

GeneWatch UK response to the Nuffield Council on Bioethics' consultation on the linking and use of biological and health data

January 2014

GeneWatch UK is a not-for-profit organisation which aims to ensure genetics is used in the public interest. GeneWatch UK has major concerns about the Wellcome Trust/Human Genome Strategy Group's proposal to build a DNA database of the whole population by including genomes as attachments to electronic medical records in the NHS.¹ The British people have never voted for this plan and its adoption by successive governments shows total contempt for any kind of democratic process and for individual choice. We are particularly concerned that data-sharing of NHS England medical records with companies like Google is already planned on the basis of "opt out" rather than "opt in" consent and that genomes may be included in this system in the future. The Nuffield Council on Bioethics' consultation is rather late in the day to have any impact on this decision or on concerted lobbying by the Wellcome Trust to weaken data protection regulation at an EU level.

The Secretary of State for Health has stated that in his view every baby should have its whole genome sequenced at birth.² This proposal was of course made by the previous government in its 2003 White Paper "Our Inheritance, Our Future". This proposal was subsequently rejected by the Human Genetics Commission (HGC) on the grounds of its excessive cost, lack of benefit to health, and concerns about ethical issues such as lack of consent and potential misuse of the information.^{3,4} If anything, the evidence has grown that genomic testing has poor predictive value for most diseases in most people and is unsuitable for use in newborn screening programmes.

GeneWatch UK has major concerns that:

- (1) Commercial exploitation of medical and genomic data will lead to the destruction of the NHS as doctors are replaced by computer algorithms, putting prescribing decisions under commercial control and massively expanding the market for preventive medication, sold to healthy people using personalised online marketing techniques and paid for "top up" payments. This will create a "demand-led" (i.e. advertising-led, rather than need-led) system which will be bad for health and add to costs for taxpayers, who will be expected to subsidise the infrastructure and follow-up. This plan is not a figment of our imaginations but was clearly laid out in Richard Sykes' book in 2000 in which he advocated retaining the NHS only as a basic service whilst using genetic screening as a means to personalise medical care and massively expand "pre-symptomatic" treatment (and pharmaceutical company profits).⁵ There is plenty of evidence that this will not benefit health but will instead lead to over-treatment of rich healthy people and a poorer service for poor sick ones. Commercial algorithms will not be independently validated and vulnerable people will be at risk of misleading claims about their risks as well as stigma and discrimination.
- (2) Government and commercial surveillance of every individual and their relatives will be unstoppable as full genomes, or genetic profiles based on SNPs, act as biometrics which can link all an individual's personal data to their physical self and also identify their relatives. It will no longer be possible for anyone to escape from intrusive surveillance or commercial advertising and this will be particularly dangerous for vulnerable people such as political dissidents, abused women and children, or women who have hidden the paternity of a child. Plans to store the data in the cloud and share with overseas companies such as Google/23andMe and BGI will mean the data is accessible to security services worldwide e.g. in the USA and China. This is a fundamental threat to human rights.
- (3) Data-mining is not science and claims that this approach will benefit health are largely based on a false premise that statistical data mining will lead to meaningful and useful predictions of individual risks. Useful, meaningful results will be swamped by spurious statistical associations and tests of poor clinical validity and utility. It is

unlikely that personalised risk predictions will meet screening criteria for use in the general population and therefore the net result for health will be bad, due to large numbers of false positives (overtreatment) as well as false negative results. In many cases an individualised approach to disease prevention (which is strongly supported by the tobacco and food industries) acts as a (deliberate) distraction from difficult political decisions such as tackling the marketing of unhealthy foods.

As individuals, members of the public must retain the right to control their own medical records, DNA and genomes, and other personal information. This does not mean that their genome should be sequenced and shared without their consent leaving them only with a say over whether they want (mis-)information fed back afterwards. People should have a say over who gets access to this data and how it is used and whether their DNA is sequenced in the first place. It is completely wrong to remove their choice about which research to take part in and to insist that the only option is to opt out completely or to opt in to a system which will allow no individual control (the care.data system). This means that legitimate researchers and medical professionals will no longer be able to make any promises to patients or research participants about how their data will be used or about their privacy (which will not be guaranteed). People will rightly become more wary of sharing information (perhaps to the detriment of their care) and there will be the risk of a major backlash against medical research. GeneWatch UK is strongly opposed to this approach and also to the idea that genomic data would automatically be added to people's electronic medical records in the future (in most cases it will not be relevant to their disease risks or their care) and also be shared without their knowledge or consent. Genomic data differs from other biomedical data because it is a biometric that can be used to track and identify individuals and their relatives. Indefinite retention of biometric data without consent is likely to contravene Article 8 of the European Convention on Human Rights.

As voters and taxpayers, members of the public should also have a say not only about how their individual data is used but also about whether the plan to invest money in this approach is in the best interests of the public and of future generations. They should not be presented with the vision of turning the NHS into a massive database as if it was the only option for the future, over which they are being given no say whatsoever. Storage and collection of data costs money, especially if this is new data such as genomes. There is already ample evidence that personal genomes are not (and will never be) relevant to most people's care so it is hard to justify such enormous expense, outside of specific research projects. It is unethical to keep lying to people about this (sometimes politely known as "genohype").

Failure to address these issues risks a massive public backlash in the future, which will harm legitimate medical research.

**Consultation question 1:
Do biomedical data have special significance?**

There are always limitations and exceptions to describing any category of data, but it is fairly clear that most people regard the content of their medical records as private and potentially sensitive. Some information may be regarded as more sensitive than others (e.g. information on sexually transmitted diseases or illegal drug use) and the sensitivity may depend on the context and uses to which this information is put (e.g. information about high LDL cholesterol might be readily shared with friends and family, but not with an employer) and the circumstances of an individual (e.g. whether they feel able to be open about a termination or not). Privacy is regarded as central to the professional relationship between a patient and their doctor, partly because failure to protect privacy could result in people not seeking medical help or not being open with their doctor if they fear stigma, discrimination or other negative consequences (damage to a relationship, loss of a job).

