future Directions Committee which has representatives from all institutions in the collaboration together with representatives of the projects and other areas of expertise.

However, while the authority of the SC/SMC is unquestioned, the number of people actively involved in the work of any of these committees appears variable, mechanisms for joining (or leaving) these groups are unclear, their responsibilities are assumed rather than stated, and lack of conflicting interests is taken on trust. Originally, the SC/SMC was comprised of four *active* Principal Investigators, one from each of the universities of Edinburgh, Aberdeen, Glasgow and Dundee; this number has gradually increased to approximately fourteen, also including specialist clinicians and information technology and e-science experts.

There was a larger group of researchers, clinicians and scientists named on the funding application. They appeared to be people who had an 'interest' in GS and gave their support but there was no evidence of direct involvement. The group was comprised of approximately fifty members, leading academics and clinicians in their fields. It was not clear if they were considered as members of the SC, but if they were then they should, in theory, have had voting rights.

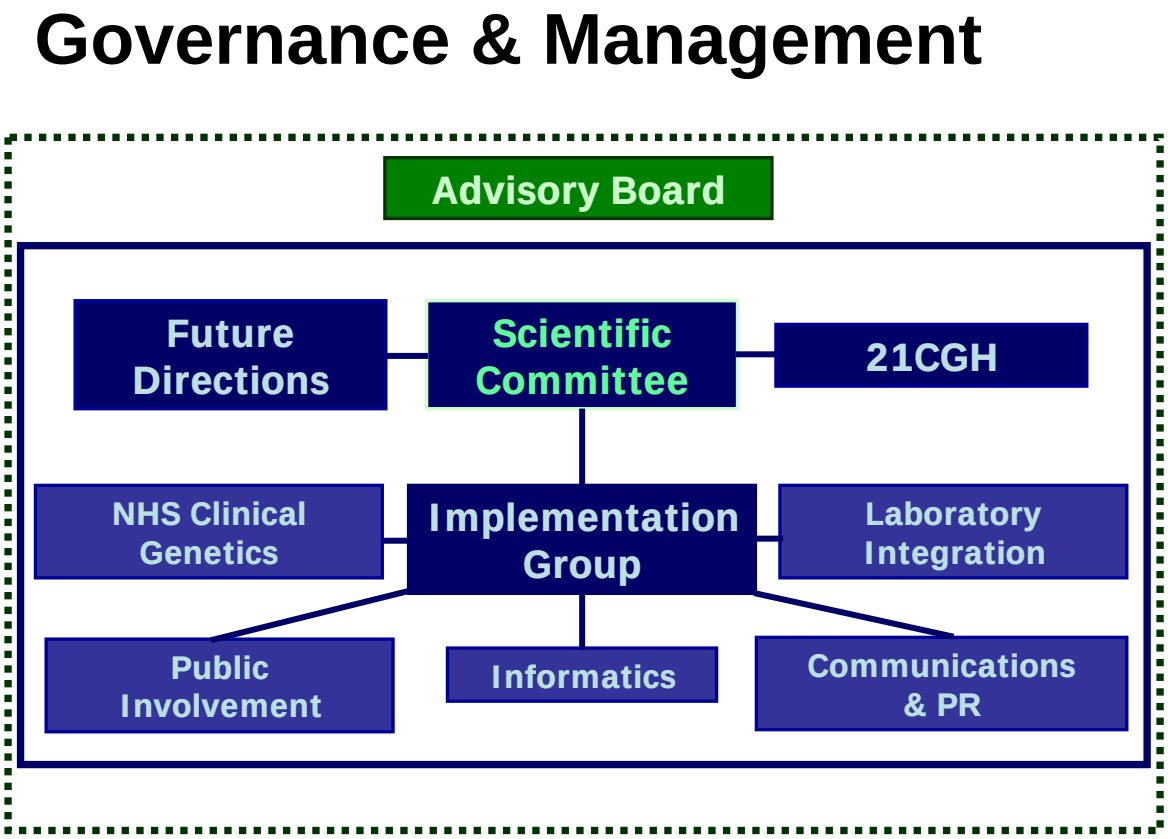


Figure 3.2 Governance and Management: PIs Presentation Slide March 2006

At the insistence of the Scottish Executive, an Advisory Board has been created, which includes both expert and lay representatives, and is appointed by public process.

> The Generation Scotland Advisory Board (GSAB) does not have executive powers, its role is to comment on and provide advice for Scottish Ministers and the GS Scientific team, on the GS project and its implications for the Scottish community. Although the GSAB will have an advisory role on ethical issues as in other matters of public interest, it will not assume a regulatory function as this is the responsibility of other bodies
>
> E-mail from the 21CGH Project Manager 10.04.06

The graphical representation of Figure 3.2 shows the structure and areas of management, but does not reveal where or by whom management and governance is carried out. It does show hierarchical relationships between various groups and activities.

Key Individuals

The 'key individuals' are regarded as those with responsibility for the execution of the management and governance. The central individuals here are the principal investigators, though interestingly, they are not all named on the graphical representation of Figure 3.3 on the next page. The slide also shows a line connecting the Scientific Management Committee to the Advisory Board, for ethical governance these two should be disconnected to avoid a conflict of interests. The title of the slide suggests a connection between all these individuals and governance. Management and governance are often conflated into a dual responsibility. Later slides show a disconnection. Changes in representation reflect the changes that have occurred as GS developed, and bodies such as the Advisory Board became established.

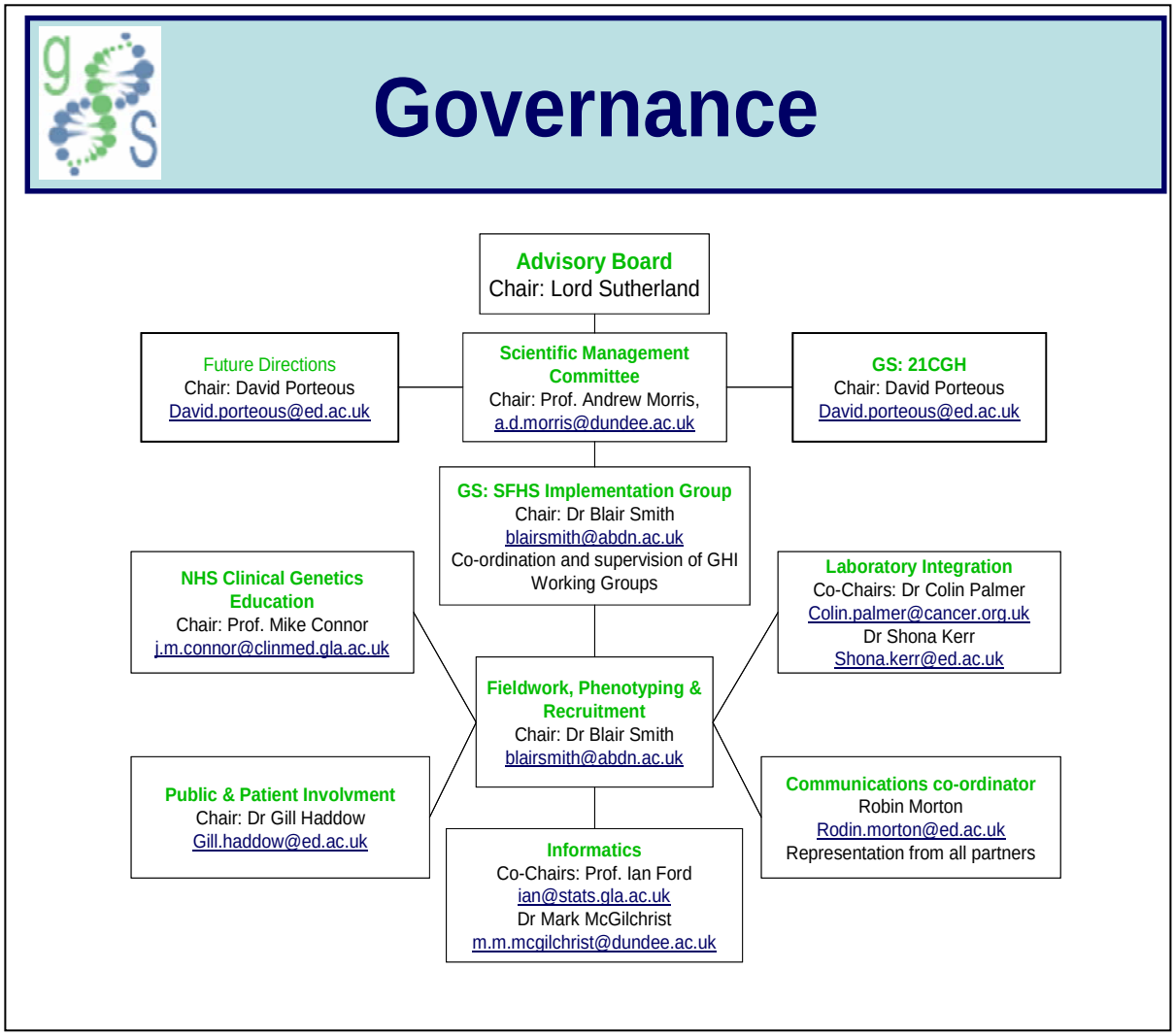


Figure 3.3. Key Individuals: PIs Presentation Slide October 2005

There is a principal investigator from each of four universities, previously shown in Figure 3.1 as the founder members and Scientific Management Committee (St Andrews not included), who represent their universities' interests and contribution to the collaboration, even if they are not included on this graphic. Others have 'joined the team' of principal investigators as GS has been developed, bringing with them areas of expertise, in particular informatics

and technology, which are crucial to the success of both, individual projects within GS, as well as the GS collaboration itself. Over time, as GS has taken a more concrete form, the principal investigators have become less visible and institutional involvement more so.

Not all of the people represented in the figure are decision-makers, nor do they all control the allocation of funding across the collaboration, nor indeed do they have a say in its governance. The management and governance of GS is dependent on a small group of individuals who serve to drive the collaboration forward. The work of these individuals is based on networks of personal relationships and these relationships, in the relatively small research community of Scotland, have been essential to the development of the organisation. Some of those included in the graphic are in fact responsible for the day to day work of GS, no less key individuals in terms of what they do, but without the power of decision-makers.

Human Resources

Generation Scotland as a Scotland-wide research project includes, as previously stated, medical schools, research institutions and other academic departments around the country producing a multidisciplinary collaboration of health professionals, information technologists, geneticists, sociologists, health geographers, statisticians, lawyers and an anthropologist.

Generation Scotland does not employ anyone *per se*, that is, there are people working *on* GS, but not *for* GS. During 2004-5 a steadily increasing number of

people came to be working on Generation Scotland, these included research fellows (4), research nurses (4) and managers (3) in the organisation funded by either 21CGH or SFHS grants. The number of people working on GS has continued to increase through 2006, particularly in the development and deployment of information technology. The monies were allocated through various university departments or research institutes to pay the ELSI research fellows, the project manager, data manager, research nurses and genetic statisticians. Those involved in the management and governance of GS may receive remuneration to cover time that they work on the projects, but they will also have commitments elsewhere. Most of the people who act in an advisory or management capacity will not have any monetary gain from doing so. Parallels can be drawn to the history of 'Scottish political elites' who traditionally work altruistically for the improvement of the Scottish nation and community.

At this point, it should be noted that prior to the 21CGH and SFHS appointments, all collaborative work had been done by people, who already had commitments to other work. It is characteristic of medical research projects that most people have multiple roles that include clinical, research and teaching commitments; to these are often added project management, planning, development, research design, writing proposals and reports, as well as papers for publication. Thus the 21CGH funding was essential in making it possible to put in the ground work of developing the project.

The integration of an ELSI team was unusual, most ELSI work on genetic research is done outside of the scientific project, funded separately, and located within social science research institutes or centres such as INNOGEN or the Oxford Centre for Ethics and Communication in Health Care Practice (Ethox). In this sense, the control of the social science research agenda and findings would usually be independent of the scientific/clinical project, and the implications of independent research not necessarily of immediate consideration or consequence. Integrating the ELSI team within the scientific/clinical project, required the research to be relevant and immediate to the project. The integration of the ELSI team with the scientific and technology aspects of the project was however somewhat offset by locating the ELSI research fellows in separate research institutes, at a remove from those working on the science and technology; generally, the separate individuals only came into contact at meetings or if seeking a particular piece of information from one another. The effect of integrating the ELSI team was unclear, a small scale study was being conducted to evaluate the effect of public engagement findings on decision-making.

Around the time the ELSI posts were taken up, genetic statisticians were appointed in Edinburgh and Aberdeen, a Data Manager was employed to develop the information technology for the project, and some time later, a Project Manager was appointed to introduce a more co-ordinated and productive approach to the development of the project. A Project Manager was central to moving the project along by maintaining networks, transfer of information and collaboration, which had previously fallen to the principal

investigators. Through 2004 and 2005, all those involved with GS constantly had to react and adapt to frequently changing proposals, plans and organisation.

Communication Systems

Communication within GS, between its various institutions, centres and departments, used a variety of mechanisms since people were dispersed across many locations. Meetings, both local and across the collaboration, formed an important part of the communication, but were often difficult to arrange so that everyone could attend, as people had other commitments. E-mail was essential to the everyday communication between all those involved in GS. Likewise, telephone communications, including tele-conferencing, also took place regularly. Within GS, 21CGH and SFHS had their own internal communication systems. GS also published a general newsletter which was circulated to collaboration and consortium members two or three times a year.

Communication beyond the organisation had mainly fallen to two of the Principal Investigators. They constituted the most public faces of Generation Scotland, both across Scotland and within international clinical and scientific research networks, as they attended conferences and were invited to speak by universities in other countries. Internationally recognised for their research work and with many contacts globally, these Principal Investigators were at the forefront of publicising and promoting Generation Scotland. Also, two members of the GHI consortium active in the creation and launch of the Scottish Family Health Study were the public/media face in Scotland. Members of the ELSI team contributed to the dissemination of information about Generation Scotland in

other fora, by attending conferences and taking part in the ethical, legal and social science discourses on population genetic databases.

Generation Scotland also created a website for communication with the public and the wider research community. The website has been redesigned twice as the shape of GS changed, and the current site was launched in February 2006, to coincide with the launch of the Scottish Family Health Study. On the website, Generation Scotland is represented through institutional and organisational logos rather than individuals. It appears that the organisation is becoming more and more anonymous, certainly more than it was a couple of years ago. Thus the authority and integrity of the project, which had initially been invested in individuals, has shifted to various institutions with a vested interest in the collaboration. The public of Scotland, who are potential participants, and researchers who may wish to use the genetic database can no longer see the individuals that drive this data collection, or who will manage and use it.

The launch of the Scottish Family Health Study was marked by a concerted effort to gain media coverage. Press releases and interviews were published in both national and local newspapers, the SFHS was featured on Scottish national TV stations as well as national and local radio programmes. The public response was several hundred volunteers interested to participate across Scotland. The TV coverage, a three minute local news item, confused GS, the SFHS and UKBiobank by mentioning them all, but not explaining the distinctions. Whilst in the newspapers SFHS eclipsed GS. The media coverage did not generate any public debate.

Power and Influence

Power and influence within the organisation were based on the same principles that underly credibility in other science and medical research: reputation, experience and a proven track record in a specialist field (Latour and Woolgar 1986). This power was held by the Principal Investigators who made the decisions, controled the funding by holding and administering the grants, and set the agenda for the future. Others brought different specialist knowledge and expertise, such as informatics, to the project that had the power to shape decisions. However, as the organisation grew, another type of power was emerging. Based on a more general knowledge, a small number of individuals with an overview and knowledge of what was happening where, influenced the co-ordination of the collaboration, and were instrumental in controlling knowledge sharing.

The scale and range of Generation Scotland means that some individuals fall into both of the above categories, and these will have most power within the organisation. However, within the collaboration, individuals act to disperse any concentration of power in one place or person. GS has been fragmented into a number of projects centred around different institutions and led by different key individuals. Thus, control of the data collection is continuously contested and renegotiated by the various collaborators. Power and influence are inherent in control of the samples and data. All blood samples for DNA extraction were going to be sent to a single genetics lab for processing and storage. Whilst the blood samples for clinical tests will be sent to four different, regionally located, labs for processing and storage. At the time of writing, Autumn 2006, a

researcher from the e-Science Centre (Edinburgh) has been employed to undertake a review of the practices and processes of each of those four labs to ensure the standardisation of sample processing and quality control, which is of vital importance if the samples and data are to be used collectively in future research projects.

Dispersal of the blood and DNA sample collection can be overcome by use of technology, which will facilitate the connection and sharing of the data, but it also means no one institution will hold or control all of it to the advantage or disadvantage of others. These disconnections create more complicated questions about access, management and use of the collection, but also ensure that there can be no unilateral decisions. This gives rise to a more complex set of legal and ethical issues since there is a shift from a single database and sample collection to several databases and sample collections.

By contrast, the collation of the questionnaires for collecting phenotypic data has raised different problems, as the aim was to bring a range of interests and questions into one document. While many studies have a common interest in some questions, for disease specific studies there are often particular questions that provide information important to that disease that may be irrelevant to another. The initial draft life style questionnaires attempted to integrate all the possible questions that specialist researchers wished to ask, which made for a very long questionnaires; the concern was that such time-consuming and possibly tedious questionnaires would put potential participants off before the blood samples and physical measurements could be collected, or would return

incomplete data, either of which would jeopardise the dataset. Consequently, the length of the questionnaire was reduced and decisions had to be made as to which phenotypic data should be collected and which omitted. I return to the questionnaire in chapter seven.

As the collection of samples grows and the research into gene identification becomes feasible, decisions about which genes should be genotyped for complex diseases could be challenged. The 21CGH body for decision making and resolving contested research issues is the Scientific Management Committee.

Intra-organisational Dynamics and Inter-organisational Politics

The lines between intra-organisational dynamics and inter-organisational politics are blurred by the collaboration that constitutes GS. It is not possible to draw a line around GS and define its parameters, in this sense, the dynamics of how the organisation works in practice are all intra-organisational. However, these often take place within, and are influenced by inter-organisational politics and the way in which the various constituent members of the collaboration relate to each other. The GS collaboration appears as a highly complex set of relationships constituted through people, places, technology and equipment. However, health research has a history of collaborative research and development that links the expertise of people and departments from a range of disciplines, i.e. most people in this field are used to working in this way and therefore regard it as either unproblematic, or the problems are known and there are mechanisms to deal with them. Nevertheless, different categories of research are geographically or spatially separated.

The appointment of individuals to particular positions was planned to integrate and embed aspects of the project within particular departments and disciplines, across departments and facilities. This led to the GS *team* being dispersed and members often isolated from their co-workers on the project. As a strategy it has its strengths and weaknesses. The strengths lie in the integration and embedding of shared resources, the weaknesses lie in the isolation of the individual members of the team, which means that they are often unaware of what other people are working on, and carry out their work without the mutual support of a shared agenda or the exchange of ideas. Support comes from those who most closely surround an individual at work, and all of us found that the location within a particular department or institute carried with it obligations of participation in the activities of that place in addition to the obligations to the project. Whilst these were not necessarily onerous, they serve to underline the split between the project and the place. This lack of centrality, of belonging to a place specific to the project, will extend as the project progresses, as more people across Scotland work on projects that come under the Generation Scotland umbrella. This means that particular *known people* become more pivotal in creating a sense of integration, cohesion and continuity that connects the network together. Another way through which people are connected is through shared objects, be that equipment, or - as is the focus of this thesis - blood samples. The need to collect and process blood samples brings diverse people into a relationship as it moves 'from arm to desktop'.

However, as some people and objects act to connect, others act to disconnect. For example, the new application for GHI funding for the SFHS took a two-

pronged approach: first, there was the blood sample and data collection; and second, this bid proposed to integrate public engagement into the project. Thus, as previously mentioned in chapter two, the sociologist working on public engagement in Edinburgh was invited to work on the proposal and run the public engagement. The work that was being developed in Edinburgh was transferred to the new bid and linked with the application led by Dundee and Glasgow. Within the remit of collaboration this was a logical step to take. However, there was some concern that the geographical focus of the work was moving and with that control of this work. Although fitting the idea of collaboration well, and in theory creating a more integrated overall project through sharing knowledge and expertise, some friction was undoubtedly caused, particularly as knowledge and information began to circulate through extended and at times different groups of people. In a collaboration the loyalty usually expected and given to an employing institution is stretched, and it is often difficult to know where it should lie - with the employing institution or the group of people you may be working with. Working between groups of people within a collaboration can be difficult, and although an assumption of shared knowledge and working within a project should be unproblematic, in reality individuals can find themselves in awkward situations and having to make decisions about what knowledge they can or are willing to share without invoking the disapproval of colleagues elsewhere. During my involvement with GS, there were occasions when I asked for information from different people in different places, only to be told that I could not have it because it was confidential. The first time this happened I was surprised, and I still hold the assumption that if we are all working together, sharing information should be

unproblematic, indeed, that it is desirable. In fact, within the collaboration there were occasionally signs of a tendency toward territoriality, some people were more inclined to collaborate than others.

Time frame

There are three time frames within GS, characterised by short term, medium term and long term goals. The first, short term goals are orientated toward getting the collection of samples and data together as rapidly as possible, and completing the work on public engagement. The second set of goals is the utilisation of the population genetic database for gene identification in the medium term. Finally, the long term goals are orientated toward commercialisation and drug development, together with, of course, the long term management of the database and its use. These are not entirely sequentially separate steps, there will be some overlapping, but generally speaking, this is how it is envisaged. The lack of specificity regarding the length of time that the samples and data will be kept and used is one of the ethical and legal issues that surround patient/participant consent.

Research Outcomes

As an organisation, GS could currently be described as organic, in that it is developing, changing form, growing and is permeable to both institutions and individuals. Different people and institutions bring new ideas and priorities, alternative perspectives on current activities and ways of doing things; new possibilities are constantly expanding and refining the development of the collaboration and the potential collection of data and applications of the genetic

database (phenotypic and genotypic data), and the technologies, both biological and informational.

Consequently, research projects are developing and shape the collection of blood samples and information from participants; and existing projects are being utilised to collect blood samples and information for the genetic database. The Scottish Family Health Study is undertaking the collection of 50,000 blood samples and phenotypic data across Scotland, which will form the core project for the creation of another genetic database. 21CGH has just started collecting 2,500 blood samples and data, which will form a baseline dataset which can be used as a comparator for future research projects. These 2,500 blood samples are being collected in groups of 500 from different centres, each centre using the infrastructure of an already existing study to tag on an additional 500 participants, whose blood sample and data will come to 21CGH. An application has also been made to the Scottish National Blood Transfusion Service (SNBTS) to provide 6,000 blood samples with some phenotypic data. The SNBTS has a Sample Governance Committee, and ethical approval is required for blood samples to be supplied to research projects. The new Donor Information Leaflet now makes it explicit that some blood will be used for research.

Multiple research outcomes are anticipated from these collections: the creation of a population genetic database as a research resource; a baseline resource to be available as a comparator for a range of future studies; the scientific and medical research outcomes in terms of genotyping, gene identification and analysis; the public engagement findings; the legal and ethical framework for governance;

the commercial exploitation for drug development; and a mechanism for benefit-sharing. The scientific and medical outcomes will be dependent on the design of genetic research projects.

Research Design

Research design is important because it predetermines what data will be useful to collect and how it can most effectively be used. The database needs to hold sufficient numbers of samples for genotyping and a wide enough range of phenotypic data to enable researchers to design research that will be robust and statistically significant. A database is designed and constructed from previous knowledge and experience, and anticipates the application of that knowledge into new or extended areas of research. The Generation Scotland database is aimed at drawing together the data required to meet that anticipation. The design of the GS database is based on established research study design that has proved effective in previous studies.

There are different ways to approach designs in genetic research. Where there is reason to believe that a gene may contribute to a trait, for example an illness, an initial design based on linkage may be used. Linkage design can be used to scan a whole genome using markers to indicate segments of the DNA. Linkage refers to two or more genes on the same chromosome, and alleles of genes on the same chromosome are more likely to be inherited together, thus serving as a marker. Many markers are already known and, by genotyping pedigrees across a number of markers, it may be possible to identify areas that suggest linkage to a particular trait. These areas can then be looked at more closely. However,

linkage studies can be inexact as there may be more than one gene in the segment marked, and it is not possible to tell which of these is associated with the trait. Scanning a whole genome is becoming less of a requirement since so many segments of the genome have now been sequenced with identified genes, that once a linkage looks positive it may be possible to use existing data to find the genes close to a marker and test those genes directly. Known as Association Design, this allows researchers to focus more closely on particular segments of DNA.

Carey (2003:93) comments on the phrase 'identifying the genes for complex diseases'; this sounds like there is a single gene for each complex disease and obscures the fact that there are multiple genes implicated in complex diseases. In the onset of a complex disease like cancer, there could be as many as one hundred different genes implicated. What makes such a disease complex? Firstly, the term cancer can refer to a range of symptoms which can affect different parts of the body and in different ways. Secondly, complex diseases have many factors that contribute to the onset of the disease; they also have many and variable symptoms, for example, a cancer tumour is rated on a scale of aggressiveness, which can vary between individuals although they have the same diagnosis; the progression of the disease may vary, with different responses to the same treatment and different outcomes for individual patients; there could be a single (possibly toxic) or multiple environmental factors (the usual suspects – diet, alcohol, smoking, exercise) which contribute to the onset of disease; and there are many possible interactions between genes and the environment. Talk about complex genetics pertain where it is thought that more

than one gene and possibly anything up to one hundred different genes are implicated in the onset of a complex disease. It would therefore be more accurate to talk about 'identifying multiple genes for complex diseases', which is probably self-evident to those working with genetics, but not necessarily to those of us who do not.

Association studies test the genes identified as having a probability of contributing to a trait or disorder. 'Association is a statistical concept in its genetic sense, it looks for association between allele frequencies and disease status' (E-mail from Statistical Geneticist 14.09.06). Association can be used to isolate a gene following on from a linkage study or where there is reason to suspect a particular gene of contributing to a trait, a candidate gene. There are two main types of association study, one is population–based, the other family based.

Population–based studies require a number of cases with a diagnosed disease and control samples randomly selected from the population. All samples are genotyped on the gene of interest, thus identifying all alleles of the gene, which can then be categorised. Statistical tests can then be applied to calculate the Odds Ratios of a particular allele, and the level of association between the genotype and the trait, or phenotype. For example, the APoE gene for Alzheimers disease has four alleles, which contribute in varying degrees to the possible risk of the onset of the disease, with APoE 4 carrying a much higher risk of onset of disease than the other three alleles. Population-based association studies are relatively straightforward to design and carry out, and are used to

calculate increased risk of disease, a concept I discuss further in chapter nine. However, there is 'a problem with population-based studies, allele frequencies are not evenly distributed across human populations' (Carey 2003:184), so the samples for the disease group and the control group need to be matched to reduce the possibility of errors in the statistical analysis. A control group is a randomly selected group from the population which is used to compare the results of the disease group with the rest of the population. This explains the particular interest of researchers in homogenous populations, such as Iceland or Scotland, for large scale studies.

'Family based studies are considered the gold standard' (Carey 2003:185) because they avoid the problem of association studies, namely population stratification, the unknown distribution of alleles across a population. Carey (2003) describes two research designs used in family studies. One aims to include affected and unaffected relatives, preferably siblings, where the unaffected genetic relatives are used as controls, which removes the need to randomly select a control group. Siblings create the closest match for comparison and 'by analysing the relationship between the gene and the phenotypes of the affected and unaffected it is possible to statistically test for an association between the gene and the phenotype.' (Carey 2003:186). The other design looks for transmission disequilibrium. This is based on pedigrees, parents and offspring, and tests the transmission of alleles from parents to offspring to see if one allele is associated with an increased risk of an affected phenotype.

Both research designs for family-based studies use association because it has a number of advantages. Firstly, there is an existing knowledge of heredity and genes to build on and it is easier to find a particular gene through association than linkage. Secondly, family based studies can use much smaller numbers of samples to achieve statistical significance. And thirdly, and this is important to the GS agenda to identify genes associated with complex diseases which have multi-factorial traits, association studies can detect genes with small effects.

'Given their versatility, population–based family studies could become a principal framework in epidemiology, and move genetics from its traditional focus on high-risk families to give a wider clinical and population health relevance' (Hopper, Bishop and Easton 2005:1397). Hopper et al discuss three types of family based study, Case-control families, Case-families with or without population controls, and Case-control-families. 'Family studies built around population-based sampling of both cases and controls can in some circumstances be more powerful and robust than a case-control or multiple-case family-based approach alone.' (Hopper et al 2003:1397). They differentiate the types of study according to the degree of participation by family members: Case-control family studies 'include data about disease status and other characteristics of relatives of cases and controls but the information is obtained only from interviews of cases and controls' (Hopper et al 2003:1397). Thus information may be unreliable. Case-family studied include cases and their relatives, data is collected directly from all those who participate, making it possible to compare cases and their unaffected relatives - 'Case-family studies allow all comparisons achievable within traditional population-based case-

control studies – namely genetic effects alone, effects of environmental exposures alone, and gene-environment interactions' (Hopper et al 2003:1398). Case-control-families 'subsumes the two previous ones' by expanding recruitment to include cases and their relatives, and controls and their relatives, this allows comparison between sets of relatives, and strengthens the data on family histories through cross-validation, with the advantage that 'Population based case-control-family studies are likely to give clearer answers than analyses of members of mutation-carrying families ascertained through opportunistic sampling from genetics clinics' (Hopper et al 2003:1400).

A great many genes have already been identified out of the estimated 30,000 or 50,000 candidates, and as more and more genes are identified the need for linkage studies is likely to decrease and the main focus will shift to association studies. Indeed, 'With the costs of SNP genotyping dropping rapidly, it is now becoming feasible to perform 'whole genome association studies', looking at thousands of markers across the whole genome, and not just in candidate regions.' (E-mail from Statistical Geneticist 14.09.06). For the foreseeable future, GS anticipates the continuing use of both research designs, so datasets are being collected to facilitate both population-based and family-based studies.

Finally, with regard to research design and databases, the members of the Cartagene programme in Canada have concluded that in order to promote the design of good epidemiological genetic research, given the probability that many of the researchers seeking to access data from their particular database will not be familiar with it, they will offer support to researchers with regard to

finding their way around the available data and assistance with study design. (Cartagene meeting, INNOGEN 24/03/06). Likewise, GS has recently started an education programme which addresses these issues and will provide training to statistical geneticists.

Conclusion

Generation Scotland emerges from the analysis not as an organisation but as a complex assemblage of institutions, places and people linked through technology, layered with connections and disconnections. The analysis also reveals GS to be a virtual representation of this assemblage with an unlimited virtual capacity to absorb digital data. The collection of the data is carried out through a range of specific projects, GS will therefore comprise not one single database, but rather many that are linked through networks and information technology platforms.

Information technology has made it possible to do three things: first, by creating a website it enables a corporate image of a disparate collaboration to be pulled together and be presented as a cohesive unified organisation; second, it facilitates a network of communications; and third, it creates information platforms, systems and ever larger databases, and links these together. Thus vast quantities of data can be utilised as a research resource; as the data are electronic, they can be stored, archived, shared and transmitted across organisational, national and international boundaries. In health research, the development of informatics allows for the combination and statistical analysis of data from a wide range of sources including phenotypic data from

questionnaires, patient records and prescription records, with data from genotyping. The development of scientific technology, in particular that associated with genetic research, has made it possible for blood samples to be collected on a large scale (many thousands), for DNA to be extracted, and to genotype DNA for research into diseases and for epidemiological purposes.

There are advantages and disadvantages in any research design. Broadly speaking, 'Linkage studies can detect effects in a segment of DNA where the genes are not known' (Carey 2003:189), this is useful to narrow down the possible segments of DNA, or the number of genes that might be of potential interest. Thus positive linkage can find new candidate genes or be used to reduce the number of candidate genes for analysis. Association studies are used to test association where the genes are already known. In effect, each design complements the other, linkage is used to find the gene and then association studies can be used to test the effects of the gene on the phenotype.

The problem of historical inconsistency was due to the approach I took to the analysis. I was at the time concerned with how to describe Generation Scotland collaboration with all its complexity, rather than the changes that were occurring. However, on reflection closer attention to the timeframe would have helped avoid inconsistency and the problem of aspects of the analysis appearing inaccurate from other perspectives.

GS is simultaneously a site, an object and a tool complicit in shifting the moral, ethical and political valuation of DNA and personal information locally and

globally. The GS assemblage is contingent, unpredictable and complex, and as such, the analysis suggests it does not have agency. Yet, in the website and representations of GS it does seem to have scientific, social and political agency. This contradiction might be explained as a further characteristic of an assemblage, in which case it will be resolved by GS either evolving into a more structured organisation or disaggregating. This has implications for the different types of public engagement envisaged by GS and its participants. It also has implications for the future management and control of the research resource and on business with the commercial sector. In the next chapter I go on to consider the collection of blood samples.

Chapter 4

'It's just routine': Collecting Blood

Moving on from the organisation of Generation Scotland, I now turn to the practicalities of constructing a human population genetic database. These practices take place at a remove from the organisation, planning and management; they are carried out in different locations, by different people, using specialist equipment. The first in a series of steps towards this end is the collection of blood samples. A blood sample is enacted as an object in a specific way; by focusing on the practice that enacts a blood sample, the object is brought to the foreground. Mol (2002) argues that if we look at objects from different perspectives, they become fixed, but by looking at the practices that enact these objects, we find that there are multiple versions of an object, that the object is no longer passive: 'objects come into being - and disappear – with the practices in which they are manipulated' (Mol 2002:5). In this, and the following chapters, I focus on the practices, the interactions of people, equipment and places that produce particular objects. These practices create new connections and disconnections between both the familiar (social) and the unfamiliar (science), through processes of transformation.

Venepuncture is the insertion of a needle into the vein of an individual for the purpose of extracting a quantity of blood, which may be used in research, or for testing for particular phenomena associated with a particular disease. The research nurses regard this as a 'routine' process and perform the task regularly. Indeed, when I initially contacted the research nurses seeking permission to

come and observe them collecting blood samples, some of them expressed surprise at my interest; 'it's just routine' (RNG) and 'it's not very interesting' (RNK) I was told. 'Venepuncture is a mundane procedure' (Pfeffer and Laws 2006:3012).

Blood samples are collected and enacted through the practice of venepuncture, the people are research nurses, the equipment consisting of needles and tubes, and the places are medical settings. Venepuncture is a 'routine' practice occurring thousands of times a day across Scotland in hospitals, clinics and general practices. Mainly for diagnostic purposes, it implies certain assumptions and expectations. Venepuncture for medical research is less common, although the practice is similarly considered 'routine'. The practice of venepuncture for the creation of a population genetic database may be a 'routine' practice, but the assumptions and expectations taken for granted with a blood test disappear, making way for a new set of ideas, moral values and intentions. A blood sample for DNA extraction is, I would argue, a new version of a familiar object. Not only is it a new version, but, as I aim to show through examples, a complex object, physically comprised of more than one single container, ascribed with diverse values, and with the potential for multiple uses. The blood sample represents the convergence of the individual participant, their family, their community, and wider public interest, with the research nurse, Generation Scotland, medical science, health policy in Scotland, and global medical technology.

Blood in the medical setting is good and bad, injury and illness, life saving and death threatening; in a genetic context, however, blood is routine, a resource that carries not symbolic meaning but utility. The intention to collect 50,000 (GS) or 500,000 (UK Biobank) blood samples is therefore unexceptional, a routine practice that requires no further discussion beyond the means of recruiting sufficient participants, the blood is taken for granted.

Research nurses have a multifaceted role in the collection of blood for research. They work at the interface between the social world of the participants and the scientific world of the lab technicians. The research nurses in Generation Scotland form a node in the network of connections that are formed between participants and the project. In their everyday work they may contact or talk to potential participants, and come face to face with the members of the public who have decided to participate. At the same time they are peripheral to the scientific work: they do not go into the labs; nor do they see the lab technicians who work on the blood and DNA; and depending on the project they may, or may not, know the researchers who designed the project and analyse the data.

The practices of the research nurses take place in particular locations, for example a medical space such as a surgery. The participants, by contrast, are not so easily fixed, as recruitment to and participation in giving blood for research is not a routine action. Many unknown people, who may be located in any number of places, become part of the process of recruitment, and, if they agree to participate and give a blood sample, come into a specific medical setting for a brief time. Blood is a powerful mechanism for connection, but for research it

must be disconnected from the body. It is at precisely that point, when blood is disconnected from the body, that it acquires multiple values: the moral value of giving, the value of relatedness through kinship and/or community, the medical or academic value for research and the economic value[1].

What is a blood sample?

The blood samples will be collected in accordance with current law and guidelines concerning the ethics and practice which define or describe blood in several ways. These versions of blood or blood samples are relevant to, but not usually enacted in medical settings; they have an institutional formality and are somewhat ambiguous, and they are enacted through the practices of law and ethics, as opposed to research nurses with needles and tubes.

In November 2004, the UK Parliament passed the Human Tissue Act. A guide to the Act by Jane Kaye states that 'when it comes to living persons the Act only regulates the storage and use of 'relevant material'' (Kaye 2005). 'Relevant material' is defined as 'material other than gametes, which consists of or includes human cells'. Excluded from this definition are embryos outside the human body, or hair or nail from the body of a living person. This definition means that a blood or tissue sample is regulated by the Act because it 'consists of or includes human cells'. However, once DNA is isolated and extracted from the blood or tissue, the DNA no longer falls under this definition as it does not 'consist of or include human cells' (Kaye 2005).

[1] Economic value may take following forms: non-monetary exchange for information (individual test results); non-monetary exchange for research results (public or community health); or potential monetary exchange through access or licensing to the commercial sector.

It is interesting to note that in Scotland the Human Tissue (Scotland) Bill, which was introduced by the Health Minister in June 2005, only refers to organ donation from living people. It does not include tissue, e.g. blood, removed from living people for other purposes, such as research. UK law regarding tissue does not apply in Scotland, apart from the sections making unauthorised taking of DNA criminal; however, it may have implications in the future for Scottish research. Both the UK and Scottish Parliaments were advised by a number of bodies including the Medical Research Council, the Nuffield Foundation and the Wellcome Trust.

The Medical Research Council (MRC) produced the 'Human tissue and biological samples for use in research: Operational and Ethical Guidelines' (2001), in which they separate human material and human tissue.

> Human material: All biological material of human origin, including organs, tissues, bodily fluids, teeth, hair and nails, and substances extracted from such material such as DNA or RNA.
>
> MRC 2001:2

> Human tissue or sample collection: Any samples of human biological material to be kept for reference, teaching or future research use.
>
> MRC 2001:2

The guidelines go on to say 'Much medical research depends on the use of samples of human biological *material*.' (MRC 2001:3), but 'We recommend that tissue samples donated for research be treated as gifts or donations' (MRC 2001:8). Thus I would interpret this as meaning the blood sample is both, human

material and human tissue depending on context; whether that is significant or not, I cannot judge, but it illustrates the ambiguity that can invade the discourse. Many such discussions are characterised by a lack of clarity and words are sometimes used interchangeably or have different meanings for different participants. The European Commission is acutely aware of this, and their '25 Recommendations on the ethical, legal and social implications of genetic testing' which included biobanks, begins by stressing the need for a global consensus on the definition of genetic data and testing terminology. The report includes a definition for 'identification of samples' but not what a sample might be, it does not define blood, tissue, or biological material (McNally and Cambon-Thomsen 2004:107). This approach implies a disconnection between embodied substance and scientific processes.

The Nuffield Council on Bioethics took a different approach to defining human tissue in a 1995 report on 'Human Tissue Ethical and Legal Issues', describing biological structures and specific cell types, their final list of human tissue consists of: 'Organs and parts of organs; Cells and tissue, Sub-cellular structures and cell products; Blood; Gametes (sperm and ova); and Embryos and fetal tissue.' (Nuffield 1995:19) Blood appears to be a particular type of tissue, 'different from other tissue in that the cells are not embedded in a solid matrix but are separated each from the other in a fluid called plasma' (Nuffield 1995:17). In their discussion of tissue banks, neither blood nor DNA are specifically mentioned and discussion of blood and blood products is focused on blood donation for transfusion. The collection and use of blood in general,

and blood samples for DNA extraction for research must be assumed to be implicit in their discussion of tissue in general.

The Council of Europe takes a broad approach in a 'Draft recommendation on research on human biological materials' (Oct 2005), and states 'This recommendation applies to the full range of research activities in the health field that involves the removal of biological materials of human origin to be stored for research use' but 'This recommendation does not apply to embryonic and foetal tissues' (COE 2005:4).

These versions of blood as material and tissue are ambiguous. They are also remarkably insubstantial and remote from embodied blood or blood in a test tube. They act to formalise the disconnection of blood from the person. They seek to reduce it to an object which can be a suitable subject for regulation. This is a process of control and audit, often about quality and safety standards, and usually concerned with the public as a whole. The individual from whom the blood, or material, was taken merely become part of that whole – disconnected from their contribution to medical science.

Research Nurses

I carried out observation of six research nurses at work on three different studies, as they collected fifteen blood samples for genotyping, over a period of four months in 2005. None of these projects came under the GS umbrella, but were ongoing studies at the time and involved the same procedures that would be used for GS. Blood and data collection for GS projects did not commence

until early in 2006. I only had permission to observe and talk to the research nurses, and will not report on anything that the participants said or did. In places, I will however mention the research nurses' perceptions of participants, their observations and opinions.

I observed the research nurses in peripheral hospital spaces, that is, while their offices or clinics are part of a hospital complex they are set apart from the everyday spaces of clinical activity, often with separate entrances indicating a separate department or unit. This points to research as an activity that is particular and separate from the mainstream provision of health care. All research nurses are fully trained NHS nurses and have held a range of jobs within the nursing profession. Why they are working as research nurses is not discussed here, but they are simultaneously similar to and yet different from the nurses who care for patients in either Secondary or Primary Care; they are similar in terms of their training and previous work experience, whereas their current work in research is distinct in so far as it is a career choice rather than an obligation or requirement of the 'caring profession'. Some research nurses continue to be NHS employees within the research setting, whilst others are employed directly by the respective project, usually within an academic setting. Here, I refer specifically to the type of contract that they may have for their work. Some of the research nurses I observed, like other clinical and non-clinical researchers, are working on fixed term contracts that were defined by the requirements of a given project. This was also illustrated by the wearing or not wearing of a nurse's uniform. The research nurses in the clinical research facility and those on one of the other studies wore uniform. Only on one study that I

observed did the research nurses not wear a uniform. These research nurses were the most remote from the mainstream activities of the hospital they were based in, and also the furthest away from the central organisation of the project they were working on. The boundaries between the NHS and a medical school, an academic medical department, or research facility are blurred and difficult to discern from observation only. There is undoubtedly a complex set of overlapping interests, eg finances, and involvement through multiple roles of individuals, between medical research, academia and the NHS. It should also be mentioned that nurses who work in research are not always attached to projects such as the ones I was observing; indeed there is a completely separate field of nurse-led and nursing research programmes within the broad framework of health research.

The research nurses are the *face* of a project, in the sense that they are who the participants meet when they come to give their blood samples and data. Research nurses act as a connection between the public and the clinical or scientific research; they also work at the intersection of interdependent interests and activities. Research nurses are usually recruited to a project once the funding has been awarded, and will only work with the project for the duration of the data collection. It may be the case that they have very little or even nothing to do with other members of a research team and their activities. However, GS: 21CGH has included research nurses in the planning stage of the project, and their expertise has helped inform the development of the project protocol and the application to the ethical committee. For instance, the GS: 21CGH research nurses write the protocols for the collection of data and blood

samples, which will then be carried out by a number of other research nurses over a wide geographic area. The protocols they write will ensure consistency in practice while allowing flexibility for adaptation to individuals' preferences, participants' vagaries and different locations. One of the research nurses writing the protocol told me, 'it is a difficult thing to write, because you need to cover everything, but don't want to make it so detailed that people can't make any decisions of their own – some protocols have too much detail so that you can't be flexible with a patient – others are so vague you have to make it up as you go along' (RNK).

The world of the research nurses is also a social one. They share space and time with the other research nurses, either on the same project or working in the same place. They share meal breaks, often sharing food and making drinks for each other. They chat when they are on their breaks and between participants' appointments, sometimes about work, but also about their families, out-of-work activities, and other interests. They support each other, through personal problems as well as in their work, on a daily basis. There was an ethos of co-operation and openness between them in each of the places where I carried out observation. Working with the research nurses, I found that they extended this openness to me, to other researchers and members of hospital staff who came into their space. I was invited to join them on their breaks, they made drinks for me and included me in their conversations, even though I was coming and going on an occasional basis. They were always very helpful and friendly, willing to answer questions, frequently offering further help, and checking that I was able to get the information that I wanted.

Research nurse – participant relationship

As the *face* of the project that was seen by participants, the research nurses were located at the interface between the public and the scientific medical research. They were probably the only people from the project that participants were ever likely to meet. The face to face contact established a relationship between the participants and the research nurses that was an integral part of the process and relationship between the public and the project. The nurses were responsible for collecting the blood and checking the questionnaire data that would form the phenotypic database; this responsibility included ensuring that the participants were informed and had consented. Trust also had to be present, as handling blood raises issues of health and safety (pollution and danger) - for the research nurses as well as for participants. Participants entrusted research nurses with their wellbeing as well as their *gift* , which I discuss later in the chapter, in the practice that transformed the status of their blood from embodied substance to an object of medical scientific research.

Where the research nurses worked on a single project, they often spoke to the participant prior to their attending an appointment, the participant knew the research nurse's name and voice, and where possible the research nurses tried to organise it so that they saw that participant when they attend. There was no research on how important this might be to the recruitment process, but the research nurses felt that it allowed them to establish both rapport and trust with the participant. This helped the participant to feel more relaxed, which facilitated the collection of the data, and also produced a positive experience of participating in research. In the studies that did not use research nurses in the

recruitment process, but other project members or even an intermediate agency, the research nurses did establish a rapport with the participants, and the lack of previous contact did not appear to be detrimental to the collection of the data; but it was a subtly different experience for both, the research nurse and the participant, and there was not the same sense of engagement.

The research nurses were all unfailingly courteous to the participants and appreciative of their participation. They were meticulous in thanking the participants for attending the appointment, both on arrival as well as departure. When collecting data including the blood sample, they kept talking using positive prompts and frequent reassurances, such as 'that's great', 'good' or 'very good', 'are you happy with that?' and sometimes 'excellent', thus confirming to the participant that they were indeed doing something good. The role of the research nurses was a highly communicative one: they communicated with research participants who came in for appointments or whom they visited; they communicated with other members of their research project team; and they communicated with each other, as well as occasional others like myself. Throughout these serial interactions they code switched between everyday informal language when chatting together, more formal and technical language when talking with colleagues and other clinical researchers about procedures or equipment, and a combination of language of research based rhetoric coupled with informal everyday language when talking with participants. By that I mean they reiterated the rhetoric, the phrases, that were associated with this type of data collection, 'it will help us to understand disease better', 'it will help us to understand the genetic component of disease', 'you are

helping other people', 'we will be able to develop new drugs' or 'we will be able to develop better drugs'. The use of 'we' implicated the participant in the research, and created a connection between the participant, the research nurse and the project to construct a collective effort.

The research nurses felt a responsibility to the project that was employing them but, also to 'their participants'. When the participants were not there, some of the research nurses expressed reservations about the collection of data for a database. They were aware of the ethical issues, particularly those surrounding recruitment, information and consent, precisely the areas with which they were dealing. One said 'I don't know what they are going to do with the data - I don't have any training in genetics - and if I don't know, how are the people who participate expected to understand?' (RNK). The lack of knowledge about what happens to blood samples following their production was also found in a study of blood tests. What health care professionals knew was 'limited to their place in the division of labour' (Pfeffer and Laws 2006:3019), for example, nurses had no idea about what practices and processes were carried out by laboratory technicians. Another research nurse told me that she was 'not entirely sure about recruitment and collecting blood samples when it's done in hospital clinics' (RND). She felt that the people she saw were not always 100% clear that they were taking blood samples and collecting information for a research database, that they confused the activities of the research nurses with those of the clinic. 'Venepuncture is understood as something 'they just do' in hospital settings. It is an everyday practice that patients must and do accept without questioning' (Pfeffer and Laws 2006:3014). However, whilst some patients

thought they were not asked if they wanted to give a blood sample for tests (Pfeffer and Laws 2006:3014), they did in fact have to sign a consent form agreeing to give a blood sample for research.