The Wellcome Trust's own research shows clearly that people are keen to take part in medical research, but only when they have been asked.^{6,7} This is an important safeguard to protect not only individual privacy but the broader public interest.

In December 2008, Connecting for Health held a consultation about the sharing of medical data for research without consent.⁸ The consultation did not mention that this would include sharing of genetic information, however the Human Genetics Commission (HGC)'s response included a large number of concerns raised by the HGC's Consultative Panel of members of the public, including concerns about sharing of data in "sealed envelopes" and the fact that "anonymisation" of data in a way that made individuals unidentifiable was likely to be impossible for rare disorders.⁹ In its response to the consultation the Wellcome Trust Sanger Centre "*encouraged the NHS Care Records Service to prepare for the integration of significant amounts of genetic and genomic information into patient records*" and argued that: "*If robust systems are in place.....the benefits of research will outweigh the risks associated with the use of identifiable information*" (including information that patients have requested to be kept confidential in 'sealed' and 'locked' envelopes).¹⁰ However, a quarter (25%) of the members of the public stated that they did not believe that it was possible to effectively anonymise data and some people were adamant that "their data" should not be shared for any purposes. There was wide concern amongst participants in the general public about the ability of the NHS to protect personal data. Concerns included risks of data loss by NHS staff, hacking and selling of data to third parties for commercial purposes, especially insurance companies and employers. The consultation revealed widely divergent views between the general public and researchers.

Thus, removing people's choice to decide which research to take part in (as is currently being done through the "opt out" system adopted NHS England's care.data scheme) is a recipe for a major loss of public trust, when people become aware that "approved researchers" (mentioned but not explained in the leaflet¹¹) means companies like Google and private healthcare companies, and that the purpose of the data-sharing will include making individualised risk predictions for personalised marketing.

In 2007, the Science Horizons project (funded by GE Healthcare)^{12,13} highlighted public anxieties about privacy and surveillance, erosion of the human dimension in services and relationship building, future employment, trustworthiness of authorities, safety, fair access to technology and the potential for technologies to be misused. The concern that technology is being developed by industry and/or government in order to make profits, rather than in response to societal needs was "*a fairly common theme*" and it was also "*widely assumed that policy-makers in government and big business are not candid with citizens*".

Overarching issues raised by the Deliberative Panel¹⁴ included:

- trust in expertise - who can be trusted?;
- concerns about the security, privacy and integrity of personal information (IT- or genetically-based);
- concerns about safeguards against abuse of technologies by authorities or by criminals;
- and fears about loss of the 'human touch' in everyday interactions, for example in relation to health, and in work.

There was a "*striking trust deficit*" and some people saw expert priorities for research investments as inevitably not the same as those of the average citizen.

Biomedical data is covered by a number of international instruments which emphasise the importance of informed consent. The Helsinki Declaration includes requirements to protect the "*dignity, integrity, right to self-determination, privacy and confidentiality of personal information of research subjects*".¹⁵ Research participants must be informed of "*the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may*

entail, and any other relevant aspects of the study” before giving their consent, preferably in writing. For consent to valid it must be fully informed and freely given.

Special protections must be accorded to people who lack capacity to give consent and account must be taken of the changing capacity of children as they grow up.¹⁶

The European Convention on Biomedicine states (Article 16): Research on a person may only be undertaken if all the following conditions are met¹⁷:

- (i) there is no alternative of comparable effectiveness to research on humans;*
- (ii) the risks which may be incurred by that person are not disproportionate to the potential benefits of the research;*
- (iii) the research project has been approved by the competent body after independent examination of its scientific merit, including assessment of the importance of the aim of the research, and multidisciplinary review of its ethical acceptability;*
- (iv) the persons undergoing research have been informed of their rights and the safeguards prescribed by law for their protection;*
- (v) the necessary consent as provided for under Article 5 has been given expressly, specifically and is documented. Such consent may be freely withdrawn at any time.*

These important rules are being abandoned under the care.data scheme because medical professionals will no longer be able to guarantee privacy or even inform patients or research subjects where their data will end up.

The Wellcome Trust has produced a plan which involves including a variant file, containing the whole genome of every person minus the reference genome, as an attachment to every medical record in the NHS in England.¹⁸ The proposal to build a DNA database of the entire population has been supported by the Human Genomics Strategy Group (led by Professor John Bell) and the Government (working with Chief Scientist Mark Walport, ex-head of the Wellcome Trust) is now putting in place the infrastructure and policies to build it, all without any democratic debate.¹⁹ Under this plan, genomic data as well as medical records would be made available to “approved researchers” (including commercial companies) for data-mining in the cloud and personalised risk assessments would be returned to individuals. Other data such as social care records and education records would be added later. The intention is to treat genomic data in the same way as other biomedical data, on the grounds that the information it contains is no more sensitive (see Section 5.8 of the Caldicott report²⁰ and the 100k Genomes Project’s so-called ethics report, which proposes that no one will be able to have genetic data collected within the NHS without it being stored and later shared with government-funded institutes and private companies worldwide without their knowledge or consent²¹). This approach is highly problematic because genomic data acts as a biometric, which links a person’s data permanently to their physical self, and also allows their relatives to be identified. Some genetic information (e.g. carrier status for a genetic disorder) may also reveal health-related information about other members of the family.