Several of the research nurses commented on the reasons that participants gave for participating in a study, 'they usually tell us why they have come' (RNB). These reasons were varied; some are motivated by a single straight forward reason, whereas for others a more complex interaction of relationships and circumstances may be involved. The research nurses observed several motivations: participants who have the disease hope that the research will lead to a cure; those who know that the disease is 'in the family' see themselves as helping family members; those who have received health care themselves see this as a way of paying back for the help they have received (this seems to be associated with acute health care, particularly a hospital experience); and those who viewed participation as an opportunity to get some sort of information on their own health. The research nurses all observed that participants were interested in feedback, either in a general way or for something specific, like their blood pressure or cholesterol level. 'A lot of patients see this [taking part in the study] as a way of getting a free health check, others think we are going to find a cure [in the case of a disease specific study]' (RNN). Overall, the research nurses felt that, 'most people see this as a way of helping others' (RNC).

Based only on the range of my own observation of the collection of blood samples, I do not have the empirical evidence to expand on this here. Further research into the experience and knowledge of research nurses as well as

participants, in the light of a Scottish population genetic database would be helpful in confirming or contradicting these possible reasons for participation. However, the above categories of participants' motives resonate with the motives of Titmuss' blood donors in *The Gift Relationship* (1970), namely altruism, reciprocity, replacement and duty. The question of motive is important in determining what a blood sample may be, that is how it is perceived and may influence recruitment and participation. So far, little research has been carried out into the methods of recruiting families, participation rates and experiences for human genetic database creation, but what has been done raises issues and concerns (Kreiger, Ashbury, Cotterchio and Macey 2001; Austin 2002; Beskow, Botkin, Daly, Juengest, Lehmann, Merz, Pentz, Press, Freidman Ross, Sugarman, Susswein, Austin and Burke 2004; Ottman, Berenson and Barker-Cummings 2005). Generation Scotland will contribute to this research as part of its ELSI agenda.

The organisation of research nurses

Two of the studies I observed were employing research nurses to recruit participants and collect data for that project; another study was based in a clinical research facility where the nurses were engaged in collecting data for several different projects. In the clinical research facility, a team of nurses work on various projects; there was a board stating the study name and listing which nurses would be working on which projects on any particular day. Participants for each project tend to be organised so that several people for a study come on the same day. As the research nurses have to know the protocol for each of the studies they work on, the nurse manager allocates particular nurses to particular

projects, so that not all nurses work on all projects, but any one nurse may be working on one or more projects.

The research nurses that were working for a single project worked on different aspects of the project at different times. There were periods of time when they were working on contacting potential participants for recruitment from an office; and other times when they were collecting the data and blood samples in a clinic, general practice surgery or in participants' homes. Thus their time tended to be organised around blocks of activity, starting with recruitment. Once a sufficient number of participants had been recruited, they organised a session for data collection where they would see several participants in sequential appointments in one place.

Research nurses who went to clinics, a general practice, or even occasionally to participants' homes, had to take all the necessary paper work and equipment with them for a session. As one research nurse observed to me as I helped her carry multiple bags from her car in to a general practice surgery - 'I feel like a bag lady' (RNB). The equipment included the needles, tubes, labels, sharps boxes, trays and other paraphernalia necessary for collecting blood samples, but sometimes also included equipment for taking physical measurements such as weighing scales and height sticks. For studies that required measurements like an ECG, the participants had to come to a clinical facility where this equipment was set up and maintained. Responsibility for transporting equipment, and in particular, the data and blood samples themselves, fell to the research nurses. Equipment and paper work, including the consent form and any questionnaire

data, were taken to the place where the project was based, while the blood samples were sent to a laboratory.

The appointment

What happened when an individual came to participate in a project? The research nurse brought the participant into the designated medical space, which could have been a ward, a general practice surgery, or interview room in the clinic. First of all, they ensured that the person agreeing to participate had read and understood the information that has been sent out to them prior to the appointment. An information sheet explaining the project is required for all clinical studies by the medical ethics committee. In most studies, but not in all, it was the research nurses' responsibility to ensure that the individual had read the information sheet, believed they understood its contents and had been given the opportunity to ask questions, *prior* to signing the consent form. The consent form was signed before the participant could proceed to give a blood sample or any data collected.

Some of the research nurses told me when they gave an information sheet to a participant they always asked them to read it; others said they asked if the participant had already read it. If required, the research nurse allowed time for a participant to read over the information sheet before asking if the participant has any questions. The research nurses agreed some participants had questions, others did not, most indicated that they were 'happy to help', but it was not always easy to determine if the participant did in fact understand the information that they have been given. In some cases, where it was clear that the

participant had not read the information or did not understand it the research nurse would go through the sheet with them. Only once the research nurse was satisfied that the participant had the opportunity to read or have the information explained to them, and any questions answered, would they move on to signing the consent form. In an ideal world, the process of ensuring that the participant has all the information they need prior to signing the consent form should take as long as required; in reality, the length of the appointment limited the time available, and there needed to be enough time to collect all the data for the study without making the appointment onerous for the participant or keeping subsequent appointments waiting. In practice, the nurses quickly made a judgement as to whether they thought a particular participant had understood the information or not, based on their experience as research nurses.

One of the research nurses told me of a meeting she had recently attended, in which the question of 'who is consenting the participants' had been discussed. Consenting in this context is used in two ways, there is *the participant who is consenting*, but consenting may also be used to refer to the person who is initiating the consent. In the research environment it is not unusual to find reference to *the person consenting the patient*, where consenting implies the person carrying out the process rather than the person giving the consent. The question under discussion at the meeting was whether the research nurses were actually *consenting* the participant or whether they were only *witnessing* the signing of the consent form, the consenting might been done by another person who had provided the information and answers to any questions from the participant.

The question of consent, and at what point it is given, with regard to collecting particular types of data, are currently under discussion within GS. With particular regard to the collection of questionnaire data, it might be assumed that a person has implied their consent to participate in providing data for the creation of the database by completing a questionnaire prior to actually signing a consent form. However, in the studies that I observed, no data were collected until after the consent form had been signed. Once the consent form is signed, the research nurses move swiftly into the data collection. The research nurses that I observed all left the collection of the blood sample till last, or close to the end. They tended to start with the questionnaire data before moving on to physical measurements, and took the blood sample last.

The Generation Scotland: Scottish Family Health Study proposes to collect blood samples from 15,000 individuals over the next two years in East and West Scotland during Phase 1 of the project. A further 35,000 blood samples will be collected across the rest of Scotland in Phase 2 over the following three years, aiming for a final total of 50,000. They plan to collect: 2 tubes of 4.7mls for tests; 1 tube of 9mls with ethylenediaminetetraacetic acid (EDTA) for DNA extraction; and 3 tubes of 7.2mls clot activator with gel separator, which will be centrifuged and the serum stored. The total blood sample will be approximately 40mls. (SFHS protocol Oct 2005). The collection of blood will be carried out by research nurses located across Scotland.

Venepuncture

The practice of venepuncture used well developed techniques and clinically tested equipment designs which aimed to ensure the safety of the participant and the practitioner and the integrity of the samples, in the studies I observed. Blood can be drawn from a vein or an artery, for the purposes of the studies I observed, all blood was drawn from a vein using venepuncture, as will be the case in the GS projects. The collection of blood samples for genotyping requires research nurses because they have the training and skills to carry out venepuncture.

Taking a blood sample first of all requires the participant to be in a comfortable position. The blood samples I observed were all taken from the arm, from a vein on the inside of the arm, roughly level with the elbow joint. The arm rested on a support, for one study this was across a pillow placed on the participant's lap, in another on the edge of a desk and in the third study, the participants were seated beside a bed with their arm resting on it. The research nurse moved a tray or box into a convenient place adjacent to the participant and within easy reach for themselves. The tray or box had previously been set out with the needle, the requisite number of sample tubes and lids, wadding and elastoplast, one tray for each participant. Next, they applied a tourniquet to the upper arm and palpated the arm to locate the vein. Then the research nurses washed their hands, some put on protective surgical gloves, most did not. They checked the vein again, swabbed the arm with alcohol, the needle was inserted and the blood sample drawn. The sample tubes were inverted several times to mix in anticoagulant or other preservative chemicals, depending on the purposes for

which a particular sample tube of blood was intended. Once the blood had been collected, the needle was removed, a wad was placed on the puncture wound in the skin and pressure was applied, the tourniquet was removed, and the research nurse removed their gloves, if worn. After a couple of minutes the research nurse checked the puncture wound for bleeding and if all was well then either taped the wad in place or stuck a plaster over it.

The studies that I observed were collecting more than one sample tube of blood. In order to be able to change the tubes, there are two types of 'closed system' equipment used for this procedure. Using a closed system means that more than one sample tube can be used without the need to transfer blood from one tube to another, thereby reducing risk of contamination to the sample, and to the research nurse handling the blood sample. The type of venepuncture equipment and sample tubes are dictated by the project protocol, so that even if the research nurses have a preference for using a particular type, they must use the equipment provided by the study.

The research nurses called one type of system the 'butterfly'. In the 'butterfly' system the needle was attached to a plastic 'butterfly' with a short length of narrow and flexible pipe going to a holder with an adapter valve. The research nurses would tape the 'butterfly' in place on the arm and then have both hands free to change the sample tubes. The tubes used with the 'butterfly' needle were 'vacutainers', that is, there was a vacuum within the tube. When a sample tube was fixed to the valve at the end of the narrow pipe, the blood was drawn out by the vacuum. Once the required amount had been drawn, the sample tube

was disconnected and the lid sealed, and then another sample tube was attached to create a further element of the sample.

The other type of closed system used a Monovette tube. The needle was attached directly to the end of the holder, a syringe-like tube (only without the plunger) which was fitted with an adapter valve directly behind the needle. The research nurses using the rigid Monovette arrangement held this in place with one hand, while drawing blood and changing the sample tubes with the other. The sample tube was inserted into the syringe case and connected to the valve directly behind the needle using a click and twist manoeuvre. A plunger was then used to draw the blood out into the sample tube. Once the required amount had been drawn, the sample tube was withdrawn from the case, the plunger was broken off, and the lid sealed. The process could be repeated with another tube until the required number for the sample had been collected.

The research nurses preferred the 'butterfly' because it could be taped to the arm of the participant, the needle was less likely to move around and cause discomfort, the short length of tube gave them more room to work, and it was easier to attach, detach and cap multiple sample tubes with two hands. However, they told me that the 'butterfly' needles were more expensive and not all projects wanted to spend that much out of their budget on needles that could only be used once.

The number of sample tubes varied between the studies that I observed. The different number of tubes per study was explained by the range of tests that a

study may want to carry out in addition to DNA extraction. One study was collecting 8 separate tubes of the blood sample, and therefore had the most tube changes. It was also the one using the type of needle fixed into the syringe tube, the one that was more likely to move and cause discomfort to the participants. Using this needle type meant that the research nurses had to do the sample tube changes and seal the lids with one hand, requiring a high level of dexterity and concentration to collect the blood into so many tubes. They worked very quickly in order to minimise any discomfort the participant might have been feeling. There were two sample tubes of 5mls ('plus a spare if you can' (RNB)) for DNA extraction, 'these are the most important ones to get for the study' (RNB). Then 5 sample tubes of 4mls, and 1 additional tube of 2mls if the participant had fasted. A total blood sample of 30mls in tubes that were of different sizes and had colour coded lids of red, purple and grey. The sample tubes were labelled, placed in plastic bags which were then closed, and packed in a 'cold' box with freezer blocks. This study was providing feedback to the participants on the results of the blood tests. Each participant was given an information sheet to take away with them, explaining what the different test results mean and was told that they would receive a letter from the study with this information. They were also advised that 'If there are any problems we will tell you in the letter, and we will also contact your GP so that you can make an appointment' (RNB). Only few studies do this, though if anything of concern is found at the research appointment, the participant's GP will be informed.

A second study was collecting the blood sample in 6 tubes. There were 2 of 9mls for DNA extraction, and 4 smaller quantities, 3 of 3mls and 1 of 5mls. The total

quantity of the blood sample - 32mls - was described to one participant as 'about 8 teaspoons of blood' (RNN). The tubes had colour coded lids. They were labelled and sorted into two different plastic envelopes according to which laboratory they were being sent, one for DNA extraction and the other to the haematology laboratory for routine tests.

The third study collected blood samples in 4 tubes. There was 1 tube of 9mls with no added chemicals, and 3 of 9mls with EDTA, a total blood sample of 36mls. In this study, the research nurse spun down the blood in the sample tubes in a centrifuge. This process separates the plasma from the heavier blood cells. Once the sample tubes had been spun down, the plasma was drawn off by pipette and transferred into 5ml tubes for storage onsite at -70°C. The tube with no chemicals was disposed of, and the 3 remaining sample tubes, now containing approximately half the contents, were sent off in batches to the study laboratory. The sample tubes were placed inside screw capped tubes in case of leakage. The research nurse then wrapped paper towels around them and placed them in a small cardboard box which was then put into a padded envelope for postage. The screw capped tubes, boxes and padded enveloped were reused throughout the course of the study.

In all studies, the lid colour indicates the chemical content of the tube, whether it is an anticoagulant, for example, EDTA, or another type of preservative, such as Heparin. The choice of chemical or preservative in the tube is determined by the use the blood will be put to in the laboratory and what type of tests carried out for the study. The presence of a particular chemical in a tube may render it

unusable for a particular test, so the chemicals in the tubes are varied according to the test that will be carried out on the blood. For DNA extraction, the use of EDTA effects the composition of the blood by reducing the coagulation and breaking down the cell structure, but this does not affect the use of the chemicals and reagents necessary, nor damage the DNA for extraction.

What the description of these studies show is 'a blood sample' from an individual is often made up of a number of tubes of blood, not just a single one. The individual might only be giving the blood once, but it is in effect several blood samples. Moreover, the blood samples are collected so that they can be used in different ways. This process of multiplying the resource for dispersal to different uses and storage is repeated in later chapters with DNA. Indeed, the storage and retrieval of samples poses one of the biggest technical challenges for both GS and UKBiobank.

Health and safety

The aspect of the research nurse collecting blood samples that was most revealing about their attitude to blood comes under the heading of health and safety. One of the things I observed was that some research nurses wore surgical gloves when they were taking the blood sample and others did not. When I asked them, it transpired that ordinarily they would not use gloves, and had only put them on because I had been there to observe them. One nurse told me that 'you don't need to worry about getting blood on yourself - it washes off with soap and water' and 'the only time you might need to be concerned is if you have a cut or break in the skin on your hand, then you would use gloves –

but apart from that you don't need to use them' (RNC). Another research nurse told me 'when we were doing the venepuncture training, they told us that gloves were optional, so it was up to us to decide if we want to use gloves of not – that might be different in another Health Board – obviously they recommend that you use them - but we don't have to, it is not insisted on - I prefer not to' (RNF). All the research nurses that I asked about gloves said that they were awkward, and that you were more likely to spill or drop something when you were wearing them. The latex gloves used to be easier to manage because they fitted better, but the new nithril gloves don't fit so well and if your gloves are loose, especially at the fingertips, you cannot feel the vein and are more likely to give yourself a needle stick injury.

Medical authorities regard needle stick injuries as the greatest hazard for the research nurses. But, one of the research nurses told me that 'I worry more about spilling some blood on someone's clothes, or if you were to drop a tube you would have to take more [from the participant]' (RNL). Another nurse said that she worried about spilling or dropping a tube of blood if she was doing a home visit, in case some went on the furniture or carpet. Spilling a blood sample, or dropping a tube is unusual when using the 'closed' system for collecting samples. Sometimes, when the blood samples were being transported, the seal on a lid leaked; or occasionally the research nurses found that a tube leaked because it was cracked. In these cases they did use gloves to clean up. But in general, the research nurses did not regard taking blood samples or handling blood as dangerous.

The research nurses were aware of the risks associated with the transmission of disease, particularly HIV and Hepatitis, but they did not regard themselves as particularly at risk from this 'routine' procedure. They felt that there was no need to worry about contracting infections as long as you made sure you washed your hands. If they had a participant they were worried about for any reason, they would use gloves, but they expected that anyone with a diagnosis, or high risk of an infection like Hepatitis would have been screened out of a study at the recruitment stage.

Blood outside the body

The participants that I observed tended to behave in a similar way to those I have seen at blood transfusion sessions. They tend to stare fixedly at a point in the distance, not looking at the research nurse, or what they may be doing with the needle or sample tubes. Pfeffer and Laws also report that 'more often than not, patients turn their face away or look elsewhere when venepuncture is performed' (Pfeffer and Laws 2006:3014). The research nurses, in the same way as the blood service nurses, tend to ask open ended questions that allow people to chat away as much or as little as they feel inclined, about the weather, traffic, summer holidays or any subject that is offered by the participant. My own experience of giving blood as a blood donor was to stare fixedly at the ceiling of whatever hall the service was using, that is until I started doing this research. Last time I went to give blood I made a point of watching everything the nurse did and was surprised to discover that there were several sample tubes of blood taken as well as the 'bag'; on enquiry I found that the sample tubes were sent to

the lab for testing, and that the results determined what my blood could be used for.

Personal experience and anecdotal evidence leads me to think that most people don't like to look at blood, either their own or other people's. Indeed, many people are known to faint at the sight of blood, or claim that it makes them feel sick to see it. Yet if we cut ourselves, or receive an injury for example from a fall, there is frequently a quantity of blood visible which does not seem to affect all people in the same way. Certainly, where children are concerned, the cleaning up of a bloody injury to the knee is commonplace, men cut themselves shaving, and knife cuts from activities in the kitchen are also common. There is no shortage of blood in the medical dramas on TV or in violent films, so it is not that we are unused to seeing blood. This implies that it is something to do with the context in which we see the blood, and possibly also with the quantity.

Blood in a medical context is different from blood in a domestic context. Drawing blood with a needle is different to bleeding from an injury. Blood in a public context signifies something has gone wrong, some sort of accident or malevolent event has occurred, for example, a car crash, a bomb, or an earthquake and more often than not is it blood that gets those images onto TV or the front page - 'if it bleeds it leads' an old newspaper saying goes. The sight of blood on the body indicates that a wound has been sustained, and evokes a response of fear or horror proportional to the quantity of blood visible. Blood in this context has strong associations with pain, and often emotional or physical trauma.

Similarly, needles also have a strong association with pain for many people. The research nurses go to lengths to reassure any participants that are anxious about the insertion of the needle into the vein that it will 'only hurt a little' (RNN) or that 'it will feel like a little scratch' (RNB). The participants assured the research nurses that they were fine, and several expressed surprise when the research nurse announced they had finished, after they 'had not felt a thing' or 'it wasn't as bad as I had expected'. According to the research nurses', all participants responded to their thanks for giving the blood sample by indicating that they 'wanted to help'.

Blood can be heroic, signifying the desire to help other people, whether that be the blood stained rescuer who has evidently risked life or injury to assist someone in distress, or the less dramatic but also altruistic act of giving blood as a donor or a research participant. The needle may not be as threatening as, for example, a collapsing building, but it does require that people make a conscious decision to overcome any anxiety that they may have and to permit someone to inflict on them what they anticipate as pain.

After the research nurse has 'taken' the blood sample, a collection of plastic tubes with coloured lids sits in a tray, box or envelope awaiting dispatch to a laboratory. The blood which was *in the body* of the participant is now *of the body* of the participant. The connection is still evident, as long as the person remains in the room, and they can still see *their blood*. The blood then continues to be connected by identity with the individual, in the sense that as long as they can see it, they can still make the claim that 'that is *my blood*'. However, after the

individual leaves the room, that connection through identity is broken. You cannot tell by looking at a blood sample who has given it; one blood sample does not look uniquely different from any other blood sample. Indeed, unless the tubes are labelled, either immediately before or after the extraction of the blood, they cannot be used in the study. The blood samples were all labelled with identification of the study and a unique identifier for the individual participating in the study. Some also included the date of birth. The blood samples and all other information collected were personal, but removal of personal details de-identified the remaining information.

Removing identifiable information and replacing it with unique identifiers works as a form of protection. Participants do not want to be identified, because they do not want private personal information about themselves, their family, or their health, to be accessible in the public domain. These systems of unique or coded identifiers intend to anonymise the blood samples and personal data. The GS projects use centralised systems of barcodes for labelling blood samples and personal information. Anonymisation serves to protect confidentiality and privacy, but the anonymisation that occurs in GS (and many other research projects) is not absolute, but rather a systems of coded identifiers that allows various different types of data to be linked. Chapter Seven returns to these issues, examining what happens when the various types of data are (re)connected.

The connection between the blood, as substance, and the body, as inhabited by a sense of self and personality, is assumed to have been broken. The blood no

longer *belongs* to the participant, but it still retains the identification of the participant. Such an assumption of disconnection is facilitated by the separation of identity from identification. The identification of the blood sample is essential to its future usefulness for DNA extraction and analysis in conjunction with phenotypic data. The blood is disconnected from the body, as in these circumstances it will not be either returned to the participant's body or given to another body (as in the case of blood donation for transfusion purposes). The information within the blood, i.e. the DNA, could be used to identify the individual participant if necessary, in the same way that DNA can be used by the police to identify an individual. The blood itself retains no sense of the personality or sense of the self that resides within the body from which it was taken, nor does western medicine include the belief that any aspect of spirit or spirituality is embedded within it and blood carries no suggestion of the intercorporeality associated with organs (Waldby 2002). It is perhaps one of the great paradoxes of blood, or rather of these blood samples, that whilst no intercorporeality, no sense of identity is attributed to them, they will be used to extract the very essence of intercorporeality, the shared substance of all humans, DNA.

Although, anecdotally, many clinical researchers assume, possibly out of convenience, that participants retain no interest in their blood or information once it has been given, there is counter-anecdotal evidence from general practitioners that patients who have taken part in clinical studies frequently enquire about the progress and results of the study. One GP told me 'you wouldn't believe the number of phone calls I get asking about research results,

and I often don't even know that this person has taken part in a research study' (Interview 04/03/02). The very categorisation of a person as participant indicates an act of taking part or sharing.

Conclusion

Research nurses are organised in varied ways, some are employed under contract to a particular study, while others are members of a clinical research facility, where they may collect blood samples from participants in different studies in the course of a day's work. Research nurses were also involved in contacting and recruiting participants, organising appointments for participants, checking that participants understand the information they have received about the research, ensuring the consent form is signed, collecting data other than a blood sample, including administering questionnaires, and taking physical or clinical measurements such as ECG scans.

A blood sample located within the particular space and time of 'giving' has the following characteristics: a quantity of substance, from 2-9mls; contained within a plastic tube; with a colour coded lid; it has a label for identification; and it will be dispatched to a laboratory. It does not belong to the 'giver'; it will not be returned to the 'giver'; and it cannot be distinguished from any other blood sample without its label. Furthermore, if it was removed from the context of the study it would have no intrinsic value of itself. Indeed, if it was removed from the specific context it would become a *dangerous* object.

Once the connections to the designated participant have been removed, the blood sample receives a new set of connections to embed it within a particular study. It has been labelled with the identification of the study thereby signifying its change from *belonging* to the individual, to belonging to the study. Anonymisation imposes a disconnection between the person and the blood or information, to protect confidentiality and privacy, but it cannot be absolute in this instance. A unique identifier is necessary because it has become one of many, a case within a study sample. A single blood sample is of no use by itself for the purposes of gene identification or genotyping, which can only take place through comparison with many others at an aggregate level. Thus there are different sets of values ascribed to the blood sample through new connections from those ascribed to it when it was embodied blood.

A blood sample is a complex object, enacted through the practice of venepuncture, which involves the interaction of people, equipment and places. The blood sample shifts from belonging to the participant to the study. The participant is drawn into the study through the 'giving' of the blood sample; at the same time, the study is connected to the social world through the participant. Research nurses are located at the node of these complex connections and disconnections.

Chapter 5

'From Nasty Dirty Blood to Nice Clean DNA'

Following the blood brought me to the next site, the laboratory, a site that is associated with a particular type of work and knowledge. There are many types of laboratory, and there are different types of laboratory study. Laboratories are the places where science is practised, be they based on physics, chemistry, biology, or combinations thereof such as biochemistry. Laboratories also tend to be highly specialised, organised, staffed and equipped for specific purposes that may focus on research, development or production of certain objects. They may be located in, for example, academic, medical, public or commercial institutions, and they may be funded in many different ways. Social scientists have taken different approaches to their research into laboratories, shaped by their agendas, funding, and intentions. Thus there have been studies that look at the culture of the laboratory or the practices therein e.g. Knorr-Cetina (1995), Latour and Woolgar (1979), Hine (2006). The results reveal different aspects of laboratories, how they work, who works in them, and what it is that they do. Each study contributes to the understanding of how laboratories and science can be studied and thought about. Inherently part of the social world, yet appearing separate, the laboratory and the science practices within it are not well understood by non-scientists. The activities of laboratories can seem strange, exotic and sometimes incomprehensible. Yet these same activities impact on the social world, often changing our view of the world we live in, or the way we think about it.

The Genetics Core laboratory is a service lab, located in the Wellcome Trust Clinical Research Facility (WTCRF) at the Western General Hospital, Edinburgh. The Genetics Core lab processes blood samples for genetic research projects and is central to the creation of the genetic databases for the GS projects. The lab is an essential element of the GS assemblage, but it is also a complex site in its own right, located within a network of connections. Three key processes are carried out in this lab to transform blood into digital data - DNA extraction, genetic sequencing and genotyping.

In this chapter the lab is described and discussed, and the story 'from arm to desktop' continues with the arrival of the blood samples in the lab, the practice of booking-in, and the extraction of DNA from the blood. The descriptions of the practices are quite disembodied, partly because by focusing on practices and objects it 'leaves the actors vague' (Mol 2002:33), and partly because this is a reflection of how things are done in the lab anyway. The lab is a tool and the people in it become part of the machinery and processes. This chapter shows that the practices enact blood samples and DNA as particular versions of these objects through the interaction of people (lab technicians), place (laboratory), and equipment. The lab technicians transform the substances and the meaning of those substances through this interaction e.g. from dirty to clean, or, from stuff in a tube to information.

The blood samples arrived at the lab, not as the bright red 'gift' of the previous chapter, but as 'nasty dirty stuff', to be handled with protective gloves on. The first morning I was in the lab, the lab manager was introducing me to the

technicians that worked there and one of them, LabTec1, said 'we will show you how we turn nasty dirty blood into nice clean DNA'. Blood samples were nasty and dirty because they had 'bugs', bacteria and viruses, which cause disease. The DNA, although it may carry the genes that predispose an individual to a disease, does not actually pose a threat to those working with it.

I observed blood samples coming into the lab from up to 10 different studies during the time I spent there, though not all at the same time. I spent a month in the genetics lab observing the lab technicians in their daily activities. They booked-in the newly arrived blood samples, extracted DNA from the blood samples, genotyped and sequenced DNA, and sent the results to the researchers working on the studies. The blood samples mainly arrived by post, and some were collected from the clinical research facility upstairs. The lab systems also managed and organised the storage of blood samples and DNA samples in the freezers.

Being in the lab

The first thing I learned in the lab was that the use of the term 'transformation' (of meaning) conveys something quite different to the people in the lab. The acting lab manager explained to me that the term 'transformation' in genetics referred to the transformation of a cell by adding, for example a cancer gene, and then growing the cell. There were also other kinds of transformation afoot in the lab at the time I was there. The lab manager post had been advertised and the new lab manager was about to arrive to take up his position, the role was at that time being filled part-time by the science manager from the Molecular

Medicine Centre (MMC) and one of the lab technicians. The Acting Lab Manager was juggling several roles, involvement in GS, temporarily managing the lab, writing grant applications, and teaching as well. There was a certain amount of anticipation and speculation about the new person coming in and what they would be like by the staff members, coupled with some reflection on the previous person who had fulfilled that role. The previous person had evidently had a particularly individual style and distinctive personality which had left a strong impression on the people who had worked with him.

The first morning I was in the lab was spent being introduced to the members of the lab staff and meeting with some of the people who were working on GS at the Western General Hospital, Edinburgh. The other GS people were located in several different places around the hospital complex, at centres associated with both research and teaching. The new Project Manager and the Data Manager came in to the lab and I joined them as they updated each other on how they were getting on with GS:21CGH and what they were doing. The Data Manager had just come back from a week away on a course and the Project Manager was beginning to find his feet and making contacts across the project. The GS Liaison Officer, was working in an office upstairs in the WTCRF, but was getting ready to leave, having secured work in London. This did not directly affect the lab but was part of the extended set of overlaps and interactions of individual GS people that worked peripherally to the lab site but centrally to GS. The GS people were there that morning not because they routinely came to the lab but because connections between the lab and other work (including GS) were embodied in the person of the Acting Lab Manager.

Learning in the lab

The lab technicians all had different roles, work that they did, that was different from that of others there. There was a sort of hierarchy of knowledge, of how to do certain things which was cumulative over time. LabTec2 had been there the longest and knew how to do everything, but worked on the DNA sequencing as no one else knew how to do it. LabTec1 did the genotyping and quality control. LabTec3 also did quality control, was learning to do the genotyping and did the peripheral blood lymphocytes (PBLs) and plasma extractions in the Cytogenetics Lab. LabTec4 was the newest of the lab staff, working on the DNA extractions as this was all she was trained to do at the time. LabTec5, the only one working in the lab who did not have a science degree, did the booking-in and some of the freezer management. They were aware that this division of labour meant that if LabTec2 was away on leave or off sick, the sequencing work could not be carried out in her absence. They were aiming to train LabTec1 up so that more than one person was able to do this task. They had a reciprocal arrangement with a technician in the Cytogenics Lab to do the PBLs and plasma extractions when LabTec3 was off.

The lab technicians trained each other by observation and application. The learner observed the experienced technician carrying out a particular process, and the experienced technician talked the learner through what they were doing. Then they swapped places and the learner carried out the process while the experienced technician watched and provided prompts to keep the learner right. Once both were confident that the learner had understood the process and the procedures she was left to get on with doing the work by herself. Every

process and procedure had a Standard Operating Procedure (SOP) which the technicians kept by them for reference so that they could check and double check that what they were doing was correct at any step of a process. For example, LabTec3 was being trained up to do the genotyping, in the lab by LabTec1, but she also went on a course run by the equipment supplier to learn more about how to set up the programming. The training was shifting from craftsmanship to technology.

When I came to the lab, the technicians understood that I wanted to observe what they did and write about it. Whilst the idea of writing about the lab and its practices did not make much sense to them, they understood the notion of observing in order to learn. So, they undertook to teach me about the processes and procedures involved in the extraction of DNA from blood samples and the other processes that the lab carried out. They expected me to do more than observe, they expected me to learn about genetics and the work carried out in the lab. My role in the lab became that of a learner: as well as explaining what they were doing, they gave me books and printed information from the internet to read. I found myself raiding the library and doing 'home work' as well as observing in the lab. I wanted to understand the science they were showing me so that I could write about it, but I was also responding to the enthusiasm with which they were trying to communicate what they do.

However, my role as learner was restricted. I was not like one of the science students or new staff who would be trained in the techniques of handling and processing the samples. I was not permitted to touch. A clear distinction was

made between the scientists and myself as a non-scientist. There appeared to be an underlying assumption that lack of knowledge and scientific skills made me somehow unsuitable or even unsafe to handle samples and equipment. This contrasts with Rayna Rapp (1999) who was able to train as a technician in a lab in the US. Rapp (1999) was able to participate in the work of the lab because it employed 'unskilled labour' and trained people up to do the work on site. It was not possible to gain the 'kinesic knowledge' of the objects in this lab, as Rapp had done, and therefore my understanding of the objects continues to be that of a non-scientist, an interested one, but still a non-scientist.

The conversations I had with the lab techs during the working sessions were mostly about the process and procedures, with one of the lab techs explaining what she was doing, and me asking questions or clarifying that I had understood correctly. But we also had snatches of other conversations, fragments about their lives and interests, my life and interests. Often in the middle of a conversation, either about the work or some aspect of life in general, whoever I was with would say 'Don't ask me any questions for a minute' or 'I have to concentrate on this so don't talk to me till I have finished'. I would wait observing in silence until they indicated that we could speak again. The partial conversation would most often be lost to the explanation of the next step of the process I was observing at the time. We talked about taste in music, nights out, holidays past and planned, previous jobs that we had, but these conversations were always brief, unfinished, glimpses of their lives outside the lab. I felt in the end I had only partial impressions of lab techs.

They also asked me questions about other people and aspects of the Generation Scotland project which they never saw. Those conversations again remained often partial and unfinished, fragmented by mechanical tasks. I found that over the time I was in the lab, and for a while afterwards, I was bringing information into the lab about the wider activities of the GS project, about the people, who and where they were, what they were doing, and the issues that these others were addressing in their research. I also took stories from the lab back to the people I was working with elsewhere. I found myself in the middle of a two-way interest between the places and people with whom I was in contact, and it felt as if my movement between places, the lab, the different offices, and the research clinics was creating connections.

Locating the lab

The genetics lab was set up as part of a clinical research facility to provide a service to genetics researchers. As a service lab it differs from a research lab with ideas about the construction of facts (Latour 1987), and from a testing lab with the construction of facts for a specific purpose (Rapp 1999). It only worked on genetics: DNA extraction, Peripheral Blood Lymphocyte and Plasma isolation, genotyping, sequencing, sample storage, and polymerase chain reactions (PCRs), (http://www.wtcrf.ed.ac.uk/genetics/ 20.01.06). It did not offer genetic testing for individual patients, this was done by the Cytogenetics Lab located by the back gate of the hospital complex.

The lab was located within a complex network shown by Figure 5.1 on the next page.

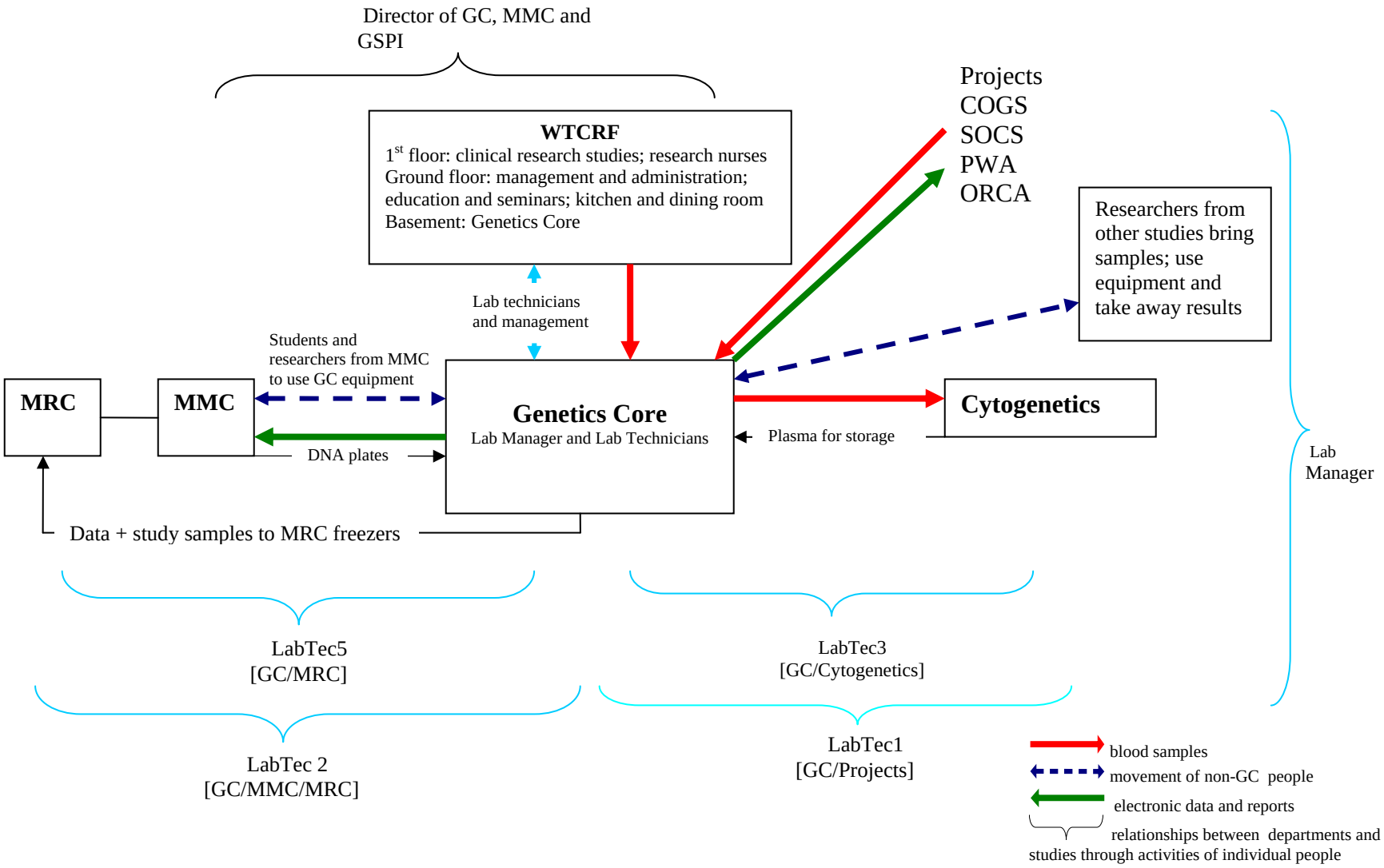


Figure 5.1 Genetics Core Network

The lab had connections with research projects, and other research centres across hospital and university departments. These connections were manifest in two ways, one was the movement and work of people and the other was through objects, for example, samples and pieces of equipment that were shared. For example, LabTec2 would go most days to the Medical Research Council (MRC) building and collect plates of DNA samples that had been prepared there and bring them back to the lab to work on them. The researchers in the MMC would complete a Loading Form, that LabTec2 left by their freezer, to tell her what they wanted done. Alternatively, students, occasionally accompanied by a lecturer, from the Molecular Medicine Centre (MMC) would come to the lab to use the equipment there for processing DNA samples. They were mostly doing small scale projects that did not have sufficient numbers or enough funding to pay for the lab's services, but were allowed to use the equipment and analysis machines if the lab techs were not running anything on them.

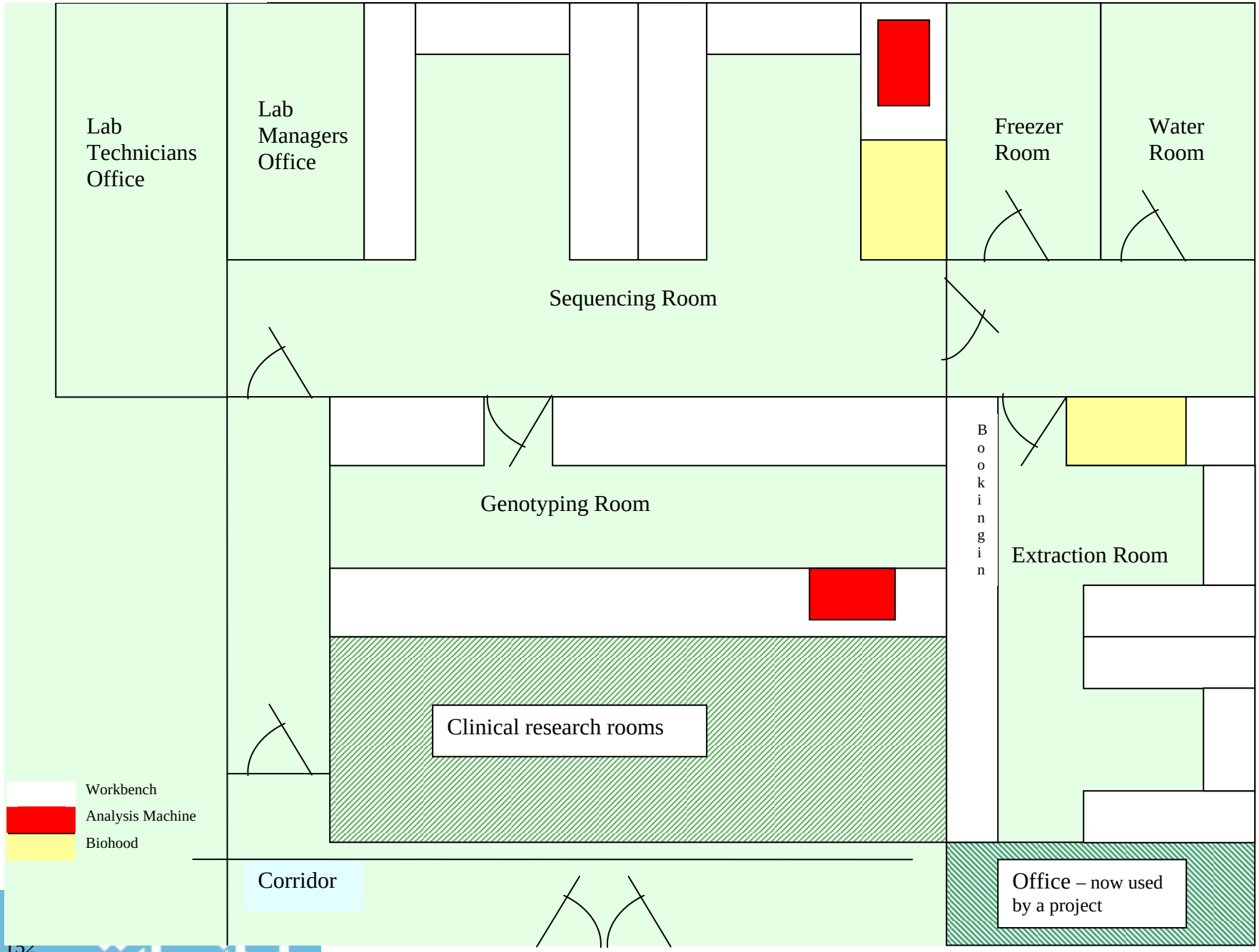
LabTec3 would take blood samples and go to do the PBLs and plasma extraction in the Cytogenetics lab. She used a bio-hood and centrifuge machine for the process, and a -80°C freezer or a liquid nitrogen tank for storage. All this equipment belonged to the genetics lab but, because there was no room for it, was situated in the Cytogentics lab on the basis that they shared the use. Sharing equipment and/or people is a common aspect of lab work (Latour and Woolgar 1986: 71).

The lab was located at a distance from the everyday work of the other people in the clinical research facility, the hospital and GS. The clinical research facility was reached from a main corridor of the hospital. The genetics lab was situated in the basement. The administrators and managers were on the ground floor, which included a reception area, interview rooms for research nurses and clinical researchers to see participants, a meeting room, the kitchen and dining room shared by the administrators and lab staff. The research nurses were on the top floor with examination rooms, clinical equipment and their own kitchen-dining room. To get to the lab from the reception area you had to go through two sets of swing doors, down a staircase, through two more sets of swing doors, along a corridor, through another swing door, another corridor and finally a security door to arrive in the lab.

The Layout

The layout of the lab was designed as a series of rooms for specific work, based on the progression of processes, a production line that the blood samples would pass through. The Extraction Room, the Sequencing Room, the Genotyping Room, and two smaller rooms, one the Freezer Room, and the other the Water Room (as shown in Figure 5.2). However, the needs of the people working in the lab had superimposed themselves on the lab design. The first space in the original design was an office, however, as the staff numbers had expanded the office had been moved to what had been a computer room for use at the end of the process at the far end of the lab areas. The original office had been reallocated to a project group and was no longer used by the genetics lab staff.

Figure 5.2 Lab Layout



Arriving into the lab through the security door, the lab technician's office was on the left, the lab manager's office straight ahead, and the Sequencing Room was on the right. Each of the areas was labelled according to the process being carried out, but the technicians had to walk through the Sequencing Room to get back to the beginning of the processes. The Extraction Room and Genotyping Room were also used by the occasional visitors, students or researchers from the MMC or MRC for preparation of samples and analysis. Walking through the lab I would encounter the technicians in reverse order of the processes. The Genotyping Room was separate from the other lab areas, so it was the one place where there were no people likely to be passing through, as it had its own door from the Sequencing Room and a sealed hatch for passing samples through from the Extraction Room to avoid contamination. The sense of a production line was created in two ways, firstly in the layout of the lab and secondly by the situating of the individual technicians at particular places within the rooms. Where each technician was located was determined by the particular process they were carrying out. On the workbenches they would have the equipment that they habitually used arranged to hand, according to their own preferences.

The lab technicians started and ended their day in the office with the computers and log books, ensuring that all records of work done on the samples were up to date in the data management files. They shared the use of the computers in similar ways to those described by Hine (2001), where once one Lab Tec has logged-in to a program if it was left running, others will then use the same computer and program without changing who was logged on. Each lab technician had their own personal log book. Log books were kept in the office

but each technician would take it with them into the lab when they began to work there.

Once a process was initiated the lab technician had to work on it until it was completed, thus they tended to time their work in blocks according to how long it would take to complete. At times, samples could be done to a particular stage and left on the workbench, but once chemicals were added, the process had to continue until the results were produced. The placing of the lab technicians around the working spaces means that they were usually working in isolation from each other, so little chat was possible for most of the day, except for the occasional exchange of a few words if someone was passing. Mostly, they worked in silence or with a background noise from equipment. There was a CD player in the Extraction Lab which was used to play music quietly sometimes. Indeed I found, once I had become familiar with this environment, a distinct appeal in the order, quiet and routine of the lab and its work.

Danger and Risk

The lab was not without moments of chaos or upset: like the morning we came in to find water all over the floor because one of the machines in the Water Room had flooded; or when one of the lab technicians discovered that they had a cut on their hand which was bleeding. On both occasions I was told to stay out of the way and not ask any questions. There was a distinct tension throughout the lab. I retreated to the office until the flood from the Water Room was cleared up and the day continued as usual. However, the cut hand was a different matter. I did not know what had happened, but the anxiety in the lab was

tangible. Once again I retreated to the office which remained empty until lunchtime. Later in the afternoon the lab technician told me that she had discovered blood on her hand while putting a tray of samples into a freezer. She didn't know how or when she had cut herself. There were evidently two problems. First, she was not sure if samples she had handled had created a risk to herself, and second, possible contamination to the samples. She had retraced her steps checking everything she had handled for blood smears, had finally concluded that it was a paper cut and that she had not contaminated either herself or any of the samples. The relief was again tangible. Blood in a sample tube was one thing, blood anywhere else sent the stress levels of the technicians rocketing up. I had never before seen such an explicit example of the notions of 'purity and danger' (Douglas 1966). Danger was in fact ever present in the lab in many forms, blood, different chemicals, and assorted equipment from 'sharps' to centrifuges.

There was a Risk Assessment for every chemical and every procedure in the lab. Every technician has to read and sign to say they have read the Risk Assessment sheets for products, such as hazardous substances and procedures for equipment use. For both activities and products a range of information was included in the risk assessment: the uses of product or equipment; risks to health; symptoms of exposure; storage and handling precautions; factors that increase risk; personal protective equipment; emergency action for spillage or fire; and first aid. Every product arrived in the lab with a Safety Data Sheet from the manufacturer. These sheets include information on: composition of the product; hazard identification; first aid measures; fire fighting measures if

applicable; handling and storage instructions; personal protective measures; and toxicological, ecological, and disposal measures.

The lab technicians each had a white lab coat for the Extraction and Sequencing Rooms, which they used most but not all of the time. The lab coat was only considered essential for 'blood work'. Usually the technicians worked without lab coats, or put them on for warmth, because it was cold in the lab.[1] There were green lab coats on hooks by the door for the Genotyping Room, however, I didn't see anyone use one of these while I was there. Visitors wore their own white lab coat in whichever room they were working. I was not asked to wear a lab coat while I was there. If the lab techs had been wearing their lab coats all the time this would have indicated that I would not be handling samples, chemicals or equipment. As it was at the time, it made me less conspicuous as a visitor rather than more so. I did not explore the significance of the lab coats, on or off, further as it appeared to be down to personal preference and at the time was not part of my 'blood' agenda.