Unlike stored data, a biometric cannot be altered because it is linked to a person’s body. The EU’s Article 29 Data Protection Working Group has warned that biometrics allow for automated tracking, tracing or profiling of persons and as such their potential impact on the privacy and the right to data protection of individuals is high.²² Biometrics that can be derived from DNA include: forensic DNA profiles (a string of numbers based on parts of the DNA called Short Tandem Repeats); genotypes based on multiple Single Nucleotide Polymorphisms (SNPs) which are single chemical letters which differ between individuals; and whole genomes (the entire string of chemical letters which makes up a person’s DNA). Forensic DNA profiles are not unique identifiers but have sufficiently low probability of being shared by chance with another person to be useful to the police; individual SNPs have low power to discriminate between individuals, but panels of multiple SNPs are sufficient (and are often used to identify body parts after disasters); whole genomes are thought to be unique. If whole genomes, or sufficiently detailed genotypes, are stored with other information this allows all that information to be connected to that individual.

Since DNA can be obtained from a person's coffee cup, for example, a DNA database allows individuals to be tracked by the police, security services or anyone who can gain access to the system. If DNA is linked to other information such as medical records, all this personal information can also be identified. In addition, relatives can be identified by searching databases for partial matches with the DNA of an individual (known as 'familial searching'). Non-paternity can be revealed if the records of a child and its supposed father can be identified and the child's DNA does not match half of the father's sequence.

Under the EU Data Protection Directive (95/46/EC) biometric data are in most cases personal data. Therefore they may only be processed if there is a legal basis and the processing is adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed. The EU's Article 29 Data Protection Working Group states that a prerequisite to using biometrics is a clear definition of the purpose for which the biometric data are collected and processed, taking into account the risks for the protection of fundamental rights and freedoms of individuals, and: "*It must be clear that such consent cannot be obtained freely through mandatory acceptance of general terms and conditions, or through opt-out possibilities*". Valid alternatives must exist for consent to be regarded as freely given (e.g. people must not be forced to seek care outside the NHS or go without treatment if they do not want their genomes sequenced, stored or shared). However, proposals to treat whole genomes as health data – including the proposed use of "presumed" or opt-out consent and the indefinite storage of genetic data by the Government – threaten to undermine these important safeguards.

The Data Protection Act's section 33 exemption allows personal data held for research purposes to be retained indefinitely. "Research purposes" includes statistical or historical purposes, provided the data are not processed to support measures or decisions with respect to particular individuals and that the data are not processed in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject.²³ Whilst retaining data for research often makes sense, under the proposed plans this would allow a DNA database of the whole population to be built by stealth, with no possibility of removal of any of the records. Storage of genomes collected in the NHS would allow every individual and their relatives to be tracked, because genomes are biometrics. This would include babies whose DNA will have been collected without them giving their own consent to the collection or retention of their genomes. This gives enormous power over individual citizens to the government, which could easily be abused (for example, to track down dissidents and political opponents using their DNA, and even to identify their children). For this reason, the Government's proposal might be regarded as breaching Article 8 of the European Convention on Human Rights (the right to privacy). In an unanimous judgment by the Grand Chamber in December 2008 in the case of *S. and Marper v. the UK*, the European Court of Human Rights found that the indefinite retention of two innocent persons' biological samples, forensic DNA profiles and fingerprints by the police in England "*constitutes a disproportionate interference with the applicants' right to respect for private life and cannot be regarded as necessary in a democratic society*".²⁴ It is therefore hard to see how building a DNA database in the NHS without consent can be regarded as compatible with human rights. However, such legal cases can take many years to be decided.

Further, data-sharing proposals (including storing data in the cloud and sharing with overseas companies such as Google and BGI) mean that overseas governments, law enforcement, border and security agencies will also get access to the data, as the revelations about the US National Security Agency's activities by Edward Snowden have clearly shown. This means that newly developed rapid DNA testing systems (e.g. developed by GE Healthcare²⁵ or Life Technologies or other companies²⁶) could be used at borders (or even on the street) to match an individual to their online records and potentially not only track their whereabouts but also retrieve linked medical records and other data (such as social care and education records) and those of their children or other close relatives.

Anyone who can hack into or otherwise access the system (e.g. by working in the NHS or as a researcher) will also be able to access this information (e.g. organised criminal gangs or individuals seeking to track down vulnerable individuals). Once this data is stored and widely shared it a system of total surveillance will have been constructed and it will not be possible to delete or destroy all copies of the data.

Privacy concerns are linked to concerns about stigma and discrimination. The Caldicott2 report proposes that multiple data sets will be linked in so-called “safe havens” and this information to be made available for research. This means employment, education, social care, tax and police records could all be connected. Since de-identification is unlikely to be preventable, and some use by insurers is anyway expected, there is a possibility that people may be refused insurance, visas or a job based on this information. There is potential for stigma and discrimination based on stored medical information (e.g. past use of drug rehabilitation or sexual health services, or unhealthy habits such as poor diets or smoking); other unrelated data (such as employment or tax histories); or genetic make-up. Fear of stigma or discrimination might make some groups of people less willing to seek medical care, especially if they fear losing their job or benefits, or having their children taken into care.

The Equality Act 2010 restricts what employers can ask about in pre-employment medical checks, so they can only ask for information that is directly relevant to the applicant's ability to carry out the work, or needed to make 'reasonable adjustments' to the workplace to enable a particular person to work there (as required by law). This largely allays fears about genetic discrimination in the workplace, and discrimination based on other medical information, because it tightly restricts the circumstances in which employers could access job applicants' or employees' medical and genetic information. However, this protection would be undermined by the plan to create a DNA database in the NHS because employers can become researchers. For example, an employer could conduct a study on their own employees by data-mining the genomes and associated medical data stored in the cloud. Because of the high potential for “deductive identification” individual workers' data is likely to be identified and could be misused e.g. to dispute a compensation claim²⁷.