By contrast they used protective gloves for everything that they did with the samples, often changing the gloves several times during a single process to avoid contamination. The gloves were purple nitrile and came in boxes, rather like a tissue box with a slot in the top for pulling them out, which were located on the workbenches around the lab. The technicians (like the nurses) did not like these gloves and found that they did not fit well, making it awkward to finely manipulate equipment. They preferred the latex gloves that had been in use

[1] The Lab Manager commented when he read this that he wished they had had their lab coats on, but conceded that I had to write what I had observed. They all wear lab coats now.

previously, as they fitted better, but nitrile is less permeable to chemicals like chloroform than latex and therefore considered safer. Chloroform was used in the extraction of DNA from blood samples.

Equipment and Chemicals

The equipment and chemicals were physical mechanisms of disconnection. Before describing the processes in the lab, some description of the chemicals and equipment (scientific technologies) commonly used in the processes might be useful. The use of all chemicals and the amounts were recorded by the lab techs for every process that was carried out. The machines all had SOPs and health and safety assessments.

The lab techs were also responsible for keeping the equipment clean. Deionised water was used throughout the lab for diluting chemicals or for cleaning equipment, LabTec2 told me that

> It is cleaner than tap water because it has had all the chemicals and bugs removed from it and the electrical charge has been reduced so it doesn't act as a conductor when it passes through the equipment – it is important because samples could be contaminated - and we would get inaccurate results if they were contaminated – and too much current affects the readout from the sequencing and genotyping equipment which are very sensitive.
>
> LabTec 2

The deionised water was produced in the lab in the Water Room, a small room at the end of the corridor just past the freezer room, which was filled with equipment for deionising the water, an autoclave which was a large machine for sterilizing equipment, also connecting pipes, containers and a sink. The

autoclave was pressurised and heated the water to steam. The technicians filled designated containers, which were kept to hand on the work benches, with quantities of ionised water as required.

Virkon solution was used throughout the lab to kill 'the bugs'. This disinfectant is used to clean equipment, surfaces, and in the disposal of the blood waste after DNA extraction. It was advertised as having low toxicity and being environmentally friendly. It was pink in colour in solution and was effective for approximately seven days, or until the pink colour faded. Virkon and deionised water were used to keep the 'dirt' from the world outside the lab at bay.

Reagents were used in processing the DNA. Reagents 'A' and 'B' were chemical mixtures that came as components of a kit for DNA extraction. The kit comprised several bottles of varying size which contained chemicals for adding at different points in the process. Reagent 'A' was a concentrate which had to be diluted with deionised water before it was used; other components of the kit were not diluted. The use of kits is important, as, like the machines, it points to the degree to which commercial suppliers determine what can be done in the lab, and how (see Kleinman 2003).

Pipettes were important for precise measuring of chemicals. The pipettes could be set by twisting part of the casing until the setting required showed the amount of chemical that the technician wished to measure into a tube. The size of the tips varied according to the quantity to be measured. The tips were set in racks and like everything else, were used sequentially from the rows in the

racks. The held part, or hand grip, of the pipette was pushed down into the tip to attach it; and there was a press release which allowed the used tip to drop off into a waste container. All the work benches and the biological cabinet had a tub or jar labelled for 'sharps' into which all the tips were dropped and then disposed of in the clinical waste bags located in the lab. The racks of tips came in assorted sizes and had lids for when they were not in use.

The pipette had a button on the top which was depressed against a spring. There were two pressure points, one about half way down and one when it was fully depressed. To collect a chemical in a pipette, the button on the top of the mechanism was depressed to the first pressure point, the tip was inserted into the chemical and then the button released to draw the chemical into the tip. The tip of the pipette was then transferred to the receiving tube, the button was depressed fully to the second pressure point, pushing all of the chemical into the tube.

The pipette work put a strain on the hands and wrists of the lab technicians, as it demanded repetitive and extremely controlled movements. Those that had worked in a lab for more than a short while had wrist straps for support when working with pipettes. One of the technicians was being treated by a physiotherapist for repetitive strain injury. There were a few electronic pipettes in the lab which required less physical effort but the same levels of control and concentration, however, most of the work with pipettes was done with the press and release ones.

Wheels were used for mixing, and centrifuges for separating. The wheels consisted of a flat disk about 20cm across with clips spaced around the outer edge. The tubes were fitted into the clips with the lid inwards and the whole rotated slowly driven by an electric motor. The centrifuge by contrast looked like a large metal box casing with a top opening lid. Inside the box was a drum and spinning mechanism with four equally spaced brackets that act as holders for four pots. The pots could be changed according to the size of the tubes you might want to spin; each pot held the same number of tubes for balance.

Booking-in

LabTec5 showed me the booking-in system, which she does most days when she comes over from the Medical Research Council (MRC) building. In her absence, the booking is done by LabTec1. LabTec5 was, as previously stated, the only person in the lab without a science degree, nor had she ever done any similar work before, but said that she really enjoyed it. This work did not require the handling of equipment or chemicals, which presumably rendered it feasible for a non-scientist with experience of working in other ways around the lab to do.

The blood samples arrived in the lab by post in padded envelopes, or were collected from 'upstairs' in the clinical research facility. These blood samples were not for GS projects but projects that were going on in 2005. LabTec5 collected the post when she came in and brought the days' intake down to the lab with her to book-in the samples. The blood that came through the post should have arrived in the lab within 48 hours of removal from the body. The samples were usually in the lab within this time but if they were delayed, for

example in Christmas post or over holidays then the lab requested replacement samples. Some envelopes contained one tube of blood, others had up to four tubes of blood from any individual participant, depending on the protocols of different studies. Each tube of blood was inside a bigger tube with a screw on lid. The tube of blood could have from 3-9mls of blood in it. The lids were red and orange: red lids went into the freezer after being booked-in, the orange lids were collected together and sent to the Cytogenetics lab in another building in the hospital complex.

A line of paper forms, one for each of the studies sending in blood samples for DNA extraction, were laid out along a workbench in the Extraction lab. This would have been located at the start of the processes in the original lab layout, and therefore a point which everyone passed on their way to different tasks. The tubes of blood were booked-in on the forms, and in LabTec5's log book. The details from the labels of each tube of blood were recorded: ID numbers, date of arrival and any other details pertaining to the blood sample that might be included by the study.

In theory all the blood and DNA samples would be anonymised by the time they reached the lab, in practice some of the studies were sending in blood samples that had personal details on the labels - the labels of one study included the name of the participant. I was surprised by this because as I understood it the blood samples would be kept separate from a participant's personal details, however, this does not appear to be the case with all studies, though it was unusual for the lab to receive this information. On the labels of another study

were the dates of birth of the participants, who all appeared to be quite elderly. When I commented on this, LabTec5 replied 'I know it's a shame, and they are sick - they are taking 25mls of blood from them and we don't need that much – we only use a little bit and the rest is put into an archive'.

A lab identifier number was allocated to every blood sample. Freezer Works was a computerised data management system which allocated the unique identifier to each blood sample that came into the lab. That identifier also designated the location in a rack or tray of samples for a given study, and the spaces where these were stored in the freezers. A strip of bar code labels was printed off so that there were labels for all tubes and aliquots of each blood sample and the DNA extracted from it. The strips of labels were fastened together in bundles and kept in a box in the office. Whenever a technician was working with the batch of samples, the bundle was taken from the office to the lab so that any new tubes could be labelled. LabTec1 told me

> Different studies want different things, some just want DNA extracted, but others want plasma and PBLs. One of the studies is sending eight tubes into us - three for DNA extraction, two for plasma extraction and storage, the two others are for PBLs, and they also get one spare – which is quite a lot from a person who is sick with cancer
>
> LabTec1

The occasions when a lab technician expressed an awareness of the participants in a study were rare. These were prompted by blood being connected to a piece of information, such as age or disease. Mainly, the lab technicains were concentrating on sorting the multiple tubes of 'a blood sample' so that they were

correctly labelled and placed. Two labels would indicate that the blood samples were to be stored; more than two would indicate that DNA was to be extracted, and possibly that there were also tubes of white blood cells or plasma to be stored. Every time something was done to the sample in the lab the identifiers were logged in the technician's log book. Thus every sample could be located and every process that it underwent could be traced.

Booking-in gave me a difficult experience as an observer. On one of the Fridays that I was observing, only LabTec4 was in the lab, the other technicians either being off or attending a symposium for the day. LabTec4 brought the post including the blood samples down to the lab, and left them on the work bench. She then proceeded to get on with the DNA extractions she was doing that morning. I asked about the blood samples that had just come in, being aware that if they were not booked-in they would not be put into the freezer before Monday by which time it would be too late for them to be viable. This could have meant contacting a researcher from the study and requesting replacement samples. I was strongly aware of the nuisance value this would have for both research nurses and especially participants, with the possibility that not all participants would agree to make another appointment and give another blood sample. I felt quite agitated particularly on behalf of the participants.

LabTec4 had been instructed to leave the blood samples as she had 'not been shown how to do the booking in', and was quite adamant that they should be left. I, on the other hand, at that stage had observed the process several times and felt that with my detailed notes for back up that I could do it. I did not

suggest that I should do it because I felt that this would be going beyond the agreement that I had with the lab technicians to observe, but I experienced a great deal of frustration, feeling that I could do or say something but that if I did, it would take me across the agreed lines by which I was allowed to be present in the lab. LabTec4 was just as aware as I was, probably even more so, of the implications of leaving the blood samples where they were but was equally unable to do anything about them, though for different reasons: she straightforwardly did not know what should be done to book them in and had been instructed not to worry about them. And so we left at lunch time, LabTec4 because she had finished for the day and I because there was no one left to observe. She told me on the following Monday morning that someone had come in late on Friday; whoever it was must have travelled across the city after the symposium had finished, and done the booking-in, because she had found all the samples in the freezer when she came in to work that morning. I never found out who it was that had come back to do the booking in as the technicians were all very busy that morning and later I forgot to ask.[2] The incident did serve to underline for me the difference between being a scientist and a non-scientist in this setting, I could observe but I could not participate. But it also underlined the restrictions placed upon the scientists themselves through their practices of training and hierarchy.

Extracting DNA

DNA extraction was done 'by hand' as the quantities and tubes are too large to use the robotic equipment, which was used later for genotyping very small

[2] When the new lab manager read this he set out to find out what had happened. Later he told me that it was LabTec1 who had come in late in the day, booked-in the samples and put them in the freezer.

quantities of DNA. This came as something of a surprise having spent months listening to talk and reading literature about the development of scientific technologies. LabTec4 was doing the DNA extractions, at the time I was in the lab, she had only been there for a week on a fixed term part-time contract. She was working mornings, doing the DNA extractions in order to complete the sample set for a study so that genotyping could begin. She told me that she was very happy to have got the job as it is difficult to find work as a lab technician; it takes an average of seven months to find a job after graduation. Although this was a temporary part-time job, it would allow her to gain experience in working in a lab which would strengthen any application for another post in the future. She also talked about her hopes to visit relatives and travel around Australia. As it turned out she booked a ticket to go for six months at the end of her contract and then a couple of weeks later was offered a full time post in the lab as they were increasing the staffing level, but she turned it down having already booked her ticket.

When I first observed LabTec4 doing the DNA extraction she was working with the SOP beside her all the time, checking that she was following the process correctly. After a couple of weeks she told me that she 'knew it by heart'. I could see a difference in the way she worked, more confident in what she was doing, assured in how she moved and handled the equipment. However, she confessed to still being worried about making a mistake, especially that she might drop something on the floor, she laughingly told me that she had nightmares of herself trying to scoop a tiny blob of DNA up off the floor. Then a few days later she told me with a mixture of anxiety and relief that she had made a mistake:

while transferring blood from the tubes they arrived in, into the Greiner tubes for processing, she had poured part of a blood sample into a tube that already contained one. The mixed blood from two samples could not be used. But, as there was enough of the second sample left for processing only the first sample had been lost. She felt bad about the mistake but the other lab technicians reassured her that even with the best will in the world and taking meticulous care, mistakes do happen. That is why they have back up systems, records and samples for every step of every process.

Preparing for DNA extraction

The blood samples were removed from the freezer, fourteen at a time, and placed into a rack to defrost. Fourteen seemed like an odd number to me, especially as this left empty spaces in the racks the tubes stand in, but none of the technicians could say why they worked with this number. One speculated about the number of tubes that can be loaded into the centrifuge machine, but that did not match; one suggested that it was to do with timing, that the rate of the reactions of the chemicals was such that once reagents had been added to the last tube by hand, the first tube was ready for the next chemical to be introduced; and another suggested that it was to do with the optimum number that a technician could concentrate on at any one time.

While the blood samples were defrosting, LabTech 4 set up a new rack with the same number of tubes all numbered, by hand with a marker pen, to match the identification numbers on the tubes being defrosted. The identification of the tubes being defrosted were also written into her lab log book with the date,

what she has done and any comments about the process. Any problems with a sample was logged and also reported.

Once the blood samples were defrosted the process could begin. The rack of blood samples was moved into the biological safety cabinet. The defrosted blood was a darker colour than fresh blood, it did indeed look dirty by comparison. All 'blood work' was carried out in a safety cabinet in case of spills, leaks or splashing. The safety cabinet was a rectangular cabinet set on the workbench, so the working area of the cabinet was at workbench height. The workbenches were of a height where the technicians could stand to work or they could sit on a high stool. The front of the cabinet was glazed from the top down but left a space at the bottom so that the technicians could get their hands and arms, up to the elbow when bent, inside the cabinet with enough room to move freely to work with the tubes and containers, but their faces would be protected by the glass, and they had to look through this to see what they were doing.

Making a backup

The first thing that the technician did was put a spot of blood, approximately 100µl (microlitres) from each sample onto a Whatman FTA card. The card has a chemically-treated fiber matrix which protects DNA from degradation. The blood spot looked just a bit larger than if you pricked your finger with a needle and then dabbed it with a tissue, about 5-6mm in diameter. The Whatman card was about the same size, and texture, as a plain postcard folded in half. One card took four spots of blood and each spot was numbered sequentially with the identifier of the tube. Some of the blood spots bubbled up and produced a clear

crystal-like bubble on the blood which crunched to a gritty dust when the technician closed the protective flap on the card. They looked quite strange, peaked rather than round, and could vary in size from tiny and barely visible, to covering almost all of the blood spot and sticking up 2-3mm. Not all the spots did this and the technician I was observing could not offer an explanation for why this happened, nor could any of the others that I asked.

Once the blood spot had dried, the cards were packed into plastic bags according to study. The Whatman cards constituted a long term archive and if it became necessary DNA could be extracted from the blood spot at some point in the future. The bags of cards were stored on a shelf in the lab until the study was completed and then they would be moved to a filing cabinet. Since I was in the lab, the new lab manager has had a fire proof safe installed. All the Whatman cards are now stored in the safe in case the samples in the freezers are lost through a fire or another accident.

At this stage there was the rack with the defrosted blood samples all in numbered sequence; a set of Whatman cards laid out and numbered in exactly the same sequence; and a rack of empty tubes also numbered sequentially exactly the same as the ones containing the blood samples. Everything was sequential, duplicated, and recorded in exactly the same order in the set of fourteen blood samples.

The process of extracting DNA

Once the Whatman cards had been completed, the remainder of each blood sample was transferred into a 50ml Greiner tube which had a flip-open and press-close lid. As stated above, all tubes had been labelled with the identifier for each blood sample. The sample tubes with 9ml of blood transfer one for one, but some of the samples came as two tubes of approximately 4ml and both the tubes of an individual sample were transferred into a single 50ml tube. Reagent 'A', a chemical mixture from the extraction 'kit', was then added to each tube, containing approximately 40mls of liquid, the tubes were press-closed and then placed in their numbered sequence on a rotating wheel for 4 minutes to mix the reagent and the blood sample thoroughly. The lab technician recorded all the quantities of chemicals and kits that she used.

The tubes were transferred to the centrifuge machine where they were spun at 2600rpm for 5 minutes. The tubes were fitted sequentially into appropriately sized wells in one of four pots in the centrifuge machine, a total of sixteen. The centrifuge must be balanced so there were two tubes with water in them to fill the blank spaces left by having fourteen samples. The tubes of blood were removed from the centrifuge and returned to the rack in sequential order. The racks for this size of tube were designed to hold twenty tubes in two rows so there were always six blank spaces, which were left as three pairs opposite each other at one end.

The centrifuge caused the contents of the tubes to separate. The top layer is called the supernatant, a clearer liquid in which the lighter molecules were

suspended; the clearly separate thicker sedimentary layer of the heavier molecules below is referred to as the pellet. The pellet was a bit less than half a teaspoon in quantity. The supernatant was discarded by carefully pouring it into a beaker of Virkon solution, while the pellet was retained in the tube. The supernatant was left in the Virkon for 24 hours to kill bacteria and viruses before being disposed of down the sink drain. Once the supernatant was removed each tube was carefully inverted onto layers of paper roll to remove any dregs.

The next step in the process was to suspend the substance in the pellet in Reagent 'B'. 2mls of liquid was pipetted into each of the tubes, then, one by one, the tubes were placed in a small vortex machine to spin for a few moments to mix the reagent and the pellet together. The technician checked each tube to make sure the pellet had dissolved in the reagent. The dissolved pellet was then transferred into a 15ml tube set out in another rack, all tubes again being numbered by marker pen in sequence to match the larger ones.

The next phase in the process involved the adding of three different chemicals to the tubes and mixing them in. The pipetting had to be done carefully; the same pipette tip could be used to add the one chemical to each tube, as long as it did not touch the sides of the tube. As soon as that happened, the tip had to be changed for a fresh one to avoid contamination of the samples. First, 500µl of sodium perchlorate were pipetted into each tube, then the tube was inverted by hand seven times to mix it in. Second, 2ml of chloroform, likewise pipetted and again each of the tubes inverted by hand seven times. Thirdly, 300µl Nucleon

resin was pipetted into each tube. The tubes were placed in the centrifuge, the pots having been changed to fit the smaller tubes, for 3 minutes at 2600rpm.

The tubes were removed from the centrifuge and placed back into a rack in their numbered order. The contents of the tubes had separated out into three layers with a supernatant and pellet as before, and a dark coloured layer between, however, this time the supernatant was kept and the pellet discarded. The supernatant, now a colourless liquid, was transferred into fresh and numbered 15ml tubes that were set out in another rack. Then the lab technician brought a bottle of ethanol from one of the freezers where it was stored at -20°C, and poured a quantity into each of the tubes in turn, inverting each one several times. When she held one up for me, I saw tiny white strands, hanging like threads through the liquid - 'there's the DNA' she told me. It was an extraordinary experience, the first time this happened - I had no idea that I would be able to see DNA with the naked eye, and looking at the DNA of another human being, the genetic information of an unknown but very real person somewhere, left me in awe. The floating whiteness created a sort of ethereal sense that does not come with seeing diagrams or even photographs of this 'stuff of life'.

The floating white strands of DNA do not remain so, within a few minutes they had 'rolled up' to form a sort of cotton wool like blob, some were quite 'fluffy' whilst others seemed smaller and more solid – maybe less aesthetically impressive but no less fascinating. The next step in the process was to catch the blob and transfer it into a small 1.5ml Eppendorf tube. The blob was hooked out

using the tip of a glass pastette and deposited into the small tube, using a new pastette for each sample, with all tubes and their lids previously numbered and set out in sequence in a rack. The Eppendorf tubes were a sort of double tube, with a conical ended tube set inside a cylindrical tube. These tubes had screw on lids rather than the flip lids of the tubes used during the process. Once all the blobs had been transferred, 1ml of buffer was pipetted into each tube, the DNA blob dispersed out through buffer creating an aliquot of the sample. The lids were screwed onto the tubes and they were then placed sequentially in a tray on the work bench to await their turn on one of the rotating wheels. The tubes were placed on the rotating wheels for between one and two weeks to disperse the DNA evenly through the buffer before being placed in their allocated rack in one of the freezers for storage until such time as they would be used for genotyping.

Once the process was complete all the used tubes were disposed of in clinical waste bags, the surfaces wiped down with paper towels and Virkon solution. The sharps containers were disposed of separately when full. Each sample had required 5 different sterile tubes, pastettes, pipette tips and an Eppendorf tube as well as the use of different equipment, a safety cabinet, centrifuge, rotating wheel, and vortex machine. The whole process from start to finish took approximately two hours.

In another time and place it could have been called magic, the sudden appearance of something unexpected, but in this scientific setting, knowing that all of it could be explained in precise scientific terms made the notion of magic

too fanciful. When I later recounted the process to colleagues in the law school, one of them teased me about having a spiritual experience, and I had to concede that I had felt something like that, in that it was profound to see something that was so uniquely personal and yet universal. However, I did not experience the same urge as Angier to grab it and 'shape it into little animals' (Angier 1988: 35). In some ways it felt like an invasion of privacy to look at someone else's DNA, to see the basis of that person's physical make up. At the same time it was looking at something that is shared with the rest of the living world. It is not something that most of us do everyday. I also wondered if I would have felt differently had it been a colour other than white, but concluded probably not. Interestingly, most of the visual representation of DNA available in the media and scientific publications show a multicoloured structure, each type of nucleotide (ATGC) being coloured to identify it, a picture of strands of white stuff suspended in liquid would convey a different type of information, create a different type of representation.

I think that much of the effect on my imagination was the consequence of seeing this fragile stuff in contrast to the power it exerts within the rhetoric of genetics in politics and policy, as well as personal lives. I could also see why the lab technicians talked about 'nice clean DNA', the white stuff suspended in a clear liquid set in contrast to the muddy brownish blood from which it had been extracted with all cellular materials, viruses and bacteria removed. For me, seeing the DNA was something remarkable, for LabTec4 it was a job done. I was surprised, whereas she was pleased. She had carried out the process and the results were correct, that is, at the end, she had extracted DNA from each of the

blood samples. When she repeated this process on a set of 14 samples everyday it became mundane and the importance lay in the success of the results, in producing a 'clean sample', in not making a mistake.

On two different occasions when I talked to scientists about DNA extraction they responded that 'you can do it at home in your kitchen with an onion', for them it was not remarkable but as ordinary as preparing a sandwich at home in their own kitchen. I thought that although DNA extracted from an onion would probably look pretty much the same as human DNA in the lab, and equally remarkable in its own way, the effect of the lab and the fact that it was human DNA from blood had a more powerful effect on my imagination and emotions. I cannot shake off the awareness of a person or persons and the social world when I think about DNA. For me it is something more than a string of molecules. Thus I found myself working with two very different perceptions of DNA, of the ordinary and the extraordinary juxtaposed, the former orientated toward a future research analysis and results, and the latter embedded in an embodied past.

Lunchtime

Lunchtime revealed the lab techs switching-off from their lab personae and reclaiming another facet of themselves, at a distance from the tasks, a reminder of other dimensions of their existence outside the lab. Having been surrounded by the practices, equipment and objects of science in the lab, lunch felt like returning to the social world. LabTec4 did the DNA extractions in the morning and then left at lunch time, the other lab technicians had lunch upstairs in the

dining room usually between 1pm and 2pm. Most people brought their lunch in with them; there was a hospital canteen nearby though the staff from the clinical research centre rarely used it, and a shop in the hospital which sold sandwiches and snacks, which they did use if they had not brought anything in with them. As well as the lab technicians the administration staff, project staff and occasionally one or more of the managers from the centre all met up in the dining room at lunch time. The nurses had their lunch up on the first floor in their own kitchen-dining area. The chat over lunch time tended to be light-hearted and included the finer points of cold pasta, trying to catch Edinburgh buses, football and motor racing. One lunch time there was a discussion about running marathons, several of the staff had competed in the Great Northern Run in the past to raise money for charity, one was planning to do it again and was looking for sponsors. On other occasions people talked about plans for holidays, or trips away. One of the administration staff was getting married and there were discussions about the plans for the hen party and the wedding. After lunch the lab technicians would work on another set of samples.

Conclusion

The practices and processes of the lab facilitate the manipulations of the blood, by creating spatial and conceptual distance between the body of the participant and the use to which the blood is put. The object 'DNA sample' has come into being, and the object 'blood sample' has disappeared. Yet many of the processes that samples undergo during the application of scientific methods were described variously by lab technicians and clinicians as mundane, routine or even boring, certainly not interesting.

The descriptions of processes in the lab that I have given become progressively disembodied as there are fewer people and more equipment than those described in the previous chapter; and the place took on a greater significance in the enactment of the object. By focusing on the object, the lab technicians became part of the practice, fading into the space and integrating with the equipment, in part through the slant of the description, and in part because they chose to.

The laboratory was a particular type of space and place, remote and unfamiliar to those who do not work within these spaces. The work, often routine and repetitive, enacted transformations: producing disconnections through separation and inscription, and connections through ordering and collation.

Chapter 6

'This strange mania for inscription': putting DNA to work.

This chapter continues in the lab and shows how DNA was put to work in the construction of a genetic database through practices of quality control, polymerase chain reaction, sequencing and genotyping. These processes aggregated individual samples into sets of samples. I was looking for the transformation from substance to digital data, and there was certainly a lot of work by the lab technicians in the preparation, but then the DNA substance was 'denatured' and disappeared into 'inscription devices'. The sequences of letters and graphical representations that came out of the devices were difficult to understand. Including definitions of the objects being transcribed was of some help, but both scientists and the public need ways to think about these objects. What I wanted to do with this chapter was to describe the objects in the lab, gene and DNA, and then map the social meaning onto these scientific objects. What I show is a juxtaposition of the both scientific and the social gene and DNA. There is another type of 'inscription device' that is commonly used to try to understand unfamiliar objects, a linguistic one - the metaphor. In the final part of the chapter, the scientific inscription devices are juxtaposed with the linguistic inscription devices showing a disconnection between the scientific and the social meanings. This shows the struggle for language to describe and understand the science. The relationship between these scientific objects and the public (and indeed scientists) is mediated and connected by metaphor. One of the problems is that it becomes increasingly difficult to know whether it is objects or concepts that are under discussion.

The DNA in this lab was one version of many possible versions of the object DNA. In other labs in other places DNA, locally and globally, is utilised for different purposes including cloning, transgenics, reproductive technology, stem cells, embryology, and immunology to name a few. Annemarie Mol argues that different versions of an object are enacted through practice, and that 'far from necessarily falling into fragments, multiple objects tend to hang together somehow' (2002:5) because although the objects differ from one practice to another there are relations between these practices. In the case of DNA, I would argue to the contrary: while it starts as a single object, the different versions of DNA do not hang together; practices force different versions of DNA apart, creating disconnections rather than connections between them. To facilitate different types of work utilising DNA, the practices are enacted by different people, in different places and using different equipment - they must be disconnected from each other to make different manipulations permissible.

The description of the work done in the lab and the processes carried out was based on lab work as it was happening at the time I was there. The blood, DNA and electronic data that I observed all belonged to different studies that were on-going. These were, however, the same processes that samples from the Generation Scotland collection projects, the Scottish Family Health Study and 21CGH, undergo. The scientific methods were standardised, routinised and repeatable so that, in describing what happened to a set of samples from one study, you can in effect describe what will happen to samples from another study that are sent to the lab for DNA extraction and analysis. The handling and

processing will be similar, any project variation will lie in specifications, for example quantities of blood samples or choice of assays.

People

DNA, or rather the inscription of DNA, increasingly took centre stage in the lab and the lab technicians faded further into the background. The work of the lab technicians, PCR, genotyping and sequencing the samples of DNA, that had previously been extracted from the blood samples, moved progressively from the hands of the lab technicians into complex pieces of computer controlled electronic and robotic equipment. It seemed at times as if the lab technicians almost melted into the lab environment and became components of the processes as their 'craftsmanship' was taken over and immersed in the technology (Rabinow 1996:117). 'Inscription devices' of increasing size, carried out the processes of inscription and transferred the resultant readouts to computer programs (Latour and Woolgar 1979). The plates of samples grew smaller, were placed into racks, and disappeared into a machine while the process was carried out, then re-emerged at the end to be transferred to another machine or as waste to be disposed of. It became harder to think about the processes in terms of some type of social interaction, of relationships that were constructed between the samples, the machines and the lab technicians. The lab technicians each had a different role in the processing of the samples, as described in the previous chapter, and there was some training going on so that the technicians could provide cover for each other. As I moved from process to process, I spent time with each of the lab technicians in turn.

Place

The PCR and genotyping were carried out in the Genotyping Room, separate and closed off from the other lab spaces, it was accessed through a door from the sequencing area. The space consisted of a rectangular room with no windows, just the door and a hatch for passing samples through from the extraction lab, the workbenches mostly filled with different large 'black boxes'. The larger machines ran robotic mechanisms and had cylinders of gas to drive the hydraulics underneath the benches. By contrast, the sequencing was done in the lab space next to the office. This meant that people were passing by all the time, but the area itself was set to one side and partially screened by shelves and one of the biological hoods. There was a large machine that did the sequencing and a computer on one workbench; on the opposite workbench was a collection of plates, pipettes, tips, bottles and paperwork. But, before these processes could begin, the samples had to go back to the extraction lab, for quality control.

Quality Control

Quality control procedures were applied to all the DNA samples prepared in the lab and were often carried out on samples that were sent to the lab for analysis. 'DNA can sometimes be grubby or tatty' (LabTec1). The DNA could be grubby if the sample had not been extracted in accordance with procedures to achieve a clean sample. It could also be tatty because it had been left lying around or it had been frozen for a long time; in which case the strands can fragment. DNA that had been prepared for analysis other than in this genetics lab and had been sent for genotyping or sequencing was sometimes grubby or tatty in which case 'it may not work so well and when you get the readouts they

may not look great' (LabTec1). The high quality of the DNA samples was a point of pride with all the technicians in this lab.

All DNA samples that had been extracted in the lab were checked for purity and concentration after two weeks of slowly spinning on the rotating wheels. The two weeks on the rotating wheels separated the 'blob' of DNA out into the buffer solution so that there was an even distribution of the DNA through the solution. In a very literal sense, the DNA curls and twists together and had to be forced apart using chemicals and mechanics. Each study specified which type of quality control was to be carried out. There were three methods practised in the lab, using a spectrometer, agarose gel or PCR, each of which checked the concentration and purity of DNA in a sample. The different methods demonstrated how, over a period of three years, the technology had advanced, become more cost effective, and changed the way in which processes were carried out in the lab. The spectrometer method required the samples to be checked one at a time by hand which was laborious and time consuming. The agarose gel could be used to check batches of up to approximately 50 samples at a time, but was also time consuming to prepare, as it involved setting up the samples by hand and photographing the results. By contrast, the PRC machine would take up to four plates of 98 samples at a time, was less time consuming, measured concentrations more accurately and supplied a read out of spectrometer results. Different studies required different 'strengths' of sample from 50 to 250 parts per microlitre. Genotyping could be done from very weak solutions, but most studies aimed for a stronger solution as this reduced the

need to PCR the samples. Any samples not meeting the required levels of concentration were retuned to the wheel for further spinning.

Storage and management of samples

The samples that had passed the quality control for a study were made up into aggregate sets of master stock for storage and working stock for use in analysis. This involved first doing a calculation for each sample based on a spectrometer read out. This showed the amount of buffer that needed to be added to each sample tube to create a master stock of samples all of the same concentration. When the calculations were done, LabTec1 pipetted the required amount of buffer into each sample. Then 200 µl of each sample was pipetted into the wells of a storage plate, as before in sequence and labelled with the sample identifiers and the details of the study. This plate was sealed and became the working stock. The remainder of the samples were retained as aliquots, in small screw capped tubes, as the master stock. The working stock and the master stock were stored in different freezers. Enough working stock had been separated for it to be used several times. Once all working stock has been used, a new plate can be made up from the original master stock. Some studies also split the master stock and store it in two different freezers, so there is always a back up supply of the samples should one of the freezers break down, or the samples be contaminated in some way.

The quality control results and stock details were recorded in the Freezer Works computer programme. More recently, the computer system has been upgraded to the Laboratory Information Management System (LIMS). This may not seem

a very interesting fact but, at the time I was there, the lab was processing samples for ten studies, with up to three thousand samples for some of them. Since then, the collection of blood samples for GS genetic databases has started with the Scottish Family Health Study, which has a target of fifty thousand; as well as the 21CGH collection of two thousand five hundred and the 3D project of six thousand samples. The lab is archiving and managing tens of thousands of aliquots of master stocks, and thousands of plates of working stocks for multiple studies all stacked on shelves in the freezers. A powerful system like LIMS is needed to know where the samples are. On my first day in the lab, the Acting Lab Manager showed me the freezers and said 'you could store DNA samples for the entire Scottish population in here' - that would be over five million samples.

Polymerase Chain Reaction (PCR)

PCR is an important process because it means that small quantities of a DNA sample can be used for genotyping and sequencing, so the sample can be used many times before it is exhausted. PCR has made it possible to collect a set of samples which can be used for potentially many future studies. The invention of PCR had an immense effect on the possibilities for processing DNA by providing unlimited amounts of genetic material 'facilitating the development of the practices and processes used in genotyping and sequencing' (Rabinow 1996:3). In the lab, I observed PCR used routinely for quality control and in preparation for genotyping or sequencing.

PCR was used to increase the concentration of the DNA in a sample for genotyping, that is, it made more DNA by replicating the strands that were to be genotyped. There were two aspects to the process of PCR, heating and replication. The PCR was carried out in machines on the workbench in the Genotyping Room. The plate of samples was placed into the PCR machine, which looked much like a large black toastie maker, then the lab technician selected the programme settings and switched it on.

To carry out PCR, the plate of samples had to be heated three times to three different temperatures, first 94°C, then 50°C and finally 64°C. Heating the DNA denatures it. Denaturing separates the DNA into single strands and straightens it out so that it comes out evenly when analysed. Assays were used to provide the necessary molecules for the replication.

The assays used in the replication of DNA during PCR and for genotyping contained a primer, a probe and restriction enzymes. The restriction enzymes recognised a specific nucleotide sequence and cut the segment of DNA that was of interest for analysis. The primer was made of single strand DNA or RNA with a known nucleotide sequence which would identify that segment of DNA in the sample. The primer would have been synthesised, copied into large amounts, and when placed in contact with denatured human DNA, it binds to the single strands of DNA that contain the corresponding nucleotide sequence, thus replicating the segment required for analysis. The probe carries a fluorescent tag to illuminate the location of the strand of DNA to which it is bound (see Carey 2005:111). In effect an assay selects and then 'unzips' a

segment of the strand of DNA, allowing molecules from the assay to attach to each side of the unzipped segment thus producing two identical segments where there had been only one before. This process was repeated in PCR so that the quantity of DNA in each sample was much more concentrated; this meant that there was a lot more of that segment of DNA to analyse. The machine carried out approximately 40 cycles and with each cycle completed the amount of DNA had been doubled.

There were two types of assay, assay 'by demand' and assay 'by design'. These were purchased from a laboratory supplier to the requirements of a particular study. Assays 'by demand' were tried and tested, known to select particular locations on a strand of DNA, an off the shelf assay from a lab supplier. Assays 'by design' had to be developed, by the lab supplier. A researcher could identify a location on a strand of DNA, a SNP, they wanted to genotype and then a request would be sent to the assay suppliers who would develop the assay and send it to the lab. But assays by design were not considered very reliable as they had not been used repeatedly, tried and tested. Because assays by design were less reliable than assays by demand the practice was to use the larger amount of 5µl working stock in the plates, rather than the usual 2µl quantity run with an assay by demand, to ensure there was enough of the right DNA segment to run the analysis.

Denaturing DNA

I find denaturing an evocative term. On the one hand it describes part of the process of PCR applied as a specific technical term in the lab, on the other it

seems to symbolise the argument that the concept of nature is changing under the onslaught of science and technology.

> Biotechnology's hallmark, it could be said, lies in its potential to get away from nature, to construct artificial conditions in which specific variables can be known in such a way that they can be manipulated. This knowledge then forms the basis for remaking nature according to our norms.
>
> Rabinow 1996:20

In Rabinow's account of the Cetus Corporation and the development of *Making PCR*, the process was described thus

> rather than trying to harness a biological process as part of a larger project Mullis decontextualised the process and by doing so conceived of a way to turn a biological process (polymerisation) into a machine
>
> Rabinow 1996:9

Denaturing may be seen as a crucial step in a process that then allows subsequent steps of manipulation, both physically and conceptually. Without denaturing the DNA molecule, the rest of the PCR process cannot be carried out, and therefore it would not be possible to use the samples repeatedly for genotyping and sequencing into digital data. Without denaturing the DNA, it would not be possible to think about it as a molecular structure rather than an embodied substance.

Denaturing becomes not just the separation of the strands of the DNA molecule, but the disconnection of an individual sample from its social life and its incorporation into the biotechnological 'machine'. Taken literally, denaturing

DNA represents 'promises' of better health and more effective drug therapies (MacIntyre 1997:1095); but taken metaphorically, denatured DNA stands for the 'concerns' and fears that surround biotechnology (MacIntyre 1997:1096). Denaturing DNA is not a complex or difficult step in itself; its significance lies in what becomes possible thereafter.

Sequencing

Before describing the practices, I think it is helpful to define the objects and what the scientific terms that are applied to them mean. The practice of sequencing was used to find and identify genes, alleles or polymorphisms. Genes, alleles and single nucleotide polymorphisms are the names of the risks [of disease] within the body. Risk of disease will be discussed in chapter nine.

A *gene* is 'a part of the DNA molecule of a chromosome which encodes for a protein' (PGH Foundation 04.05.06), or put another way, 'A gene is usually defined as a section of DNA that contains the blueprint for a polypeptide chain. The term locus is a synonym for gene; it carries with it the implication that a gene has a fixed location on a chromosome' (Carey 2003:68). The human genome is made up of approximately 30 billion base pairs of nucleotides, 90% of these have been considered 'junk', which leaves the remaining 10% to contain an estimated 25-30,000 genes (estimates have gone as high as 100,000 but these are considered over-estimates). Genes may vary in size from about 10,000 base pairs up to around 2 million. They are 'not simply spatial units in the sense of a continuous sequence of base pairs, they are regions of DNA made up of spans called 'exons' interspersed by regions called 'introns'' (Rabinow 1992:237).

A *Single Nucleotide Polymorphism* (SNP) is 'any polymorphic variation at a single nucleotide' (Strachan and Read 2004:641), or 'A DNA sequence variation that involves a change in a single nucleotide' (PHG Foundation 04.05.06). SNPs are useful because they are easily identified in a set of samples when they are sequenced. 'Humans differ by one nucleotide per every thousand or so nucleotides' (Carey 2003:118) which makes them useful in genotyping and searching for polymorphisms. Indeed the identification of SNPs was regarded as so useful that the Human Genome Project set up a group to work just on SNP identification. Locating and identifying a SNP helps to divide up the human genome, making it possible to 'mark' off sections which narrows down the search for a specific polymorphism on any given strand of DNA.

Alleles are 'variant forms of the same gene' (PGH Foundation 04.05.06), or it 'may be defined as a 'spelling variation' at a gene (ie a difference in the position and ordering of the A, T, C. and G along that stretch of DNA)' (Carey 2003:68).

Preparation for sequencing

LabTec2 showed me the sequencing. She was the most experienced lab technician, the only one working on sequencing, and had been with the lab since it opened. There were plans to train LabTec1 to carry out the process so that sequencing could continue if LabTec2 was away.

Most of the work of the lab technician was in the preparation for sequencing. The process itself was carried out within a large machine with a computer beside it. Echoing Latour and Woolgar's 'strange mania for inscription'(Latour

and Woolgar 1979:48) first, a Loading Form had to be completed with the plate name, identification, date and operator, the sample IDs were written into a grid corresponding to the location of a sample in the plate. The same details were also written onto the plate and copied in the lab technicians log book. The details of the Loading Form were then typed into the computer connected to the sequencing machine.

Once the form was completed the samples were transferred to corresponding wells on a plate. 'Big Dye' was added and then the plate was centrifuged to make sure the contents were all at the bottom of the wells. 'Big Dye' contained four different fluorescent tags, one that would 'stick' to each of the different types of nucleotide, A, T, C, or G. This meant each nucleotide was distinctively 'marked' so that when the molecules passed through the laser beam in the sequencing machine they could be recorded into a computer programme. The transfer of samples and the addition of the dye were done by hand using a pipette. A plate of samples could be transferred by rows from a working stock plate to an analysis plate using an eight tip pipette. The quantities used for sequencing were so small that it was difficult to see if a sample was at the bottom of a well, so the plates were routinely centrifuged. The plate was then heated to denature the DNA. Once the heating had been completed the plate could sit in a tray of ice until ready to be put into the robotic machine that did the sequencing.

The sequencing work I observed was of samples that had been sent to the lab by researchers, mostly postgraduate students, from other labs. Some people sent

plates with prepared samples to the lab for sequencing, others sent DNA that needed to be prepared by adding the 'Big Dye'. One set of samples sent in for sequencing had too much dye to produce good results, and needed some work as the fluorescence was too high and would have made the results uncertain. The sample and dye needed to be the right dilution for 'good results'. If the sample was too concentrated, LabTec2 added deionised water to dilute the sample, as too much dye 'pulls' the molecules so that the results would have been 'messy', making them blurry and unclear to read. The quality of the samples could also vary and LabTec2 told me that she spent time on preparing, checking and modifying concentrations or adding the Big Dye. Checking samples was time consuming and could become a problem if the number of samples from paying studies increased, she would not have much time to check and correct the quality of samples coming in from students.

The Sequencing Process

The sequencing machine was an inscription device, a large 'black box' with a window at the front so that you could see what was inside. Not that that was particularly informative about the actual process, but it was possible to see plates, racks, hair thin tubes, and pipette tips being moved around using robotics. The machine took plates which were stacked into a rack at one side of the machine prior to turning it on. Each plate slotted into a tray and had a lid fitted over the top with a hole above each well in the plate. Once the plates were inside the machine, they were moved by robotic mechanisms run by a computer programme.

Inside the machine the set of extremely fine tubes with very fine tips were inserted into the samples in the plates row by row. The molecules in the sample were drawn by a weak electrical current through the tube containing a gel and passed through a laser beam which read off the sequence of molecules as they pass. The Big Dye attached to each of the different nucleotides so that each one showed up clearly as G, T, A or C. The sequence of the nucleotides was recorded into the computer file set up with the sample IDs. It was possible to watch the sequence appear on the array viewer on the computer screen as it was recorded by the data collection programme.

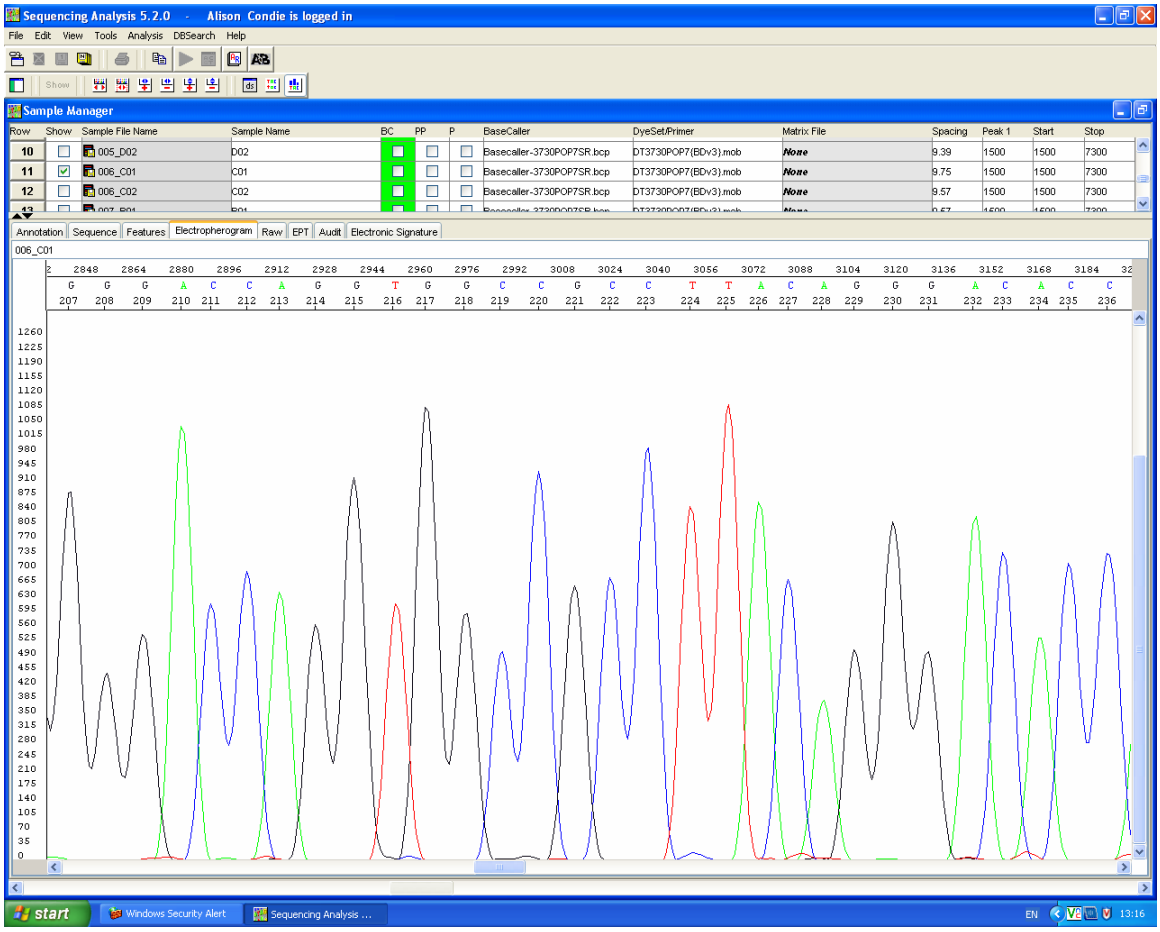


Figure 6.1 A Sequencing Trace

The sequence appeared on the computer screen as 48 multicoloured bars, or a four line trace, where every G, T, A or C molecule showed up as a different colour bar or peak in a sample sequence. Each sequenced plate produced three files of results on the computer, a chromogram, a text file and a graph, as shown in Figure 6.1. The files came out at around 200kb per sample so a full plate produced about 18.5Mb of data which was then sent to the researcher. The data collection programme also recorded data on the sample quality and showed how many of each G, T, A or C molecules it had recorded for each sample, the ideal is somewhere between 200 and 2000 parts per microlitre. Less than 200 meant there was not enough DNA in the sample, while over 2000 indicated that the concentration was too high. Once the sequencing was completed the plates were removed from the machine, stacked beside it, and kept for several weeks in case they were needed again. The quantity of sample in the plate did not change during the process because it was molecules that were being 'sucked up' through the gel in the tubes and not the substance in the well. Sequencing one plate took approximately two hours.

Once the sequencing had been completed, the lab technician checked the results by eye and 'tidied up' the sequences. The sequencing showed on the computer screen as a graph with four differently coloured lines that formed waves and troughs. The lab technician showed me a single nucleotide polymorphism (SNP) and pointed out that this was what the researchers would be looking for, a point on the graph where two of these lines came to a peak of roughly the same height and at the same place.

Toward the end of the sequencing process the chemicals started to run out, the waves on the graph flattened out and conveyed no useful information. The graph could run on for some time after the relevant data had been recorded, and it was this data that was cut or 'tidied up'. Then the data files were saved to disk, to shared server space or e-mailed to the researcher. The sequenced data was a linear representation, of a selected segment of DNA, along which a researcher could hunt for genes, alleles or polymorphisms.

Genotyping

The practice of genotyping was used to analyse genotypes and polymorphisms. Before describing the practices, again, it is helpful to define what the objects are, and what the scientific terms applied to them mean. Genotyping sorts people by difference, classifying them by alleles or variants into groups. They become statistical variables that can be connected to other information for analysis.

A *genotype* is 'the genetic constitution of an individual' (PGH Foundation 04.05.06; Carey 2003:68), but when scientists or health professionals talk about a genotype they may be referring not to the whole genotype of an individual but to one particular gene from many individuals, or in some instances several genes. The genotyping being carried out in the lab involved looking at one, or more, genes across many individuals. The results showed the distribution of different polymorphisms of a gene across the dataset, which might be for samples of hundreds or thousands of individuals.