There is currently a voluntary agreement between the insurance industry and government which means the industry does not use predictive genetic test results to determine insurance premiums, with one exception for high value policies.²⁸ However, there is no legislation to prevent genetic discrimination by insurers in the future. Insurers might be able to access data surreptitiously whilst working as researchers but more likely they would simply require the release of genomes or genetic risk information as a prerequisite to obtaining coverage. Although most diseases in most people are poorly predictable from people's genes, there are exceptions (such as mutations in the BRCA1/2 genes and breast cancer risk) which could be of interest to insurers. It is also likely that this requirement would apply to all individual risk assessments, whether they are based only on a person's medical records or also on their genome. If this happens it will be a major expansion of current underwriting practices which (whilst also discriminatory) use a relatively small subset of information.

Health risks are not the only risks that might be calculated from a person's genome. Although studies of genetics and behaviour have to date delivered very little (none of the many statistical associations made between genes and intelligence²⁹ or personality³⁰ have been definitively confirmed) it is also possible that people might be treated differently in future based on studies linking their genes with high or low intelligence or criminality, even if these links are spurious. The lack of any evidence does not appear to dent some ministers' enthusiasm for the idea of sorting people based on their supposed genetic intelligence.³¹

**Consultation question 2:
What are the new privacy issues?**

New privacy issues are being driven by a strong commercial desire to generate profits from data-mining medical records and other personal data; by governments' interest in public-private partnerships to use personal data (especially medical data) to generate investment, income and economic growth; by governments' interest in gaining access to data for surveillance purposes; and by increased technical capacity to generate and store large quantities of data (at costs which have been reducing significantly overtime).

The Wellcome Trust's plan involves two aspects:

- (1) Routine sharing of personal medical records containing all the data collected during the course of an individual's care.
- (2) Collection, storage and sharing of genomic data which is largely irrelevant to the individual's care and which also acts as a biometric.

The plan is to facilitate commercial exploitation of this data and gain commercial and venture capital investment as well as massive public subsidy to build the infrastructure. This means that instead of allowing people to choose whether or not they want to give their DNA to a company like 23andMe and pay for it to be tested, this decision will be made for them and subsidised by taxpayers money so that all the costs are sunk upfront before anyone has time to find out that whole genome sequencing is largely useless for most people's care. Investors will of course expect a return on their investments, which they will expect to recoup through personalised marketing i.e. feeding back risk predictions to individuals online and encouraging them to buy medication, supplements, functional foods or further tests and treatments. In addition to the discussions that have already taken place with Google-funded 23andMe³², investors already include Asia's richest man^{33,34} and the Chancellor has already promised to share the data with China.^{35,36,37} Patrick Chung, a 23andMe board member and partner at the venture-capital firm NEA 23andMe told Fast Company: "...23andMe will make money by partnering with countries that rely on a single-payer health system. *"Let's say you genotype everyone in Canada or the United Kingdom or Abu Dhabi," he says, "and the government is able to identify those segments of the population that are most at risk for heart disease or breast cancer or Parkinson's. You can target them with preventative messages, make sure they're examined more frequently, and in the end live healthier lives, and the government will save massive expenses because they halted someone who's prediabetic from getting diabetes. 23andMe has been in discussion with a bunch of such societies"*. Yet there is not in reality a scrap of evidence that this approach is good for health (see further discussion below) because genomic tests have limited clinical validity or utility so there is no health benefit to targeting segments of the population in this way. Anyone can write an algorithm but most algorithms will be neither valid nor useful, an important fact that is oft ignored (and has led to the current withdrawal of 23andMe's gene tests from the market due to intervention by the FDA).

Quite apart from any value attached to exploiting the content of people's genomes (their genetic information) having a unique biometric identifier (a 'genetic fingerprint') is a gold mine to commercial companies who may wish to link separate data sets to a single individual to monitor their behaviour in detail, and to totalitarian governments who will then be able to track every individual and their relatives. This is similar to the New Labour Government's proposed ID card scheme, except that genomes, rather than fingerprints and iris scans, will be the biometrics, allowing a person's relatives to also be identified. Many companies want to use biometrics (and also "deductive identification" by pooling the data they have about individuals) to create a world in which "the computer knows who you are".^{38,39} In GeneWatch's view this should not be treated as the inevitable consequence of new technology but as a democratic decision in which members of the public can have a say. Nor should the Government continue to pretend that this kind of de-identification is impossible when in reality it could become routine.

Even advocates of whole genome sequencing acknowledge that privacy can no longer be protected if a person's whole genome is stored.⁴⁰

Storing an individual's DNA sequence linked to their name and other identifying information allows a form of biological tagging or "biosurveillance" which can be used to track them or their relatives.⁴¹ There is widespread agreement that the creation of such databases raises human rights concerns. Because of these concerns, the Coalition Government has adopted the Protection of Freedoms Act and deleted more than 1.7 million DNA profiles belonging to innocent individuals or children from the police National DNA Database and destroyed more than 7.7 million DNA samples containing sensitive biological material.⁴² However, under the Wellcome Trust's plan for the NHS, a searchable genetic database would be created in which the individual genome of each person in the NHS is stored as an attachment to their electronic medical record. Once complete, this would allow anyone who can obtain a DNA sample from an individual to search their genetic profile against the database and use biometric matching (or partial matching with close relatives) to:

- identify that individual and their relatives, if they can obtain access to the linked identifying information;
- obtain linked personal health information that has made available for research purposes, even if they cannot access identifying information. This may or may not be sufficient to identify the individual (and/or their relatives) via "deductive identification".

This proposal differs significantly from the current situation in which genetic test results focused on a specific gene or genes are stored in the health records of a relatively small number of patients. Whilst such relatively limited genetic data can lead to the deductive identification of an individual (e.g. based on knowledge that they have a rare disease, combined with other information), its collection and storage in specific cases based on clinical need does not amount to the creation of a biometric database. In addition, medical records are currently shared mainly for a person's care or for research which they have consented to take part in, for which a single named researcher is responsible. This will be transformed under the new plan into a system where pseudo-anonymised data will be made readily available to any approved organisation wanting to perform any statistical analysis.

An access order granted by a court can allow police access to samples from existing collections held by other parties, including the NHS.⁴³ But until, now this power has been used only in rare cases. If a searchable database of genomes existed in the NHS this could be used to identify individuals from their DNA routinely, in the same way that the police National DNA Database is used now. The National DNA Database is based on data from parts of people's genomes, but the planned NHS database would contain the whole sequence, once this becomes affordable, or individuals' genotypes consisting of thousands of SNPs. Searching for partial matches with relatives could also allow the police or security services to identify relatives of a suspect (a process known as 'familial searching'). There would be a danger of misuse because the police or government could use such a system to track down political opponents or other 'undesirables' in exactly the same way (for example, by taking DNA from coffee cups left at a political meeting and looking for a match on the DNA database, which then reveals the name and medical details of the individual). Anyone who can infiltrate the system (for example, organised criminal gangs) might also use the database, perhaps to track down victims or expose witnesses on protection schemes or undercover police officers. Unlike the National DNA Database, which is accessible by a small number of people, an NHS database would be accessible to large numbers of NHS workers and researchers all over the world, and would be much harder to keep secure.

GeneWatch UK strongly disagrees with the Wellcome Trust's repeated attempts to argue that the requirement for fully informed consent only applies to interventional research (such as clinical trials) and not to data-mining. Lobbyists, including the Wellcome Trust, its former Director Chief Scientist Sir Mark Walport, and an unprecedented number of commercial

companies (including Google), are trying to use two different mechanisms to remove people's right to be asked for their consent for the sharing and use of their medical records and genetic data in non-interventional research (i.e. research which does not involve medical interventions such as taking drugs). One is the Caldicott Review and the other is the revision of data protection laws in the European Union and the associated negotiations under the Transatlantic Trade and Investment Partnership (TTIP). Google has been in negotiations with the Department of Health about access to NHS medical records and genomes since at least 2008 (as reported by Professor Sir John Bell to the House of Lords Science and Technology Committee⁴⁴) and are building a shiny new HQ in London in anticipation of data-mining all this information. To date, approved registered researchers for UK Biobank include commercial researchers in the USA's Silicon Valley (probably the Google-funded gene test company 23andMe) and Europe, and public institutions across the world, including in several in China.⁴⁵ Approved researchers to be given access to NHS medical records for research already include private healthcare company Bupa. The Wellcome Trust and MRC have already funded joint research with the Google-funded gene testing company 23andMe^{46,47,48} so there is little doubt that Google will become "approved researchers". Whether the general public would approve is of course another question: their right to decide who gets their data is presumably being removed because they probably would not. It is fundamentally unethical to allow the general public to believe that "approved researchers" means scientists trying to cure cancer when in fact it will include companies using the data for market research and personalised marketing. Failure to allow any choice over who gets access to this data (i.e. to allow "opt in" to specific research projects rather than a blanket all-or-nothing "opt out" approach) will undermine legitimate research due to loss of public trust.

The Government expects that medical records with the names of individuals removed will count as "anonymised" or (more likely) as "pseudo-anonymised" data. However, as databases become larger it has become clear that individuals' identities can often be deduced from combining information such as age, postcode, medical history and occupation, even if the names have been stripped off. Genetic information alone, especially whole genomes, can be sufficient to identify an individual and the idea of "anonymised" data can therefore rapidly become meaningless.⁴⁹ Several studies have shown that whole genomes cannot be reliably anonymised and individuals' identities and those of their families can be deduced either from their genome alone or from other stored data, such as details about medications, diseases, age and postcode.^{50,51,52} The use of "pseudo-anonymised" data adds to the risks (in comparison to using anonymised data) because data stored with the NHS number can be re-identified by matching it back to personal details (including name and contact details) stored in the NHS Personal Demographic Service (PDS), to which more than 800,000 NHS staff have access. Because genomes are biometrics this adds further risks, since data could also be retrieved by matching all or parts of a person's DNA sequence (obtained from their coffee cup or toothbrush, for example) to the one stored in the record.

The Government plan to collect and share all this data without people's knowledge or consent has major implications for the doctor-patient relationship and for vulnerable people who may be frightened to seek medical care if their privacy cannot be guaranteed.

Under the Government's plans, doctors will be required to consent on behalf of their patients to transfer of their practice's data to the Health and Social Care Information Centre (HSCIC). But they will have no idea where their patient's medical records will end up or how they are going to be used. The Caldicott Review (Section 7.4) suggests that GPs will be required to explain to anyone who objects to sharing of their medical records that they will only be used for research to improve medicine and the health service and identities will be protected. This is despite the fact that GPs themselves will not be fully informed of how the data will be used and cannot realistically make any guarantees about anonymity. Further, passing data to the Information Centre is likely to be necessary in order for the GP to obtain payments under the Government's Quality and Outcomes Framework (QOFs)⁵³ and for commissioning purposes.

Thus, the GP's advice may not be disinterested and the consent of the practice to data-sharing cannot be regarded as freely given, since there will be no option to opt-in to selected data uses and out of others.