A *polymorphism* is the 'variation in a region of DNA sequence among different individuals; *the variation should be present in at least 1-2% of the population to be considered a polymorphism*' (PHG Foundation 04.05.06 [my emphasis]), putting it another way, 'the existence of two or more variants (alleles, phenotypes, sequence variants, chromosomal structure variants) at significant frequencies in the population' (Strachan and Read 2004:639). While 1-2% of the population does not sound like very many, in a population of, for example, 5 million (the Scottish population) that amounts to 50-100,000 people. It is thought that if these groups can be identified they could be helped to prevent, or lower the chances of, onset of a particular disease such as cancer.

Preparation for Genotyping

The first step was to transfer a quantity of the samples to be genotyped from a working stock of samples into a plate for analysis. Calculations were done in the log book to work out how much of the assay, buffer and deionised water to add to the samples. The preparation of the plates was very much as previously described for sequencing, pipetting chemicals and deionised water into the samples row by row. Then PicoGreen was added, a fluorescent dye that 'sticks' to the molecules making them easier to see for measurement by spectrometer. Once everything had been added, the plate was covered by heat sealing a thin layer of clear plastic onto it, to prevent the assay from evaporating when heated in the PCR machine. The concentration of the samples was increased using PCR. Later the concentration of the samples was checked to ensure they were adequate for genotyping.

I was shown this process by LabTec3; she had recently learned it by observing, and then being observed by LabTec1. LabTec3 had been off sick the first week I was in the lab, missing my introduction and explanation of why I was there. As a result, we had not really spoken and she was evidently suspicious of me. Observing her work gave us the opportunity to talk and the first thing she did was ask me why I was there and what I was doing. Once I had explained that I was there to learn she responded 'That is alright then, I thought you were here checking up on us'. Her suspicion prompted me to explain further, that as an anthropologist it was not my role to be judgemental. She then told me she had previously worked in an animal research lab, experienced protest activists, and been subjected to people 'sounding off' at her about using animals in research. She said 'you would think that people would want research on animals - if it would make them well'. She found that people didn't comment or ask questions about working on human genetics.

After that conversation she was as helpful and enthusiastic as the others to show me what she did in the lab, but it did highlight the importance of the introduction. I had been introduced to the other staff by someone they trusted, and so had been accorded, at least a degree of, trust by these people. The lack of comment on human genetic work LabTec3 had experienced, in contrast to the response to animal work, could be a consequence of the inherent assumption that all medical research is good. Alex Plows found that opposition to medical genetic research was difficult to justify and that activist and opposition groups struggled to find a way to mobilise objections (Plows 2004).

The addition of PicoGreen dye and the PCR was carried out the day before genotyping. Once the PicoGreen was added to the samples, the plates were wrapped in foil because it is light sensitive. The dilution of each sample was checked against a normal curve, the normal range for dilution was 100-200 parts per microlitre. If all the samples were within this range, the plate was left over night. One of the PicoGreened plates showed two samples with low dilution. As there were only the two samples that required a correction to their concentration, the lab technician did this by hand.

On another occasion, a plate of samples from a study that was sent into the lab for genotyping was prepared using PCR with PicoGreen and the concentration results were poor. When LabTec3 repeated the PCR, the concentration results were slightly better but not good enough for analysis. The lab technicians discussed this and thought it was probably caused by poor quality samples being sent. LabTec3 was to contact the study to let the researchers know that the problem could not be fixed in this lab.

Repetition was a characteristic of lab work. Where there were a large number of samples outside the optimum range in a plate, preparation processes were repeated to try and bring the samples within the range. If the results of a process did not 'look' right or were not as expected, then the process was repeated, as many times as required until it did work. Sometimes repetition is not enough.

Working with large numbers of samples

Genotyping works with large numbers of samples, and the work was carried out using plates for 394 samples at a time. These plates have very small wells and only use small quantities of chemicals. Where more than one plate, up to eight at a time, was being genotyped, the robotic equipment was used for adding assays and dye to the samples for speed and efficiency. I also observed the preparation of a single plate done by hand, but this was unusual. Generally, preparation for genotyping was not carried out until a large number of samples could be done at the same time.

The fact that multiple processes were carried out simultaneously in the genetics lab evoked a sense of industrialisation. The lab work and the spaces were organised as a production line and there was a division of labour. This contrasts with the lab that Latour and Woolgar (1979) observed, where a single researcher moved between the work bench and the desk in the production of knowledge; and the lab described by Rapp (1999) where a single lab technician carried out all the sequential processes for an individual test. In these other labs there was a sense of connection, through continuity, of some sort of relationship between the researcher or lab technician and the samples they were working on that was not present in the genetics lab. There were also contrasts in the number of samples that the lab techs were working with, for example, in the genetics lab the techs were working with hundreds, even thousands of samples of blood and DNA; whereas the lab techs in Rapp's lab worked with only 5-6 at a time. The time frame for the work also differed in Rapp's lab the work was done over a short period of a few consecutive days, whereas the lab technicians in the

genetics lab were working with samples intermittently over a period of months, or possibly years.

The Genotyping Process

The biggest inscription device in the lab was the 'black box' that did the genotyping. Unlike the PCR machine with a lifting lid that enables you to see the plates go inside, or the sequencing machine with a glazed panel in the front, this machine was a 'black box' for inscription (Latour and Woolgar 1986:242). The processes to be carried out within the box were programmed through a computer on the workbench beside the box. A file, in the form of a table, was set up on a computer programme with details of the samples, assays and dyes. The plates went into the machine on a sliding drawer, similar to the way a disk goes into a DVD player. That was all that was to be seen, until the results appeared on the computer screen.

Inside the box, molecules from each sample were drawn through hair thin tubes by running a weak electrical current across the gel inside, a process called electrophoresis. A fluorescent dye from an assay served to show up the molecules as they passed up the hair thin tubes and through a laser beam which registered and recorded the results into the file that had been set up. The data could be seen compiled on the computer screen, represented as a table and a scatterplot. If there were enough samples the computer program would automatically colour code the plot, as shown in Figure 6.2 over the page.

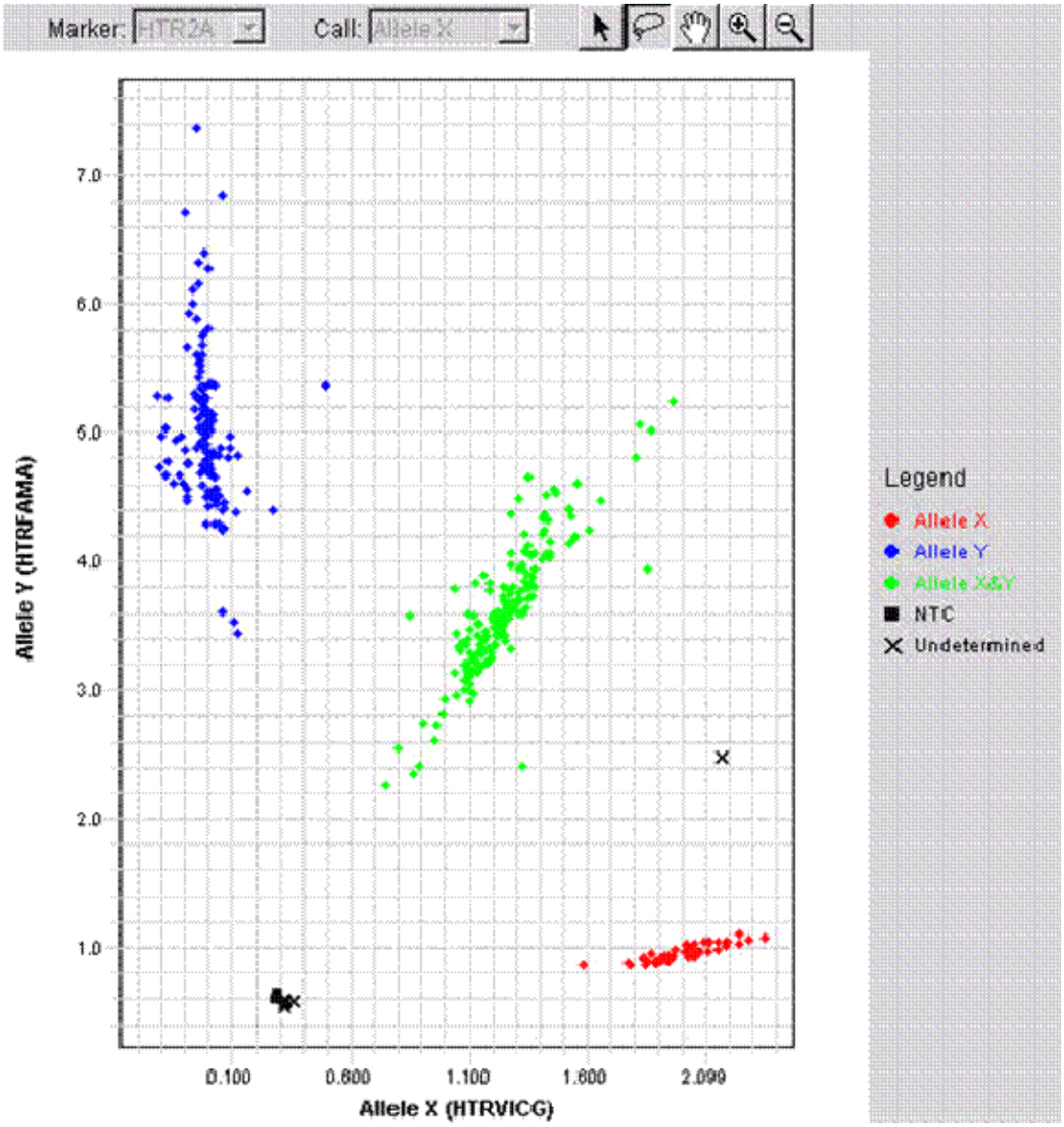


Figure 6.2 Genotype Scatterplot

The plot showed three distinct groups of points, referring to the possible variations of homozygous and heterozygous pairs of nucleotides. The plot for each genotype was checked, then it was double checked and through discussion a decision made about tidying up any doubtful outlier readings. Rapp also observed that the production of test results involved 'strict scientific regulation

and interpretive freedom' (Rapp 1999:192). Finally, an inscription of a set of samples, in the form of an electronic text file of the results, was sent to the study.

Inscription devices

Genotyping and sequencing produced inscriptions of DNA, or segments of DNA, in the form of digital data. These inscriptions were created using machines, 'inscription devices' (Latour and Woolgar 1987:51). Much of the work described in the studies of Latour and Woolgar (1987), and Rapp (1999) involved detailed and careful inscription by lab technicians and researchers. Indeed, 'the strange mania for inscription' (Latour and Woolgar 1987:48) was also an integral part of the work in the genetics lab. I saw inscription as occurring twice, creating two layers: first, there was the inscription of 'anonymisation', a coded identification done by hand or bar code on every sample, tube, lid, plate, work sheet, and technician's log book; the second layer was created by the 'inscription devices', where the inscriptions were made in and by the machines (Latour and Woolgar 1987:51).

Throughout this chapter, I have been describing scientific objects using mainly scientific terms, but both scientists and non-scientists often find these objects are more easily talked or written about when they are inscribed by another type of device, a linguistic device, the metaphor. Susanne Knudsen (2003) examines the proposal by Boyd (1993) that there are two fundamentally separate categories of scientific metaphor – the 'theory constructive' metaphor and the 'pedagogic' or 'exegetical' metaphor. Theory constructive metaphors represent original scientific thought and terminology, while pedagogic metaphors describe or

explain existing knowledge (Knudsen 2003:1249). Thus, over time, the constructive metaphor may become pedagogic in scientific discourse. Tracing the use of the metaphor of DNA as 'code-script', Knudsen categorises this metaphor as originating as a pedagogic metaphor. As time passed, the pedagogic metaphor became central to theory and was transformed into a constructive metaphor. Scientists also incorporated associated metaphors of 'letters', 'words', 'alphabet' and 'language' in their theory construction. Then, as research progressed and more was learned about DNA, the metaphor became a descriptive term for the concept, and thereby closed the metaphor. However, a metaphor is never entirely closed.

Scientific metaphors do not remain within scientific discourse. Discoveries and successes are announced to the public, and become open to non-scientific discourse. The audience does not have the same knowledge as the scientist, concepts have to be explained, and where the concept is difficult to explain in non-scientific terms, one way to do this is to relate the concept to more generally familiar and accessible metaphors, thus the closed metaphor is re-opened. Furthermore, Knudsen continues, the relational nature of a metaphor in different discourses, genres and contexts means that meaning is transformed, because those things which are the referents of the metaphor are different. She also suggests that when scientific metaphors are used in non-scientific discourse they lose status as scientific concepts (Knudsen 2003:1260). However, I would not agree that 'A democratisation has taken place rendering all metaphors equal regardless of origin or scientific status.' (Knudsen 2003:1260). All metaphors are not equal but rather they are accorded status by their referent. The metaphor of

'code-script' is attributed to Schrodinger from his lectures entitled 'What is life?', subsequently published in 1944 (Knudsen 2003:1251). He used the metaphor in his discussion of chromatine, which in due course, turned out to be the wrong chemical entity. Nevertheless, the idea that DNA somehow contained a code for life was captured by his metaphor and has continued to be influential in how both scientists and non-scientists think and talk about DNA.

The use of the metaphors, such as 'code-script', 'blueprint' or 'book-of-life', may be one of the reasons that DNA is not talked about in the same familiar everyday way as the gene is. By adopting the 'code-script for life' metaphor, DNA became inscribed with powerful, magical and sacred referents and associated with the big questions about 'life' (Nelkin and Lindee 1995). The status of the metaphor then became dependent on the beliefs of a group or an individual. Nelkin and Lindee (1995) carried out a study of the place of DNA and genes in popular American culture. They found that DNA could assume a mystique and genes had a propensity for becoming iconic.

> DNA in popular culture functions, in many respects, as a secular equivalent of the Christian soul. Independent of the body, DNA appears immortal. Fundamental to identity, DNA seems to explain individual differences, moral order, and human fate. Incapable of deceiving, DNA seems to be the locus of the true self, therefore relevant problems of personal authenticity posed by a culture in which the 'fashioned self' is the body manipulated and adorned with the intent to mislead. In many popular narratives, individual characteristics and the social order both seem to be direct transcriptions of a powerful, magical and even sacred entity, DNA.
>
> Nelkin and Lindee 1995:2

These ideas of DNA as powerful, magical and even sacred were evoked in the metaphors of DNA as a 'code-script' or 'book of life'. However, the metaphors for DNA were not confined to the sacred. Nelkin and Lindee also found ideas related to social order, about family, health, race, and crime. The 'iconic genes' were put to work in ways that supported political agendas, indeed, 'genetic essentialism can serve many different social agendas' (Nelkin and Lindee 1995:194). For example, the 'molecular family' reduced family relationships to a biological parent and child dyad that could be placed in opposition to the socially constructed and more inclusive family. Nelkin and Lindee argued that biology was given precedence over emotional and social bonds and 'popular interest in genetic connections coincides with the increasing visibility of - and public discomfort with – unconventional family arrangements' (Nelkin and Lindee 1995:78). Likewise, ideas about DNA and 'bad genes' appeared to influence views on crime and deviant social behaviour as predetermined 'Genetic explanations of behaviour and health appear to locate social problems within the individual rather than in society' (Nelkin and Lindee 1995:194) and genetic predispositions of health and disease, whilst promising medical breakthroughs, also worked to exonerate individual responsibility for future outcomes.

What concerned Nelkin and Lindee in 1995, and more than ten years on continues to be a cause for concern, was the application of political and social ideas that were underpinned by genetic essentialism. They found that advocates for particular causes of social concern were able to appropriate the findings of genetic research and relate them to existing cultural beliefs and values in the US,

highlighting in particular the application of prevailing right wing political ideologies to what they regard as 'the service of narrow or socially destructive ends' (Nelkin and Lindee 1995:199).[1] By comparison, in Scotland prevailing ideologies are liberal and left wing and would, I think, resist the application of a genetic essentialism that clashed with cultural beliefs and values. That said, even the most left wing politicians have not been able to resist the attractions of the predictive powers of DNA, particularly when they can be applied to planning, risk assessment and decision-making in public health.[2]

Knudsen concludes, and I would agree, that the distinct categories of metaphor as theory constructive and pedagogic are not clear cut because they are dependent on function, context and purpose (Knudsen 2003:1261). However, I think her analysis is helpful in underlining the notion that metaphors and concepts only retain their referents within historically and culturally specific contexts and that once they are relocated they become open to the possibility of new referents and meanings.

But, metaphors can also become obsolete, and as the knowledge and understanding of DNA expands it has become clear that DNA is not as fixed as its textual metaphors and their referents would suggest. Indeed the relationship between DNA, proteins and enzymes is dynamic, not only can DNA replicate itself but it also has processes for checking and repair. This requires more

[1] By contrast in Rabinow's *Making PCR* (1996) the US scientists when asked about their background all professed to holding liberal or left wing ideologies.

[2] A recent TV programme *The Killer in Me* (ITV 9pm 08.11.07) about genetic testing and disease prevention has led to controversy over the predictive accuracy of these tests and the information that was conveyed to the public.

interactive metaphors for theory construction and it is hardly surprising that those metaphors are emerging from the technologies with which genetic research is so closely associated. Thus new metaphors to be tried and tested relate to information technology systems and networks. The use of metaphors from this field create a relationship between DNA and technology which renders it increasingly open to technological manipulation and control (Fox Keller 2000) and at a further remove from the more 'organic' social world.

DNA is layered with inscription. There are the inscriptions of identifiers and labels which work to anonymise the DNA. There are the inscriptions carried out using the inscription devices in the lab to create digital data. There are the inscriptions of metaphors with various referents. But DNA is also embodied, it is inscribed by and inscribes a particular body, a person. Scientific inscription works to override the naturally occurring inscription by disconnecting the DNA from the individual who has donated the blood sample. One of the problems of working with DNA is that the information it contains is always connected to a specific individual. This problem extends further because much of this personal information will be shared with and connected to their relatives. The complex legal implications of how this information can be used for medical research in the public interest and at the same time protect an individual's privacy is one of the issues addressed by Laurie and Gibson (2003). In practice in GS this will done through robust data management.

The gene

Metaphors can be problematic in that they can confuse a concept with an object. The gene is a good example of the problem of whether people are talking about an object or a concept. The gene does not occur as a sample or material. The gene is a concept, used to describe a unit of transmission, a carrier of information. The word gene appeared in 1909 and is attributed to Wilhem Johannsen. 'A little word, perhaps – but a remarkably powerful one nonetheless' (Fox Keller 2000:2). It was used to capture the idea that the characteristics of an organism are specified in the gametes. Also conceived of as a unit, element or determinant, scientists were searching for a way of explaining the mechanism by which characteristics or traits were transmitted, and for some time there was little agreement between scientists as to what a gene was. Over time the idea of the gene as a unit of heredity became expanded 'to become the foundational concept capable of unifying all of biology' (Fox Keller 2000:3). The Human Genome Project not only intended, and achieved, the mapping of the human genome, but in doing so, also aimed to identify genes. It was thought that once the structure of the genome was mapped it would be possible to decipher the code and 'read' it like a program, and discern its component parts as distinct elements.

A gene has a molecular structure and configuration as a physical entity, but it is the DNA that constitutes the molecule. It is the DNA that is ascribed with physicality, as sample or biological material. This is interesting since in law it is the gene that is patented, not random lengths of DNA. Describing a sequence is not enough, function must be shown. This implies the patenting of a physical

entity, yet conceptually the gene is information. The gene is material in the sense that it is made up of nucleotides, and this series of nucleotides is patentable, but it is also in the scientific sense a concept, and an obsolete one at that (Fox Keller 2000). The gene has been described both as an entity and information by both scientists and lawyers. What I find interesting is the extent to which the law and regulatory bodies are influential in defining what DNA and a gene are, and the ways in which they can be used.

Sequencing the structure of DNA gave little understanding of the function of the gene, and genetic research entered a new phase (Fox Keller 2000). The Generation Scotland genetic database will contribute to research into the function of genes and the dynamic relationship between genes, phenotypes and environment. Functional genetics is moving away from the reductionist perceptions of genetic essentialism into a much more complex and nuanced view of genes through epigenetics, and theories based on information technology networks and systems.

Data and Information

Historical accounts of the theories, and consequently the metaphors, surrounding DNA and the gene given by Moss (2003), Fox Keller (2000) and Garcia-Sancho (2006) have differing point of departure, Aristotle, Bateson and Information Theory respectively. They each trace the transitions through concepts that changed the way scientists and Western society attempted to understand DNA. They all outline a separation into broadly two schools of thought, one focusing on the structure of DNA as information, the other

focusing on the function of DNA and the relationship between it and other elements of the cell, epigenesis. Metaphors such as 'message', 'text', 'information', 'code' and 'translation' were also being used in genetic research, by perhaps less well known names than Schrodinger mentioned above, for example, Gamow (Moss 2003:64). Following on from WWII Information Theory a gene was imagined as a 'message'. The idea of a message had two aspects of interest, the content and the transmission. Content was problematic because a message with a high content of information could have multiple interpretations, whereas transmission reduced the gene to 'a mathematical notion of information, not taking into account content, but rather the accuracy of transmission' (Garcia-Sancho 2006:17). This resulted in an information concept 'demarcated from meaning and expressed in the form of a quantity unrelated to the material basis of the message' (Garcia-Sancho 2006:18), which led to a shift from the notion of genetic information as a message to genetic information as a text. By the end of the 1980s it was clear that there were limitations to understanding DNA and genes in this way. The shifts in the way the referents of the metaphors for DNA and genes reflect to some extent the confusion with regard to these objects being treated as genetic data or genetic information.

The distinction between data and information, more broadly, is sometimes lost and the terms conflated. One dictionary definition of data is 'known facts used for inference or reckoning'; whereas information is defined as 'knowledge; what is told; news'. Data contains information but it is assumed to be in a passive state until something is done to it; but data can also be derived from information. The terms data and information are often used as though they are

interchangeable. Whatever the detailed definitions, they represent a spectrum ranging from single facts, numbers etc (data), to more meaningful concepts that contain facts (information) in a given context.

Data may be characterised as

- Facts, statistics used for reference or analysis
- Numbers, characters, symbols, images etc., which can be processed by a computer
- Data must be interpreted, by a human or machine, to derive meaning
- Data is a representation of information
- Latin 'datum' meaning 'that which is given'

Information may be characterised as

- Knowledge derived from study, experience (by the senses), or instruction
- Communication of intelligence
- Information is any kind of knowledge that is exchangeable amongst people, about things, facts, concepts, etc., in some context
- Information is interpreted data

It appears that context is particularly important in making the distinction, for data in one context can be information in another, and vice versa. It is data, not information, that is central to the construction of the GS databases. However, the data of the database, as I show in the next chapters, are often derived from what constitutes personal information in another context. Information is transformed into data through disconnection, anonymisation and aggregation. The connection between information and data is thereafter blocked through multiple and diverse mechanisms such as, the legislation of the Data Protection Act, security measures, and governance.

Conclusion

There were three types of product from the work of the lab technicians when the processes were completed: first, DNA samples, consisting of master stocks in racks and working stock in plates, which were all stored in freezers; second, sequenced data, in the form of electronic files of graphs and text readouts; and third, genotypic data, in the form of electronic files of scatterplots and tables. The practices were based on already existing knowledge and technology deployed to inscribe segments of DNA which could then be used in analysis. The practices in the lab manipulated the DNA samples, so that other versions of DNA appeared and were inscribed. Polymerase chain reaction (PCR) was the process through which the substance, DNA, was denatured and amplified. Sequencing and genotyping produced other versions of DNA as digital data. Polymorphisms and genes were not new objects, there was an already existing knowledge of genes in the broad sense. Here the practices and technology were used to process the DNA in order to seek and identify genes associated with specific diseases.

In the Genetics Core lab, this DNA was collected and processed for a particular purpose, to produce human genetic data. Purpose of use appeared to be a crucial factor in disconnection and the possibility of new connections when working with DNA. The purpose of the DNA in this setting lent itself to the deployment of particular referents and metaphors which serve to reinforce and permit the particular set of practices that worked to produce the desired objects. These disconnections created a tension between the different versions which were maintained through practice and by allying them to referents which

supported the disconnections. Here, for example, the law was a particularly powerful ally. For example, the law will treat these 'research objects' as the property of those who created them and not as the property of the person(s) who supplied the original material.

DNA is not a simple substance, and indeed there are many versions of DNA, for example, a bodily substance, a molecule, information, tissue, a biological mechanism for the transmission of hereditary traits, a code-script, a book, an icon, and in the case of a genetic database a resource. Taking DNA as referring to a single long molecule, it lies between ideas about the genome (a complete set of chromosomes for an organism), and genes which are characterised as segments, locations or parts of a strand of DNA. Graphical representations show these as linear structures, but in fact DNA molecules left to their own devices roll, curl, or fold themselves into a 'blob' rather than taking a linear form, as previously described. DNA has to be disconnected, pulled apart and straightened in order to become accessible to inscription through computerised digital representation in a linear format. The object, blood, that I started to follow from arm to desktop had divided and disappeared in different directions: the blood went down the drain in the last chapter; the DNA samples had gone into storage in the freezers to create a tissue archive or biobank; and the digital data files had left the lab heading for desktops elsewhere.

> This living information is happily compatible with the data management capacities of computers, so that various computer-based technologies can be used to read, analyse, store and synthesis molecular information.
>
> Waldby 2006: 1

What I found in the lab were practices and processes that created a series of disconnections between the substance that I had started following and the end result, digital genetic data. DNA was put to work, denatured and inscribed both literally and figuratively in order to make digital data and other objects possible. Genotyping is one type of work that DNA can do, but in order for this to occur we have to think about it in a particular way, as data, and in doing so we disconnect it from other versions of DNA.

Many different mechanisms were used in this process, from chemicals and mechanics to metaphors and legislation, collectively and separately, so that DNA was straightened, pulled apart, reconfigured and inscribed. Once an inscription was completed all the steps which made its production possible were forgotten or taken for granted as being merely technical matters (Latour and Woolgar 1987:63). But in order to understand the objects that are being used to create the population genetic database, the importance of the practices of people, place and equipment should not be underestimated. The objects are constituted and enacted by the practices neither DNA nor genetic data would exist without them.

The disconnections from the participant's body to digital data are traced through the progressive processes and practices of the lab; but the unique identifier attached to the data can be used to connect back from the digital data through the samples to the personal details, and ultimately the person. Genotyping and sequencing complete the extraction of digital data from embodied substance, rendering it accessible to computer analysis and the

connection to other data from other sources. There were diverse types of inscription constructed here, records, data management, DNA sequences and genotypes, and language. In the next chapter the inscription continues as bodies, relationships and experience are inscribed as phenotypic and environmental data.

Chapter 7

Phenotype and Environment: Turning Families into Data.

Phenotype and environment are terms used in population genetic research and they have specific meanings within this context. The meanings of the terms derive here explicitly from the relationships to genotypes. Applying the terms disconnects selected social factors from the complexity of their meaning in the social world and reduces them to manageable discrete statistical variables. They are terms of both disconnection and connection. The terms disconnect a potential myriad of social factors from the social world in which they are embedded and relocates them in a medical science setting where they are conceptualised as connecting to genotypes in specific ways. However, the disconnection is more apparent than real, it can never be absolute, because the derived facts always refer to the individual participants.

High quality phenotypic and environmental data are an essential requirement for the databases to be useful resources. In this chapter I describe how phenotype and environmental information are collected and inscribed as data. These data are being collected in the form of questionnaires, physical measurements and cognitive tests. The chapter examines the tensions that are produced through the disconnections that are enacted in the construction of these data and through the practices of collection. Further tensions arise from the interdisciplinary nature of Generation Scotland, where science, law and anthropology compete for different viewpoints and consequently meaning. For instance, phenotype and environment can be scientific objects or legal facts,

whereas questionnaires can also be seen as cultural artefacts. The differences between the ways which these disciplines assign meaning and produce knowledge are significant: where medical science often collects and orders information through the practices of disconnection, anthropology and to a certain degree also law create meaning through building connections and through context; in comparison, the scientific view can easily look reductionist.

A practice of disconnection central to medical science is anonymisation. Anonymisation, which happens to protect privacy, a central concern for law, takes place at the interface between the medical science world and the social world. Anonymisation is achieved through the routine removal of particular pieces of information, for example, name, date of birth or address, from all forms and labels. The removed information is replaced by the same unique personal identifier code that was previously allocated to the blood samples. The data can then be processed through a series of disconnections that render them into discrete statistical variables for analysis. The 'personal' information that has been removed is retained in a separate database. Importantly, through the unique identifiers of the DNA samples in the genetic database, the discrete statistical variables can be connected to health record data. These connections will be taken up in the next chapter.

Phenotype and Environment

Phenotype and environment are enacted here as objects through the practices of people, places and equipment. These practices occur in conjunction with the collection and processing of the blood samples. The data are disconnected from

the participant by the research nurse at the point of collection. The phenotype and environment data are then disconnected from the blood samples, only the blood goes to the lab. What happens to the phenotype and environmental data will be described later in the chapter. The practice of collecting and processing phenotypic and environmental data creates versions of these objects that have a particular purpose in population genetic research. However, these data overlap with and diverge from already existing notions of what these objects are, how they may be measured and what they can usefully be interpreted to mean, both within medical science and across disciplines. At the time of writing the collection and management of these data were still being developed within the GS projects. The standard operating procedures are still being written for some areas of the projects. The protocols, information leaflets and consent forms are regularly amended. 21CGH, as previously stated, has a target of 2,500 participants across five regions.

> The Genetic Profile of Scottish People
> GS:21CGH will recruit 2,500 individuals, 500 each from 5 different regions of Scotland. Basic physical measurements, a blood sample, and lifestyle information will be collected from each participant. The aim of 21CGH is to build control cohorts that are representative of Scotland's sub-populations.
>
> Generation Scotland website 15.05.07

The SFHS has a total target of 50,000 participants across Scotland, with a target of 15,000 to be recruited to the current Phase 1.

> Clinical Study
> The goal of GS:SFHS is to build up a large, intensively phenotyped, family-based cohort with which to study the genetic basis of common complex diseases and response to treatments. Recruitment began early

> in 2006 with the aim of signing up 50,000 family members over 5 years. Individuals aged between 35 and 55 are being approached through their GP, and these participants, together with any consenting first degree relatives (aged 18+), are invited for a detailed clinical assessment. During the clinic appointment, blood and urine samples are taken, cognitive tests are performed and lifestyle information is collected. The health of recruits will be monitored indefinitely by accessing their medical records.
>
> Generation Scotland website 15.05.07

The data will come from different sources and has to be quantified or coded in order to be able to use statistics for calculating relationships, interactions and occurrence in a population. Mainly derived from what is usually referred to as the 'lifestyle' questionnaire, phenotype and environment are taken to be observable and measurable, and include many diverse factors that may impact on an individual's life. The rest of the data comprises physical measurements and test results using scales or counts that can be recorded as numeric digital data.

Phenotypes may have common factors e.g. disease symptoms, but these can be as idiosyncratic as any individual who has a complex disease. Also, the number of combinations of causal environmental factors could be almost infinite. Pragmatically, all this potential data has to be narrowed down to manageable workable quantities. More than that, the data collected from participants, whether in blood, on paper or through physical measurement, must be kept confidential.

Great care is taken in the collection and disconnection of all the different types of data. The data are anonymised, separated, processed, quantified, checked and

then archived in separate databases. People, places and equipment are used separately and in combination to achieve these disconnections. What is paradoxical, although the data may be carefully disconnected, it always remains connected in the databases, through the unique personal identifier, but as importantly, through the individual that gave their blood, physical measurements, and lifestyle information. The disconnections are important in the creation of the database, but there must also be common factors so that connections can be made for future research. Systems, networks and datasets need common factors if they are to be linked together to share and transfer data. The standardisation of data collection and processing is advocated in the interest of 'harmonisation', in Europe (European Commission 2004) and elsewhere, for example the Public Population Project in Genomics (P3G),[1] to facilitate the sharing of data and extending the utility of such databases for research globally.

Phenotype

Like other objects that I have discussed previously, phenotype is problematic because it exists in more than one version. Not only are there multiple versions enacted by different people, places and equipment but there is an inherent ambiguity between those different versions. The use of the generic term 'phenotype' glosses over these ambiguities and introduces the possibility of any

[1] The Public Population Project in Genomics (P3G) is a non-for-profit international consortium to promote collaboration between researchers in the field of population genomics. It has been launched in order to provide the international population genomics community with the resources, tools and know-how to facilitate data management for improved methods of knowledge transfer and sharing. Its main objective consists in the creation of an open, public and accessible knowledge database. (http://www.p3gconsortium.org/ 06.04.07)

combination or configuration of factors without requiring an explanation. Phenotype is a classification with infinite possibilities, which can include the colour of your eyes, the symptoms of a disease, and which can even be extended to include the type of house you live in.

A phenotype may be defined in two ways: as the observed characteristic or trait of a genotype (Carey 2003: 68); or as the result of genotype-environment interaction. Starting with phenotype as an observed trait or characteristic, this is not straightforward as there is ambiguity in the ideas of what is observed and how it is observed. Just as genotype can refer to a whole genome, or part thereof, so phenotype can variously refer to the physical, biochemical or physiological make up of an organism. A phenotype may then be a whole organism such as a human being, or may refer to a part, for instance, the bone cells of this particular human being, or indeed a genetic disease affecting the bones, such as Marfan Syndrome. Furthermore, phenotypes can be problematic when it comes to observing, measuring or recording. For example, physical traits such as weight can be observed, but the weight of an individual can fluctuate; mental traits can be observed in behaviour but are difficult to measure; and in the case of many disease symptoms, the phenotype may in fact not be visible but may require specialist equipment, for example an x-ray, in order to be detected.

A phenotype can also be the result of genotype-environment interaction. For example, in the case of common complex diseases there may also be environmental factors external to the body that affect phenotypes. Public health

research and policy has historically been based on the affect that environment has on the health of a population. This view has been strongly supported by evidence over a long period of time. The connections between phenotype, genotype and environment are taken further in the next chapter. For now I want to consider briefly - what is environment?

Environment

Environment can in a commonsense way refer to an infinite number of factors that affect people's lives, from big issues such as global warming to the miniscule house dust mite. The range of environmental factors can include living conditions, occupational hazards, family relationships and so on. Clearly, every individual can lay claim to a unique set of environmental factors. However, SFHS is interested in families and 21CGH in geographically located communities. Families are assumed to share environmental factors as well as genotypes. Geographically located communities are of interest because it is thought they are likely to share common environmental factors, but not necessarily the same genotypes.

'Lay understandings' of environment and its effect on health have not to my knowledge been researched, but there is anecdotal evidence to suggest that people do have lay, or local knowledge about environment and its influence on health and illness. There have been media stories relating high incidences of illness, especially cancers, in communities living next to nuclear power stations, overhead high voltage electricity cables and mobile phone transmitters. During interviews for previous health studies, participants have referred to the local

environment as having effects on health. For example, one woman interviewed about participation in UKBiobank described the illnesses of her neighbours on the street where she lived and the surrounding area (Marsden, Sullivan, Duffy and McLaren 2002). Many of the families in this community had experienced one or more cases of cancer. The interviewee ascribed this to the local coal mining industry, in particular to the idea that there were mine shafts running under the area where the houses were built. 'How else do you explain why so many people in one place are sick? It must be something to do with where the houses are - something about the environment we live in' (Interview 10.03.02). Whether this was in fact accurate or not, she described a local knowledge and understanding of this feature of the local environment as having an adverse affect on the health of the people who lived there. The coal mining activities had ceased years before, but the mines were still regarded as dangerous to local health 'You know there might be gas seeping through or something' (Interview 10.03.02). The interviewee perceived the prevalence of cancer among families living within a few streets of each other as being environmental. In this instance, local knowledge of coal mines invoked environment as causal in explaining a cluster of ill health within a community. People make their own connections between environmental factors and illness that do not necessarily coincide with scientific explanations.

Environment is not confined to what is external to the body, in genetic research there is also an internal environment. This internal environment can refer to the environment of the DNA within a cell, as well as to the environment that surrounds the cell. The study of this environment takes place at the molecular

level. Physical measurements, blood and urine samples can serve as quantifiable indicators of the internal environment, for example cholesterol levels. The GS projects will collect some measurements in conjunction with the 'lifestyle' questionnaires. All these different data can be connected to each other in aggregated datasets and also to the genetic data.

Lifestyle Questionnaires

Questionnaires can be viewed in different ways, as a mechanism for the collection of facts or as a culturally constructed artefact. The questionnaires for SFHS and 21CGH have been constructed within a specific medical science cultural setting. The different versions of the questionnaires tell us something about aspects of the medical science and the way in which it constructs the social world. The questionnaire is simultaneously located within the science and within the social world. The reductionism of the questions in the scientific version creates a disconnection between the two.

The product of the questionnaires – digital statistical data - are constituted as objective indicators of phenomena, but can also serve as indicators of the GS organisation's priorities and practices. The questionnaires are based on underlying assumptions that the participants share the knowledge and values on which they are based so that they can i) understand what the questions refer to, and ii) know the answers.

Phenotypic and environmental data are being collected under the broad term 'lifestyle questionnaires'. Many of the questions are the same as are routinely

asked in a medical consultation, others have their origins in studies of public health. The lifestyle questionnaires are composed of a mixture of questions that include family history, medical history, medication, family health, habits such as smoking, alcohol consumption, diet, exercise, levels of education, income and occupation. The classification of these categories of question to phenotype or environment is however not clear cut. For example, education, income and occupation in public health are taken as a composite variable for social status but, depending on how they are viewed, can also refer to environment or phenotype. Distinctions have to be made in terms of structuring the data and databases, but this affects how the data can be used and blocks off the many factors that contribute to, or underlie, the answers to these questions, for example family histories, which will be discussed later in the chapter.

The practicalities for GS of collecting these data also had to be taken into consideration; while great quantities of phenotypic and environmental data are desirable for analysis, excessively lengthy questionnaires were thought likely to deter people from participating. Many of the clinicians and geneticists involved in the development of both projects had ideas about what data they wanted to see collected and early drafts were extremely long. Pre-existing clinical and genetic research needed to be taken into account, so that the data to be collected would fit as easily as possible with data already collected, for example, in studies on hypertension, colon cancer, arthritis or schizophrenia. Understandably, the data the clinical researchers wanted to include reflected their own areas of research. Particular types of data could be included to enhance pre-existing data to extend the scope of projects or provide new

dimensions to previous work. What questions to ask and what to leave out was important not only in terms of being able to link to existing data collections, but especially because the GS databases will serve as a baseline for future research. The design of future research projects will therefore be shaped by the structure and data of the GS databases.

Questions and measurements had to be worked out and approved by the Scientific Management Committee with future uses of the databases in mind. Meetings to discuss the questionnaires and other data to be collected went on during 2004-2005. For example, in family history, how many generations should be included with regard to place of birth or cause of death, allowing that this sort of information is often incomplete or inaccurate? Is there room in the questionnaire for questions about aunts, uncles, cousins? Is it better to ask participants who smoke or drink alcohol about their habits on a daily basis or a weekly basis? And how should the answers be quantified so they can be used in statistical analysis? Once the questions had been agreed upon they had to be worded, in as detailed a way as was thought necessary and as unambiguously as possible. There were particular questions about family and diseases that were important to ask from the genetic perspective of heredity, and there were others which pertained more to current health status and environment. Initially, 21CGH and SFHS both developed their own questionnaires, but in order to facilitate connection it was decided that the 21CGH questionnaire would be a subset of the SFHS questionnaire.

A comparison of the SFHS (Appendix 1) and 21CGH (Appendix 2) lifestyle questionnaires shows that there are four shared questions. The order of the questions varies between the two, but they both ask questions about family history, family health, smoking, and education. In the previous chapters, the family had become disconnected from the data through the practices in the lab, but the shared questions bring the family back into focus. The questions frame the family in a particular way. Seeking to establish the biological version of a family, the questions strip away the social versions of complex relationships and emotional ties that exist within 'cultures of relatedness' (Carsten 2000).

> Relatedness between persons is traced and symbolised in ways that might be described as belonging to the realm of the biological and the realm of the social, they are both given and forged elements in kin relations.
>
> Edwards 2000:28

The answer boxes make no provision for the many and complex versions of who a family might be. This has created a tension between the reductionist biological version of families and the great variety of families that may actually exist in the population of Scotland.

Family History

The questions on family history ask: where were you born, where were your parents born, and where were your grandparents born? This information will facilitate research on the population and the possible association of genotypes with particular geographic areas. As previously stated, the questionnaires are based on underlying assumptions that the participants i) understand what the questions refer to, and ii) know the answers. The options for answering are

'country', and if in Scotland, 'council area' and 'town'. Country appears to be relatively clear, but for Scotland the response could be given as either the UK, Great Britain or Scotland. Town is a problem if you were born anywhere outside of one. The answer option 'council area' may not make any sense at all to participants. This is first, because births are registered in districts; second there are different types of council, regional and district; and thirdly, the recruitment age range is 35-55, i.e. these people were born between 1952 and 1972 (with relatives possibly earlier or later) and local government boundaries have changed over time.

Births, deaths and marriages in Scotland have been registered with the General Register Office of Scotland since 1854. Births are registered by district. The districts replaced parishes which had previously acted as registers. Parish boundaries lay within counties. Prior to 1974, Scotland was divided into 33 counties and the larger cities were governed by city corporations. Following the Local Government (Scotland) Act 1974, a two-tier system of local authorities was set up. There were 9 Regional Authorities which were broken down into 53 District Councils, and 3 unitary Island Councils. In 1995, local government in Scotland was reorganised again, this time into 29 unitary authorities, following the Local Government (Scotland) Act 1994. The effects were not uniform throughout Scotland. For example, the region of Tayside was broken up into three district councils, whereas other regions were left intact whilst their districts disappeared, as was the case in Fife (Gateway to Scotland 2007).

There is no equivalence between register districts and regional or district local government councils. A quick look at the current Directory for Registers shows that registrars are located in many places, some are in council offices but most are not. They can be found in health centres, post offices, town halls, libraries, and many other places, for example if you want to register a birth in Durness, you go to the Highland Tourist Information Office. It is therefore likely that people will find it difficult to understand and answer a question about which council area they were born in. This could affect the quality of the data, and in turn its usefulness in analysis.

The family history of where you were born, where your parents were born, and where your grandparents were born, are pieces of information. They are also elements of a wider narrative of who an individual might be, how they think about themselves and the people that surround them. For an individual, their place of birth is part of their own history and identity. It is also information that relates to a particular group of people, their sense of family and place within a particular group of related (biologically or otherwise) other individuals. This information is rooted within a particular group of people in a set of very specific relationships, and these in turn are historically contingent and geographically specific.

> Kinship embraces connections, people trace to each other through notions of shared substance, be it blood, genes, flesh, or bone; at the same time it places a greater or lesser emphasis, at different historical moments and in different parts of the world, on the creation and maintenance of social relationships through intimacies of care and effort.
>
> Edwards 2000:27

Where a person was born is just one element of a complex narrative of identity and belonging. Jeanette Edwards explores and elucidates English kinship through the idiom of 'born and bred' (2003). In order to belong to a place, a person has to be born and bred there. But having been born in a certain place does not necessarily mean one belongs there, or is recognised as belonging there by others, 'one also needs to be brought up in a particular way' Edwards (2003:28). The questionnaire assumes that geography is the significant feature of family relationships and hereditary disease occurrence. The wider narratives of identity and belonging have been stripped away.

The social relationships of the family or individuals within the family may be threatened by specifically biological versions. The collection and ordering of this information in this way may have little or no equivalent in the everyday social, economic or political orderings of a family. If this is the case, the scientific genetic ordering has the potential to contradict and disrupt the everyday understandings of who the family is and where they come from. There is also potential for the assumptions, that family specific relations are biologically based, to be proved inaccurate or wrong, for example, paternal discrepancy. Within the collection of data and samples for a population genetic database, this is an area of particular sensitivity.

It has been estimated that in the UK the number of paternal discrepancies in the population could be between 10% and 30%. These estimates are not reliable as they are not supported by published evidence (Macintyre 1991:870). Paternal discrepancy, or non-paternity, occurs where a father raising a child is not the

biological progenitor of that child, as previously assumed. This could be known, suspected or unknown to the father and/or the child and/or the biological progenitor. A review of studies that revealed rates of paternal discrepancy carried out in 2005 shows that figures can vary from 1.4% to 30% (Bellis, Hughes, Hughes and Ashton 2005:751). The actual number of families affected by paternal discrepancy might be less than estimated, as it is probable that paternal discrepancy will cluster in some family groups (Bellis et al 2005:752). Whatever the figure, it raises serious issues in relation to health and relationships. Paternal discrepancy, often associated with infidelity, has complex consequences for families, including the break-up of long term relationships and even violence.

Generation Scotland will carry out paternity tests as a quality control measure, but the results will not be disclosed. This anticipates cases of paternal discrepancy occurring in the data. Information is available in the Frequently Asked Questions (FAQs) page of the website but not included on the information leaflet

> **Will Generation Scotland perform a paternity test?**
> Paternity testing is a sensitive issue. It is a scientific process that can positively identify relationships between people from their DNA. As part of the Scottish Family Health Study, researchers will perform tests to check that family members are genetically related, because this is essential for the success of the study. The researchers who carry out these tests will not know, or be able to find out, the identities of the people who gave the samples. Generation Scotland will not pass the results of family testing back to families. Requests from participants for paternity testing will not be considered.
>
> GS website 02.06.07

The Scottish Family Health Study and other genetic health studies will reveal paternal discrepancy if it is present in the data. 'More recently, investigations of familial patterns of disease inheritance have identified paternal discrepancy and led to further estimates of its prevalence.' (Bellis, Hughes, Hughes and Ashton 2005:750). This is of increasing concern to researchers in the conduct of genetic testing, screening and clinical studies, and may be a potential problem for GS. At present research studies do not inform those affected. There is an argument for disclosure to children in the interest of their health. Genetic information is increasingly used in informing health care, diagnosis and therapeutics which strengthens the case for children to know. It could be significant that 'recent development in assisted fertility (for example in Sweden and the UK) now place the child's right to know their biological father above that of the donor (biological parent) to remain anonymous' (Bellis et al 2007). If children have a right to know about their paternity, and it is known that GS will as a matter of course carry out paternity tests, might there not come a time when GS is required to disclose this information if requested to do so by an individual?

The numbers affected by paternal discrepancy are unclear, but the possible numbers affected by maternal discrepancy are unknown. Yet there is a folk knowledge of women raising children as their own when they were born to daughters, sisters, or other people, and of other offspring believing they are their brother or sister. This knowledge appears to be even more closely hidden and viewed as potentially disruptive to family relationships. Story lines from fiction and documentaries presented on the television, in magazines and newspapers point to the great lengths that people will go to conceal such facts

from family members, and the potential distress that disclosure can bring. There are of course also inevitable implications not just for the immediate family concerned but also for the biological progenitor (Bellis et al 2005:752). Attitudes are shifting as reproductive technologies change the way that families are viewed and constructed, for example 'My sister is my son's mother' a story reported in *The Guardian* (13/05/03). But long concealed family secrets tend to cause shock and distress when they are revealed.

The period during which the family members being recruited to GS were born, saw an increase in the recorded numbers of unmarried mothers, divorces, and variously configured families. None of these was socially acceptable at that time and many people did go to great lengths to conceal their 'mistakes'. Viewed in this light a 30% rate of paternal discrepancy might not be an overestimate. Studies like the Scottish Family Health Study may prompt family members to reveal hidden family relationships or biological relationships with unknown others.

If the rate of paternal discrepancy in Scotland were 10%, of the 50,000 participants in the SFHS study one would expect to see 500 cases. It is, however, doubtful whether SFHS will give an accurate picture of the scale of paternal discrepancy in Scotland for two reasons. First, it is unlikely that members of families with hidden paternal discrepancy will participate. Second, it is thought that paternal discrepancy is more prevalent in families from lower socio-economic groups. People categorised as belonging to lower socio-economic groups are thought less likely to participate in GS. There are measures in the

protocol that aim to reach these groups, but it will be some time before it is clear if they have been successful.

Cultural Background

A question on cultural background is followed by questions about where people come from. Answers are given in tick boxes of fixed options. The answer options are often found elsewhere under the heading of ethnicity. There are two classifications within the answer options to which people are expected to ascribe: skin colour and geographic location. Thus the first options include black, white, Asian or mixed. The second option is about countries, regions of countries, or regions of the world. The conflation of 'cultural background' as a heading with fixed categories of ethnicity (or race) as answer options confuses geographic origins and colour of skin with many things that have little to do with either. Cultural background can refer to many aspects of people's lives, for example religious beliefs, political systems, economies, education, food, clothes, houses, land, art, literature, cosmology, war, cattle, kinship, languages, or playing the bagpipes to name but a few. The list suggests that anyone who ticks a box shares many unspecified things with any other person who ticks the same box.

> Commentators have criticised the terminology used for the classification of ethnic and racialised groups in health research for a number of years. The shortcomings of fixed-response categories include the reproduction of racialised categorisations, overemphasis of homogeneity within groups and contrast between them, and failure to offer terms with which people identify and which can express complex identities.
>
> Bradby 2003:5

The SFHS and 21CGH questionnaires are laid out differently. The SFHS question is divided into two parts: section A gives the options of white, black, Asian or mixed; section B offers a list of geographically related options Scottish, English, Welsh, Irish, N.Irish, Pakistani, Indian, Bangladeshi, Chinese, African, Carribean, or other (Appendix). Thus it is possible in this format to identify a person in a variety of combinations for example, Black Scottish.

The 21CGH questionnaire by contrast lists the possible combinations as white-Scottish, white – other British, white - Irish, white – any other white background, mixed - any mixed background, Asian – Indian, Asian – Pakistani, Asian – Bangladeshi, Asian – Chinese, Asian – any other Asian background, Black – Carribean, Black – African, Black – any other Black background, Any other ethnic background (Appendix). You can be white and Scottish / Irish / English / Northern Irish / Welsh, but not Asian. Likewise, Asians can be Indian / Pakistani / Bangladeshi / Chinese, but not Scottish or African, and if you are Black you can be Caribbean or African, but not Scottish or Pakistani. These options assume particular 'cultural backgrounds' in the Scottish population.

Both questionnaires are based on a set of assumptions about people, places and cultures, e.g. all Africans are black and the entire continent shares the same culture, that Asians come from different countries and therefore have different cultural backgrounds, or that all Scottish people are white and share the same culture. There are options to self identify as 'mixed' implying not pure, or 'other – specify' indicating marginalised categories. The assumptions that underlie this order are not necessarily held by the people themselves, and thus create

disconnections between the people that make up the Scottish population. It is a specific culturally constructed ordering of the social world, which perpetuates the idea that ethnic and racial differences are discrete, scientifically recognised, and potentially biologically constituted (Bradby 2003:7).

Family Health

With regard to 'family health', the participant is asked to indicate whether 'your father, mother or any brother, sister or grandparent has been affected by any of these conditions', followed by fifteen *diseases*. The list reflects the priority given to particular diseases in Scottish population health research, e.g. chronic heart disease, cancer, and mental health, but also includes diabetes, asthma and arthritis. Family health is measured by diseases.

The answer options are laid out in a grid, with columns for family members, and rows for each disease. The kin terms are used as classifications. The use of the kin terms denotes relationships that anticipate the existence of biological facts rather than social relations, which occur 'after the fact' (Strathern 1992). Family in this setting is applied as a biological term.

The family health questions on both the SFHS and 21CGH questionnaires differ only in one point, namely the participant's own health. While the SFHS questionnaire includes the participant as a family member, 21CGH asks a separate question. In other words, the SFHS question makes the participant an integral element of family health and the 21CGH treats the health of the participant as separate from the rest of the family.

There is a final question after the list 'Any other serious illness that runs in your family?' The final question renders explicit what has already been implied by the format and layout of the list and tick boxes, and reiterates one of the central premises of the GS database, that complex diseases are linked through heredity. In this way, relationships are constructed through shared disease rather than through shared substance.

The metaphor 'runs in the family' refers to blood. This metaphor of blood as shared substance is proportionally quantified (half from parents, quarters from grandparents and so on) and with it comes knowledge of how much one has within. But genes and disease do not flow like blood, 'the fact that each individual contains two sets of genes, inherited from the connecting of persons, replicated through cells of the body, does not prompt an intrinsically relational metaphor' (Strathern 1992:80). Genes are not transmitted proportionally, but as unique configurations of diverse particles, 'we derive an image of a different order, of an individual who shows in her or himself a unique combination of heterogeneous particles' (Strathern 1992:81). What emerges is therefore random, unique combinations that supersede the relational and proportional flow of blood.

The questionnaire version of family health strips away the family knowledge and narratives of events and reduces them to digital data. The problems and tensions that surround family history and family health entail certain implications for the collection of phenotypic data for the database. The knowledge that family members have may be partial or inaccurate, which

affects the quality of the data. Even more important than inaccurate knowledge may be accurate but hidden knowledge, this could act as a powerful deterrent to participation for some families. Genetic studies that might reveal parental discrepancy do not currently offer any support or advice services to participants (Bellis et al 2005:752). The collection of phenotypic data could affect how people view their relatives and the relationships that exist. The disclosure of hidden knowledge is potentially disruptive, even dangerous to family members, and the reductionist phenotypic version of family health can construct a diseased family.

Smoking

Smoking questions appear on both the 21CGH and SFHS lifestyle questionnaires. Why include these questions and not other ones? The selection could be arbitrary, but it is generally assumed in public health research that it is a strong indicator of healthy or unhealthy lifestyle behaviors, and incidence of illness. Smoking tobacco has been targeted by public health policy as an important factor in relation to illness and disease. The smoking questions aim to discover the extent of the habit, and exposure to secondary smoking among participants. The SFHS and 21CGH questionnaires quantify it as number per week, although the 21CGH question is slightly confused, first asking how many per day but demanding an answer per week. The question about 'giving up' smoking anticipates that there have been changes in people's smoking behaviour, prompted by public health drives to reduce smoking in the population, recent changes in legislation and health and safety policies.

Education

In public health studies, levels of education and health are associated with each other. People with less education, who, by implication are in lower socio-economic groups, are viewed as being at higher risk of illness and bad health behaviour. Research on recruitment, together with assumptions made by clinical researchers, suggests that the most likely participants for projects such as GS and UKBiobank will be middle class, educated, and scientifically literate. It is ironic that the groups most likely to participate are those thought to be at least risk from the diseases, such as chronic heart disease, due to their environment and resources to expend on the maintenance of their health. By contrast, poorer families, whose health is most at risk, are assumed to be least likely to participate. It will be interesting to see if this is borne out by the data.

The system of education and educational qualifications, like the changes in council areas, has changed over time. Questionnaire answer options are based on the current organisation of examinations and qualifications, offering Standard Grade or 'O' level, Higher Grade, University Degree and Other professional or technical qualification or diploma after leaving school.

The Scottish Certificate of Education Standard Grade was introduced in 1984, with the first awards in 1986. Until 1986, examinations taken at the end of S4 (fourth year at secondary school) were for the Ordinary Grade (no to be confused with 'O' level), which was introduced in 1962 for the top 30% of the ability range. In practice, far more than 30% took the Ordinary Grade, and by 1985, 75% of all school leavers held at least one award at Ordinary Grade.

However, a substantial minority of school leavers still held no nationally recognised certificates recording their achievements at school (The Scottish Office Fact Sheet 5: 1996). The 'O' level belongs to a completely different exam system, with its own history of changes over time, and is not commonly used in Scottish schools. The Higher has existed in various forms since 1888, most recently revised in 1987.

Such changes are not confined to secondary education, but have also occurred in the areas of vocational training and prompted the emergence of Further Education. The range of qualifications awarded in this sector has changed considerably over past decades - as has the need for recognisable qualifications; there is a wide difference between the expectations of an employer in the 1950s and today. In the questionnaire, underlying the question for 'Other professional or technical qualification or diploma after leaving school' is the assumption that these lie outside the area of academic achievement - even though ticking this option could equally refer to a typing certificate as well as to the professional qualifications of a surgeon.

SFHS only questions

The longer SFHS questionnaire has more detailed lifestyle questions and includes a set of clinical questions that are not part of the 21CGH subset questionnaire. The lifestyle questions relate to alcohol consumption, dietary intake, physical activity and occupation. The clinical questions pertain to chest pain, chronic pain, fractures, medication and operations. The 21CGH section on the participant's health makes provision to indicate medication or operations in

relation to the list of 15 diseases, but without the detail or additional SFHS questions. Although the data being collected by 21CGH is a subset of the SFHS questionnaire, it should be sufficient to establish control cohorts for future studies. The greater detail in the intensively phenotyped SFHS will support the future study of the genetic basis of common complex diseases.

Dietary Intake

The question on 'Dietary Intake' asks 'In general, how often do you eat …..?' followed by a list of food types: fresh fruit; green leafy vegetables; other types of vegetables; oily fish; other types of fish; chicken, turkey or other poultry; liver; other types of meat; eggs; dairy products; and brown bread. It could be interpreted as a list for 'a healthy diet', but that assumes it is recognisable as such, that the people who look at this list know what 'a healthy diet' is supposed to be. A 'healthy diet' is both medically and culturally constructed, and this question points to the way in which health has become a cultural preoccupation. It is indicative of a shift from ideas about illness and the provision of care for the sick, to a view of health as not just desirable from an individual perspective but that it has become a social, political and economic issue.

The knowledge on which this section and list of possible answers is based may not be held by participants. They may not share the eating habits of those who compiled the questionnaire, nor conceptualise the food they eat as 'dietary intake'. Food and its consumption are culturally variable. The social activities of

preparing and consuming food are at least as much about relationships as they are about dietary intake.

The list is weighted by categories so there are two questions on vegetables and two on fish, one specific to liver, and then all other types of meat which would include pork, lamb and beef. It appears that the categories overlook such staple Scottish meals as mince and tatties, which the answer would lump together with bacon rolls. It may be difficult to discern differences because of the categorisation of the food types, for example, what does 'chicken' refer to – the answer could be 'twice a week' - but it is not clear, whether the participant means fresh organic skinless fat free chicken breast or chicken nuggets from MacDonalds. In 'Other types of vegetable' - the answer could be 'twice a day' and refer equally to a raw carrot or chips. Advice on healthy eating would lead to the conclusion that eating a raw carrot twice a day is considered better for you than eating chips twice a day. The statistics derived from this question may be misleading if the inferences made are based on assumptions about healthy eating.

Questionnaire 2: Participant Questionnaire Booklet

One area that has been specifically identified for inclusion in the GS database is mental health. In addition to the 'lifestyle questionnaire', the SFHS and 21CGH projects include a second questionnaire specifically orientated to mental health research. The questions and test are identical in both questionnaires, and consist of extracts from several cognitive function tests, which are validated and recognised as useful tools for these types of measurements. These data will be

used in conjunction with the genotypic data to understand mental illness better. The questionnaire will be completed by participants at the clinic visit. The questions and tests were designed to measure personality, intelligence, memory and emotional states. A score is calculated at the end which will be entered into the database.

Physical Measurements

In addition to the questionnaires and blood samples, a range of physical measurements is collected. These measurements include height, weight, blood pressure, ECG scans for heart rate, and from blood samples sugar levels, cholesterol, kidney and liver function. This list is not complete but shows measurements that are taken as a matter of routine in a medical setting. In order to ensure the quality of the data, the same equipment must be used in the same way to standardise the measurements as closely as possible. This involves not only using similar equipment, but calibrating it to standard scales and using it in exactly the same way by following standard operating procedures (SOPS). Bodies are increasingly measured and evaluated within defined parameters that are considered acceptable or not acceptable in medical, and increasingly, social terms.

This particular version of the physical environment of the body blurs with phenotype. Some of the measurements pertain to the physical attributes of the body, whilst others are derived from tests that will be carried out on blood and urine samples. The measurements and counts quantify the body into data which, like the genotype, can be digitised and put into a database. Collectively,

these measurements produce another version of participants, a medicalised picture of an individual or a family. In this version the data on weight, heart rate and cholesterol levels could combine to portray a person with chronic heart disease that is disconnected from the body of the person who may or may not experience symptoms. Furthermore, this version of an individual is constructed by researchers, not by the person themself, who also has no say or control in that construction. These data are nevertheless intensely personal. Anonymisation does not render the individual or their data non-unique, but simply less easily identifiable. However, personal as these data are, they are not considered the same as personal details from which an individual can be readily identified.

Personal Details

Personal details have to be disconnected from all research data. Personal details include name, date of birth, address, telephone numbers and e-mail addresses. These are the details used to contact potential participants at the recruitment stage of the projects. Only at the early stages of contact and recruitment will personal details be in any way accessible, and will only be seen by a small number of people, who will be in effect separate from other GS people. All the data collected are anonymised using unique identifiers.

The SFHS recruitment works either from a list generated using Community Health Index (CHI) numbers at a central point or, from an already existing cohort study. The lists for different general practices are passed to a third party,

who then contacts individuals with a letter signed by their GP. The SFHS Protocol states:

> The research group will identify an independent party to assist with the approach of potential participants. This party will be staff of a university research group, and will handle patient identifiable data, but will not participate in the main study in either an administrative or research capacity. They will have an (honorary) NHS contract, conform to NHS rules and regulations, and will be isolated from the researchers on the study team. The independent party, will work with Practitioner Services Division of ISD, and generate a list of eligible people, according to agreed criteria, from the Scottish NHS register (known as the CHI). An internal letter management system will then generate letters on a per-practice basis on practice headed note-paper, which will be delivered to the practice for signature. These letters will be returned to the independent party and then dispatched by post.
>
> SFHS Protocol

The 21CGH project uses a combination of approaches, in one area it shares the SFHS methods while elsewhere it draws on lists from pre-existing studies and contact individuals either through their GP or through a clinic they are already attending. Once people, for either project, have agreed to participate, their details will be passed to one of the research nurses who will contact them, have information, consent forms and questionnaires (in the case of SFHS) sent out, and make a clinic appointment. All the data collected will be anonymised but given a unique identifier, and kept separate from the personal details. Personal details will be stored in a separate database. No researchers using the GS phenotypic and genotypic data will ever have access to personal details.

Collecting questionnaire data

The original intention was for both the SFHS and 21CGH projects to use identical methods and collect identical data; this has not proved feasible, though they are as similar as possible, given the different aims of the collections. Because they are funded as different projects, different teams are working on data collection in different areas of Scotland. The teams from both projects have worked separately with different sets of priorities, and consequently they have taken different approaches to the data collection. The 21CGH team first developed the technology to manage the data, so that the data could be entered directly into the database. The SFHS project started with data collection using paper forms, while developing the database technology simultaneously; the reason for this is that SFHS is dealing with a considerably larger volume of data, and in order to complete the collection of 15,000 samples within their time frame, time consuming tasks like recruitment and collection had to start early.

21CGH developed an electronic method of collecting the questionnaire and measurement data using a tablet computer with a datacard that links into a virtual private network (VPN), which is similar to a mobile phone network. This is used for the 'lifestyle' questionnaire, the mental health questionnaire and the physical measurements. The questionnaires are completed by the participant at their clinic visit. The questions (or tests) appear on the tablet screen and answers can be given by touch pen, clicks or typing. Once the questionnaire, tests and measurements are completed, the data is submitted. The program encodes the answers into a document, which it then attaches to an e-mail that it generates to a designated e-mail address. The datacard connects to a VPN and the e-mail

including the attachment is sent to a mail server. There, a mail daemon picks up the arrival of the e-mail and initiates a program which picks up the e-mail attachment and automatically puts the data from the form into the database. It is possible to track the e-mails online to see how many forms have been submitted at any time and where they have come from. This has only been possible because the database was set up before data collection began.

Designing databases

These data have to be organised and managed so that they are useful and accessible hence the database design is of crucial importance. Designing a database is a complicated and highly detailed task. The population genetic database is not one single database, but rather, comprises several databases which can be linked.

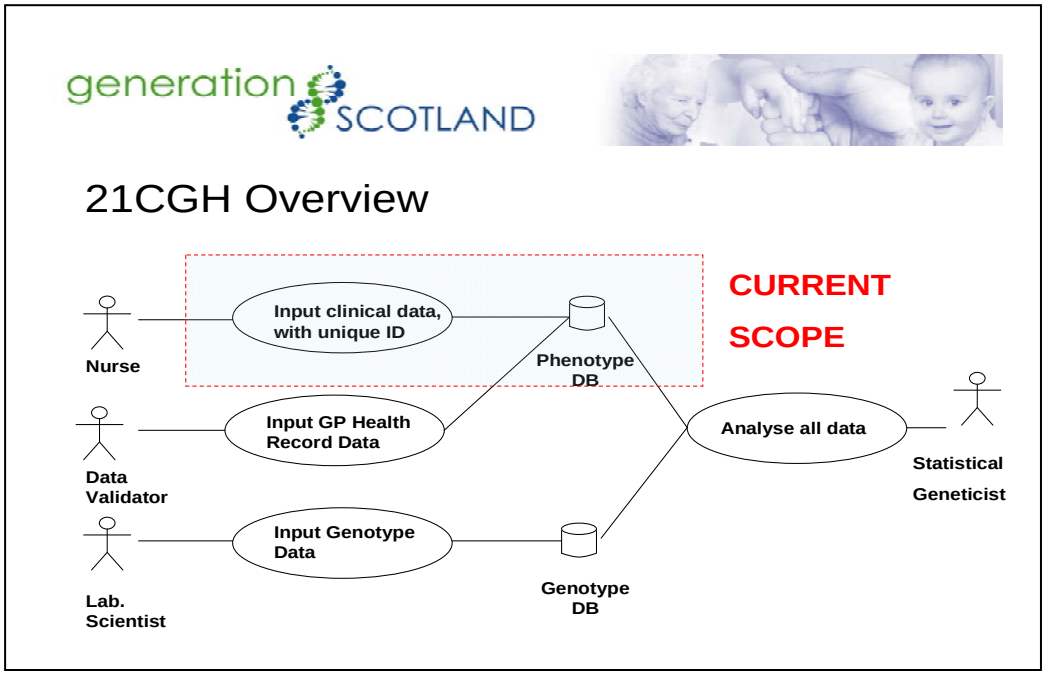


Figure 7.1 21CGH Overview: Data Manager Presentation slide 2004

The use and development of technology is integral to the collection and connection of the data for the databases, if it is to work in the way envisaged, as shown in Figure 7.1. The 21CGH overview shows how the databases are envisaged as separate, while facilitating data connection. The SFHS will use a similar approach. The GS databases need to work so that different types of data can be either kept separate, related to each other, or merged with other datasets in other databases.

The SFHS database is still being developed. As mentioned earlier, SFHS data collection began earlier than 21CGH, but using paper questionnaires. The SFHS administers the 'lifestyle' questionnaire prior to the clinic visit, and the second mental health questionnaire at the time of the clinic visit. The 'lifestyle' questionnaire is sent out to people once they have agreed to participate, but before they have their clinic visit. The intention is that they have time to complete the questions before coming in to the clinic, which cuts down on the appointment time and gives the individual plenty of time to answer the questions. The participants then bring the completed questionnaire give it to the research nurse. The research nurse checks the form with the participant for missed answers, or clarification. The results of the mental health tests and the physical measurements are recorded on paper forms. Answers and results from both questionnaires together with physical measurements are then sent to a research centre where the data will be transferred from the paper versions to the database using a double entry system. Double entry means it will be done by two people to ensure accuracy.

Databases for all types of data are determined by and constructed around the way information is asked and given. The format in which types of answers, results and measurements, are entered into fields, orders the data in a particular way. A study by Hine of databases as 'scientific instruments' is helpful here (Hine 2006:269). She examined the way in which the construction and use of databases orders scientific work. The database offered a means of characterising a set of objects and fixing a set of relationships between them in order to create a resource (Hine 2006:277). The design of the database was based on working practices and existing knowledge, it did not create things but held representations of what were already there. Therefore the relationship between the DNA samples and the database was central to its future use as a resource. Without samples, the database is only 'a record of work', whereas with the DNA samples it becomes a tool embedded in working practices (Hine 2006:279).

The database in Hine's study was not developed in the lab, but by computing specialists based elsewhere (Hine 2006). One of the things that Hine observed was that the computer service specialists learned a good deal about the science in order to be able to write the necessary programs and algorithms, but the scientists did not learn an equivalent amount about computers and programs - 'the balance of need to understand is, however, one-sided' (Hine 2006:282); indeed the scientists often used the computers in the lab in ways that were unexpected by the computer specialists. Hine found that

> The 'digital ordering' represented by the database is thus highly contingent, representing the upshot of lengthy negotiations between the collaborators over the nature of the natural objects involved and of the scientific culture in the workplace
>
> Hine 2006:288

There is one important difference between Hine's study and Generation Scotland. In Hine's study the database came into the laboratory space, whereas in Generation Scotland the data leave the laboratory, the database is located in a different space, adjacent to a set of data management desktops. Generation Scotland is using the services of computer specialists, located in different places, to create the database(s). Collaboration between them is overseen by the Scientific Management Committee. The combined efforts and expertise of the collaboration work within the assemblage in which the database(s) are being created. For example, the database for the genotypic data to be stored and managed is still under discussion. At the time of writing, the lab creates files of raw data which are saved into separate files and stored on disks or shared drives, or sent to the projects that have had the work done. The transfer of data creates the impression that the scientists, the geneticists, have left the laboratory and moved into offices, where they have computers to run complicated and lengthy statistical analyses of data. However, Hine argues that, while 'databases, may provide occasions for developing new work practices in science, and may lead to the exploration of new areas of knowledge' (Hine 2006:291) they do not necessarily produce a radical transformation of scientific practice. Lab work and computer work may appear to be distinctively different, but they connect in the creation of a database.

Conclusion

The initial funding for Generation Scotland came from SHEFC to develop the infrastructure for collaboration between various Scottish research institutions and the creation of a population genetic database. This work is done by

computer scientists and technology experts working within the fields of health informatics, science and technology, along with the Data Manager, through practices enacted by people, places and equipment integral to the Generation Scotland collaboration. At the level of information technology and genetic research, the systems and networks that support the collection, management, analysis and transfer of data to create and maintain disconnections between participants, researchers and information are an essential part of the infrastructure.

The genetic database will contain genetic data from more than one project, yielding different datasets, but constitutes only one of several databases that will be used to store and manage data from different sources; for example, there will also be a database to store the lifestyle data collected from questionnaires. Programming is being developed to facilitate data from different databases or datasets to be merged and analysed rapidly. A network and management system will facilitate the transfer and merging of data from different databases for analysis at the desktop.

A particular set of scientific practices are at work in the construction of phenotypes and environment as digital data for a database. Measuring and counting phenotypic and environmental data changes personal information, that is, information about a person, into data. As with the DNA, these data are not about the content of the information but the transmission of the text. By that I mean these data do not tell the narrative of people's lives and experiences, but quantifies their bodies and their lived environment. Phenotypic data does not

reveal 'who' an individual participant is, but 'what' they are, in phenotypic terms tall or short, well or ill, and so on. Similarly, environmental data does not show 'how' an individual participant lives, but to a limited extent 'what' their environment comprises, based on an historically given set of assumptions.

It is essential that projects collecting data adhere to ethical and legal requirements that protect the rights and privacy of the participants and their data. Locating phenotypes, environment and genotypes as objects of research disconnects these data from the social world and obscures the fact that these data comprise personal and intimate information about the individuals, and the families that participate in these projects. The Generation Scotland projects have been working out the problems and practices of data collection and the management of databases as research resources in this light. Central to the protection of privacy is the process of anonymisation. Anonymisation is used to disconnect identifiable personal details from all other information held about an individual. Its purpose is two-fold: firstly (and obviously), it protects the privacy of individual participants, and secondly, it facilitates connections between diverse data.

It is not possible to know a person from their data, but it is possible to know about this person, indeed, in the case of genotyping, it is probably possible to know things about someone that are unknown to themselves. The data are disconnected from the individual who has given it in such a way that it *should not* be reconstructed as recognisable information about that person. The databases are however designed in a way that allows the data to be connected

for analysis. In the end, the layers of disconnections appear to be a sleight of hand - scientific, organisational, technological and even ethical manipulations which serve to create an illusion of disconnection between participants, the social world and the science of medical research. The ultimate goal of GS databases lies in the connection of data. Connecting data will be explored in the next chapter.

Chapter 8

Making Connections: Genotypes, Phenotypes and Health Records

Generation Scotland creates connections. The GS collaboration connects people, disciplines, universities and research institutions, building an infrastructure for future genetic research in Scotland. Having discussed organisational connections and infrastructure in chapter one, I now want to consider what and how some of the other connections are being made, or will be made, in order to show how connections and reconnections are layered throughout the different dimensions of science, technology, the organisation, and the data. Notwithstanding previous attempts to disconnect these layers, they overlap and infiltrate each other thus creating a web of connections. The connections that GS is making are global in the sense that, as Collier and Ong argue, 'Global forms are able to assimilate themselves to new environments, to code heterogeneous contexts and objects in terms that are amenable to control and valuation.' (Collier and Ong 2005:11). Genes and digital bits of information are global forms, only limited by technical infrastructures, administrative apparatus or value regimes.

The Generation Scotland databases will have value in themselves but they do not stand alone. The utility of the genetic database can be extended when it can be related to other sources of data in other databases. The scientific, health and commercial value of GS lies not only in connecting the genotypic, phenotypic and environmental data held by GS, but also in its ability to connect to other data, in other databases. The scientific and technical connections of all these data

have ethical, legal and social implications, and it is particularly issues of confidentiality and privacy that come into focus at this point.

Confidentiality and privacy are key ethical and legal concerns underlying all these different types of connection and disconnection. I would argue that GS faces inherent problems with regard to preserving the confidentiality and privacy of participants. First, confidentiality is based on relationships of trust between people, but this is stretched beyond recognition when it is extended across organisations such as GS, the NHS, by data sharing and transfer. Second, the connection of information from diverse databases could render individual participants or families identifiable. The more information you connect together about an individual, or especially a family group, the more distinctive they become. The connection of phenotypic and environmental data might not be conclusive, but the connection of these data with unique genotypic data could be. This creates an inherent tension between the interests of the individual and medical science research. These are routinely dealt with by medical science research through anonymisation and the aggregation of data. Researchers can use aggregated anonymised data without consent from the individuals from whom the data is derived. It is assumed that the researchers have no interest in the data of a particular individual.

I begin the chapter with another source from which data will be collected into the GS phenotypic database: health record data will be extracted in an ongoing way from computerised health records over time. Health records are embedded in already existing layers of information technology within the health service. It

will be possible to connect those with GS because health record data are increasingly digitised and located within databases. Linking GS to health records to extract data raises issues of privacy and confidentiality. Different from the one-off collection of the blood sample, the questionnaire and physical measurement data, participants will not know when or what data is extracted. From the connections being made in this chapter, two types of risk emerge: risks of disease and threats to confidentiality and privacy. I then go on to examine the scientific reasons why it is useful to connect these data; and finally arrive at the end of the story 'from blood to data' with data analysis at the desktop. To reiterate my previous discussion, the aims of GS include identification of the genes associated with complex diseases, in order to contribute to a better understanding of the aetiology of disease and of genetic epidemiology. The disparate statistical data are (re)connected at the desktop for analysis of the relationships between genes, genotypes, phenotypes and environment .

Health Records: Access

Participants are being asked to consent to GS accessing their health records, past, present and future, and collecting data from these records. These data are needed to feed into the database in order to provide enough data from which meaningful statistical analysis can be drawn. The SFHS consent form asks participants to allow 'researchers access to information about my health and treatment from my past and future medical records', and the website states that 'The health of volunteers will be followed for up to 30 years by accessing their medical records.' (GS website 09.09.07). The 21CGH form, too, states that 'Consent will also be sought for access to and subsequent extraction of

anonymised information from patient records.' (Protocol 2006), and the GS website states

> 21CGH
> Another key aim of 21CGH is the development of data management systems, which will ensure that participant data are collected efficiently and stored securely. These systems will be designed so that the individual records can be automatically updated from health records in the future while maintaining participant anonymity.
>
> GS website 19.03.07

The statements on the website and on the consent forms make this access appear straightforward. But these brief statements belie both the complexity and the controversy of such practice within ethical and legal discourse with regard to confidentiality, privacy and indeed consent.

Health Records are regarded as highly confidential. Many people believe that only their doctor (GP or consultant) has access to these records, and that they keep these records in trust on behalf of their patients. In reality, there are many people within the NHS, other than the doctor seen at an appointment, who have access to these records. Any member of the health care team in Primary or Secondary Care can access health records if required. All these people are bound by confidentiality not to expose or discuss a patient's records with anyone not directly involved in their health care.

Almost all current health records in Scotland are computerised. Over the past ten years, practices and clinics have gradually shifted their record keeping onto computers. The health records of GS participants will therefore have been turned into electronic versions at some point during the last decade. In Scotland, the health records system most commonly used in general practice is the

General Practice Administration System for Scotland (GPASS), but a very small number of general practices have gone completely paperless using the Egton Medical Information System (EMIS). One would only expect to find a complete health record in electronic form in a paperless practice. Most practices have opted for a straight switch, leaving older records in paper form and storing only new information in electronic form. This, combined with the age range of the participants, e.g. in SFHS 35-55 years, means that the electronic health records of the participants are likely to be incomplete in many practices, which could have implications on the quality of data for GS.

Under the Data Protection Act it is permissible for patients to access their own health records. In the interest of investigating health records, I viewed my own primary care health record, on the grounds that it was likely to have similarities to other health records, not so much in content as in overall format. In order to do this, I had to submit a letter of request to the general practice I am registered with. Permission was granted and I was requested to make an appointment to view my records, and to pay a £10 fee. The fee would have been waived if I had had an appointment with the doctor in the previous 40 days, but as this was not the case, I paid the fee.

I was met by the practice manager, who showed me to an empty table in the administrative area of the practice. There, I was given a folder and asked 'will you want more than ten minutes?'. I replied that I thought I would want at least half an hour, which obviously surprised her, but she passed no comment. She then told me that a GP would be available if I had any questions and requested

that I inform her when I had finished, and return the folder. I did not see, nor was offered access, to the electronic version of my health records.

What is in a health record?

The folder contained seven pages of notes, a number of letters and a recent summary sheet, as well as results slips from laboratories, dating back to my initial registration with a GP in the UK. There was an assortment of different types of record that reflected different types of forms used by different practices (as I had moved around the country) and the transition from A5 sized envelopes to A4 sized folders to contain records. A summary sheet had been printed off and placed in the folder. There was also a summary sheet for medical investigations and operations (removal of appendix); and a record of health surveillance tests, immunisations, and blood tests. There was no record to show if my records had been used in research or for audit in any of the general practices I had been registered with.

My health records contained registration details (name, address, date of birth), NHS number, CHI number, as well as the clinical notes. The clinical notes comprised dates where I had attended for an appointment with a doctor, diagnosis and treatment, with an occasional comment. There was a note that my father had died of prostate cancer. There were details of pregnancy tests, anti-natal care, and the births of my children. Within the ethical and legal discourse on personal information, it is an important point that my record also includes information about other people.

Thinking about the GS projects and their request for consent to access participants' health records it was not obvious to me, exactly what type of data they might want to extract from my past records. All of my medical records were in paper form up to the year 2000, and it would require a researcher to go into the practice and source any data from the paper records prior to that time, which would be time consuming and labour intensive. If GS aims only to collect data from electronic records, the data will be restricted retrospectively to the point at which any general practice switched from paper records to electronic ones.

I asked the practice manager about research and researcher access to health records, and again I can only report the policy of that particular practice, from which others may vary to some degree. A researcher can request data for a study from the practice, the assumption being that it is clinical research. There are different ways that data can currently be collected from health records for research. A study can provide the GP with a specific question. The GP then extracts the data from the health records of patients in the practice and sends the data to the study. No personal information is ever collected or disclosed by the GP. Alternatively, a research nurse from a study can come into the practice and go through patient records to extract data, again no personal details are collected. The practice manager said that they had never been asked to hand over patient records to a study, the practice has sole custody of the records they hold on behalf of the patient. There is no external electronic access to patient records, these are kept on a secure closed system within the practice. The practice would only permit access to the health records of their patients with

explicit signed consent from the patient. As in this instance, I could therefore choose what I disclose and what I want to keep private. There are a couple of things regarding past medical events that I have chosen not to tell here, but I have no way of knowing if these pieces of information have been or might be collected from my records for research. Certainly no personally identifiable data should have been used, but I have no idea how far my confidentiality might be stretched.

Health Records: Uses and Safeguards

Health Records are held in primary and secondary care, this includes general practice, out-patient clinics and hospitals where a patient has been treated or admitted. The NHS in Scotland has a system linking and accessing electronic records for the transfer of test results. Referrals to Secondary care from general practice are traditionally made using letters, and discharges from hospital to general practice are accompanied by letters but there are moves towards the use of e-mail. There certainly has been a massive increase in the use of e-mail for communication between different health practitioners and clinicians that contain health information about patients, and patients themselves have begun to use e-mail to communicate with their doctors (Neville et al 2004).

The closed electronic health record systems in general practice means that they are only accessible within practices. Methods for collecting data from electronic health records have been devised for health research and auditing. This has involved a researcher going into the system to extract data. One of the ways this has been done was to write a computer program designed to collect data from

the electronic files. This could be copied onto a disk which could then be taken by a researcher, or sent to a practice, where the disk would be loaded into a practice computer and the program run to extract the data. This kind of data extraction is routinely done without the knowledge or consent of the patient. The anonymised data is used within the NHS or academic medical departments in evaluating provision of health care and practice performance.

The lines tend to be a bit blurred between NHS staff and medical researchers, but generally, clinical researchers (health professionals such as doctors and nurses) can access confidential data in a way that non-clinical researchers cannot. It is assumed that all health professionals are bound by confidentiality, but that non-health professional researchers are not. The differences lie in what data can be accessed and how. There is a distinction between research with identifiable data and that of anonymised data. Doctors act as the gatekeepers to patient records and information, and any non-clinical researcher is likely to have restrictions on access to any data. As in the case of the general practice previously mentioned, the doctor may extract the data and pass limited anonymised information on to a researcher. Generally, you would expect data to be anonymised before non-clinicians, such as statisticians, gain access to it.

Anonymisation assumes the removal of personal details so that an individual is not identifiable. However, the removal of personal details such as name, address and date of birth may not be sufficient, other distinctive features may be recognisable, or certain factors put together, may either reveal identity or make it possible to trace personal details. Anonymisation is a condition for disclosure

that does not require consent of the patient, and it assumes that any information that could allow the identification of a patient by any means has been removed. The disclosure of anonymised data assumes that the recipient of the data does not have access to a 'key' that could link to the identity of the patient.

Personal health information, and the uses it can be put to, are protected by the Cauldicott Guardians. The Guardians' key responsibilities are to oversee how NHS staff use personal health information and ensure that patients' rights to confidentiality are respected. Every Health Board appoints a Cauldicott Guardian, this is normally a senior health professional, for example the Medical Director. The Cauldicott Guardians are also responsible for overseeing the implementation of the new information management and technology strategies in NHS Scotland, in particular, the protocols that enable the sharing of data and the networked systems that are being developed across Primary and Secondary care.

Over recent years, a range of programs have been developed including Electronic Patient Records, Electronic Health Records, Electronic Clinical Communications, Telemedicine, Scottish Care Information, and Community and Preventive Care Systems. All of these are part of an overall strategy to utilise information technology to raise the levels of communication and information transfer and sharing across a range of clinical teams and health services. The CHI number, is the unique 'tag' for NHS Scotland, and facilitates the linking of patient information. The new systems of electronic information and technology would in theory make it easy for GS to collect personal medical

data. GS will hold personal details of its participants, including the CHI number which creates a link to electronic health records. Most of these changes are currently under development, with some under trial in different health boards across Scotland.

The custodianship of medical records and the numbers of people that have access to this information has become progressively more blurred. Traditionally, patients were familiar with the idea that 'their' doctor kept their medical record, that access to that record was restricted and that all information in the record was confidential. Increasingly, large numbers of health professional and other people, including for example information technology experts, have access to these data and this is set to expand as the new systems are networked across Scotland. The traditional paper record isolated from the wider world on a shelf in the local general practice is becoming a thing of the past. The medical information of individual patients will be accessible to anyone working in NHS Scotland. The NHS Scotland National Strategic Programme for Information Management and Technology 2001-2005 aims to exploit modern technology to establish secure electronic sharing of information in support of patient care, providing information and the development of an infrastructure. Health records don't feel private anymore.

The plans and development of extensive connections within the NHS system are also going to extend beyond the NHS, making connections with other agencies and organisations. There are plans to create links with, for example, social services where the sharing of personal patient data will be subject to agreed

information sharing protocols. Information may also be disclosed to organisations outside the NHS, for instance, pharmaceutical companies, using anonymised data. The Cauldicott Guardians are responsible for overseeing the disclosure of information and the use of personal information; however, the NHS Management Executive is setting up a Working Group to develop a national framework for the sharing of information between the NHS Scotland and other agencies.

> NHS Integration
> The Information Services Division of NHS National Services, Scotland and the National e-Science Centre (NeSC) are working together to create a research platform for emerging technologies to link health informatics and genetic research.
>
> NHS website 16.09.07

The Information Services Division is the custodian of NHS data, with very careful procedures to protect patient privacy in place.

All this information sharing may be good for patient care, but it also turns NHS Scotland into a huge networked resource for research. The GS (and UKBiobank) collection will be able to utilise this resource. NHS Scotland supports the GS database. There are also health professionals who work within the NHS and are collaborators in GS. However, GS is not part of the NHS nor does the NHS have any say over the running of GS, they exist as two separate organisations. GS has its own separate governance.

What this means is anonymisation of the data collected for the GS databases will not be absolute, and by implication protection cannot therefore be taken as

certain. Health record data will be collected by GS for years to come. In order to collect this data it will be necessary to access the personal details of participants which will have been stored in an isolated database. Linkage will be made between the GS databases using a coding system, which will act as a 'key', to access personal details. The personal details will be used to link to health records so that follow-up data from can be collected, and added to the phenotypic database. Who will hold the key is still under discussion. While it would be convenient for an individual within each project to hold the key, some of the collaborators feel that it should be held by an independent person.

Confidentiality and Privacy

Data sharing and transfer has raised questions about confidentiality and privacy, what they are and how they should be maintained. The concepts of confidentiality and privacy are used by GS, and medical research in general, to protect participants. One of the ways in which this can be done is by anonymisation of the data, by removing an individual's personal details. But, anonymisation of the data collected for the GS databases, as shown above, will not be absolute, and by implication protection cannot therefore be taken as certain.

Lowrance argues confidentiality is experienced as a relationship of trust between people

> Confidentiality is the respectful handling of information disclosed within relationships of trust, such as health care relationships, especially as regards to further disclosure. Confidentiality serves privacy.
>
> Lowrance 2002:8

The key feature of confidentiality here is the relationship of trust between people, not a person and an organisation or project. This refers to the role of the GP or clinician in recruitment and the crucial part that research nurses play in the collection of blood samples and data. These are the people engaged in the social interaction with recruitment and participants - the faces and names that participants trust. It is not that people are unaware that they are engaging with, for example one of the GS projects, or that their information is going to the project, but they do not form a social relationship of trust with an abstract idea of an organisation, or the many people that work within an organisation. Some organisations are deemed more trustworthy than others, for example, those that are readily identifiable as 'concrete' institutions such as hospitals and universities. But, in this setting, public engagement has shown a tendency to express mistrust of organisations such as pharmaceutical companies and insurance companies to act in people's best interests (Haddow et al 2004; Weldon 2007:68).

Confidentiality implies some sort of restriction on disclosure of information, that information will not be divulged, nor access to it permitted. Yet the rationale for creating the GS databases as a research resource is that many people, projects, and organisations will have access to and be able to use the GS data. The notion of confidentiality as a social relationship of trust between people is further confounded by the fact that GS projects, as described in chapter three, are an assemblage constructing a virtual space. An assemblage is not anything like as 'concrete' as an institution. Furthermore, virtual space is generally designed to be as accessible as possible, the only thing you can do is

ring-fence it with security measures that mark it as private. But, private to whom?

> Privacy is a status of information about aspects of a person's life over which he claims control and may wish to exclude others from knowing about. Stated as a right, privacy is the right of a person to control the collection, use, or disclosure of data about himself.
>
> Lowrance 2002:8

Privacy is then about power and a person's control of their information. This is however not straightforward or self evident, privacy is relative

> Privacy claims may or may not be conceded by others or guaranteed by laws. Privacy is a relative status, and claims to it must be negotiated against countering claims such as rights of others or collective societal goods.
>
> Lowrance 2002:8

As stated previously, the more information about individuals or, especially, a family group is linked, the more distinctive they become, and therefore more identifiable. Within the GS data collection projects, there are ethical practices and security measures to safeguard the confidentiality and privacy of participants. However, once the databases are 'open for business' as resources for research, there will be no way of knowing what information might already be held by researchers or projects. The ELSI team have some concern that projects and researchers who use GS data could already have data that contains personal details. When their data is linked to GS data, GS participants can become identifiable. Even without a coding key, 'identifiability' is a matter of degree (Lowrance 2002:29). Health data is particular to individuals, for example events such as operations, accidents, childhood illnesses like chicken pox,

pregnancies, and unusual tests or diagnoses all combine to create a distinctive picture of an individual. This health information, if connected with, for example, ethnicity, occupation, place of birth, NHS identifiers or postcodes, could make it possible to identify an individual. The interlinking of databases and powerful search engines makes identification easier (Lowrance 2002:29). The risks to privacy and confidentiality are taken very seriously by GS and strict measures have been put in place to protect these data.

Why connect the different types of data?

Connection refers to the way that data from an individual participant has been separated into different databases, and then linked back together for analysis. The digitisation of the different sorts of data into numerical or coded values facilitates connection for statistical analysis. In previous chapters, I focussed on the ways in which blood is disconnected from the body, DNA is then disconnected from the blood, and finally how data is disconnected from DNA to produce inscriptions of genotypes. As previously stated, knowledge based on genotype alone is limited, and it is necessary to connect the genotype to other data if we are to begin to understand the relationships between genes and complex disease. This is why GS draws on diverse types of data, on phenotype and environment, using a variety of sources, i.e. questionnaires, tests, measurements and, over time, health records. A brief explanation of the science is helpful to understand why Generation Scotland projects aim to connect genotypes, phenotypes and environment.

Connecting genotypes and phenotype

The relationship between genotype and phenotype is not straight forward. There is not always a simple one to one correspondence between a genotype and a phenotype. For example, the phenotype of the ABO gene is observed as blood type, but two different genotypes, AA and AO, both have the blood type A as phenotype.

The relationship between genotype and phenotype can be affected by the concept of penetrance. Penetrance refers to the probability that an individual exhibits a phenotype, given that the individual has the genotype for that phenotype, for example, the probability of developing a disease if an individual has the genotype for it. Single gene disorders are described as having complete penetrance, for example, an individual with the gene for Huntingdon's disease will develop the disease. Many disease genotypes, however, have incomplete penetrance, this means that it is less certain whether a person with a specific genotype will develop the respective disease (Carey 2003:69). In addition to penetrance, the relationship between genotype and phenotype is also affected by pleiotropy and variable expressivity. Pleiotropy occurs where a single gene can influence more than a single phenotype, producing multiple and different symptoms of a disease. For example, a hypothetical disease of the nervous system might affect the brain, producing symptoms of mental disorder, and at the same time affect control over muscles, producing motor disorder. Variable expressivity refers to a single gene producing a range of phenotypic values for a single trait, so that the phenotype of a particular disease might be expressed in

varying degrees of effect, and some individuals might have worse symptoms than others, even though they have the same genotype (Carey 2003:70).

Thus far, the relationship between genotype and phenotype has been confined to single genes. However, many diseases, such as cancer, chronic heart disease and mental illness are the effect of multiple genes which, when combined together, produce a single disease phenotype. The genes themselves do not combine, rather each gene produces a product, this could be a protein, an enzyme or a peptide, and it is the combination of these products that produces the phenotype. These phenotypes are categorised as complex or multi-factorial diseases, as previously mentioned in chapter three in relation to research design.

Multi-factorial diseases are also sometimes called oligogenic and polygenic; oligogenic refers to a case where only a few genes, maybe twenty, are implicated in a disease, while polygenic refers to diseases that involve many genes, possibly as many as fifty, in some cases more than a hundred. The analysis of multiple genes and their relationship with a phenotype is complicated, e.g. if three genes interact to produce a disease phenotype, and each gene were to have two alleles (variants), all possible combinations produce a total of twenty seven different genotypes in a population.

The discovery of the BRCA 1 and BRCA 2 genes for breast cancer was important, but they only account for 3% of all breast cancer cases. The EMSY gene is also implicated in breast cancer, and recently a fourth gene, BRIP 1, was reported as having an effect. Researchers expect that more genes are involved in

this disease (Cancer Research website 09.09.07). As the number of genes implicated in a disease phenotype increases, possible combinations increase exponentially. Furthermore, as is the case with BRCA genes, it is likely that not all genotypes will have an equal probability of manifesting the disease, some will produce a higher risk than others. It is extremely difficult to identify all the genes that may be involved in producing a complex disease, particularly those with smaller effects.

In order to determine the effects of multiple genes, especially weak effects, large numbers of genotypes need to be analysed together. This is one of the main aims of the GS database and sample collection: to make it feasible to carry out analysis on large numbers of genes to determine how many and to what extent they contribute to a particular disease. The connections between multiple genes in complex diseases, multiple genes and phenotypes, and multiple genes with environments require researchers to work with huge amounts of data and variables. This is one of the reasons why the research that will be done using the databases may in some cases take years of work before results are certain.

Predisposition and Heritability

Predisposition refers to a state of health where an individual shows no phenotypic symptoms but is at risk of developing a complex disease. The onset of complex diseases often does not occur until people are middle aged or older. The risk of an individual developing a complex disease might be calculated on the basis of a statistical model of liability. This is where linking to health records

will be useful. Once predispositions of risk have been calculated, surveillance of health records over time will reveal if participants develop the disease.

Heritability is defined as 'the proportion of phenotypic variance attributable to or predicted by genetic variance' (Carey 2003:290). Calculating a correlation coefficient between genotypic values and phenotypic values gives a value for heritability. This is taken as a quantitative index of the importance of genetics for individual differences in a phenotype. The correlation coefficient is the most common statistic used to quantify, for example, the similarity of relatives for a continuous genotype-phenotype and the extent to which one can be used to predict another. In other words, if there are two members of a family and one has a complex disease but the other does not - can the risk of developing the disease in the non-symptomatic individual be predicted from knowing genetic information about their relative? To work out the risk large numbers of cases are needed for analysis to be able to say anything of statistical significance.

Connecting Genes and Environment

Another aim in the creation of the Generation Scotland databases is understanding the interaction between complex genes for disease and environment. Environmental factors, as described in the previous chapter, can be problematic to qualify; and there could be an infinite number of environmental factors at work. It is difficult to identify causal environmental factors, and they are often not easy to measure and quantify. Similar to much phenotypic data, this makes it difficult to ensure good quality data, standardised practices, and consistency of data collection for compatible data

sharing or transfer. The UKBiobank project is designed to focus on environmental factors more extensively than GS.

As seen above, there may be multiple genes implicated in a given disease genotype, but the genotype may not manifest the expected phenotype, as the phenotype can be influenced by environmental factors. Complex genetic disorders are thought to be the effect of multiple genes and, possibly multiple, environmental factors, hence they are described as multi-factorial or complex diseases, sometimes represented as

genotype + environment = phenotype

The notion of interaction may be used both in a broad sense and in a specific, statistical sense. Broadly, both genes and environment contribute to a phenotype, so their relationship is always important. However, in the statistical sense, based on the concept of interaction in the analysis of variance, 'the actual relationship between the environment and a phenotype depends on the genotype, or equivalently, the actual relationship between a genotype and a phenotype depends on the environment' (Carey 2003:293); just how important this interaction is statistically is, however, generally unknown.

The interaction between genes and environmental factors may not be well understood, but is thought to be crucial in determining why an individual who has predisposing genes does or does not develop a particular complex disease. The effect - environmentability - may be calculated using the correlation coefficient between environmental values and phenotypic values, where

environmental values are quantified from environmental experiences. Environmentability is defined as 'the proportion of phenotypic variance attributable to or predicted by environmental variance' (Carey 2003:290). Complex disease may have many predisposing genes which can be affected by the environment of the cell and/or the organism.