Some possible scenarios show what loss of medical confidentiality could mean in practice:

- A person's employer or a pharmaceutical company could be classified as a "researcher" and thus gain access to data about individuals who suffer from a workplace-related illness or an adverse drug reaction: they are likely to be able to use "deductive identification" (based on the occurrence of a rare event with other information) to work out who these individuals are. They could try to look for data that might allow them to blame the condition on a person's genes, or for unrelated personal data (e.g. sexual health or use of drug rehabilitation services) that might be used to discredit that individual should they make a claim against the company.
- A person's DNA can be obtained easily from a beer glass, coffee cup or toothbrush. Anyone who could get that DNA sequenced could search it against stored variant files and identify the individual, either directly (if they have access to the medical record in the NHS or the de-identifying system) or indirectly by the clues stored in their public records. They could also look for partial matches to identify that person's relatives (including paternity and non-paternity). This process could be used by the police or state to track individuals who have not committed any crime (creating a "surveillance society") or it could be used by criminals to track undercover police officers, witnesses on protection schemes, and potential victims (including women and children fleeing abuse).
- The same process could be used to find out what personal medical information is linked to a particular genome, including e.g. use of medical services, including sexual health, or specific information about a disease or carrier status for a genetic disorder. This might be of interest to the press, private detectives, parents, neighbours, or insurance companies. Unscrupulous charities might even use the data to seek donations from the relatives of anyone with cancer.

An individual inherits half their DNA from their mother and half from their father. Hence DNA can be used to identify familial relationships, including non-paternity. Identification of non-paternity can already occur in some NHS screening programmes for recessive genetic disorders (disorders which require two copies of a mutation to be inherited, one from the mother and one from the father)⁵⁴, such as the Sickle Cell and Thalassaemia screening programme. Current guidance states that the risk of non-paternity needs to be handled carefully if relationships and families are not to be disrupted. The Guidance states it is not in the interests of anyone to cause a division in the relationship by revealing this information, that the situation must be discussed with the mother alone (but only when necessary) and the possibility of errors fully considered, and that results must be carefully documented and communicated only to those professionals who need the information to support the family.⁵⁵ It is difficult for physicians to know the consequences of their actions (which may go way beyond issues related to a diagnosis) if they reveal such information. Apart from family breakdown (which may not be in the best interests of the child or other members of the family) there is a risk that routine exposure of such information might drive some women away from seeking appropriate care for themselves or their children, or, in some cases, could put the woman and/or child at risk of domestic violence or even so-called "honour killings". Whilst such situations are always difficult, they will not be made any easier by breaching confidentiality.

The number of families in this situation will be significantly increased if sequencing of the whole population's DNA is allowed to go ahead. In addition, the possibility for revelations about non-paternity to disrupt family relationships, or for health related information to have implications for other members of the family, cannot be discussed prior to testing if such testing is conducted on the basis of "presumed consent". As Professor Sir John Sulston has noted, if everyone has their whole genome sequenced and stored in the NHS "*There will be*

no secrets about paternity anymore".⁵⁶ Even if an individual's name is removed from data made available for research in the cloud, it is likely that relationships will be identifiable using a process of "deductive identification" based on information that is accessible.⁵⁷ Anyone who has access to linked data (i.e. genomes associated with names and personal identifying information) will be able to identify paternity and non-paternity with a simple search. This adds significantly to the concerns outlined above that collecting, storing and sharing such data when it is not necessary for a person's care may not be in the best interests of them or their families.

There has already been a sharp increase in concerns reported by GPs about separated parents seeking children's medical records.⁵⁸ Although this is often well-intentioned, in some cases men are thought to have been seeking access to the information as a way to find out where their former partner is living or whether she has a new partner, which can be a major concern if the mother has been physically abused and living in a place of safety. The NHS Personal Demographic Service (PDS) has already raised concerns because contact details and addresses have been made widely available to NHS staff.⁵⁹ These concerns would be exacerbated by the creation of a DNA database in the NHS because a child's DNA can often be obtained quite easily e.g. from their toothbrush or hairbrush. Anyone with access to the database could search for a match and find the child's address and medical records, even if they have changed their name. This means that vulnerable people who have been abused, victims on witness protection schemes, and even undercover police or security officers will have no place to hide, because they cannot change their DNA.

If the Wellcome Trust's plan is implemented, large numbers of people (including vulnerable women and children) may be forced to seek care outside the NHS to avoid being identified or to keep hidden family relationships from being exposed. Examples of particularly private information that people (including young people) may not wish to be revealed to family members, employers or others is likely to include: treatment for sexually transmitted diseases, sexual abuse, drug addiction, alcoholism, stress and psychiatric disorders. There will be a danger that some people do not seek care when it is needed because of fears that this will lead to stigma or discrimination in the future, particularly as all records will be retained indefinitely and there will be no possibility of erasing records at a later date.

International transfers of data also raise concerns because data that is sent overseas may not be secure. For example, in the US, medical data can be bought and sold and there is increasing concern about the implications for people's civil rights.⁶⁰ As the Snowden revelations showed, all such data can also be accessed by the National Security Agency. In other countries, such as China, it is unclear how personal and genetic data might be used, including whether the government or police might be able to gain access: it seems highly unlikely that they will not. In some countries, particularly in the Middle East, discovery of non-paternity could have very serious consequences for women and their families.⁶¹ Discovery of other personal medical information, such as use of sexual health or drug rehabilitation services, or HIV status, might also have more serious consequences in some countries, or lead to travel restrictions such as the refusal of visas.

Undermining ethical standards such as those in the Helsinki Declaration means medical professionals might in future be put under pressure to build a DNA database for a dictatorial regime, by undertaking similar analysis of "spare" biological samples without seeking fully informed consent. The surreptitious collection and/or analysis of DNA from adults and babies at birth by medical professionals could readily be abused to build databases allowing the police or security services to track individual political opponents and their relatives. Identification and exposure of linked health data could also be used to target dissidents or other minorities (e.g. homosexuals or particular ethnic groups).