The GS projects include the collection of some environmental data of the organism, i.e. the participants. Only some types and quantities of variables can be selected so as not to exceed manageable parameters. The selection tends to rely on the utility of already known causal factors, such as smoking, as seen in chapter seven. It is necessary to limit otherwise huge amounts of data, even though such a pragmatic reduction will restrict the variables available for research and the possible connections that can be made between genotype, phenotype and environment.

At the Desktop: connecting data

In the final stage of the story 'from arm to desktop', data arrives at the desktop of a researcher for statistical analysis. Whatever its source - genotyping in the lab, questionnaires, test results or medical records – the place where all this data ultimately connects together for analysis is at the desktop. Leaving all activities of collecting, processing, purifying and digitising the data behind, all there is now is the data set and the desktop computer. All the previous work funnels into this point.

This was the most difficult point on my journey from blood to data. Currently, genotypic data from the lab arrives at a study's desktop as electronic files of data by e-mail or disk. In addition to different ways of analyzing data, there are also different ways of managing these electronic files and keeping them secure. In order to seek permission to observe work at the desktop, I approached the principal investigators of several studies that were sending blood samples to the genetics lab at the time. My approaches were unsuccessful, not because the researchers were unwilling to talk about their work, but either because they were too busy, or, as was often the case, the data were still being collected and analysis had not yet started. Eventually, I was able to speak to two people who were working on analysing genetic datasets, and had the opportunity to have a brief observation with one of them. The first person, CG, was a clinical geneticist I met at a GS meeting, he worked at a different university hospital; the second, SG, was a statistical geneticist who had just joined the GS team.

At the time, they worked in different spaces. CG worked in an open-plan office in a research centre, and SG in an office adjoining a lab. Both offices were located in buildings on the grounds of a university hospital. CG's location emphasised the research into health care, and SG's location highlighted the scientific connection with a lab, until he subsequently moved to an office suite in another part of the hospital building. The researchers come to the data rather than the data going to the researchers, as the data was held on servers which were located within secure departments. The researchers had to enter through security systems of passes to get into the buildings, and security doors which required codes to gain access to different areas within the buildings. Both CG

and SG are nodal, in that they connected the variously collected and processed data. They also connected through networks: within the GS organisation, across the NHS and academic institutions, to other organisations, and to other research projects and databases.

The idea of observing someone work at the desktop seemed strange to them, which was indeed understandable, since there was not a great deal to 'see', as the work takes place either in the researcher's head or in the computer. As described by Latour and Woolgar (1987) piles of documents and publications were stacked on each side of the desks. But, instead of journals and published articles, these piles consisted of bulky manuals on computer programs and statistics, with only a few of scraps of paper with scribbled notes. Latour and Woolgar's informants were 'readers and writers of neuroendocrinological literature' (1987:56); by contrast, my informants, at the time I was there, were reading user manuals and not concerned with writing.

CG (the clinical geneticist) was also seeing patients, and as this took up most of his working day, he had to fit his research around clinic times; he either came into the office before the clinic started, or after the clinic was finished. This made finding a time when I could observe him difficult, because he was unable to predict when he might next get a chance to work on the datasets. After many attempts, I only managed to meet up with him twice. At that time, he was working on a dataset to validate findings from a study that had been carried out by a deCode research group in Iceland. He told me that the deCode group had 'found something they are very excited about and have asked us here in

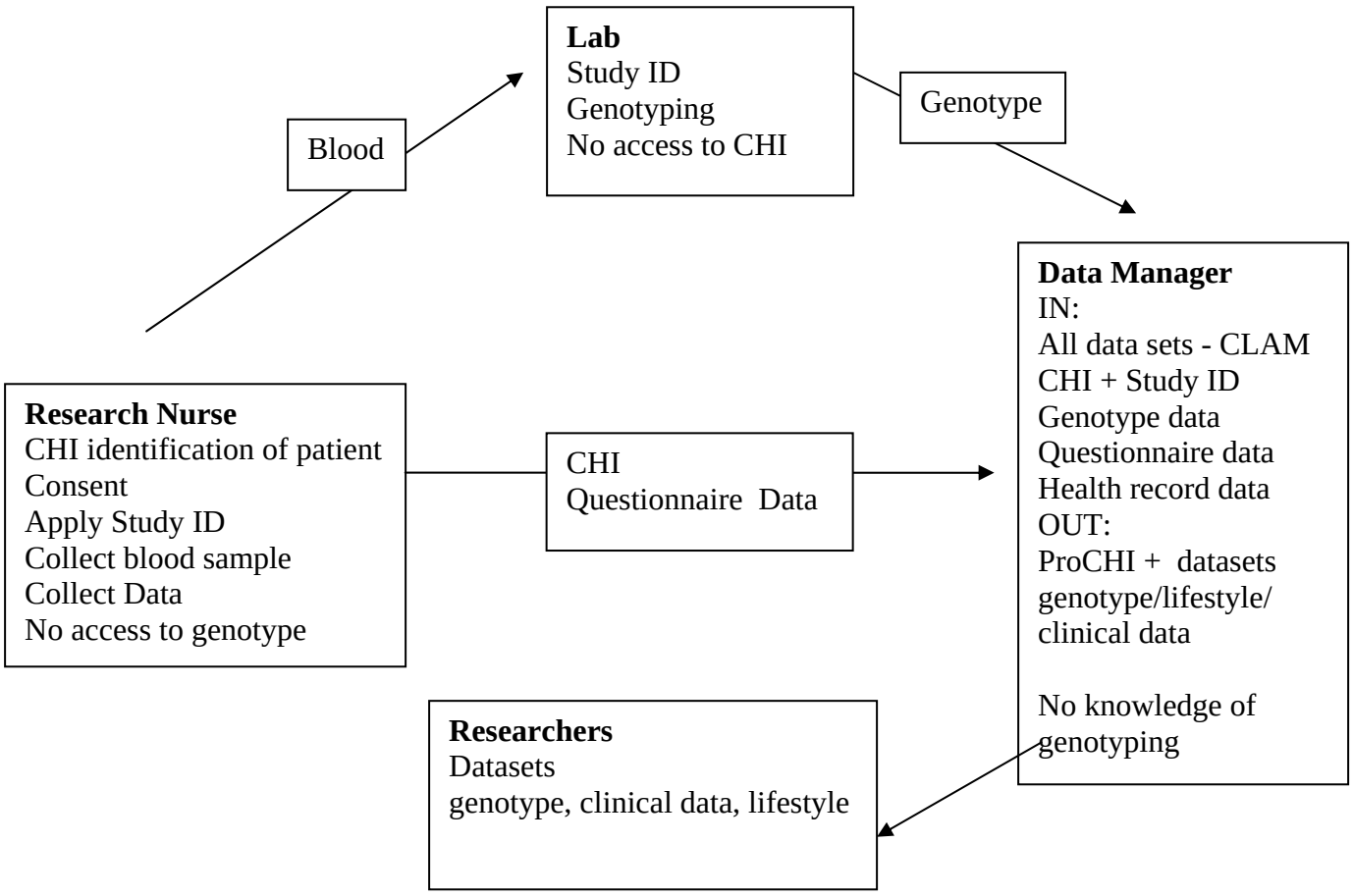
Scotland to run analysis on the same genetic marker to see if we come up with the same results'. CG had been using the Statistics Package for Social Scientists (SPSS), a general statistics package for analysis, but had recently changed to STATA (a brand name for a statistical analysis tool), a command-line computer program. He told me, 'I still need to work with the user guide - the instructions have to be typed into a dialogue box', as opposed to selecting a process from a drop down menu. 'STATA is a more powerful program, it can run the analysis of a greater number of cases - and faster than SPSS'. Dataset files were 'imported' into different computer programs for analysis.

CG explained how the datasets were managed at the institute where he did his research: the datasets were held by the data manager, who occupied a separate secure office, and was the only person with access to the servers that held the data. There were 25 datasets from different studies, each of which could potentially be merged for analysis. Anyone wanting to carry out analysis on a dataset was required to request the dataset from the data manager.

A system called CLAM was used on new data coming into the datasets, the raw data was checked and then – CLeaned, Anonymised and Mapped. Data came in with the previously discussed CHI number as its identification code; CHI numbers contain personal information, and were therefore removed by the data manager and instead, a ProCHI number was applied to each case. With the help of a scrap of paper, CG explained the data management system which was organised so that personal details and individual data were separated to ensure that no researchers ever had access to the personal details of participants. Figure

8.1, on the next page, is a copy of the diagram CG drew on a scrap of paper for me as he explained the system.

Figure 8.1 Data Management System (not GS)



CG was working with genetic markers which could be statistically analysed for association with many different variables or combinations of variables. He also worked on longitudinal data where changes could be graphed or plotted over time for markers that vary in rapidity of clinical changes or effects.

When I spoke with SG (the statistical geneticist), he had just joined the GS team and not actually started work on a human genetic dataset. I made the mistake of calling him a 'genetic statistician', but he sees a difference between being a statistical geneticist and genetic statistician: a statistical geneticist is a geneticist who uses statistical tools, as opposed to a genetic statistician, statistician who works on genetic datasets. At the time he thought it was likely he would be working on data from a study that was carried out by one of the principal investigators because there were as yet no GS datasets.

SG had just completed his PhD, which had involved working on a dataset of virtual mice, an idea that was as strange to me as my idea of observing people working at a desk was to him. He pointed out that it was more accurate to say his work had been done on virtual chromosomes, just as strange, but for me a less evocative image. He had used the virtual chromosomes to create and analyse computer simulations of genetic populations, and these data were then related to mouse data. He anticipated that 'working on human data, real data, would feel different to working with virtual data' - in part, because it would be human data, but also, because it would be 'real data' and he would have 'real results'. The processes of the statistical analysis, whether virtual or real, were identical, but he made a connection between real data and real people, and had the feeling that this work could 'help people'.

The work of this statistical geneticist 'involves running lots of analysis looking for relationships between genotypic and phenotypic data', that is, looking for gene variants and their possible associations with environmental or lifestyle

factors. 'Environmental/lifestyle factors may influence the phenotype, and we'll have to account for them in the analysis. To drop into statistical parlance, genotypes are independent variables, phenotypes are dependent variables, environmental/lifestyle factors are covariates' (E-mail from SG 14.09.06).

Switching between or conflating scientific and statistical terms while explaining something is not unusual. The digitisation of the data facilitates the mapping of statistical terms and analysis onto scientific data. In doing so, the data is further abstracted and disconnected from its origins in the social world. Theory and practice of both science and statistics require language and vocabularies that are highly specialised and specific to particular knowledge. Genetics and statistics share the desire to predict outcomes and the concept of probability. SG, like many other scientists, first had to learn to 'speak' statistics. I found it an intriguing, and also a confusing, fact that so much of the science of genetics, genetic research and genetic testing should in fact be expressed in the language of statistics.

SG liked the office in the lab where he was temporarily based, because it meant he could talk to the people working in a genetics research lab and see what they were doing. He had done biochemistry as an undergraduate but had not been near a lab since then. Working on genetic statistics had separated him from the lab, all his work was now done on a computer. He regarded working with statistics as a 'necessary evil', a side effect of having to work with large quantities of data in order to arrive at any conclusion; he did, for instance, tell

me that it is possible to run an analysis of a genotype with an environmental variable in one hour.

SG and CG are nodal both to the local and the global connections of data. The results of the analyses they both work on will potentially go on to create new connections. From the funnel point at the desktop, their results can spread out into public health policy, health care provision, the development of genetic tests and therapeutics within the health care sector and the commercial sector in Scotland and further away. There is the potential for many more people like SG and CG to work on analyses at a desktop almost anywhere through access to the databases or data transfer and sharing.

Conclusion

Connections are being made by people, technology and objects that have global forms. These have value for health, research and possible wealth creation of the population of Scotland. But, as the infrastructures and technologies develop, expand and increasingly link together, the confidentiality of an individual person is clearly stretched beyond recognition and privacy cannot be guaranteed.

The GS database, as a global assemblage, is configured through a set of connections that are based on existing knowledge. The configuration is structured through the mechanism of the database and the scientific ordering of the objects genotype, phenotype and environment. The structure imposes order

on a massive quantity of data and frames the possible relationships between these multiple data.

These connections are a central focus of legal and ethical attention. The ethical and legal concerns underpin the mechanisms that disconnect the various data in order to preserve the privacy of participants. The mechanisms to protect individual privacy have been formalised beyond the usual assumptions of confidentiality of data in a research setting.

Still under discussion is the question of whether there should be one centralised phenotypic/environmental/lifestyle database containing all the data or several in different research centres. There could be centralised databases with linked access for researchers to all the data, or alternatively several databases in different places that could be accessed and linked through a network. In the next chapter I continue to examine connections in a wider context.

Adapting to new assemblages, new connections, and the linking of diverse information takes time and is problematic. It has generated debates and contested areas both within and across perceptual boundaries as researchers, practitioners (from many disciplines) and the wider public (here in Scotland and globally) grapple with the possibilities and implications of population genetic databases. Information once digitised, is highly amenable to crossing boundaries, be they geographic or conceptual, and connecting with other information to create new spaces and assemblages for thinking about, for example, in this case health and illness.

Chapter 9

Making Connections Beyond Generation Scotland

Generation Scotland is creating a set of connections that extend outward beyond the organisation. It is a node, not of a single network but of diverse scientific, political, economic and social networks whose interests connect or conflict within and through GS. These connections are an integral part of the way in which ideas of health and illness have changed, and continue to change, in the light of genetic research both locally and globally. Genetic risk is central to the way in which ideas about health and illness have changed. Once the genes are identified, genetic risk can be calculated. Gene identification and a better understanding of how genes work, also opens up the possibility of new drug therapies. GS is therefore seeking to make connections with pharmaceutical companies in the commercial sector. Furthermore, GS is one of a growing number of projects internationally that are orientated towards similar goals of improving both population health and economic wealth. The success of these goals depends on the support of the public. GS needs the support of the Scottish public.

I begin this chapter by identifying the ways in which GS aims to connect with the Scottish public. The research is designed to identify genes that predispose people to particular diseases and calculate the risk of developing these diseases. I review what we already know about genetic risk and possible responses in the light of previous studies. The chapter then goes on to examine connections between GS and the commercial sector. The final section locates GS

internationally through a brief comparison with other population genetic database projects in other countries which shows that there are similarities and differences between these projects.

GS connections to the Scottish public

Part of the Generation Scotland project is to make connections with the public in Scotland. The key interfaces for this are through recruitment, participation, public engagement, benefit sharing and governance. Participation and interaction with research nurses, was discussed in Chapter Four. The GS proposal explicitly invokes the history of medical research in Scotland and a national pride to connect itself to the Scottish public. The GS website, which acts as a central forum for communicating with the public, suggests that a partnership exists between the organisation and the public:

> **Partnership with the people of Scotland**
> Generation Scotland is a partnership with all the people of Scotland and volunteers for the Scottish Family Health Study are welcome whatever their cultural, ethnic or social background. Every group has its own mix of genetic and lifestyle factors, and because the aim of the project is to work out how these factors affect a person's chances of getting certain diseases, Generation Scotland is encouraging people from every community to get involved.
>
> GS website 15.02.07

This statement indicates an inclusive approach, but groups and communities are not easily defined and notions of belonging are constructed in complex ways (Anderson 1991; Edwards 2003). The classification of people by researchers as group or community is problematic, indeed the statement itself implies

assumptions of difference. I examined this problem in Chapter Seven, in the collection of questionnaire data.

Involvement of the public is not confined to participation, but also includes a programme of public engagement to initially address people's concerns and reservations about participation.

> **Public Consultation Study**
> Public involvement is essential for the success of GS:SFHS and therefore an early and sustainable public consultation programme is a key component of the project. The aim of the programme is to understand and explain the public reaction to a wide range of relevant issues including genetics in healthcare, the use of bioinformation, and concerns surrounding consent and confidentiality
>
> GS website 15.02.07

The results from the initial phase of engagement (Haddow, Cunningham-Burley 2004) were used in shaping the Patient Information Leaflet (PIL), the Consent form, and concerns raised went on the Frequently Asked Questions (FAQs) page of the website. There is ongoing research into how GS is communicated, organised and governed.

The connection between GS and the public is modeled as a two way street. GS is demonstrating an active agenda of going out to engage the public, but also offers a range of ways in which the public can come to engage with GS. These ways of connecting are presented as straightforward, but once any of them are taken up, the resulting connection between an individual and the GS project will have implications for the individual and their families. Participation by giving blood and information places the individual and their family inextricably within

the matrix of samples and databases. Even though withdrawal is permissible, it might not be as clean and easy as it sounds, for example, if the data or sample has been used by a study they cannot be extracted from either calculations or results.

Participating in public engagement through interviews, focus groups or panels draws people in as a representative of themselves, their family, community and the public. The connection through governance is still being worked out. It is unlikely there will ever be representatives of the public on the Scientific Management Committee, but there should be representatives in the separate governance group.

In the UK, and specifically in Scotland, participation in health research is entirely voluntary. This brings us to one of the questions that arise from the development of this type of research – does the population that receives health care from the state have an obligation to participate in or contribute to health research? Rose and Novas discuss a notion of biological citizenship where 'Different ideas about the biological responsibilities of the citizen are embodied in contemporary norms of health and practices of health education.' (2005:440). They argue that, increasingly, states regard citizens as potential resources, that 'specific characteristics of the genes of groups of their citizens may potentially provide a valuable resource for the generation of intellectual property rights, for biological innovation and the creation of what Catherine Waldby calls '*biovalue*' (Rose and Novas 2005:441). The biological citizen appears to have multiple citizenships which may be a local Scottish citizenship, a broader European

citizenship, and a global citizenship. Each one implies a moral obligation to a community. Moreover, the biological citizen is caught at the interface between public spirited, altruistic giving in the interest of her fellow citizens and the commercial exploitation of her private DNA and information.

Genetic Epidemiology and Risk

Genetic epidemiology has implications for public health and risk from genes that predispose people to various diseases. The concept of risk of disease from predisposing genes connects the outcomes of the GS population genetic database research to the social world. Predictions of risk are likely to be one of the initial and principal outcomes of research using the databases, greatly increasing the number of risks of disease perceived as coming from within the body. In due course, other outcomes will follow from the gene identification, the development of genotype tests, the efficacy of drugs for particular genotypes, and drug development, to name a few.

It is commonly held that complex diseases are the result of an interaction between predisposing genes and environment, that these genes can be identified, and that the association between predisposing genes, and between genes and environmental factors can be statistically found. This information will facilitate the calculation of risk of complex diseases in the population, and will be used to design and promote preventive measures. The concept of 'risk' and risk analysis is problematic. The disclosure of information about genetic risk to the public, individuals and families involves complicated issues that are not completely understood.

Historically, public health has been concerned with external factors that affect health and illness, however, genetics offers a new source of 'risk' from within the body. Environmental variables are being connected by geneticists with genotypes to better understand the way that genes work, so that external factors (to the body) are being used to explain internal phenomena. Conversely, in genetic epidemiology, internal phenomena (to the body), namely genes, are being connected to external social and environmental variables to understand better disease aetiology and its implications for public health. Perceived sources of risk appear to have shifted from factors external to the body to phenomena within the body.

There are concerns within public health research that genes have a 'tendency to exclude consideration of social determinants of disease in epidemiological thinking' (McDermott 1998:1189). Exclude may be too strong a term, but certainly a focus on genotypes as causal phenomena in diseases can obscure or devalue social factors. Giving priority to the genotype has lead to a reductionist concept of causation which 'blames' the individual, a position that has been critiqued as 'black box' epidemiology (McDermott 1998:1192). The concern was that 'black box' epidemiology failed to distinguish between the 'causes of cases' and the 'causes of incidence', and the emphasis was placed on the risk of biological factors within the individual (McDermott 1998:1193). This had certain attractions in terms of explanation because it opened up the possibility of medical intervention, and by implication avoided the need to address more complex social or environmental factors.

This source of 'risk' from within the body opened up a different view of health and disease, which has led to concerns about the 'geneticisation' of individuals and the social world (Lippman 1992; Finkler 2000). In a deterministic view, the body became the carrier of the genes, a mere vehicle in which genes were of central importance, making the body but the phenotypic product of those genes. However, studying the interaction of genes and the environment links, or reconnects, the biological with the social. This suggests that a previously deterministic view of genetics is being superseded by a more holistic approach.

Population based studies constitute the field of genetic epidemiology, the 'core science' of public health (Shostak 2003:2327). The intersection of genetic research and public health has given rise to new models of 'gene-environment interaction in the production of human health and illness' redefining 'both genetic and environmental 'risks' and their potential implications for public health practice' (Shostak 2003:2327). Shostak argues that by focusing attention both 'inward, to the gene', and 'outward, to particular places', public health has created a tension between body and place 'through which culture and biology form a locally and historically situated dialectic and [this] raises important questions about the production of health and illness' (Shostak 2003:2327). Risk of disease may be calculated to find a statistical value. These statistical values have become increasingly important to current public health research and policy making, and are considered useful predictive indicators of complex disease occurrence in the population.

Shostak (2003) breaks the intersection of genetics and public health into three types: genetic epidemiology is concerned with the assessment of gene-environment interactions in the aetiology and progression of disease; molecular epidemiology uses biomarkers to assess exposure and effects of toxic substances; and toxicogenomics, the study of gene expression and gene products important in adaptive responses to toxic exposures

> The current iteration of environmental genetic research attempts to assess two different categories of risk, those posed by characteristics within the body (eg individual genetic susceptibilities) and those posed by the environment (eg chemicals or ionised radiation), and how they interact in the production of health and illness
>
> Shostak 2003:2329

In terms of the current GS agenda: gene identification is of initial interest for genetic epidemiology; molecular epidemiology supports the research into, for example, pharmacogenetics and drug development; and proteomics and cell lines will contribute to toxicogenomics.

The predictive power of genetics and statistics supports the remit of public health to target the prevention of disease in the population. Public health practice is directed to the modifiable risk factors for disease that interact with genetic variation and that may be used to help target prevention by interrupting the interaction of environmental cofactors with human genetic variation (Shostak 2003:2330). Identifiying risk in the population then facilitates public health practice through the development of interventions that encompass a range of techniques to target the population in general, or particular groups within the population that are considered to be at a higher risk.

The use of genetic technologies introduces the idea of 'at risk individuals'; through these mechanisms 'prevention then, is the surveillance not of the individual but of likely occurrences of disease, anomalies, deviant behaviour to be minimised and healthy behaviour to be maximised' (Rabinow 1992:242). However, genetic epidemiology and public health research, and indeed the Generation Scotland projects, do not provide feedback to individual participants on genotype or risk.

Risk of disease

The largest collection of GS blood samples and other personal information is being carried out, as previously stated, by the Scottish Family Health Study (SFHS). As a family-based study, the emphasis is on heredity. Heritability of disease is analysed from genetic information and family history. Risk can be constructed from both. People already have knowledge of traits and disease that 'run in the family', but how can genetic risk in the population be understood? Understanding genetic risk or family risk is, as studies have shown, problematic. These studies, some of which I will discuss in the following section, point to the ways in which risk can be interpreted and understood. Not just genes, but also family history becomes a risk factor for disease. I think that some of the issues and problems raised in these studies indicate problems that might be encountered in dealing with risk in the population, and might also be found in the recruitment of families, particularly in the role of the proband. The proband is the affected individual through whom a family with a genetic disorder is ascertained, and can find themselves in the position of being asked to recruit other family members. There are problems for both individuals and

families in generating genetic information, understanding the risk, and then disclosing this information.

Two Scottish studies have examined how families view heredity: the Midspan study of chronic heart disease (CHD) explored responses to constructing a family history by looking at lay knowledge and understanding of inherited risk (Hunt et al 2001; Emslie et al 2003); a study of family communication about genetic risk (Forrest et al 2003) examines how genetic information was shared in relation to genetic testing for disease. Both studies identify some of the features that concern families in Scotland about genetic information, family history and risk and lend themselves as a starting point for research into understanding the experiences of families that participate in GS - and also of families that don't.

The aim of the Midspan study was to understand if, or how, people view family history. This links to my discussion of the way in which the GS questionnaire asks for information in Chapter Seven. The Midspan study asked if people see themselves as having a family history of chronic heart disease or not. Kate Hunt, Carol Emslie and Graham Watt found that the way in which people perceived the risk and its relationship to family history can differ between health professionals and lay people where 'a reported family history is related to sex, social class, and parental deaths' (Hunt, Emslie, and Watt 2001:1168). There were also different versions among participants, some thought they had a family history of disease, some did not, and others were ambivalent. 'When deciding whether they had a family history [of disease], respondents considered the number of relatives affected, their age, and their relationship to the

respondents' (Hunt et al 2001: 1169). At times, participants had incomplete knowledge of family health and causes of death. They also differentiated between family risk and personal risk, 'respondents frequently made a distinction between inherited risk within their family as a whole and for themselves personally' (Hunt et al 2001:1170). The participants identified differences in their lifestyle from affected relatives or thought that they 'took after the other side of the family' (Hunt et al 2001:1170).

Families discuss family and relationships, locating individuals in terms of membership - 'a family history of disease is related to sex, social class and parental deaths' from a particular disease (Hunt et al 2001:1168). Participants tended to relate themselves and relatives to paternal or maternal sides of the family, based on a set of attributes. The common knowledge of heredity comes under three headings: discrete physical attributes; constitution; and personality and behaviour (Hunt et al 2001). These attributes are perceived as passed down and blended from both sides of the family and are related to assessments of illness and disease. The study also found some evidence to support the assertion that women are the 'kin keepers' of the family (Richards 1996; Pilnick 2002; Emslie et al 2003; Forrest et al 2003). Also, women were more likely to use the word 'genes' or 'genetics', 'Genes were used as a general term for biological transmission between generations, rather than understood as a physical entity with a specific location on a chromosome' (Emslie et al 2003:61). Genes were often perceived in terms such as 'stronger' or 'more of', when explaining particular attributes or the risk of disease.

The second study examined 'Family communication about genetic risk' and was carried out in the North East of Scotland, drawing on people diagnosed with either Huntingdon's disease, breast or ovarian cancer. The researchers found that 'Telling family members about genetic risk was generally seen as a family responsibility' and that 'family structures, dynamics and 'rules' influenced disclosure decisions' (Forrest et al 2003:319). Likewise, responsibility toward younger generations of the family was considered important. Parents were deemed to have the 'primary responsibility to pass on information to any offspring (adult or younger)' (Forrest et al 2003:320). There appeared to be a hierarchy of authority for passing genetic information, with women often taking the role of 'telling'. Telling was seen by the researchers as a process rather than an action and 'fell into two categories which we have termed pragmatism and prevarication' (Forrest et al 2003:321). Pragmatists had a more practical and active approach whereas prevaricators sought the 'right time' or the 'right opportunity'. Men were less likely to 'tell' but when they did, they took a more proactive approach, informing family members and initiating action. Family members were also found to 'trade' information with kin other than close family, e.g. cousins. Decisions about sharing genetic information, about what and who to tell, were influenced by ' the nature of pre-existing relationship, patterns of interaction, tension and rifts which act to promote or hinder communication about genetic (and any other) information' (Forrest et al 2003: 324).

Generating and disclosing risk

There is a problem with the collection of family history data in that it is 'part private, part public, possibly not entirely known or accurate' (Sachs 2004:28). The collection of genealogical data in relation to health and causes of mortality can lead to a view of the family as 'diseased'. Sachs reports that when participants, of a study on a new genetic consultation service in Sweden, were questioned about their relatives the information was recorded graphically in the form of a diagram showing kinship linkages. This type of representation had a powerful effect on some of the participants creating a perception of the family that had not previously existed - a 'picture' of a diseased family. The suggestion is that through the medicalisation of society this perception of family becomes dominant and creates a disruption to the continuity of the narrative of the family. Clearly the women interviewed were worried by the notion of 'carrying' a disease gene, and therefore being responsible for the transmission of the genes that may cause disease in their offspring and their descendants (Sachs 2004). The diseased family constructed relatives as risks.

Generating genetic information and whether (or how) to disclose this information to relatives raises ethical issues (Hallowell et al 2003:74). A study of the motives of women diagnosed with breast or ovarian cancer to undergo genetic testing, and on the experiences of disclosing information to relatives found that participants worked to balance autonomy with responsibility (Hallowell et al 2003:74). The study found that 'women view their role in generating genetic information for their relatives as less ethically contentious than disclosing this information to their kin' (Hallowell et al 2003:76). It appears

in this study that women felt an obligation to be tested in order to generate information that would be useful to other family members, at the expense of their own autonomy. These women 'conceive of themselves as selves in relation' (Hallowell et al 2003:76). The obligation was engendered in the light of their own diagnosis, these women were not 'healthy', and the connection made by 'selves in relation' overruled their own autonomy. Genetic testing was not seen as helpful to themselves, but rather appeared to be protective of family members, especially daughters to whom they may have passed the disease gene. These women had particular reasons for being tested. Genetic information was regarded as empowering in this study, because it could provide other individual family members with information that would arm them against the disease and enable them to make choices. It is unlikely that these factors will pertain to healthy individuals participating in GS. The possibility of individual or family empowerment is absent from the creation of the GS database as the genetic information will not be available to the participants.

Hallowell and co-authors make a connection between genetic testing and family history - genetic information is both 'information derived from molecular genetic testing and/or pedigree analysis' (Hallowell 2003:29). They go on to suggest that 'It is the familial nature of genetic information that distinguishes it from other types of medical information' (Hallowell et al 2003:29). This places this type of information in direct conflict with the assumption of autonomy on which consent to research, medical diagnosis and treatment are based. Autonomy of the patient/participant is taken to be the norm in the practice of health care and in medical research. Genetic information has, however,

implications for kin as well as for the individual. There is an ongoing debate as to whether genetic information should be seen as different from other medical information or not (Laurie 2000, EU Commission 2003:8, Nuffield Council on Bioethics, Pharmacogenetics 2007). The connection between genetic information and family history creates a contradiction between personal autonomy and family members' rights to know. This creates dilemmas for patients and health professionals, with regard to preservation of privacy and confidentiality versus the need to inform those at risk.

Understanding risk

The Risk Evaluation and Education for Alzheimer's Disease (REVEAL) study by Margaret Lock and others, shows that the knowledge of genetic risk does not necessarily affect people as much as might be expected (Lock et al 2006). The REVEAL researchers interviewed people who had received educational sessions on multiple causation of risk (including genes); half of the participants were then told their risk of late onset of Alzheimers disease due to the presence of ApoE gene. There are four ApoE alleles, with the $ApoE_4$ being the allele which shows the most susceptibility for late onset of Alzheimers disease. The effect of the ApoE gene is uncertain and people with this gene may not develop the disease; conversely, people who do not have the ApoE gene can also develop the disease. Those who were told their risk had grasped the idea of multiple causation and therefore appeared to be either relatively unconcerned or uncertain about how much risk they were at from the gene. Follow-up interviews with participants showed that those who had been told about their genetic risk had often forgotten what their level of risk was, and even which

gene contributed to their risk. The participants 'nested' genetic information into their existing knowledge (Lock et al 2006:290). Their existing knowledge came from diverse sources, such as experience, health professionals, support groups, a range of literatures, and even from the Internet (Lock et al 2006). The idea of 'blended inheritance' (Lock et al 2006:288) is consistent with the Scottish studies previously mentioned, a range of traits are taken into consideration when a disease is thought to run in the family. Can what we know about genetic testing and risk be useful in thinking about population risk?

Calculations for a population are based on relative risk, which can be uncertain. Lock et al (2006) identify several problems with calculations of relative risk: relative risk can be uncertain because in many studies the baseline is unclear; different studies produce different results; the risk can appear to have different ranges; and clinical studies can produce higher probability of risk due to the nature of the population sample (Lock et al 2006:282). The studies discussed above refer to single genes, risk from multiple genes (as they are identified) in complex diseases could give rise to even more confusion and uncertainty. Considering the uncertainty of who is at risk, and from what, based on single genes, such as the BRCA or APoE genes, it is a paradox that health care policy and practice appears to rely increasingly on genetic risk information as it pursues policies in preventive health care. If people feel uncertain and confused by information on genetic risk when undergoing genetic testing, it is likely that they will be even more confused and uncertain about population genetic risk. One of the possible responses to the confusion and uncertainty about risk is the formation of interest groups around particular genes or genotypes.

Biosociality

Paul Rabinow (1992) has described the emergence of 'biosociality' as the forging of collective identities under the emergent categories of biomedicine, 'new groups and individual identities and practices arising out of new truths' (Rabinow 1992:244). He has suggested that the groups or networks that were constructing these identities would most likely emerge not face to face, in the social spaces of community, but in the virtual spaces of the World Wide Web, in chat rooms and websites.

Disease groups and charities are showing a trend towards an increasing political activism and interest in having a voice in genetic research, for example, PXE International promotes research into pseudoxanthoma elasticum, an inherited disorder that effects the skin, and can also effect the eyes and blood vessels. Groups can form around 'orphan' diseases, those who feel at a disadvantage when only few people are affected, and these groups are forming alliances such as the Genetics Interest Group (GIG) in the UK. GIG is 'a national alliance of patient organisations with a membership of over 130 charities that support children, families and individuals affected by genetic disorders' (GIG website) and in the USA, Genetic Alliance is

> [A]t the crossroads of the genetics community—we provide a rich nexus of advocacy and community organizations, government, industry, and private entities. These interactions accelerate translational research; improve the climate for the development of technologies; encourage cohorts for clinical trials; increase the availability of linked, annotated biological resources; and ultimately lead to improved human health.
>
> Genetic Alliance website 22.06.07

Genetic research and a population genetic database is after all highly political, not just in terms of public health and policy making, but in the control and use of the resources and information, and eventually commercial exploitation.

There are however several factors which mitigate against biosociality in Scotland. First, organizations and support groups that already exist; for instance, many cancer support groups, specific and general, or the British Heart Foundation, are founded on the experience of disease and providing support for people with the diagnosis and their carers - 'endemic to this tradition is the valorization of experience and the (often justified) scepticism of the availability or good will of public agencies to solve what are widely perceived to be intractable problems' (Rapp 2000:286).

Secondly, a study of people in Scotland with asthma, which is a chronic disease, showed that they often resist medical categories and interventions. Patients, because they had been diagnosed with a chronic condition, received prescribed medication and had been advised about a management regime. The people interviewed in the study did a number of things: they rejected the categorisation of patient; they were insistent that they lived a normal life; and often, they were non-compliant in taking their medication (Steven et al 2004). Patients with chronic disease are notoriously bad at adhering to medication regimens and management plans, despite the best efforts of the health professionals that work with them (Steven, Morrison and Drummond 2002). Of course it is unreasonable to make sweeping generalisations about illness and disease because, as was found with the asthma study, there were 'degrees of' acceptance and resistance.

I would argue that if people with symptoms resist a categorisation of illness (Adam, Pill and Jones 1997; Steven et al 2004), then people without symptoms are likely to resist the categorisations of illness and disease even more. Researchers who work in the field of chronic disease would not anticipate that pre-symptomatic testing that identifies susceptibility-conferring genotypes would lead to the formation of biosocial groups in Scotland.

Thirdly, there is the question of health inequality. Biosociality suggests not only access to computing and other resources, but to a knowledge and confidence about being able to mobilise these. The health statistics for Scotland quite clearly show that the people that are most ill, and most at risk of disease, are located within particular geographic areas, and are categorised in the lowest socio-economic categories. In other words, they have the least access to resources and are least likely to be in a position to mobilise biosocial or political activity.

Finally, there is a phenomenon in Scotland known as the 'Scottish Effect' (Scottish Council Foundation 1998, Hanlon et al 2001). Scotland has a history of poorer health than England (and other Western European countries), measured by excess mortality rates. This has been attributed to higher levels of deprivation, but deprivation in Scotland has been reduced across the whole population over the past twenty years. Analysis based on census data for 1981, 1991 and 2001 shows that while deprivation could account for the differences in 1981, by the 1990s this was no longer the case, and by 2001 it accounts for less than half the difference (Hanlon et al 2005:199). The excess mortality in Scotland has remained unexplained. The Scottish Effect is not well understood, but it has

been suggested that there could be a cultural 'stoicism' in the way people view their health and illness. If so, it may well be related to the idea of resistance discussed above. What it suggests is that there could be a particular cultural interpretation of health in Scotland which produces an unwillingness to either report illness or be categorised as having a disease, and this is linked to the mortality rates.

How or whether biosociality emerges as a phenomenon in Scotland through the forging of new identities or groups in the light of genetic information remains to be seen. If it does, it is unlikely to be spread evenly, but could turn out to be a further indication of health inequalities or cultural preferences.

Connecting Family History and the wider world

There is also the potential for genetic research to connect to family narratives in a different way; not via health connections, but through the 'origins' of the population. Generation Scotland could produce information about origins through connecting genetic information to geographical information. There was a proposal to use the GS database to trace populations historically, but it was not intended to provide direct feedback to participants. Some studies have already used genotypes to trace populations and their movement, such as the work done by Cavalli-Sforza (2000). The Human Genome Diversity Project, a large scale global study of genetic diversity, met with mixed responses across the world. While some areas, particularly in Europe, welcomed the project, in other countries it met with vehement opposition. The attitude in Europe is attributed to a popular and widespread interest in genealogy, and ideas about

tracing ancestors. By contrast, in other countries, indigenous peoples were concerned to protect their cultural knowledge. Tutton examines the European attitude to 'historical sleuthing' (Tutton 2004:106) that constructs representations of genetic connections 'uncovering secrets about the past, resolving long debated questions about origins, or tracing continuity between people living today and their ancestors' (Tutton 2004:106). Families who have a history of diaspora, have a tradition of tracing their family roots back to the country of origin. Conversely, there is also an increasing awareness, from media representations, that those who stayed 'at home' may not be who they think they are. Genetics traces back further and in a generic way to origins which might not be accessible through records and documents. The awareness that there have always been waves of migration across land masses points to origins which had not previously been imagined. Tutton (2004) concludes that this interest in the past is widespread, as evidenced by TV documentaries, websites and increasingly accessible methods of searching old records often facilitated by technology.

In his study in the Orkneys, Tutton (2004) found that this can lead to an imagined exotic past, or an expectation of confirmation of family knowledge. Existing knowledge could be connected to genetic information in new ways to produce new narratives of relatedness, history and place. There is, however, a potential problem with 'finding out'. Patterns of exclusion or inequality may be reinforced, or new genetic communities may be created that could undermine known identities (Tutton 2004:116), creating new connections can disrupt existing connections.

Connecting to the Commercial Sector

Back in 1996, a strategy was laid out by Scottish Enterprise and the Royal Society of Edinburgh that 'aims to increase the contribution Scotland's science base makes to the economic wealth of Scotland' (1996:3), in a document entitled *Technology Ventures: Commercialising Scotland's Science and Technology*. Scottish Universities were identified as having a strong scientific research base, but this resource was under-exploited in terms of contributing to economic development through commercialisation. 'Commercialisation is defined as the process of converting this knowledge into marketable products and industrial processes' (Scottish Enterprise and The Royal Society of Edinburgh 1996:3). The document proposed that science and technology should contribute more to Scotland's economic development and included a vision of how this contribution could be increased, identifying initiatives and providing a framework for the next 10 years (Scottish Enterprise and The Royal Society of Edinburgh 1996:3). Envisaged as an increase in formal licensing deals, spin out companies, and the expansion of existing companies with the hope that 'one or two commercialisation projects turn into global winners' (Scottish Enterprise and The Royal Society of Edinburgh 1996:4).

What does this have to do with the Generation Scotland database, DNA or the analysis of genotypes? Scottish Enterprise and The Royal Society of Edinburgh have been involved in the development of the project and have supported the establishment of Generation Scotland as one prospective initiative. Thus Generation Scotland is committed to the process of commercialisation, through patents, licensing deals and the possibility of spin-off companies sometime in

the future. This blurring of distinctions, or building of bridges between two systems traditionally kept separate is echoed elsewhere in the world in the establishment and commercialisation of population genetic databases, as in Estonia, Iceland, Tonga, and Framingham. These have not all met with the anticipated success and reasons for this may vary.

The document may be a 'vision' but it has some concrete recommendations to make about the funding that is required, the establishment of links between academia and industry through said funding, the need to bring staff with commercial experience into academia and to train academics in commercialisation. The idea of 'visions' and 'dreams' is a recurrent theme in the prospecting for commercial 'products' or 'bioproducts' associated with molecular biology and genetic research. In Paul Rabinow's work DNA is seen as a valuable resource through which wealth can be created, a member of staff at Applied Biosystems Inc is quoted as saying 'The company's dream is to be the world's supplier of synthetic DNA' (Rabinow 1992:239).

The visions and dreams in these enterprises seem to be focused on the creation of wealth. In defence of scientific, technological, and medical researchers, I should point out that the creation of wealth seems to be a political aspiration in Scotland. The previously mentioned document, *Technology Ventures*, contains a set of tables summarising research findings; one is entitled 'Academic Motivation' and rates staff objectives, where 'Increasing my material wellbeing/income' scored the lowest – these people are not in it for the money, nor do many aspire to 'undertaking applied research with business'.

The connection between government funded academic research and the commercial sector, generally characterised as biotech and pharmaceutical companies in this instance, is not without controversy. The commercialisation of population genetic databases has raised ethical issues. The first project to raise these issues was DeCode in Iceland, and it has consequently received more attention than any other subsequent projects thus far. One of the concerns was that commercialisation of a population genetic database serves the interests of the commercial company more than it serves the interests of the public (Merz, McGee and Sankar 2004).

Further controversy arose over the issue of consent, which continues to be a focus for ethical debate. Consent covers different aspects of the research; here I want to single out consent for commercialisation. GS will generate income through Intellectual Property rights, patenting and licensing the use of datasets and samples. GS also regards the involvement of pharmaceutical companies as necessary for drug and treatment development although it will not accrue financial benefits directly from the marketing of new drugs, these will be retained by the pharmaceutical company. The SFHS Consent Form explicitly asks for consent to

- Allowing transfer and storage of a specimen of my samples for analysis in medical research on the assumption that it is free of any legal claim on my part (such as there may be) and without expectation of personal financial gain.

- The information or samples that I provide being used for the making of patent applications relating to the results of genetic studies. This will be without any payment to me, or my heirs, and without any individual acknowledgement of my contribution.

- The information or samples that I provide being used by the investigators and funding bodies (which have supported this study and which may have a financial interest in using the results of this study to develop diagnostic tests, new treatments or to target existing treatments more effectively) to develop collaborations which may involve commercial companies.

SFHS Consent Form 2007

The GS participants will be aware not only that there might be commercial involvement but are asked explicitly to consent to this. During the first phase of the GS public engagement, focus groups were asked for their views on commercialisation. While a good deal of ambivalence was expressed, there were some who felt that working with the commercial sector was necessary for the development of new drugs and treatments (Haddow et al 2004). GS has made drug development and understanding the relationship between drugs and genotypes one of the aims of the database. Connecting to the commercial sector through the exploitation of, for example, intellectual property have been integral to GS from the beginning.

GS has a different organisational model from the Icelandic population genetic database, which is being developed by a commercial company, and from other models, where a database belongs to the public sector but has a separate biotech company set up to exploit the resource, e.g. the Medical Biobank of Umea and UmanGenomics in Sweden. The GS database will be held in custodianship for the public, but it will also retain control of the commercial exploitation of the resource through the collaborating universities. A Memorandum of Understanding is being drawn up in order to manage this. Should these connections become commercially profitable a further connection would be

made to include the Scottish population through the mechanism of benefit-sharing.

GS is working to develop a model of benefit sharing so that, should financial profits accrue from exploitation of the database, there will be a mechanism to return a proportion to the Scottish population. Members of the Generation Scotland ELSI team have been working on this and have recently published a paper 'Tackling community concerns about commercialisation and genetic research: A modest interdisciplinary proposal' (Haddow, et al 2007). They argue that a model of 'benefit-sharing' that combines a sociologically informed model with a legal framework would work in Scotland. This would connect any financial benefits derived from the database and GS research projects to the Scottish population.

International connections

There are wider connections internationally between population genetic databases. These have occurred in two areas of research, namely the scientific and the ethical, legal and social implications. In the area of science, many Principal Investigators (PIs) know each other and collaborate in international research programmes, for example, the European Twin Study Network on Schizophrenia (EUTwinsS Project). PIs can make requests to each other to validate results, and they are likely not only to continue to collaborate but also to share data. PIs have a shared knowledge of the construction and management of databases, and many of the databases share similar features of design and organisation. For example, with regard to research design and databases, the

members of the Cartagene programme in Canada aim to promote the design of good epidemiological genetic research. The Cartagene programme is aiming to offer support to researchers finding their way around the available data and assistance with study design, as few of the researchers wanting to access their data are likely to be familiar with the way their database is organised (Cartagene meeting, Innogen 24.03.06). The European Union has also been actively involved in the promotion of genetic research, the development of guidelines and legislation which encourage the connection of genetic research and the sharing of knowledge across international boundaries. The moves towards standardisation of governance and quality of data aim to facilitate these connections.

The connections may produce similarities, but there are also differences in the models for creating and managing a population genetic database. These are the consequence of local socio- historical, political and economic influences. Each project is adapted to fit particular systems of organizing research and health care, particularly reflected in the way government policy is orientated and funding is allocated.

Generation Scotland is distinctive as a broad collaboration across institutions and organisations. Other population genetic databases are structured in different ways. The structure of each organisation is shaped by the sources of funding, level of government involvement and level of commercial involvement. Four projects are reported as having been the subject of government legislation in Iceland, Estonia, Latvia and Tonga. However, in the

case of Tonga the proposed legislation was withdrawn in the light of the public opposition to the deal the government was purported to have made with Autogen (Austin, Harding and McElroy 2003). The proposed Tongan database and the Icelandic database are the only ones that have been entirely in the commercial sector. Thus far there appear to have been three approaches to the creation of a population genetic database:

i) government funded and based in academic and medical departments of universities that anticipate co-operation with commercial enterprise e.g. the Scottish Family Health Study through Generation Scotland, the Medical Biobank of Umea in Sweden through Uman Genomics, and the Estonian Genome Project Foundation through EGeen

ii) government funded institutes created for genetic research with no expected commercial involvement e.g. The Genome Institute of Singapore, Cartagene in Canada, and Biohealth Norway

iii) commercial sector funded companies, backed by pharmaceutical companies, and licenced by government e.g. DeCode in Iceland and Autogen in Tonga

Each database works slightly differently, and most are associated with commercial enterprise at some point; the difference lies in the level of commercialisation and in the structure of its organisation. In Sweden, the genetic database was already in existence, created from the collection of samples and data for studies on cardiovascular disease and diabetes by the University of Umea and the County Council local health authority. Uman Genomics was set up as a company to commercialise the blood samples and data held on more than 100,000 individuals. This arrangement was predicated on a pre-existing data collection and there are many other collections in many countries which

could potentially be exploited in this way; that said, most are not of sufficient scale for the required statistical significance of population research. Most proposals for the creation of a population genetic database require the collection of new blood samples and data in sufficient numbers and specific to the purpose of creating a research resource.

Generation Scotland, which spans the NHS, universities and medical schools, is publicly funded, and essentially in the public sector. This is different from other population genetic databases in other countries which are being set up by genomics companies. A *Google* search for genomic companies produces a huge list of websites, with more than two hundred companies world wide on any one website. These include the companies that are creating population genetic databases - like deCode in Iceland, UmanGenomics in Sweden, Center for Information Biology and DNA Data Bank of Japan (CIB-DDBJ), The Estonian Genome Project Foundation, a non-profit organisation founded by the Government, to name only a few. In Estonia and Sweden, the commercial companies of EGeen and UmanGenomics have been set up to handle the enterprise aspects of the databases, while the samples and data will be collected and held in the public sector in a collaboration between the health service and university research centres. Biohealth Norway appears to have most similarities with GS in that it is comprised of the health service and several universities. However, there are two main differences: Biohealth Norway is run by the Department of Public Health, so firmly in the health sector, and there is no stated intention of commercialisation. It also aims to use databases that are already in existence from previous studies in addition to newly collected data.