The Government's plan removes the right for individuals to decide who uses their medical records for research and for what purposes and may even allow whole genomes to be sequenced and shared without an individual's knowledge or consent. Even if an opt-out is allowed, this is likely to be a blanket out-out from any medical research. This means that individuals will not be able to decide whether some research projects are more legitimate and useful than others, or whether they trust some researchers, but not others, to maintain confidentiality and act in the public interest. This could be highly damaging for medical research because people will not be able to opt-in to a specific study without also allowing all their data to be shared with Google, Bupa and other companies or governments. Researchers will no longer be able to make promises regarding how people's information will be used or guarantees about its confidentiality.

Although most researchers dislike red-tape and some approval processes could be streamlined, some statisticians have argued strongly against abandoning informed consent and questioned the value of data-mining large data sets without consent as a means to do research.⁶²

There are particular concerns about sequencing or genotyping DNA from babies or children who are unable to give informed consent, unless this is a necessary part of their care. One possible source of DNA is the blood spots taken from every baby 5 days after birth to perform a few specific health tests. These blood spots could be genotyped or sequenced, perhaps with the consent of parents, or perhaps not, but obviously not with the consent of the individual baby. Millions of babies' blood spots have been stored within the NHS.^{63,64} All newborn blood spots collected in the NHS are retained for a minimum of five years as part of quality management, but some hospitals have policies to retain the bloodspots indefinitely or until adulthood. There is no explicit national policy for destruction of the blood spots. Guidelines were published in 2005 and incorporated into a Code of Practice which states the blood spots can be used for research where the samples have been anonymised and the research project has ethical approval, as outlined in the Human Tissue Act.^{65,66} Under current guidelines, parents can be re-contacted and asked to allow the use of the blood spots for research, provided they agreed to this when the blood spot was taken (the blood spots are stored with a barcode so they can be linked back to the individual child). Currently, newborn screening laboratories may not sell, or grant exclusive access to, residual newborn blood spots to commercial organisations. However, the Code of Practice is now under review and the new version has not yet been published for public consultation. Storage and use of babies' blood spots without consent has proved highly controversial in other countries.^{67,68,69,70}

Sequencing or genotyping babies' DNA may sometimes be necessary or useful for their care (for example to identify an undiagnosed genetic disorder) but it is not necessary to store this data indefinitely or to share it for other (non-clinical) uses to which the child cannot give or refuse consent. It is sufficient to share scientific results (such as the identification of a new mutation linked to a particular disorder) without retaining the entire genome of the child for data-mining by unspecified researchers. The principle of data minimisation should apply until the child is old enough to decide whether or not to take part in research or to take any further genetic/genomic tests relevant to adult-onset diseases. Research involving children should require the fully informed consent of their parents for specific purposes.

Consultation question 3:

What is the impact of developments in data science and information technology?

The main impacts of developments in data science and information technology are:

- (1) There are a lot more IT companies lobbying for business and trying to convince governments to spend vast sums of public money on databases that are rarely completed, usually insecure, never meet their budgets and rarely improve services;

- (2) The technical capacity exists (and is growing) to collect and store vast quantities of data (such as whole genomes) whether it is useful to do so or not;
- (3) There is considerable interest in exploitation of personal data for personalised marketing or for government surveillance, including using public-private partnerships to establish databases of biometrics such as genomes or genotypes which can track every individual and their relatives;
- (4) An ideological change is taking place in science which tends to abandon the idea of hypothesis-driven science and replace it with mindless data mining, giving rise to an overwhelming number of false positive results in the scientific literature (including in, but not limited to, genetic epidemiology);
- (5) There is an irrational faith in the idea that computers make better decisions than human beings (as, for example, in the Atos disability assessments).

One important consequence of data-sharing without consent will be that people will no longer be able to check whether there are conflicts-of-interest involved in research being undertaken using their stored medical records and genetic data, before deciding whether to take part. Commercial data-mining, aimed at personalised marketing, is not the same as scientific research, conducted in the public interest. In this context it is important to think about the definition of “research” – which according to the Data Protection Act includes any statistical analysis (including, for example, market research) - and about who is an “approved researcher”. This will be dominated by those who have the money to do the statistical analysis of stored data (or who can pay others, perhaps in universities or other public institutions, to do it). It is likely to include:

- Researchers working for Web 2.0 companies, such as Google (and the Google-funded gene testing company 23andMe) which aims to use personal data for personalised marketing;
- Researchers working for private healthcare companies, such as Bupa and GE Healthcare, who wish to sell more healthcare products and services to people deemed to be at risk of becoming ill in future;
- Researchers working for companies with products to sell based on personalised marketing using individual risk assessments, such as: pharmaceuticals, nutraceuticals, functional foods, supplements or other products;
- Researchers based overseas, in any commercial or government-funded institution in any country.

Making data widely available is unlikely to remove problems with bias in medical research, which can arise from many different sources, including commercial bias caused by conflicts-of-interest in the outcomes of the study.^{71,72,73,74,75,76,77} The history of genetic research is riddled with conflicts of interest involving industries seeking to blame diseases on individuals’ genes rather than their products or pollution, starting with Sydney Brenner’s secret meetings with the tobacco industry which led to the funding of the Human Genome Project.^{78,79} Whilst it is possible that the tobacco industry might not be granted legitimate researcher status needed to gain access to NHS data, other companies – including pharmaceutical, food, chemical, nuclear and private healthcare companies - are unlikely to face restrictions. A major area of interest will be the personalised marketing of products and services based on unregulated predictions of people’s future health.