The genomics companies have not always been successful in securing target database resources or populations. For example, Framingham Genomics was set up with the intention of obtaining an exclusive licence to commercialise the Framingham Heart Study database. The licence was not granted by the funding agency, the National Heart, Lung and Blood Institute (NHLBI), and Framingham Genomics was abandoned. Also, as previously mentioned, Autogen did not get the agreement that it wanted from the Tongan government due to public pressure and abandoned its proposal to create a population genetic database there.

Only two of the genetic databases, Iceland and Estonia, were the subject of government legislation when they were set up. Other databases have more indirect government involvement through funding, the health sector, or governance. Health sector involvement varies, much depending on how health care is provided in any particular country. Countries that have a government funded national health service provision for the delivery of care and research, have been involved in the creation of population genetic databases; for example in the UK, for both UKBiobank and Generation Scotland, NHS practitioners and resources are involved in recruitment; in Norway the data has been collected and is held by the department of public health; and in Singapore the Genome Institute is funded by the Singapore Economic Development Board. Cartagene in Canada is somewhat different in that it is not responsible for the collection of samples or data but has direct involvement of a government committee for governance.

Scotland, Iceland and Estonia have database proposals that explicitly link the creation of a population genetic database to the creation of wealth for the country, although it may be taken as implied through the involvement of the Economic Development Board in the case of Singapore, and the setting up of UmanGenomics in Sweden and Biohealth Norway. The creation of wealth is linked to the research through the concept of intellectual property, a legal concept that has come progressively more to the fore since the 1980s in association with the expansion of information technology and biotechnology.

Conclusion

Generation Scotland is creating a set of connections that extends outward beyond the organisation. These connections reach into the everyday world to effect, for example, families, economics and politics. This chapter has touched on just some of the connections that exist, or are under construction. GS is constructing connections with the Scottish population in diverse ways, and it fits with a political agenda that envisages the generation of health and wealth for Scotland. GS is also located within an international network of population genetic research.

Gene identification and calculations of genetic risk construct, and can potentially disrupt, connections within families. People already have a sense of characteristics and disease that 'runs in the family' or 'in the blood' that is part of the family narrative. Understanding risk is complicated. The knowledge that there is disease 'in the genes' can alter the perception of the family, and may even threaten its future continuity. Not just genes, but family history has

become a risk factor for disease. Moreover, people make their own connections between the scientific information and their experiences, and knowledge, of health and disease. A growing number of groups aim to influence research agendas.

All of these connections between population genetic research projects, between science and ELSI, between the public and the GS projects, between individuals and the research database, between the commercial sector and academic research, create a wider set of connections. The connections reach into the social, economic and political framework of Scotland and map onto, for example, different government policies, or the health of people in particular socio-economic categories. GS is likely to influence the shape of health politics and the expectations of the population with regard to health care in the next decades.

Conclusion

This thesis is about practices and processes, and shows how the formation of the Generation Scotland assemblage is the producer of disconnections and connections. The disconnections and connections are creating a framework of new referents between health and illness, identity and relationships in a way that rearticulates the body and the population. It traces the transformation and aggregation of heterogeneous elements which will become fixed in the population genetic database through scientific ordering and relationships which will be rendered immutable by the technology.

I began this project with two very basic questions. First, what is it that Generation Scotland is asking the public to engage with? And second, but related to the first - how do you transform blood into digital data? I see the two questions as related because GS is collecting thousands of blood samples from the population, both individuals and families. This is a particular form of engagement. The practices of collecting blood samples and the processes that then follow in the lab, as the samples are transformed into electronic data, reveal a series of disconnections. The phenotypic and environmental data also undergoes disconnection through collection, processing and management. The disconnections are achieved through manipulations that separate the blood and data from each other and a wider set of connections that exist within and beyond GS.

Researching and writing about a scientific project that does not yet exist but is still under construction proved complicated. I have pieced diverse aspects of this project together. I substituted the collection of blood samples from other studies. The lab work will be the same as GS but the samples I observed also came from other studies. I coupled these with the GS projects SFHS and 21CGH questionnaires as they were at the time, but may have been revised since. I extrapolated future health record data collection from my own health records and what I could find out about the changes underway in the NHS. These field data were woven together with i) the scientific rationale for GS, explanations and critique of the more generally used concepts, and ii) central ELSI issues. Together they create a complicated and multifaceted piece of work. My objective in bringing these various aspects of the database together is to demonstrate that engagement and/or participation in a population genetic database is not a simple matter of 'helping other people'. There are complex scientific concepts, power relations, and social and political issues all at work here simultaneously.

This thesis contributes to the overall knowledge of population genetic databases, and understanding of these as both social and scientific entities. It contributes to knowledge specific to

i) the collection of blood samples and the participation event as a point of connection between the project and the social world
ii) the collection of blood as the point of disconnection of substance from the body; the collection of other data as a point of disconnection from the social body, relationships and environment
iii) an understanding of the practices and processes that transform blood into digital data, and the laboratory setting in which these are carried out

iv) an understanding of what phenotype and environment mean in the context of a population genetic database
v) a better understanding of the technology, construction and design of a population genetic database, and the issues these collectively raise pertaining to ethical and legal matters such as confidentiality and privacy

More generally, the thesis contributes to knowledge of methodological and multi-disciplinary approaches to the topic of population genetic databases, and one way of crossing the 'divide' between science and the social world. It demonstrates the utility of the concept of assemblage, especially the way in which an assemblage accounts for the production of connections and disconnections. It also contributes to the notion of biosociality, by showing a version where it is the scientists who are having to come to grips with the social world in order to explain scientific objects. The thesis also demonstrates that attention to 'practice' in this setting was useful for understanding not only what is going on in the construction of a population genetic database, but also the way in which scientific 'objects' are 'produced' so that different versions can be disconnected in order to render them available for manipulation both in the lab and as digital statistical data.

Chapter One was the literature review, which showed that the literature on population genetic databases draws on a range of disciplines that includes science, medicine, bioethics, law and sociology. Population genetic databases are constructed and critiqued within a multidisciplinary environment that is still under development. It is a relatively small literature that is a side branch of a wider literature on the new genetics. I also sought a single authoritative

theoretical framework in the literature in the belief that this would give me a solid base to work from, but I could not find one that offered this across the diverse aspects of my project. Instead, I found a range of authors who had taken different approaches to a field that is changing and evolving by the day.

I was throughout both an individual researcher and a team member, and in Chapter Two I showed how I could not rigorously separate these roles, they intersected and often happened simultaneously. I discussed some of the constraints and tensions created by working within a multi-disciplinary project and the problems I encountered in trying to construct a synergy of perspectives across the 'divide' between science and social science.

In Chapter Three the analysis of GS as an organisation revealed a complex organisation with an unlimited cyberspace capacity to absorb digital data and develop new connections that could be national or international. The data is collected through different projects, initially SFHS and 21CGH. GS will therefore comprise not one single database but many, linked through an information technology platform. The design of genetic research shows how large numbers of cases are needed for research on populations that could inform public health policy and planning for a range of complex diseases. The results could in turn be used in the future to develop individual testing, diagnosis and personalised drug treatments.

Blood is a powerful mechanism for social connection, but as I show in Chapter Four, for research purposes it must be disconnected from the body both

physically and conceptually. The language of recruitment and participation within GS and more broadly in medical ethics invokes the concepts of the gift, obligation, reciprocity and altruism. These concepts derive from Titmus's work on blood donation in the UK and have a resonance for the collection of blood as samples, but this blood is not about saving a life it is about collecting information. Research nurses have a multifaceted role in the collection of blood and data for research. They work at the interface between the social world of the participants and the scientific world of the lab technicians. The research nurses in Generation Scotland form a node in the network of connections that are formed between participants and the projects. Research nurses have responsibility for checking that participants have understood the information they have been given about the project, and ensuring that they sign the consent form. The research nurses see names, dates of birth and personal information. They carry out the practice of venepuncture, label and dispatch the blood to the lab. They also collect data by administering tests, questionnaires and taking physical measurements. It is the research nurses that apply the unique identifiers that anonymise the blood samples and all other information.

Chapter Five moved the story into the lab. I began by asking what is a blood sample? The answer is unclear as to whether it is tissue, material or cells. The story continues with the arrival of the blood samples in the lab, the practice of booking-in, and the extraction of DNA from the blood. The blood samples collected by the research nurses arrive in the lab, not as a bright red gift, nor as the life saving substance of the blood transfusion service, but as *nasty dirty* stuff that has to be handled with protective gloves. The blood is moving away from

the people who have given it and extending the disconnection. The lab is located at a distance from the everyday and is a place where a particular type of work is undertaken. The practices of people - lab technicians, place - laboratory, and specialist equipment enact particular versions of the objects blood and DNA. The remains of the blood and the bugs are stripped away and the substance purified in order to produce nice clean DNA. These versions of blood (for information) and DNA (for inscription) have a value that is specific to the database as a resource.

Chapter Six examined how DNA is manipulated, put to work, to create digital genetic data. The practices are based on already existing knowledge and technology deployed to inscribe segments of DNA which are then used in analysis. DNA is put to work, denatured and inscribed both literally and figuratively in order to make digital data and other objects possible. Other versions of DNA appeared and were inscribed within 'black boxes'. Polymerase chain reaction (PCR) is a process through which the substance, DNA, is denatured and amplified. Sequencing and genotyping produce other versions of DNA as digital data. The polymorphism is one version, appearing as a new object from the process of sequencing. The genotype is another version produced through the process of genotyping. In the lab, practices and processes create further disconnections between the substance that I had started following and the end result, digital data. In social science there is an ongoing debate about genetic exceptionalism, whether DNA and the information derived from it should be treated in the same way as other information held in medical records, or not. There is no similar discussion about blood, whether it is or

should be exceptional blood. Giving blood for DNA extraction is about giving information, not just about an individual self but about related people too.

In Chapter Seven I turned to the ways in which phenotype and environmental information are being collected and inscribed as data. These data are collected in the form of questionnaires, physical measurements and cognitive tests. The chapter examined the tensions that were produced through the disconnections enacted in the construction of these data and through the practices of collection that turn individuals and families into data. But the chapter is also underpinned by a tension created for me by the pull in different directions according to disciplinary origins of science and anthropology. For example, phenotype and environment are taken as scientific objects, but the questionnaires could at the same time be viewed as cultural artefacts. The scientific practice of disconnection through collecting and ordering information in a particular way appeared reductionist in the light of my cultural understanding of the same information. A layer of disconnection was imposed through the anonymisation of the data at the time of collection. Anonymisation is the central concept in the processes of disconnection. It operates within medical science and between the science and the social world. Anonymisation is assumed to protect privacy, and disconnects the person from their (personal) information.

Chapter Eight shows how Generation Scotland creates many connections, a web of connections. The connections and reconnections are layered in the science, the technology, the organisation, and the data. The most important of these is the (re)connection of the genetic data with the phenotypic and environmental data,

the linking of vast amounts of data. The utility of the genetic database can be extended when it can be related to other sources of data in other databases, for example patient health records. The scientific, health and commercial value of GS lies not only in connecting the genotypic, phenotypic and environmental data held by GS, but importantly, in the ability to connect to other data, in other databases. The other aspects of connection include the connections between people, disciplines, universities, research institutions and the NHS, creating a Scottish infrastructure for future genetic research. The scientific and technical connections of all these data have ethical, legal and social implications. Here, in particular, the issues of confidentiality and privacy come to the fore.

Generation Scotland is creating new connections that extend outward beyond the organisation. Chapter Nine examines some of these wider connections. First, GS is actively seeking to connect with the Scottish public following moves across various fields of science to engage the public with the work of science. Second, the aims of GS include plans for the improvement of health and wealth of Scotland, but they also bring new ideas about risks that reach into the everyday social world. Third, GS is also connected within Scotland to a political agenda that combines health and wealth. Finally, GS is connected historically to other projects with similar aims and objectives that are happening simultaneously in other countries. All of these connections between population genetic research projects, between science and ELSI, between the public and the GS projects, between individuals and the research database, between the commercial sector and academic research, create a wider global web of connections.

Disconnections and Connections

Generation Scotland is constructing an infrastructure based on scientific expertise and connected through technology and funding. It is an assemblage, and as such it brings together heterogeneous elements into a global organisation. It works across traditional boundaries of place and people to create new disconnections and connections that separate information (genetic and personal) from the social world.

The disconnections are achieved through manipulations that are specific to this setting, they are more conceptual than real. The disconnections can only be created and maintained by the technology that makes up and encloses the virtual space of GS. But virtual space is infinitely porous and connections made to it can link the data within to other databases in other virtual spaces that are potentially global in every sense of the word. The risks to privacy mean no ultimate reconnection of the data to the individual is allowable. The disconnections will have to be maintained.

The disconnection of the data is achieved in two ways, through anonymisation and technology. But, it is not an absolute disconnection. There are two threads that keep it connected, one scientific the other social. The scientific thread of connection is to the health records of each participant in order to collect further data over time. This thread means that there is a mechanism for connection to another database domain. The other thread of connection is to the individual participant, and that thread is their DNA. This information only refers to one person; DNA is unique to every individual. The disconnections that were

constructed by the technology only act as a barrier between the data and the person. The information exists simultaneously in the scientific domain and the social world. The disconnection is only constructed by and in this setting, it is not fixed.

As a complex assemblage of institutions, places and people linked through technology, Generation Scotland is one of an emerging type of organisation in genetic research. The disconnections in the GS organisation are layered in two ways. The first is a matter of geography, there are people working on the GS projects in different places. The second is epistemological, the multidisciplinarity of GS personnel ranges across geneticists, lawyers, general practitioners, sociologists, research nurses, consultants, molecular biologists, public health researchers, IT specialists, and this social anthropologist, to name a few.

The disconnections created by the processes in the lab are mirrored in ideas about the disconnections between the science and the social world, for example, that the public(s) are disconnected from understanding the science; in kinship, genetics disconnects family members that are not biologically related; and in multidisciplinary work there are epistemological disconnections. The notions of disconnection and connection are layered through the organisation, the practices, and the public(s) of Scotland.

The metaphor of a parallel universe comes to mind because back in the 'real' world the connections between for example, families, participants, GS, and

health records, continue to exist. Any seepage from the virtual into the real world has the potential to create problems. Disconnections and connections are about more than cleaning up nasty dirty blood to get nice clean DNA and turning it into data. They are an integral part of the power relations and responsibilities of researchers to protect participants, especially their privacy, that exist or are being created through the work of GS within the field of medical scientific research, and political and economic policy.

Management and governance of these data will have to work to maintain the disconnection between the scientific data and the social world. This will become more complex as the possibilities increase for linking data with ever growing numbers of other databases that could be medical, academic, or commercial, both nationally and internationally. The wider the scientific connections are made, the more important it will be that connections to an individual cannot be made.

The connections are also layered: firstly, through the science and the linking of data; secondly, through the organisational collaboration; thirdly, by people moving around; and fourthly, by the technology. An agreement to co-operate in the sharing of funding, knowledge and resources, the collaborative connections mainly occurs at the planning, organisation and management levels. There are individuals moving between the layers creating connections. These individuals visit other places to share knowledge, solve problems, report progress or plan ahead but in unstructured ways. The people working 'on the ground' are the

least connected. The collaboration is also connected through technology, the core of which is the database(s).

The database resource is enacted differently in different settings, sometimes as an object in its own right, and sometimes as complex sets of data from diverse sources. It also has a coding system which will act as a key to link to health records for the continuing collection of data in the future. The lab produces and archives material samples which are at once a tissue or bio-bank and part of the database. Thus it is a complex set of connected data and samples rather than a discrete object. It has required those involved in the creation of the database to adapt existing knowledge or to develop new practices in the use and understanding of information technology. The data are processed into a scientific ordering but the data are not as fixed as they might appear, indeed the objects from which information are extracted are contingent on this setting and time. The data and the database are constructed by people who interpret them and give them meaning.

The combination of laboratory techniques/technology and information technology now make it possible to process samples and produce electronic results in files and formats that are easily transferable. Computer programs have been developed based on statistical methods for the analysis of large quantities of genetic (and phenotypic) data, and a new field of expertise is emerging, genetic statisticians, people with knowledge of molecular biology and trained in statistical analysis. The knowledge of molecular biology is necessary for the interpretation of the results, the statistical skills to manipulate the data. The

database creates the potential for many unknown connections to be made in cyberspace that may have implications across time and space. Donating blood, information, and consenting to access to health records in the creation of a database as a research resource means the data can be used repeatedly and the possible uses are infinite, unknown and ongoing.

The thesis serves as a starting point for future research about, for example

i) the way in which health, illness and medical research are increasingly shaped by statistical data and not knowledge of the body
ii) risk, understanding risk, and especially how risk in the population can or should be understood
iii) cell lines in this setting
iv) changes and modifications to the data collection since it has started across the three GS projects
v) whether the GS assemblage is becoming more structured or disaggregating, and how this might be happening
vi) recruitment rates and in particular non-participation
vii) how the 'partnership' between GS and the people of Scotland is working

and further into the future

viii) if there is an ongoing relationship between participants and GS, what form that might be taking
ix) the effects of research outcomes, such as, risk on those participating and non-participants
x) the final configuration of the database(s) and its utility
xi) early users of the database and biobank of samples - who, what for, and how they find accessibility
xii) commercial involvement, set up and uses of the database
xiii) practices and processes of benefit-sharing
xiv) and since the notion of biosociality in its broader sense is not developed in this thesis, future work could take this thesis as a starting point to investigate the utility of the concept as a mechanism for crossing the 'divide' in this setting

More broadly the thesis could serve as a starting point for comparative studies of other population genetic databases globally, the progress of their development and uses, with particular interest in the connections, or potential for connections between them, as well as more in-depth studies of data transfer and sharing especially of phenotypic data, environmental data and data from medical records.

As a new story this is a difficult one to read, it is complicated and crosses 'boundaries' in unfamiliar ways. The story lacks unity and wholeness, but that is because it reflects the features of an assemblage, which is defining new material, collective and discursive relationships. Generation Scotland is a site where the forms and values of collective and individual existence are at stake.

Bibliography

Adam, S., Pill, R. and Jones, A. 1997. Medication, chronic illness and identity: the perspectives of people with asthma. *Social Science and Medicine.* **45**, pp.189-201.

Anderson, A. 2002. In search of the Holy Grail: media discourse and the new human genetics. *New Genetics and Society*. **21**(3), pp.327-337.

Anderson, B. 1991. *Imagined Communities*. London: Verso.

Anderson, B. and Arnason, E. 1999. Iceland's database is ethically questionable. *British Medical Journal*. **318**, pp.1565.

Angier, N. 1988. *Natural Obsessions: The Search for the Oncogene*. Boston: Houghton Mifflin Company.

Anscombe, N. 2003. Venturing out. *New Scientist* , **178**(2397), pp.54.

Appadurai, A. 1986. Introduction: commodities and the politics of value. *The Social Life of Things: Commodities in Cultural Perspective.* Cambridge: Cambridge University Press.

Arnason, V. 2004. Coding and Consent: Moral challenges of the database project in Iceland. *Bioethics.* **18**(1), pp.27-49.

Arnason, G., Nordal, S. and Arnason, V. 2004. *Blood and Data: etical, legal and social aspects of human genetic databases.* Reykjavik: University of Iceland Press and Centre for Ethics.

Arnason, A. and Simpson, B. 2004. Refractions Through Culture: The New genomics in Contemporary Iceland. *Ethnos*. **64**(4), pp.533-543.

Atkin, K. 2003. Ethnicity and the Politics of the New Genetics: Principles and Engagement. *Ethnicity and Health*. **8**(2), pp.91-109.

Atkinson, P. 1995. *Medical Talk and Medical Work: The Liturgy of the Clinic*. London: Sage.

Atkinson, P., Parsons, E. and Featherstone, K. 2001. Professional constructions of family and kinship in medical genetics. *New Genetics and Society.* **20**(1), pp.6-24.

Austin, M. 2002. Ethical Issues in Human Genome Epidemiology. *American Journal of Epidemiology* Vol. 155, No. 7 : 585-592

Austin, M., Harding, S. and McElroy, C. 2003. Genebanks: A Comparison of Eight Proposed International Genetic Databanks. *Community Genetics.* **6**, pp.37-45.

Barbour, V. 2003. UK Biobank: a project in search of a protocol. *The Lancet*. **361**, pp.1734-1738.

Bates B., Lynch, J., Bevan, J. and Condit, C. 2005. Warranted concerns, warranted outlooks: a focus group study of public understandings of genetic research. *Social Science and Medicine.* **60**, pp.331-344.

Bell, D. and Bennett, B. 2001. Genetic Secrets and the Family. *Medical Law Review*. **9**, pp.130.

Bell, J. 2004. Predicting disease using genomics, *Nature*. **429**, pp.453-456.

Bell, N. K. 1996. Responsibilities and rights in the promotion of health: differing positions of the individual and the state. *Social Science and Medicine.* **43**, pp.775-782.

Bellis, M., Hughes, K., Hughes, S. and Ashton, J. 2005. Measuring paternal discrepancy and its public health consequences. *Journal of Epidemiology and Community Health*. **59**, pp.749-754

Bentley, D. 2004. Genomes for medicine. *Nature*. **429**, pp.440-445.

Berg, M. and Mol, A. (eds.), 1998. *Differences in Medicine: Unraveling Practices, Techniques, and Bodies*. Durham, NC: Duke University Press.

Beskow, L., Botkin, J., Daly, M., Juengst, E., Soleymani, L., Lehmann, J., Merz, R., Pentz, N., Press, L., Freidman Ross, J., Sugarman, L., Susswein, S., Terry, M., Austin and Burke, W. 2004. Ethical issues in identifying and recuiting participants for familial genetic research. *American Journal of Medical Genetics*. **130A**, pp.424-431.

Bradby, H. 2003. Describing ethnicity in health research. *Ethnicity and Health*. **8**(1), pp.5-13.

Burley, J. (ed), 1999. *The genetics revolution and human rights.* The Oxford Amnesty International Lectures. Oxford: Oxford University Press.

Busby, H. 2004.Blood donation for genetic research: what can we learn from blood donors' narratives?. In Tutton, R. and Corrigan, O. (eds), 2004. *Genetic Databases: socio-ethical issues in the collection and use of DNA,* London: Routledge.

Carey, G. 2003. *Human Genetics for the Social Sciences*. London: Sage.

Calsamiglia, H. and Van Dijk, T. 2004. Popularization discourse and the knowledge about the genome. *Discourse and Society*. **15**(4), pp.369-389.

Carlson, C., Eberle, M., Kruglyak, L. And Nickerson, D.2004, Mapping complex disease loci in whole –genome associated studies. *Nature*. **429**, pp.446-452.

Carsten, J. (ed), 2000. *Cultures of relatedness: new approaches to the study of Kinship*. Cambridge: University of Cambridge Press.

2004. *After Kinship.* Cambridge: Cambridge University Press.

Cattell, V. 2001. Poor people, poor places, and poor health: the mediating role of social networks and social capital. *Social Science and Medicine.* **52**, pp.1501-1516.

Cavalli-Sforza, L. 2000. *Genes, Peoples and Languages.* London: Penguin Books.

Clarke, A. and Parsons, E. (eds), 1997. *Culture, Kinship and Genes: towards cross-cultural genetics.* Basingstoke: Macmillan Press.

Clarke, A., Richards, M., Kerzin-Storrar, L., Halliday, J., Young, M.A., Simpson, S., Featherstone, K., Forrest, K., Lucassen, A., Morrison, P., Quarrell, O. and Stewart, H. 2005. Genetic professionals reports of nondisclosure of genetic risk information within families. *European Journal of Human Genetics*. **13**, pp.556-562.

Clayton, D. and Mackeigue, P. 2001. Epidemiological methods for studying genes and environmental factors in complex diseases. *The Lancet.* **358**, pp.1356-1360.

Clayton, E. 2003. Ethical. Legal and Social Implications of Genomic Medicine. *New England Journal of Medicine*. **349**, pp.562-569.

Cohen, A. 1985. *The Symbolic Construction of Community*. London:Routledge.

Collier, J. and Yanagisako, S. (eds), 1987. *Gender and Kinship: essays towards a unified analysis.* Stanford: Stanford University Press.

Condit, C. 2004. The meaning and effects of discourse about genetics: methodological variations in studies of discourse and social change. *Discourse and Society*. **15**(4), pp.391-407.

Cook, G., Pieri, E. and Robbins, P. 2004. The scientists think and the public feels: expert perceptions of the discourse of GM food. *Discourse and Society*. **15**(4), pp.433-449.

Corbett, M., Chambers, S., Shadbolt, B., Hillman, L. and Taupin, D. 2004. Colonoscopy screening for colorectal cancer: the outcomes of two recruitment methods., *Medical Journal of Australia*. **181**, pp.423-427.

Cox, S. and McKellin, W. 1999. There's this thing in our family: predictive testing and the construction of risk for Huntington disease. In P. Conrad and J. Gabe, (eds), *Sociological Perspectives on the New Genetics.* London: Blackwell Publishing.

Crary, J. and Kwinter, S. 1992. *Incorporations.* Zone 6. New York: Urzone Inc.

Csordas, T. (ed), 1994. *Embodiment and experience: the existential ground of culture and self*. Cambridge: Cambridge University Press.

Cunningham-Burley, S. and Kerr, A. 1999. Defining the social: towards an understanding of scientific and medical discourses on the social aspects of the new genetics. *Sociology of Health and Illness*. **21**(5), pp.647-668.

De Witte, J. and Ten Have, H. 1997. Ownership of Genetic Material and Information. *Social Science and Medicine*. **45**(1), pp.51-60.

Donovan, C., 2005. The governance of social science and everyday epistemology. *Public Administration*. **83**(3), pp.597-615.

Douglas, M. 1966. *Purity and Danger: an anlysis of the concepts of pollution and taboo.* London: Ark Paperbacks.

Edwards, J. 2000. *Born and Bred: Idioms of Kinship and New Reproductive Technologies in England*. Oxford: Oxford University Press.

Edwards, J. 2002. Taking public understanding seriously. *New Genetics and Society*. **21**(3), pp.315-325

Edwards, J. and Strathern, M. 2000. Including our own. In J. Carsten (ed) *Cultures of Relatedness; New Approaches to the Study of Kinship*. Cambridge: Cambridge University Press.

Emslie, C., Hunt, K. and Watt, G. 2003. A Chip Off the Old Block? Lay understandings of inheritance among men and women in mid-life. *Public Understanding of Science*. **12**, pp. 47-65.

Eriksson, S. and Helgesson, G. 2005. Potential harms, anonymization, and the right to withdraw consent to biobank research. *European Journal of Human Genetics*. **13**, pp.1071-1076.

Evans, W. and Relling, M. 2004. Moving towards individualized medicine with pharmacogenomics. *Nature*. **429**, pp.464-468.

Everett, M. 2003. The social life of genes: privacy, property and the new genetics. *Social Science and Medicine*. **56**, pp.53-65.

Finkler, K. 2000. *Expereincing the New Genetics: family and kinship on the medical frontier.* Philadelphia: University of Pennsylvania Press.

Finkler, K., Skrzynia, C. and Evans, J. 2003. The new genetics and its consequences for family, kinship, medicine and medical genetics. *Social Science and Medicine* **57**, pp.403-412.

Forrest, K., Simpson, S., Wilson, B.J., van Teijlingen, E.R., Mckee, L., Haites, N. and Matthews, E. 2003. To tell or not to tell: Barriers and facilitators in family communication about genetic risk. *Clinical Genetics*. **64**, pp.317-326.

Fortun, M. 2001. Mediated speculations in the genomics futures markets. *New Genetics and Society*. **20**(2), pp.139-156.

Fox Keller, E. 2000. *The Century of the Gene.* London: Harvard University Press.

Franklin, S. 1995. Science as Culture, Cultures of Science. *Annual Review of Anthropology*. **24**, pp.163-184.

2001. Biologization Revisited: Kinship Theory in the Context of the New Biologies. In S. Franklin and S. McKinnon, (eds). *Relative Values: Reconfiguring Kinship Studies*. London: Duke University Press.

2002. The Anthropolgy of Science. In J.MacClancy, (ed). *Exotic No More: Anthropology on the Front Lines.* Chicago: University of Chicago Press.

2003. Re-thinking nature-culture: anthropology and the new genetics. *Anthropological Theory*. **3**(1), pp.65-85.

Franklin, S. and McKinnon, S. (eds), 2001. *Relative Values: reconfiguring kinship studies.* London: Duke University Press.

Garcia-Sancho, M. 2006. The Rise and Fall of the Idea of Genetic Information. *Genomics, Society and Policy*. **2**(3), pp16-36.

Gellner, D. and Hirsch, E. (eds), 2001. *Inside Organisations: anthropologists at work.* Oxford: Berg.

Generation Scotland, 2006. Generation Scotland: Donor DNA Databank (3D). Study Protocol. 12.06.06

Genetic Health in the 21st Century (21CGH). 2003. Project Proposal.

Gertz, R. 2004. An analysis of the Icelandic Supreme Court judgement on the Health Sector Database Act. *SCRIPT-ed,* [online]. 1(2) [Accessed 28.06.04], pp241-258. Available from: <http://www.law.ed.ac.uk/ahrc/script-ed/index.asp>

Glasner, P. (ed), 2004. *Reconfiguring Nature: issues and debates in the new genetics.* Aldershot: Ashgate.

Glasner, P. and Rothman, H. (eds), 2004. *Splicing Life? The new genetics and society.* Aldershot: Ashgate.

Godard, B. 2004. *The Cartagene Project: A Partnership Approach with the Public and Partcipants.* Project Outline. Unpublished.

Gold, R.E., 1996. *Body Parts: Property Rights and the Ownership of Human Biological Materials*. Washington: Georgetown University Press.

Good, B. 1994. *Medicine, rationality and experience: an anthropological perspective.* Cambridge: University of Cambridge Press.

Goodman, A., Heath, D. and Lindee, M. S. (eds), 2003. *Genetic Nature/Culture: Anthropology and Science Beyond the Two-Culture Divide*. Berkeley: University of California Press.

Haddow, G., Cunningham-Burley,S., Bruce,A. and Parry,S. 2004. Public and Stakeholder Views from Focus Groups and Interviews.[online]. [Accessed 04.06.04]. Preliminary Generation Scotland Consultation Exercise 2003-4. Available from: www.generationscotland.org.

Haddow, G., Laurie, G., Cunningham-Burley, S. and Hunter, K. 2007. Tackling community concerns about commercialization and genetic research: A modest interdisciplinary proposal. *Social Science and Medicine.* **64**, pp.272-282.

Halliday, J., Collins, V., Aitken, M., Richards, M. and Olsson, C. 2004. Genetics and public health - evolution or revolution? *Journal of Epidemiology and Community Health*. **58**, pp. 894-899

Hallowell, N., Foster, C., Eeles, R., Arden-Jones, A., Murday, V. and Watson, M. 2003. Balancing autonomy and responsibility: the ethics of generating and disclosing genetic information. *Journal of Medical Ethic.* **29**, pp.74-79.

Hallowell, N., Arden-Jones, A., Eeles, R., Foster, C., Lucassen, A., Moynihan, C. and Watson, M. 2005. Communication about genetic testing in families of male BRCA1/2 carriers and non-carriers: patterns, priorities and problems. *Clinical Genetics.* **67**, pp.492-502.

Hanlon P., Walsh, D., Buchanan, D., Redpath, A., Bain, M., Brewster, D., Chambers, J., Muir, R., Smalls, M. and Willis, J. 2001. *Chasing the Scottish Effect – Why Scotland needs a step-change in health if it is to catch up with the rest of Europe.* Glasgow: Public Health Institute of Scotland.

Hanlon, P., Lawder, R., Buchanan, D., Redpath, A., Walsh, D., Wood, R., Bain, M., Brewster, D. and Chalmers, J. 2005. Why is mortality higher in Scotland than in England and Wales? Decreasing influence of socioeconomic deprivation between 1981 and 2001 supports the existence of a Scottish Effect. *Journal of Public Health*. **27**(2), pp.199-204.

Haraway, D.J. 1991. *Simians, Cyborgs and Women: The Reinvention of Nature*. London: Free Association Books.

1992. When Man Is on the Menu. In Crary, J. and Kwinter, S. (Eds). *Incorporations.* New York: Zone Books

1997 *Modest_Witness@Second_Millenium. Female_Meets_Oncomouse*. London. Routledge.

Häyry,M., Chadwick, R., Árnason, V. and Árnason, G. (Eds), 2007. *The Ethics and Governance of Human Genetic Databases: European Perspectives.* Cambridge: Cambridge University Press

Hedgecoe, A. 2000. The popularization of genetics as geneticization. *Public Understanding of Science*. **9**, pp.183-189.

2001. Ethical boundary work: Geneticization, philosophy and the social sciences. *Medicine, Health Care and Philosophy*. **4**, pp.305-309.

Helman, C. G. 1984. *Culture, Health and Illness.* Butterworth-Heinmann, Oxford.

Hine, C., 2001. Ethnography in the Laboratory. In Gellner, D. and Hirsch, E. (eds), 2001. *Inside Organisations: anthropologists at work.* Oxford: Berg.

2006. Databases as scientific instruments and their role in ordering of scientific work. *Social Studies of Science*. **36**(2), pp. 269-298.

Hjorleifsson S., Strand R. and Schei, E. 2005. Health as a genetic planning project: Enthusiasm and second thoughts among biomedical researchers and their research subjects. *Genomics, Society and Policy*. **1**(3), pp.52-65.

Hoeyer, K., Dahlager, L. and Lynoe, N. 2005. Conflicting notions of research ethics: The mutually challenging traditions of social scienctists and medical researchers. *Social Science and Medicine*, **61**, pp.1741-1749.

Holy L 1996, *Anthropological Perspectives on Kinship*. London: Pluto Press.

Hopper J., Bishop, T. and Easton, D. 2005. Population-based family studies in genetic epidemiology. *The Lancet.* **366**, pp.1397-1406.

Hunt, K., Emslie, C. and Watt, G. 2001. Lay constructions of a family history of heart disease: potential for misunderstandings in the clinical encounter? *The Lancet.* **357**, pp.1168-1171.

Irwin, A. and Michael, M. 2003. *Science, Social Theory and Public Knowledge.* Maidenhead: Open University Press.

Jackson, M. 2002. Familiar and foreign bodies: a phenomenological exploration of the human-technology interface. *Journal of the Royal Anthropological Institute*. **8**, pp.333-346.

Jasanoff, S., Markle, G., Petersen, J. and Pinch, T. (eds), 1995. *Handbook of Science and Technology Studies.* London: Sage Publications.

Jones, S. 1996. *In the Blood*. London: Harper Collins.

Kaye, J. 2005. A Guide to the Human Tissue Act. The Ethox Centre [online]. [Accessed 12.11.05]. Available from: http://www.ethox.org.uk/education/teach/HTAguide.htm

Kegley, J. 2004. *Challenges to Informed Consent* . European Molecular Biology Organization (EMBO). Report 5.

Kerr, A., and Cunningham-Burley, S. 2000. On Ambivalence and Risk: Reflexive Modernity and the New Human Genetics. Sociology, **34**, pp.283-304

Kerr, A., Cunningham-Burley, S. and Amos, A. 1998. Drawing the line: an analysis of lay peoples discussions about the new genetics. *Public Understanding of Science*. **7**, pp.113-133.

1998 The new genetics and health: mobilizing lay expertise. *Public Understanding of Science.* **7**, pp.41-60.

Kleinman, A. 1995. *Writing at the Margin: discourse between anthropology and medicine.* London: University of California Press.

Kleinman, D. 2003. *Impure Cultures*: *University Biology and the World of Commerce.* London: University of Wisconsin Press.

Knight, C. 1991. *Blood Relations.* London: Yale University Press.

Knorr-Cetina, K. 1995. Laboratory Studies: The Cultural Approach to the Study of Science. In S. Jasanoff, G. Markle, J.Petersen and T.Pinch, (eds). *Handbook of Science and Technology Studies.* London: Sage Publications.

Knudsen, S. 2003. Scientific metaphors going public. *Jounal of Pragmatics*. **35**, pp.1247-1263.

Konrad, M. 2003. From Secrets of Life to the Life of Secrets: Tracing Genetic Knowledge as Genealogical Ethics in Biomedical Britian, *The Journal of the Royal Anthropological Institute*. **9**(2), pp.339-358.

Kraljevic, S., Stambrook, P. and Pavelic, K. 2004. *Accelerating drug discovery*. European Molecular Biology Organization (EMBO). Report 5.

Kreiger, N. 1994. Epidemiology and the web of causation: has anyone seen the spider? *Social Science and Medicine.* **39**(7), pp.887-903.

Kreiger, N., Ashbury, F., Cotterchio, M. and Macey, J. 2001. A qualitative study of subject recruitment for familial cancer research. *Annals of Epidemiology***. 11**, pp.219-224.

Laderman, C. 1994. The embodiment of symbols and the acculturation of the anthropologist. In T. Csordas, (ed). *Embodiment and Experience.* Cambridge: Camebridge University Press.

Latour, B. 1987. *Science in Action: how to follow scientists and engineers through society* Cambridge Ma: Harvard University Press.

Latour, B. and Woolgar, S. 1979. *Laboratory Life: The Construction of Scientific Facts.* Princeton: Princeton University Press.

2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Laurie, G. 2000. Genetics and insurance: is it in the public interest to involve the law?. *SCRIPT-ED* [online]. [Accessed 20.06.07]. Available from: http://www.law.ed.ac.uk/script/newscript/GLGenandIns.htm

2002. *Genetic Privacy.* Cambridge: Cambridge University Press.

Laurie, G. and Gibson, J. 2003. *Generation Scotland: Legal and Ethical Aspects.* [online]. [Accessed 10.10.03]. Available from: http://129.215.140.49/gs/publications.htm

Levitt, M. and Weldon, S. 2005. A well placed trust? Public perceptions of the governance of DNA databases. *Critical Public Health.* **15**(4), pp.311-321.

Lewis, R. 2002. Culture and consent: Human tissues collected by anthropologists provoke increasing controversy. *The Scientist*. **3**(1), pp.1203-1204.

Lewontin, R.C., 1991. *The Doctrine of DNA: Biology as Ideology*. London: Penguin Books.

Lindee, M. S., Goodman,A. and Heath,D. 2003. Anthropology in the age of genetics: practice, discourse, and critique. In A.H. Goodman, D. Heath and M. S. Lindee (eds). *Genetic Nature/Culture: Anthropology and Science beyond the Two-Culture Divide*. Berkeley: University of California Press.

Lindenbaum, S. and Lock,M. (eds), 1993. *Knowledge, Power and Practice: The Anthropology of Medicine and Everyday Life*. London: University of California Press.

Lippman, A. 1992. Led (astray) by genetic maps: the cartography of the human genome and health care. *Social Science and Medicine*. **35**, pp.1469-96.

Lippman, A. 1998.The politics of health: geneticisation versus health promotion. In S. Sherwin (ed). *The Politics of Womens Health: exploring agency and autonomy*. Philadelphia: Temple University Press.

Lock, M. 2005. Eclipse of the gene and the return of divination. *Current Anthropology*, **46**, Supplement, December, pp.47--70.

Lock, M., Freeman,J., Sharples,R. and Floyd,S. 2006. When it runs in the family: putting susceptibility genes in perspective. *Public Understanding of Science* **15**, pp.277.

Lock, M., Young, A. and Cambrosio, A. (Eds), 2000. *Living and Working with the New Medical Technologies: Intersections of Inquiry.* Cambridge: Cambridge University Press.

Lowrance, W. 2001. The promise of human genetic databasaes, *British Medical Journal*. **322**, pp. 1009-1010.

2002. *Learning from Experience: Privacy and the Secondary Use of Data in Health Research.* The Nuffield Trust, London.

McDermott, R. 1998. Ethics, Epidemiology and the Thrify Gene: Biological Determinism as a Health Hazard. *Social Science and Medicine*. **47**(9), pp.1189-1195.

Macintyre, S. and Sooman, S. 1991. Non-paternity and pre-natal genetic screening. *The Lancet.* **338**, pp.869-871.

Macintyre , S. 1995, The public understanding of science or the scientific understanding of the public? A review of the social context of the new genetics., *Public Understanding of Science*. **4**, pp.223-232.

1997. Social and psychological issues associated with the new genetics. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **352**(1357), pp. 1095-1101.

Marcus, G. 1995. Ethnography in/of the world system: the emergence of multi-sited ethnography. *Annual Review of Anthropology*. **24**, pp.95-117

Marks, J. 2003. 98% chimpanzee and 35% daffodil: the human genome in evolutionary and cultural context. In A.Goodman, D.Heath and M. S. Lindee (eds). *Genetic Nature/Culture: Anthropology and Science beyond the Two-Culture Divide*. Berkeley: University of California Press.

Marsden, W. and Gertz, R. (with publisher). Of Law, Anthropology and Genetic Databases. In Tomlinson, K., Greene, A. and Garner, A. (Eds). *Anthropologists Working with Others: Translations, Transgressions, Transformations.* Oxford: Berg

Marsden, W., Steven, K., Sullivan, F.M., Neville, R.G., Hoskins, G. and Drummond, N. 2001. Goals
for living with asthma: balancing the options for a normal life. *European Respiratory Journal.* **18**, pp.2725.

Marteau, T. and Richards, M. (eds), 1996. *The Troubled Helix.* Cambridge: Cambridge University Press.

Mascarenhas-Keyes, S. Understanding the Working Environment: Notes Towards a Rapid Organisational Analysis. In Gellner, D. and Hirsch, E. (eds), 2001. *Inside Organisations: anthropologists at work.* Oxford: Berg

Mason, J., McCall Smith, S. and Laurie, G. 2002. *Law and Medical Ethics*. Sixth Edition. London: Butterworths LexisNexis.

Mauss, M. 1990 Translation [1954]. *The Gift.* London: Routledge.

Meade, T. 2003. The future of Biobank, *The Lancet*, **362**, pp.492.

Medical Research Council (MRC), 2001. *Human tissue and biological samples for use in research: Operational and Ethical Guidelines*. MRC Ethics Series.

Merz, J. F., McGee, G. and Sankar,P. 2004. Iceland Inc.?: On the ethics of commercial population genomics. *Social Science & Medicine.* **58**(6), pp.1201-1209.

Mittra, J. 2006. Genetic exceptionalism and precautionary politics. *Science and Public Policy.* **33**, pp.8.

Mol, A. 2002. *The Body Multiple: Ontology in Medical Practice.* Durham, NC: Duke University Press.

Mol, A. and Berg,M. 1998. Differences in medicine: an introduction. In M. Berg and A. Mol (eds). *Differences in Medicine: Unraveling Practices, Techniques, and Bodies*. Durham, NC: Duke University Press.

Morrow, J. 2001. Data protection and patients consent. *British Medical Journal*. **322**, pp.549.

Moss, L. 2003. *What genes can't do*. London: The MIT Press.

Nader, L. (ed), 1996. *Naked Science: anthropological inquiry into boundaries, power and knowledge.* London: Routledge.

Nash, C. 2004. Genetic Kinship. *Cultural Studies*. **18**(1), pp.1-33.

Nelkin, D. and Lindee, M. S. 1995. *The DNA Mystique: the gene as a cultural icon.* New York: W.H. Freeman and Company.

Nelson, S. and Hollingham, R. 2003. *How to Clone the Perfect Blonde: making fantasies come true with cutting edge science.* London: Ebury Press.

Nerlich, B., Dingwall, R. and Clarke, D. 2002. The book of life: how the completion of the Human Genome Project was revealed to the public. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*. **6**(4), pp.445-469.

Nerlich, B., Dingwall,R. and Martin,P. 2004. Genetic and genomic discourses at the dawn of the 21st century. *Discourse and Society*. **15**(4), pp.363-368.

Neville RG, W Marsden, C McCowan, C Pagliari, H Mullen and A Fannin. 2004. E-mail Consultations in General Practice. *Informatics in Primary Care* **12**; 207-14

Nilsson Sojka, B. and Sojka, P. 2003. The blood-donation experience: perceiving physical, psychological and social impact of blood donation on the donor. *Vox Sanguinis.* **8**, pp. 120-128.

O'Neill, K. 2003. A vital fluid: risk, controversy, and the politics of blood donation in the era of mad cow disease. *Public Understanding of Science*. **12**, pp. 359-380.

Ong, A. and Collier,S. (Eds) 2005. *Global Assemblages*. Oxford:Blackwell.

Ottman, R., Berenson, K. and Barker-Cummings, C. 2005. Recruitment of Families for Genetic Studies of Epilepsy. *Epilepsia*, **46**(2), pp. 290-302.

Palsson, G. 2007. *Anthropology and the New Genetics*. Cambridge: Cambridge University Press.

Palsson, G. and Hardardottir, K. 2002. For Whom the Cell Tolls: Debates in Biomedicine. *Current Anthropology*. **43**(2), pp.271-301.

Palsson, G. and Rabinow, P. 2005. The Iceland Controversy: Reflections on the Transnational Market of Civic Virtue. In Ong, A. and Collier,S. (Eds). *Global Assemblages*. Oxford:Blackwell.

Parry, B. 2004. *Trading the Genome: Investigation of the Commodification of Bio-Information*. New York: Columbia University Press.

Petersen, A. 2001. Biofantasies: genetics and medicine in the print news media, *Social Science and Medicine.*, **52**, pp.1255-1268.

2003. *The new genetics and citizenship.* London School of Economics BIOS, Vital Politics Conference I, 5th, 6th 7th September 2003, London. Available from: http://www.lse.ac.uk/collections/BIOS/vital_politics_papers.htm

2005. Securing our genetic health: engendering trust in UK Biobank. *Sociology of Health and Illness*. **27**(2), pp.271-292.

Pfeffer, N. and Laws,S. 2006. Its only a blood test: What people know and think about venepuncture and blood. *Social Science and Medicine.* **62**, pp.3011-3023.

Pickering, A. (ed), 1992. *Science as Practice and Culture.* London: The University of Chicago Press.

Pilnick, A. 2002, *Genetics and Society.* Buckingham: Open University Press.

Plows, A. 2004. Strange bedfellows and usual suspects: mapping emergent complexity of social movement society engagement with human genetic technologies. Seminar paper. Innogen, Edinburgh. 15.03.04

Plows, A. and Boddington,P. 2006. Troubles with biocitizenship? *Genomics, Society and Policy.* **2**(3), pp. 115-135.

Popham, F. 2006. Is there a Scottish Effect for self reports of health? Individual level analysis of the 2001 UK census. BMC Public Health, Open Access [online]. [Accessed 08.08.07]. Available from: http:// www.biomedcentral.com/1471-2458-6-191

Rabinow, P. 1992. Artificiality and enlightenment: from sociobiology to biosociality. In J. Crary and S. Kwinter (eds). *Incorporations*. New York: Zone Books.

Rabinow, P. 1996. *Making PCR: A Story of Biotechnology.* Chicago: Chicago University Press.

1999. *French DNA: Trouble in Purgatory.* Chicago: University of Chicago Press.

2000. Epochs, Presents and Events. In Lock, M., Young, A. and Cambrosio, A. (Eds), *Living and Working with the New Medical Technologies: Intersections of Inquiry.* Cambridge: Cambridge University Press.

2003. *Anthropos Today: reflections on modern equipment* . Woodstock: Princeton University Press.

Rapp, R. 1999. *Testing Women, Testing the Fetus: The Social Impact of Amniocentesis in America.* London: Routledge

2000. Extra Chromosomes and Blue Tulips: Medico-Familial conversations. In A. Cambrosio, M. Lock and A. Young (eds). *Living and Working with the New Medical Technologies.* Cambridge: Cambridge University Press.

2002. Genealogical Dis-ease: Where Hereditary Abnormality, Biomedical Explanation, and Family Responsibility Meet. In S. Franklin and S. McKinnon (eds). *Relative Matters: Reconfiguring Kinship Studies*. London: Duke University Press.

Rapport, N. and Overing, J. 2000. *Social and Cultural Anthropology: the key concepts.* London: Routledge.

Reardon, J. 2001. The Human Genome Diversity Project: A Case Study in Coproduction. *Social Studies of Science*. **31**(3), pp.357-388.

Rich E., Burke, W., Heaton, C., Haga, S., Pinsky, L., Short,P. and Acheson, L. 2004. Reconsidering the Family History in Primary Care. *Journal of General Internal Medicine*. **19**, pp. 273-280.

Richards, M. 1996. Lay and professional knowledge of genetics and inheritance, *Public Understanding of Science*. **5**, pp.217-230.

1996. Families, Kinship and Genetics. In T.Marteau and M.Richards, (eds). *The Troubled Helix.* Cambridge: Cambridge University Press.

Ridley, M. 1999. *Gemone: the autobiography of a species in 23 chapters*. London: Fourth Estate.

2003. *Nature via Nurture: genes, expereince and what makes us human.* London: Fourth Estate.

Rose, N. and Novas, C. 2005. Biological Citizenship. In Ong, A. and Collier,S. (Eds). *Global Assemblages*. Oxford:Blackwell.

Sachs, L. 2004. The new age of the molecular family: an anthropological view on the medicalisation of kinship., *Scandinavian Journal of Public Health*. **32**, pp. 24-29.

Savitz, D. 1994. In defense of black box epidemiology. *Epidemiology*. **5**(5), pp.550-552.

Schneider, D. 1984. *A Critique of the Study of Kinship.* Ann Arbor: The University of Michigan Press.

Shakespeare, T. 1998. Choices and Rights: eugenics, genetics and disability equality. *Disability and Society*. **13**(5), pp.665-681.

1999. Losing the plot? Medical and activist discourses of contemporary genetics and disability, *Sociology of Health and Illness*. **21**(5), pp.669-688.

Shostak, S. 2003. Locating gene-environment interaction: at the intersections of genetics and public health. *Social Science and Medicine*. **56**, pp.2327-2342.

Simpson, B. 2000. Imagined genetic communities. *Anthropology Today.* **16**(3), pp.3-6.

2003. (ed). Between Macroethics and Microrealities. Association of Social Anthropologists of the Commonwealth Conference Paper. Wellcome Trust.

Singer, E., Corning, A. and Lamias, M. 1998. Trends: Genetic Testing, Engineering, and Therapy: Awareness and Attitudes. *Public Opinion Quarterly*. **62**(4), pp. 633-664.

Singer, P. 1993. *Practical Ethics.* Cambridge: Cambridge University Press.

Smith, B., Watt, G., Campbell, H. and Sheikh, A. 2006. Genetic epidemiology and primary care. *British Journal of General Practice*. **56**, pp. 214-221.

Steven, K., Marsden, W., Neville, R. G., Hoskins, G., Sullivan, F. M. and Drummond, N. 2004. Do the British Guidelines for Asthma Management facilitate concordance? *Health Expectations.* **7**, pp.74-84.

Steven, K., Morrison, J. and Drummond, N. 2002. Lay versus professional motivation for asthma treatment: a cross-sectional qualitative study in a single Glasgow general practice. *Family Practice.* **19**, pp.172-177.

Strachan, T. and Read, A. 2004. *Human Molecular Genetics 3*. London: Garland Science.

Strathern, M. 1992. *After Nature: English Kinship in the Late Twentieth Century*. Cambridge: Cambridge University Press.

1995. Future Kinship and the Study of Culture. *Futures*. **27**(4), pp.423-435.

Titmuss, R. M. 1997 [1970]. *The Gift Relationship: From Human Blood to Social Policy.* New York: The New Press.

Triendl, R. 2003. Japan launches controversial Biobank project. *Nature Medicine*. **9**(8), pp. 982.

Trostle, J. and Sommerfeld, J. 1996. Medical Anthropology and Epidemiology. *Annual Review of Anthropology*. **25**, pp. 253-274.

Tutton, R. 2002. Gift relationships in genetics research. *Science as Culture.* **11**(4),pp. 523-542.

2004. They want to know where they came from: population genetics, identity, and family genealogy. *New Genetics and Society.* **23**(1), pp.105-120.

Tutton, R. Smart, A. Ellison, G. Martin, P. And Ashcroft, R. 2004. Ambivalence and the Biopolitics of Race and Ethnicity: Debates about the Standardization of Human Difference. (Unpublished).

Tutton, R. and Corrigan, O. (eds), 2004. *Genetic Databases: socio-ethical issues in the collection and use of DNA,* London: Routledge.

Twine, R. 2005. Biotechnology and human meaning. *Social Studies of Science*. **35**(1), pp.165-168.

Waldby, C. 2001. Code Unknown: Histories of the Gene. *Social Studies of Science.* **31**(5), pp. 779-791

2002. Biomedicine, tissue transfer and intercorporeality. *Feminist Theory*. **3**(3), pp.239-254

Waldby, C. 2006. The Human Genome Project: Information, Embodiment and Experience.[online]. [Accessed 22.09.06]. Available from: http://www.ssn.flinders.edu.au/soci/staff/riaz/genomeproject/8.htm

Waldby C., Rosengarten, M., Treloar, C. and Fraser, S. 2004. Blood and bioidentity: ideas about self, boundaries and risk among blood donors and people living with Hepatitis C. *Social Science and Medicine*. **59**, pp.1461-1471

Wagner W., Kronberger, N. and Seifert, F. 2002. Collective symbolic coping with new technology: Knowledge, images and public discourse. *British Journal of Social Psychology*. **41**, pp. 323-343.

Watson, J. 2000. *A Passion for DNA: Genes, Genomes and Society*. Oxford: Oxford University Press.

Weldon, S. 2007. United Knigdom. In Häyry,M., Chadwick, R., Árnason, V. and Árnason, G. (Eds), *The Ethics and Governance of Human Genetic Databases: European Perspectives.* Cambridge: Cambridge University Press

Williams, G. 2005. Bioethics and large-scale biobanking: individualistic ethics and collective projects. *Genomics, Society and Policy.* **1**(2), pp.50-66.

Williams, G. and Schroeder, D. 2004. Human genetic banking: altruism, benefit and consent. *New Genetics and Society*. **23**(1), pp. 89-103.

Wilson, C. 2003. A healthy investment? *New Scientist*. **178**(2394), pp. 25.

Wynne, B. 1995. Public Understanding of Science. In S. Jasanoff, G. Markle, J.Petersen and T.Pinch, (eds). *Handbook of Science and Technology Studies.* London: Sage Publications.

Reports

Chief Medical Officer. 1999. Annual Report.

Chief Scientists Office. 2004. *Guidance for Applicants to the Genetics Healthcare Initiative.* Department of Health, Scottish Executive.

Commission of the European Communities, 2003. *Towards a Strategic Vision of Life Sciences and Biotechnology.* European Commission, Brussels.

Commission of the European Communities, 2003. *Life Sciences and Biotechnology – A Strategy for Europe.* European Commission, Brussels.

Council of Europe, 2005. Recommendations on research on human biological materials. Strasbourg.

Cragg Ross Dawson. 2000. Public Perceptions of the collection of human biological samples.[online]. [Accessed 03.10.03]. Wellcome Trust and MRC. Available from: www.wellcome.ac.uk

European Molecular Biology Organization (EMBO). 2004. Report 5. *Genomics and Drugs*.

European Molecular Biology Organization (EMBO). 2004. Report 5. *The gene and its place*.

European Federation of Biotechnology. Task Group on Public Perceptions 2003. *Who should communicate with the public and how?* European Commission, Brussels.

Healthy Public Policy Network. 1998. *The Possible Scot*. Scottish Council Foundation.

McNally, E. and A. Cambon-Thomsen 2004, *25 Recommendations on the ethical, legal and social implications of genetic testing.* European Commission, Brussels.

Marsden, W., Sullivan, F., Duffy, R. and Maclaren, S. 2002. *Report on the recruitment of two Scottish Primary Care Trust areas to Biobank UK.* Unpublished.

Nuffield Council on Bioethics 1995, *Human Tissue Ethical and Legal Issues*, Nuffield Trust.

People, Science and Policy. 2002*. Biobank UK: A question of trust*. Wellcome Trust and Medical Research Council. London.

Public Health Institute of Scotland. 2001. *Chasing the Scottish Effect.* Information and Statistics Division of the Common Services Agency.

Royal Society of Edinburgh. 2002. *A Strategy for Health and Wealth Creation Through Gene Knowledge.*

Scottish Enterprise and the The Royal Society of Edinburgh. 1996. *Technology Ventures.* The Royal Society.

Scottish Council Foundation. 1998. The Scottish Effect?. Occasional paper to *The Possible Scot.* Report. Health Public Policy Network.

Task Group on Public Perceptions. 2003. *Who should communicate with the public and how?*, European Federation of Biotechnology.

Electronic Sources

Asthma Genetics Group, 2005. [online]. [Accessed 04.03.05]. Available from: http://www.well.ox.ac.uk/asthma/Current%20Projects.shtml .

Avon Longitudinal Study (ALSPAC), 2005. [online].{Accessed 04.03.05]. Available from: http://www.alspac.bris.ac.uk/welcome/index.shtml

Bioetisk Forskergruppe. 2006. Biobanking and health surveys worldwide. [online]. [Accessed 27.03.06]. Available from: http://www.bioethics.ntnu.no/biobanks/printable.php?id=biobanks

British Association for the Advancement of Science (The BA), 2004. [online]. [Accessed 20.04.04]. Available from: http://www.the-ba.net/the-ba/

Cancer Research UK, 2007. [online]. [Accessed 09.09.07]. Available from: http://www.cancerresearchuk.org/

Department of Health, 2003. Our Inheritance, Our Future: Realising the potential of genetics in the NHS. White Paper June 2003. HMSO, Norwich.

Estonian Genome Project Foundation. 2001. Estonian Genome Project.[online]. [Accessed 05.12.04]. Available from: http;//www.geenivaramu.ee

Gateway to Scotland. 2007. Local Government, Department of Geography, University of Edinburgh. [online]. [Accessed 01.06.07]. Available from: http:// www.geo.ed.ac.uk/home/scotland/localgovt.html

Generation Scotland. 2004. Newsletter 2. [online]. [Accessed 01.07.04]. Available from: http://www.generationscotland.org

Generation Scotland. 2008.[online]. [Accessed 14.04.06,15.11.06, 15.02.07, 18.05.07, 02.06.07, 09.09.07]. Available from: http://www.generationscotland.org

Genetic Alliance, 2007.[online]. [Accessed 22.06.07]. Available from: http://www.geneticalliance.org/

Genetic Health in the 21st Century (21CGH). 2008.[online]. Available from: http://www.generationscotland.org/21CGH.htm

Genetics of Human Diabetes - The Warren 2 Type 2 Diabetes study, 2005. Online http://www.well.ox.ac.uk/warren2 04.03.05

Genetic Interest Group (GIG). 2007. [online]. [Accessed 01.06.07]. Available from: http://www.gig.org.uk/

Genewatch UK. 2002. Biobank UK - A Good Research Priority? [online]. [Accessed 26.11.04]. Available from: http://www.genewatch.org

Human Genetics Commission, 2003.[online]. [Accessed 20.06.04]. Available from: http://www.hgc.gov.uk

INNOGEN. 2008. [online]. Available from: http:// www.innogen.ac.uk

ISIS and PROCARDIS sibling-pair studies of genetic determinants of MI, 2005. [online]. [Accessed 04.03.05]. Available from: http://www.ctsu.ox.ac.uk/projects/procardis.shtml

MRC BRIGHT Study. British Genetics of Hypertension, 2005. Recruitment Procedures. [online]. [Accessed 04.03.05]. Available from: http://www.brightstudy.ac.uk/info/sop_menu.html

MRC Leading Science for better health, 2005. [online]. Accessed 05.03.05]. Available from: http://www.mrc.ac.uk/txt/index/public-interest/publicnews_centre/publicpress_office/publicpress_releases_2000/public13_november_2000.htm

National Blood Service, 2005. [online]. [Accessed 03.05.05]. Available from: http://www.blood.co.uk/

National Health Intelligence Service, 2007. [online]. [Accessed 16.09.07]. Availbale from: http://www.nhis.info/

Nuffield Council on Bioethics, Pharmacogenetics 2007.[online]. [Accessed 28.09.07]. Available from:http://www.nuffieldbioethics.org/go/browseablepublications/pharmacogenetics/report_16.html

PHG Foundation. 2006. Foundation for Genomics and Population Health [online]. [Accessed 27.05.07]. Available from: http://www.phgfoundation.org/pages/resources/glossary.htm

Scottish Consumer Council, 1999. Confidentiality of personal health information. Response to NHS consultation. [online]. [Accessed 05.05.05]. Available from: http://scotconsumer.org.uk

Scotland's Health on the Web (SHOW). 2007. [online]. [Acccessed 10.04.07]. Available from: http://www.show.scot.nhs.uk/

Scottish Health Survey, 1998. [online]. [Accessed 21.04.04]. Available from: http://www.sehd.scot.nhs.uk/scottishhealthsurvey/sh8-00.html

Scottish National Blood Transfusion Service, 2005. [online].[Accessed 03.05.05]. Available from: http://www.scotblood.co.uk/

Scottish School of Primary Care, 2006. [online]. [Accessed 12.11.06]. Available from: http://www.sspc.ac.uk/

Wellcome Trust. UKBiobank. 2008.[online]. Available from: www.wellcome.ac.uk/en/1/biovenpop.html

Wellcome Trust Clinical Research Facility, 2006. [online]. [Accessed 23.07.06]. Available from: http://www.wtcrf.ed.ac.uk/

Appendix 1

Appendix 2

A. Personal Details

1. What is your age? ⌊__⌋⌊__⌋

2. What is your sex? (Please tick one answer only)

☐ Male

☐ Female

☐ Intersex

☐ Not specified / Intermediate

3. Where were your parents living at the time of your birth?

Mother

Country ⌊__⌋

If Scotland, what council area and town (if known)?

Council Area ⌊__⌋

Town ⌊__⌋

Father

Country ⌊__⌋

If Scotland, what council area and town (if known)?

Council Area ⌊__⌋

Town ⌊__⌋

4a. Where were you born?

Country ⌊__⌋

If Scotland, what council area and town (if known)?

Council Area ⌊__⌋

Town ⌊__⌋

4b. If you were born outside the UK, what year did you come to live here? ⌊__⌋⌊__⌋⌊__⌋⌊__⌋

5. What is your cultural background? (Please tick one answer only)

☐ White – Scottish
☐ White - Other British
☐ White – Irish
☐ White - any other white background (please specify below)
☐ Mixed - any mixed background (please specify below)
☐ Asian – Indian
☐ Asian – Pakistani
☐ Asian – Bangladeshi
☐ Asian – Chinese
☐ Asian – Any other Asian background (please specify below)
☐ Black – Caribbean
☐ Black – African
☐ Black - any other Black background (please specify below)
☐ Any other ethnic background (please specify below)
☐ Not Known
☐ Not Disclosed

Please Specify: _____

_____

B. Family Health

1. Have you ever been diagnosed with any of the following medical conditions?

Condition	Please Tick	Age at first diagnosis	Any treatment required (please specify) None	Drug Treatment	Other Treatment	Operation
a. Heart Disease	☐	__|__	☐	☐	☐	☐
b. Stroke	☐	__|__	☐	☐	☐	☐
c. High Blood Pressure	☐	__|__	☐	☐	☐	☐
d. Diabetes	☐	__|__	☐	☐	☐	☐
e. Alzheimer's disease	☐	__|__	☐	☐	☐	☐
f. Parkinson's disease	☐	__|__	☐	☐	☐	☐
g. Severe depression	☐	__|__	☐	☐	☐	☐
h. Breast cancer	☐	__|__	☐	☐	☐	☐
i. Bowel cancer	☐	__|__	☐	☐	☐	☐
j. Lung cancer	☐	__|__	☐	☐	☐	☐
k. Prostate cancer	☐	__|__	☐	☐	☐	☐
l. Hip fracture	☐	__|__	☐	☐	☐	☐
m. Osteoarthritis	☐	__|__	☐	☐	☐	☐
n. Rheumatoid Arthritis	☐	__|__	☐	☐	☐	☐
o. Asthma	☐	__|__	☐	☐	☐	☐

1p. Have you ever been diagnosed with any other serious illness?

i) ______________________________

ii) ______________________________

2. Please tick the box if your father, mother or any brother, sister or grandparent has been affected by any of these conditions

	father	mother	brother	sister	grandparent
a. Heart Disease	☐	☐	☐	☐	☐
b. Stroke	☐	☐	☐	☐	☐
c. High Blood Pressure	☐	☐	☐	☐	☐
d. Diabetes	☐	☐	☐	☐	☐
e. Alzheimer's disease	☐	☐	☐	☐	☐
f. Parkinson's disease	☐	☐	☐	☐	☐
g. Severe depression	☐	☐	☐	☐	☐
h. Breast cancer	☐	☐	☐	☐	☐
i. Bowel cancer	☐	☐	☐	☐	☐
j. Lung cancer	☐	☐	☐	☐	☐
k. Prostate cancer	☐	☐	☐	☐	☐
l. Hip fracture	☐	☐	☐	☐	☐
m. Osteoarthritis	☐	☐	☐	☐	☐
n. Rheumatoid Arthritis	☐	☐	☐	☐	☐
o. Asthma	☐	☐	☐	☐	☐

2p. Any other serious illness that runs in your family?

i) _____

ii) _____

C. Family History

It is known that some health problems run in families. We have a family history section to help us to find out more about this. If you are adopted or if your parents remarried it would be better to know about your biological family (i.e. blood relations) for both your parents and your brothers and sisters.

1. When was your father's date of birth? ☐ Only Year Known

☐ Full Date Known

☐ Not Known

2. Where was your father born?

Country

If Scotland, what council area and town (if known)?

Council Area

Town

3. Is your father still alive? ☐ Yes ☐ No ☐ Don't Know

3a. If he has died, what was the date of his death? ☐ Only Year Known

☐ Full Date Known

☐ Not Known

3b. If he has died, what was the cause of his death?

4. Where was your father's father born?

Country

If Scotland, what council area and town (if known)?

Council Area

Town

5. Where was your father's mother born?

Country

If Scotland, what council area and town (if known)?

Council Area

Town

6. When was your mother's year or date of birth? ☐ Only Year Known

☐ Full Date Known

☐ Not Known

7. Where was your mother born?

Country

If Scotland, what council area and town (if known)?

Council Area

Town

8. Is your mother still alive? ☐ Yes ☐ No ☐ Don't Know

8a. If she has died, what was the date of her death? ☐ Only Year Known

☐ Full Date Known

☐ Not Known

8b. If she has died, what was the cause of her death?

9. Where was your mother's father born?

Country

If Scotland, what council area and town (if known)?

Council Area

Town

10. Where was your mother's mother born?

Country

If Scotland, what council area and town (if known)?

Council Area

Town

D. Smoking History

1. Have you ever smoked tobacco?

☐ Yes, currently smoke **(GO TO QUESTIONS 2-3)**

☐ Yes but stopped within past 12 months **(GO TO QUESTIONS 2-5)**

☐ Yes but stopped more than 12 months ago **(GO TO QUESTIONS 2-5)**

☐ No, never smoked **(GO TO SECTION E)**

2. What age were you when you started smoking? ⌊__⌋⌊__⌋ years old

3. What is the maximum number you have smoked per day for as long as a year?

⌊__⌋⌊__⌋⌊__⌋ cigarettes per week

⌊__⌋⌊__⌋⌊__⌋ packets of tobacco per week

⌊__⌋⌊__⌋⌊__⌋ cigars per week

IF YOU HAVE STOPPED SMOKING, GO TO Q4, IF YOU CURRENTLY SMOKE, GO TO SECTION E

4. How long since you gave up smoking?

⌊__⌋⌊__⌋ years ⌊__⌋⌊__⌋ months ⌊__⌋⌊__⌋ days

5. Why did you give up smoking? (please tick one answer only)

☐ On doctor's advice

☐ Family Influence

☐ Financial reason

☐ Due to illness

☐ Health reasons

☐ Prior to or during pregnancy

☐ Personal Decision

☐ Other reason (Please specify) ⌊_____⌋

⌊_____⌋

E. Exposure to Tobacco Smoke

1. Are you regularly exposed to other peoples tobacco smoke?

	Yes, a lot	Yes, some	Yes, a little	No, none at all	Not Applicable
a. at work	☐	☐	☐	☐	☐
b. in your home	☐	☐	☐	☐	☐
c. in other places (e.g. social groups)	☐	☐	☐	☐	☐

2. On average, for how many hours per week are you exposed to other people's tobacco smoke?

└─┴─┘ hours per week

3. Do you live with anyone who smokes? ☐ Yes ☐ No ☐ Don't Know

F. Educational and Occupational History and Clinical Notes

1. What is the highest educational qualification you have obtained?

☐ None

☐ School Leaving Certificate

☐ Standard Grade or 'O' Level

☐ Higher Grade

☐ University Degree

☐ Other professional or technical qualification or diploma after leaving school

2. Additional Clinical Questionnaire Notes

THANK YOU FOR TAKING THE TIME TO COMPLETE THIS QUESTIONNAIRE

The contents of this questionnaire will be considered as medically confidential and will be covered by the Data Protection Act 1998.

(i). Standard Phenotype Measurements

1. Blood Pressure & Heart Rate

Ask the subject to sit quietly for 5 minutes before recording BP/P.
Take recordings 2 minutes apart.

☐ tick here if Blood Pressure and Heart Rate not obtained

Measure	Date (dd/mm/yy)	Time (hh:mm) In 24 h format	Blood Pressure (mmHg) SBP / DBP	Heart Rate (BPM)
1	⌊__⌋⌊__⌋⌊__⌋⌊__⌋	⌊__⌋ : ⌊__⌋	⌊__⌋/⌊__⌋	⌊__⌋
2	⌊__⌋⌊__⌋⌊__⌋⌊__⌋	⌊__⌋ : ⌊__⌋	⌊__⌋/⌊__⌋	⌊__⌋

Are you are on any medication for Blood Pressure? (*Please tick*)

☐ Amlodipine
☐ Atenolol
☐ Bendrofluazide
☐ Bendroflumethiazide
☐ Losartan
☐ Enalapril
☐ Felodipine
☐ Lisinopril
☐ Ramapril
☐ Other, please specify below:

⌊__⌋

⌊__⌋

2. Height (to the nearest 0.5 cm) ⌊__⌋⌊__⌋⌊__⌋.⌊__⌋ cm OR ☐ Not obtained

3. Weight (measure to 1 decimal place) ⌊__⌋⌊__⌋⌊__⌋.⌊__⌋ kg OR ☐ Not obtained

(ii). Laboratory Blood & Urine Tests

1. Were the following samples taken?

1 x 9 ml Potassium EDTA blood tube	☐ Yes	☐ No
1 x 9 ml ACD-B blood tube	☐ Yes	☐ No

1-2. What date and time were the samples taken?

└┴┘└┴┘└┴┴┴┘ (dd/mm/yy) └┴┘ : └┴┘ (hh/mm in 24 h format)

1a. Were the following samples taken?

Blood spots on a Whatman FTA card ☐ Yes ☐ No

Buccal Cell Mouthwash ☐ Yes ☐ No

1a-2. What date and time were the samples taken?

└┴┘└┴┘└┴┴┴┘ (dd/mm/yy) └┴┘ : └┴┘ (hh/mm in 24 h format)

1b. Will blood be collected at a later date? ☐ Yes ☐ No

2a. Was the following sample obtained?

1 x 50 ml (approx.) Midstream Urine ☐ Yes ☐ No

2a-2. What date and time were the samples taken?

└┴┘└┴┘└┴┴┴┘ (dd/mm/yy) └┴┘ : └┴┘ (hh/mm in 24 h format)

The type of Urine sample ideally required for this study is second void, fasting (nothing to eat or drink except water for at least 4 hours), midstream urine. Should this not be possible, a random midstream urine sample is acceptable

(i) How long is it since you last ate or drank anything apart from water? └┴┘ hours.

(ii) How many times have you previously emptied your bladder today? └┴┘ times.

2b. Reagent Strip Results

Test	Result
Glucose	
Bilirubin	
Ketones	
Specific gravity	
Blood	
pH	
Protein	
Urobilinogen	
Nitrite	
Leukocytes	

(iii) Cognitive Function Testing

1. Eysenck Personality Questionnaire	**N** Total ⌊__⌋/12	OR	☐ Not obtained
	E Total ⌊__⌋/12	OR	☐ Not obtained
2. Logical Memory Test Immediate	Total correct ⌊__⌋/25	OR	☐ Not obtained
3. Digit Symbol Coding	Total correct ⌊__⌋/133	OR	☐ Not obtained
4. Verbal Fluency	**C** Score ⌊__⌋	OR	☐ Not obtained
	F Score ⌊__⌋	OR	☐ Not obtained
	L Score ⌊__⌋	OR	☐ Not obtained
5. Mill Hill Vocabulary	Total correct ⌊__⌋/44	OR	☐ Not obtained
6. Logical Memory Delay	Total correct ⌊__⌋/25	OR	☐ Not obtained

(iv) Additional Phenotype Questionnaire Notes

⌊_____⌋

⌊_____⌋

⌊_____⌋

⌊_____⌋

⌊_____⌋

⌊_____⌋

⌊_____⌋

generation SCOTLAND

SCOTTISH FAMILY HEALTH STUDY

PRE-CLINIC QUESTIONNAIRE

FOR OFFICE USE ONLY

Researcher Name _____

_____

Researcher Code ☐☐☐

Date ☐☐ ☐☐ ☐☐☐☐
d d m m y y y y

Place barcode sticker here

SF ☐ Geographical Site

Funded by Scottish Executive Health Department, Chief Scientist Office

SCOTTISH FAMILY HEALTH STUDY

Generation Scotland

Instructions to help with completion of questionnaire

- Complete using a black ballpoint pen if possible.
- Please complete as much of the form as possible.
- Enter numbers clearly inside the boxes.
- Enter a cross (X) inside appropriate boxes.
- Write all entries clearly using block capital letters when writing text.
- If you make a mistake and want to change an entry, please cross through the original and write the correct entry above or to the side.
- Please write only in designated areas.
- Please ignore the little review box on each page. This is for office-use only.

Pre-Clinic Questionnaire

We would be grateful if you would answer the questions on this form before you attend the GENERATION SCOTLAND clinic. Don't worry if you cannot answer all the questions, any information you can provide will be helpful. You should bring your completed form when you attend your clinic appointment.

A. Personal Details

1. Where were your parents living at the time of your birth?

If not known cross here

Mother

_____ (town or council area) *OR* ☐

_____ (country) *OR* ☐

Father

_____ (town or council area) *OR* ☐

_____ (country) *OR* ☐

2. Where were **you** born?

_____ (town or council area) *OR* ☐

_____ (country) *OR* ☐

3. Were you born outside the UK? ☐ Yes ☐ No *OR* ☐

If born outside the UK, when did you come to live here? ☐☐ / ☐☐ / ☐☐☐☐ *OR* ☐

d d / m m / y y y y

4. What is your cultural background?
(Mark an **X** in one box from each of A and B)

A	B
☐ White	☐ Scottish
☐ Black	☐ English
☐ Asian	☐ Welsh
☐ Mixed, specify _____	☐ N. Irish
☐ Other, specify _____	☐ Irish
	☐ Pakistani
☐ Not disclosed	☐ Indian
	☐ Bangladeshi
	☐ Chinese
	☐ African
	☐ Carribean
	☐ Other, specify _____
	☐ Not disclosed

For Office Use Only

Place barcode sticker here

Review ☐

If not known cross here

B. Family History

1. What is/was your **father's** date of birth? d d / m m / y y y y OR ☐

2. Where was your father born?

_____ (town or council area) OR ☐

_____ (country) OR ☐

3. Is your father still alive? ☐ Yes ☐ No OR ☐

If No, what was the date of his death? d d / m m / y y y y OR ☐

4. If he has died, what was the cause of his death?

_____ OR ☐

5. Where was your **father's father** born?

_____ (town or council area) OR ☐

_____ (country) OR ☐

6. Where was your **father's mother** born?

_____ (town or council area) OR ☐

_____ (country) OR ☐

7. What is/was your **mother's** date of birth? d d / m m / y y y y OR ☐

8. Where was your mother born?

_____ (town or council area) OR ☐

_____ (country) OR ☐

9. Is your mother still alive? ☐ Yes ☐ No OR ☐

If No, what was the date of her death? d d / m m / y y y y OR ☐

10. If she has died, what was the cause of her death?

_____ OR ☐

11. Where was your **mother's father** born?

_____ (town or council area) OR ☐

_____ (country) OR ☐

12. Where was your **mother's mother** born?

_____ (town or council area) OR ☐

_____ (country) OR ☐

FOR OFFICE USE ONLY

MCD _____

FCD _____

Review ☐

C. Medications

At the clinic you will be asked about the medications or supplements you REGULARLY take. These include:

- prescribed medicines from your GP or hospital,
- over the counter medicines bought from a chemist or shop
- supplements, vitamins, complementary or alternative medicines (eg evening primrose oil)

Don't forget to include contraceptive pills or injections; hormone replacement therapy; and inhalers (eg Ventolin)

I. Name of Prescribed or Bought Pills or other Oral Medication *OR* **Cross here if none** ☐

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____

II. Name of Prescribed or Bought Cream/Ointment or Other Topical Preparation (such as patches) *OR* **Cross here if none** ☐

1. _____
2. _____
3. _____
4. _____
5. _____

III. Name of Prescribed or Bought Inhaler or Nasal Spray *OR* **Cross here if none** ☐

1. _____
2. _____
3. _____
4. _____
5. _____

IV. Name of Prescribed or Bought Injection or Suppository *OR* **Cross here if none** ☐

1. _____
2. _____
3. _____
4. _____
5. _____

Review ☐

D. Operations

Please give details below of any operations you have had and what age you were when you had them.

OR If you have had no operations cross here ☐

If age not known, cross here

1. _____ ☐☐ age *OR* ☐
2. _____ ☐☐ age *OR* ☐
3. _____ ☐☐ age *OR* ☐
4. _____ ☐☐ age *OR* ☐
5. _____ ☐☐ age *OR* ☐
6. _____ ☐☐ age *OR* ☐

E. Family Health

1. Please mark an **X** in the box if you, your father, mother or any brother, sister or grandparent has been affected by any of these conditions:

	You	Father	Mother	Brother(s)/ Sister(s)	Grand parent
a. Heart Disease	☐	☐	☐	☐	☐
b. Stroke	☐	☐	☐	☐	☐
c. High Blood Pressure	☐	☐	☐	☐	☐
d. Diabetes	☐	☐	☐	☐	☐
e. Alzheimer's disease	☐	☐	☐	☐	☐
f. Parkinson's disease	☐	☐	☐	☐	☐
g. Depression	☐	☐	☐	☐	☐
h. Breast Cancer	☐	☐	☐	☐	☐
i. Bowel Cancer	☐	☐	☐	☐	☐
j. Lung Cancer	☐	☐	☐	☐	☐
k. Prostate Cancer	☐	☐	☐	☐	☐
l. Hip Fracture	☐	☐	☐	☐	☐
m. Osteoarthritis	☐	☐	☐	☐	☐
n. Rheumatoid arthritis	☐	☐	☐	☐	☐
o. Asthma	☐	☐	☐	☐	☐

p. Any Other Serious Illness which runs in your family..... (i) _____

(ii) _____

Review ☐

F. Chest Pain

1. Do you ever get pain or discomfort in your chest? ☐ Yes ☐ No

   IF NO, GO TO SECTION G

2. Do you get this pain or discomfort when you walk uphill or hurry? ☐ Yes ☐ No

   IF NO, GO TO SECTION G

3. Do you get it when you walk at an ordinary pace on the level? ☐ Yes ☐ No

4. When you get pain or discomfort in your chest what do you do?

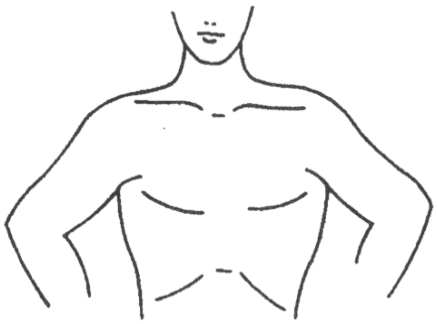
   ☐ Stop ☐ Slow Down ☐ Continue at same pace

5. Does it go away when you stand still or sit down? ☐ Yes ☐ No

   If so, how soon? ☐ 10 minutes or less ☐ More than 10 minutes

6. Where do you get this pain or discomfort?
   *(Mark the place(s) with an **X** on the picture)*

Right



Left

7. Have you ever had a severe pain across the front of your chest lasting for half an hour? ☐ Yes ☐ No

   If Yes, what was the cause? _____

G. Musculoskeletal History

1. Have you ever suffered a fracture? (broken bone) ☐ Yes ☐ No ☐ Don't Know

   If Yes, please complete below:

Site of broken bone (e.g. leg, hip, arm)	**What age you were**
a. _____	☐☐ yrs
b. _____	☐☐ yrs
c. _____	☐☐ yrs
d. _____	☐☐ yrs
e. _____	☐☐ yrs
f. _____	☐☐ yrs

2. Have you been diagnosed as suffering from Osteoarthritis (wear and tear arthritis)? ☐ Yes ☐ No ☐ Don't Know

   If Yes, please tell us which joints are affected:

	Age first affected		**Age first affected**		**Age first affected**
a. Neck	☐☐ yrs	e. Hands	☐☐ yrs	i. Ankles	☐☐ yrs
b. Shoulder	☐☐ yrs	f. Back	☐☐ yrs	j. Feet	☐☐ yrs
c. Elbows	☐☐ yrs	g. Hips	☐☐ yrs	k. Other	☐☐ yrs
d. Wrist	☐☐ yrs	h. Knees	☐☐ yrs		

If Other, specify _____

3. Have you been diagnosed as suffering from Rheumatoid Arthritis? ☐ Yes ☐ No ☐ Don't Know

Review ☐

H. Chronic Pain

1. Are you currently troubled by pain or discomfort, either all the time or on and off? ☐ Yes ☐ No

IF NO, GO TO SECTION I

2. Have you had this pain or discomfort for more than 3 months? ☐ Yes ☐ No

IF NO, GO TO SECTION I

3. Where is this pain or discomfort?
(mark **X** in the box for each question)

	Yes	No
a. Back pain	☐	☐
b. Neck or shoulder pain	☐	☐
c. Headache, facial or dental pain	☐	☐
d. Stomach ache or abdominal pain	☐	☐
e. Pain in your arms, hands, hips, legs or feet	☐	☐
f. Chest pain	☐	☐
g. Other pain	☐	☐

If Other pain, please specify _____

4. Which **one** of these pains or discomforts has bothered you the most in the past three months?
(mark **X** in <u>one box only</u> to indicate your response)

- a. Back Pain ☐
- b. Neck or shoulder pain ☐
- c. Headache, facial or dental pain ☐
- d. Stomach ache or abdominal pain ☐
- e. Pain in your arms, hands, hips, legs or feet ☐
- f. Chest pain ☐
- g. Other pain ☐

(Please circle the number on the scale below, where 0=No Pain and 10=Pain as bad as could be)

In the <u>past 3 months</u>,	***No Pain***										***Pain as bad as could be***
5. How intense was your worst pain?	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
6. How intense was your usual pain?	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

In the <u>past 3 months</u>,

7. How many days have you been kept from your usual activities (work/housework) because of this pain?
(mark **X** in the box to indicate your response) ☐ 0-6 days ☐ 7-14 days ☐ 15-30 days ☐ 31 or more days

(Please circle the number on the scale below, where 0=No Interference and 10=Unable to carry on activities)

In the <u>past 3 months</u>,	***No Interference***										***Unable to carry out activities***
8. Has the pain interfered with your daily activities?	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

(Please circle the number on the scale below, where 0=No Change and 10=Extreme Change)

In the <u>past 3 months</u>,	***No Change***										***Extreme Change***
9. How much has this pain changed your ability to take part in recreational, social and family activities?	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
10. How much has this pain changed your ability to work (including housework)?	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Review ☐

I. Physical Activity

Please record number of hours and minutes in the space provided. If you DO NOT SPEND ANY TIME on the activity record 0

For how long do **you usually**....

1. Work in paid employment each week? ☐☐ hours ☐☐ mins

2. Do housework each week? ☐☐ hours ☐☐ mins

3. When **working** (including housework) for how long are you **usuall**y....

   a. Very active each week? ☐☐ hours ☐☐ mins
   (such as heavy lifting or carrying, hurried walking, going up stairs and ladders, digging, heavy housework)

   b. Moderately active each week? ☐☐ hours ☐☐ mins
   (such as light lifting or carrying, walking at slightly increased pace, light housework, shopping, painting, decorating)

   c. Inactive each week? ☐☐ hours ☐☐ mins
   (such as sitting, standing, light arm movements, unhurried walking, driving))

4. When **working** (including housework), how often are you physically active for at least 20 minutes during which time you become short of breath or perspire? (mark **X** in the box to indicate your response)

   ☐ never ☐ less than once a week ☐ once a week ☐ 2-3 times a week ☐ 4 or more times a week

5. During your **non-working time** (including going to and from work) for how long are you **usually**....

   a. Very active each week? ☐☐ hours ☐☐ mins
   (such as competitive sports, football, hockey, squash, badminton, hill walking, cycling, swimming, running, aerobics, heavy gardening, windsurfing)

   b. Moderately active each week? ☐☐ hours ☐☐ mins
   (such as moderate walking, golf, light gardening, cricket, dancing, bowls, playing pool, sailing, taking a shower or bath, getting dressed and undressed)

   c. Inactive each week? ☐☐ hours ☐☐ mins
   (such as sitting, standing, watching TV or films, listening to music, cooking, drinking, eating, piano playing, card playing, driving)

6. During **non-working time** how often are you physically active for at least 20 minutes during which time you becomeshort of breath or perspire? (mark **X** in the box to indicate your response)

   ☐ never ☐ less than once a week ☐ once a week ☐ 2-3 times a week ☐ 4 or more times a week

I. Physical Activity cont'd

7. **During the past 12 months**, has the level of your physical activity.....?
(mark **X** in the box to indicate your response)

☐ increased ☐ stayed the same ☐ decreased

If it has changed, for how long has your physical activity been at its current level? ☐☐ months ☐☐ weeks

8. How many hours a day do you **usually** spend in bed ....? on work days ☐☐ hours *OR* Not applicable ☐

on non-work days ☐☐ hours *OR* Not applicable ☐

J. Dietary Intake

1. **In <u>general</u>, how often do you eat.....?**

	Number of times eaten	(please mark **X** in one box only) day	week	month	year	If '0 times', what age were you when you last ate this
a. Fresh fruit	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
b. Green leafy vegetables	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
c. Other types of vegetables	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
d. Oily fish (eg sardines, mackeral, salmon, herring)	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
e. Other types of fish (cod, tinned tuna, haddock)	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
f. Chicken, turkey or other poultry	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
g. Liver (including liver pate and liver sausage)	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
h. Other types of meat (including bacon, sausages, ham)	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
i. Eggs (including eggs in quiche, cakes and omelettes)	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
j. Dairy products (milk, yoghurt, cheese, butter)	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years
k. Brown bread	☐☐☐	per ☐	or ☐	or ☐	or ☐	☐☐ years

K. Alcohol consumption

1. Have you ever had an alcoholic drink?
(mark **X** in the box to indicate your response)

☐ Yes, currently drink	**GO TO QUESTION 2**
☐ Yes, but stopped within past 12 months	**GO TO QUESTION 4**
☐ Yes, but stopped more than 12 months ago	**GO TO QUESTION 4**
☐ No, never drank	**GO TO SECTION L**

2. **During the past week**, please record how many units of alcohol you have had: ☐☐☐ units
(If you have had no units, please enter 0)

To help you calculate the number of units of alcohol you have had, the following is given as a guideline.

A unit of alcohol is:	**Approximate Units**
1 pint of ordinary beer, cider or lager	2
1 bottle/can of ordinary strenght beer/lager	2
1 bottle/can of extra strength beer /lager	4
1 can of cider	2
1 litre of cider	9
1 small glass of wine (125ml)	1
1 bottle of wine (75cl)	9
1 bottle of fortified wine	10
1 litre of fortified wine	14
1 small glass of sherry	1
1 bottle of sherry	12
1 pub measure of spirits (25ml)	1
1 bottle of spirits (75cl)	30
1 bottle of alcopops	2

3. How does this compare to what you usually drink in a week? ☐ More ☐ Same ☐ Less

CURRENT ALCOHOL DRINKERS GO TO SECTION L

4. Why did you stop drinking alcohol? ☐ On doctor's advice ☐ Other reason

If Other, specify _____

_____

Review ☐

L. Smoking History

1. Have you ever smoked tobacco?
(mark **X** in the box to indicate your response)

☐ Yes, currently smoke	**GO TO QUESTION 2**
☐ Yes, but stopped within past 12 months	**GO TO QUESTION 2**
☐ Yes, but stopped more than 12 months ago	**GO TO QUESTION 2**
☐ No, never smoked	**GO TO SECTION M**

2. What age were you when you started smoking? ☐☐ years old

3. What is the maximum number you have smoked per week for as long as a year?
☐☐☐ cigarettes per week ☐☐ packets of tobacco per week
☐☐ cigars per week

CURRENT SMOKERS GO TO SECTION M

4. How long is it since you gave up smoking? ☐☐ years ☐☐ months ☐☐ days

5. Why did you give up smoking?
(mark **X** in the box to indicate your response)

☐ On doctor's advice
☐ Personal decision
☐ Other reason If Other, specify

_____

_____

M. Exposure to Tobacco Smoke

1. Are you **regularly** exposed to **other people's** tobacco smoke.....? (mark **X** in the box to indicate your response)

	Yes, a lot	Yes, some	Yes, a little	No, None at all	Not Applicable
a. at work.....	☐	☐	☐	☐	☐
b. in your home.....	☐	☐	☐	☐	☐
c. in other places (eg social groups).....	☐	☐	☐	☐	☐

2. **On average**, for how many hours per week are you exposed to other people's tobacco smoke? ☐☐ hours per week
(If exposed for no hours per week, please enter 0)

3. Do you live with anyone who smokes? ☐ Yes ☐ No

Review ☐

N. Educational and Occupational History

1. How many years altogether did you attend school or study full-time? ☐☐ years

2. What is the highest educational qualification you have obtained?
(mark **X** in the box indicating your response)

- ☐ University degree
- ☐ Other professional or technical qualification or diploma after leaving school
- ☐ Higher Grade
- ☐ Standard Grade or 'O' Level
- ☐ No Qualification
- ☐ Other qualification

3. What is your employment? (if currently unemployed, give details of your last job)

_____

OR **Cross here, if never worked** ☐

4. If you live with a spouse or partner please give details of his/her job

_____

OR **Cross here, if spouse/partner never worked** ☐

If you do not live with a spouse or partner, cross here ☐

5. What is your and your spouse's/partner's current employment status?
(mark **X** in the box indicating your response)

	YOU	**SPOUSE/PARTNER**
a. Self-employed employing others	☐	☐
b. Self-employed not employing others	☐	☐
c. Paid employee supervising others	☐	☐
d. Paid employee not supervising	☐	☐
e. In unpaid employment	☐	☐
f. Housewife/homemaker	☐	☐
g. Retired	☐	☐
h. Full-time student	☐	☐
i. Unemployed, sick or disabled	☐	☐
j. Unemployed, seeking work	☐	☐

OR **Cross here if no spouse/partner** ☐

i. If you are unemployed, please state for how long ☐☐ years ☐☐ months

ii. If you are employed, what best describes the type of work your job mainly involves?
(mark **X** in the box indicating your response)

- ☐ Sedentary, spend most of time sitting down (eg office worker)
- ☐ Standing, spend most of time standing or walking (eg hairdresser)
- ☐ Manual, involves physical effort (eg plumber)
- ☐ Heavy manual, involves vigour effort (eg miner)

iii. If you are employed, how many hours in a typical week would you work in the evening/overnight between 7pm-7am? ☐☐ hours

6. What is the average total income before tax of your entire household?
(mark **X** in the box indicating your response)

- ☐ less than £10,000
- ☐ between £10,000 and £30,000
- ☐ between £30,000 and £50,000
- ☐ between £50,000 and £70,000
- ☐ more than £70,000
- ☐ prefer not to answer

Review ☐

O. Household History

1. Including yourself, how many people live in your household? (record number) ☐☐

2. Are you living with anyone in your household as a couple? ☐ Yes ☐ No

3. What type of accommodation do you live in?
(mark **X** in the box indicating your response)

- ☐ House or bungalow
- ☐ Flat or apartment
- ☐ Hostel
- ☐ Mobile or caravan
- ☐ Sheltered house
- ☐ Homeless
- ☐ Other (please specify) _____

4. What is the status of the accommodation in which you and your household live?
(mark **X** in the box indicating your response)

- ☐ Own outright
- ☐ Own with mortgage
- ☐ Rent from local authority
- ☐ Rent from private landlord or agency
- ☐ Pay part rent and part mortgage
- ☐ Live rent free
- ☐ Other (please specify) _____

5. How many cars/vans are available to you and your household? (record number) ☐☐
(if no cars/vans available, please enter 0)

P. WOMEN ONLY (This section should only be completed by women)

If not known, cross here

1. Have you ever had a period? ☐ Yes ☐ No

IF NO, GO TO QUESTION 3

2. a. What age were you when you had your first period? ☐☐ years old *OR* ☐

b. Are you still having periods? ☐ Yes ☐ No

c. If you no longer have periods, what age were you when they stopped? ☐☐ years old *OR* ☐

3. Have you had a hysterectomy? ☐ Yes ☐ No

If Yes, what age were you? ☐☐ years old *OR* ☐

4. Have you had your ovaries removed? ☐ Yes ☐ No

If Yes, what age were you? ☐☐ years old *OR* ☐

Review ☐

WOMEN ONLY (This section should only be completed by women) continued.....

If not known, cross here

5. Are you currently taking oral contraceptive pills? ☐ Yes ☐ No

If yes, age first started? ☐☐ years old *OR* ☐

From that age, for how long in total have you taken them? ☐☐ yrs ☐☐ mths *OR* ☐

If not currently taking oral contraceptive pills, did you previously take them? ☐ Yes ☐ No

If yes, age first started? ☐☐ years old *OR* ☐

age finally stopped? ☐☐ years old *OR* ☐

During this time, for how long in total did you take them? ☐☐ yrs ☐☐ mths *OR* ☐

6. Do you currently have regular contraceptive injections? (eg Depo-Provera) ☐ Yes ☐ No

If yes, age first started? ☐☐ years old *OR* ☐

From that age, for how long in total have you had them? ☐☐ yrs ☐☐ mths *OR* ☐

If not currently having regular contraceptive injections? (eg Depo-Provera), did you previously have them? ☐ Yes ☐ No

If yes, age first started? ☐☐ years old *OR* ☐

age finally stopped? ☐☐ years old *OR* ☐

During this time, for how long in total did you have them? ☐☐ yrs ☐☐ mths *OR* ☐

7. Are you currently taking hormone replacement therapy (HRT)? ☐ Yes ☐ No

If yes, age first started? ☐☐ years old *OR* ☐

From that age, for how long in total have you taken it? ☐☐ yrs ☐☐ mths *OR* ☐

If not currently taking hormone replacement therapy (HRT), did you previously take it? ☐ Yes ☐ No

If yes, age first started? ☐☐ years old *OR* ☐

age finally stopped? ☐☐ years old *OR* ☐

During this time, for how long in total did you take it? ☐☐ yrs ☐☐ mths *OR* ☐

8. Are you currently taking any other hormone therapy? ☐ Yes ☐ No

If yes, age first started? ☐☐ years old *OR* ☐

From that age, for how long in total have you taken it? ☐☐ yrs ☐☐ mths *OR* ☐

If not currently taking any other hormone therapy, did you previously take it? ☐ Yes ☐ No

If yes, age first started? ☐☐ years old *OR* ☐

age finally stopped? ☐☐ years old *OR* ☐

During this time, for how long in total did you take it? ☐☐ yrs ☐☐ mths *OR* ☐

Review ☐

MANY THANKS FOR TAKING THE TIME TO COMPLETE THESE QUESTIONS

PLEASE BRING YOUR COMPLETED QUESTIONNAIRE WITH YOU WHEN YOU ATTEND YOUR APPOINTMENT WITH THE GENERATION SCOTLAND, SCOTTISH FAMILY HEALTH STUDY

Q. Additional comments

Please use this space, to make additional comments on any of the questions you have been asked

Contact details

If you would like to find out more information about the study please find details below:

Web address: www.generationscotland.org

Glasgow

Study Co-ordinator - 0141 330 8357

Dundee

Study Co-ordinator - 0800 027 0466