Another issue for researchers is the extent to which the idea of collecting and storing such vast amounts of data is really a good research priority. The proposal relies on data-mining (the computational process of discovering patterns in large data sets), a sub-discipline of computer science, and collection and storage of very large amounts of data (known as “Big Data”). Enthusiasts of Big Data in healthcare see the main objective as identifying correlations between genotype and phenotype (the physical characteristics of a person).⁸⁰ The use of large data sets and sophisticated statistical techniques increases the statistical

power to detect weak correlations such as those between SNPs and common, complex diseases. However, predictions based on multiple correlations can have low predictive value and can be misleading for a variety of reasons, particularly when the effect size of each SNP is expected to be small. Risk assessments may be difficult to validate and/or may not be useful to improve people's health. Researchers from disciplines such as evolutionary theory and psychiatry have highlighted the enormous difficulties in making sense of all the information.^{81,82}

Like all science, Big Data or "hypothesis free" science is based on hidden assumptions that define a paradigm: for example, an emphasis on using biological data (particularly genomic data) to predict individual risks, rather than environmental or social data (although the latter may be integrated at some stage in the future); the treatment of genetic variants such as SNPs as fixed risk factors, rather than context-dependent ones (a standard assumption although it is known to be wrong); an assumption that identification of future genetic variants will increase the utility of personalised risk assessments sufficiently for their use to improve health outcomes (contradicted by a vast swathe of publications in the scientific literature); and a focus on individuals and individual actions (lifestyle changes or medical interventions) rather than population-level policy responses to improve public health (such as stricter regulation of medicines to prevent adverse drug reactions; or measures to restrict the marketing of unhealthy foods).

Ideas promoted in Silicon Valley, which assume that everything can be predicted by computers⁸³ (although contradicted by the laws of physics!), underlie the concept of Big Data. Companies are expected to extract commercial value by data-mining large sets of data, often using the results for personalised marketing. Science itself is being re-defined to fit this idea with concepts such as "hypothesis-free" science and "4th paradigm science". This allows companies to pretend they never need to check whether the assumptions they make when analysing the data are correct, or to show the predictions they make are fit for the claimed purpose (known as "validation"). This is the same approach that led to the collapse of the financial markets.⁸⁴ Further, it remains rooted in the assumptions made by the eugenicists that free will does not exist and all human characteristics and behaviours can be predicted from a complete knowledge of all the relevant environmental and genetic factors (determinism).

There will, in effect, be an infinite number of variables in the proposed database and an infinite number of models (i.e. computer algorithms) that could be fitted to the data: thus even a database of infinite size could lead to multiple possible interpretations and misinterpretations.^{85,86} This (not lack of data) is the main reason why different companies give different genetic risk predictions and will inevitably continue to do so: not a helpful outcome for patients or consumers.⁸⁷ Lack of any prior hypothesis appears to undermine the scientific value of this type of approach.⁸⁸ Unlike search engine algorithms, which can be improved by feedback about the information people want and whether or not their search has been successful, algorithms predicting individual health risks will be impossible for the recipient to verify.

Current applications make a clear distinction between genetic testing (for people who have symptoms or a strong family history of a particular disease) and screening (for the general population). Criteria for the use of screening exist which are intended to ensure that the overall benefits to the population outweigh the harms. However, the "Number Needed to Treat" to prevent one case of a disease in screening programmes is nearly always high: many people identified as at risk will not have developed the condition (known as "false-positive" results) and may be treated unnecessarily (as a result of "over-diagnosis"). Use of genetic tests – especially whole genome scans - in population screening will increase false-positive and ambiguous test results, over-diagnosis, and incidental findings.⁸⁹ As more independent tests are added to screening panels, the overall number of false positives

(people informed they are at risk of a disease they are never going to get) inevitably goes up.

Implementation of the Wellcome Trust plan means abandoning any attempt to weigh up the benefits and risks to an individual or to the population, in favour of screening the whole genomes of the whole population. This idea involves the “creative destruction” of health services, to create new systems which revolve around information stored in electronic medical records, with the addition of genotypes or whole genomes. Under this scenario, each patient would receive a personalised risk assessment based on the information stored about them and this would form the basis of their future care. The primary commercial purpose of screening everyone’s whole genome is to make each person in the population a patient “from the cradle to the grave”, instead of only when they develop symptoms or regard themselves as ill. This allows people to buy health products – or be prescribed them by the NHS - based on their (or their baby’s) predicted risks. In the pharmaceutical industry’s view patient care would be improved by earlier treatment which would at the same time expand the market.⁹⁰ However, others have expressed concern about the creation and treatment of a new type of patient, i.e. the person ‘genetically at risk’, and the resulting ‘biomedicalization’ of health and illness, which involves the privatisation of research and a focus on health surveillance as a moral obligation.⁹¹ Earlier treatment of more people has the potential to significantly expand the market for medication and other health products such as functional foods because the “at risk” group is always significantly larger than the number of people who actually develop a disease.

One of the drivers behind the Wellcome Trust’s plan is to create a market for whole genome sequencing (WGS) by claiming that genomic research is ready to be translated into clinical practice. This involves blurring the line between researchers’ interpretations of an individual’s data (including those made by commercial companies) and clinical interpretations (which normally require a process of assessment to determine how reliable and useful they are for improving people’s health).

The Caldicott 2 report defines **data** as “*qualitative or quantitative statements or numbers that are assumed to be factual, and not the product of analysis or interpretation*” and **information** as “*the output of some process that summarises interprets or otherwise represents data to convey meaning*”. However, it does not discuss the difficulties in interpreting data or conveying meaning, nor does it define **misinformation**, which can also result from (mis-)interpreting data. Whole Genome Analysis (WGA) is the term often used to describe the interpretation of whole genome sequencing (WGS), but WGA is not a simple process of reading out the meaning of the genetic code.

To date, sales of genetic tests by commercial companies have been controversial for several different reasons. Firstly, investigations (including by the US Government Accountability Office (GAO);^{92,93} academic researchers⁹⁴; and GeneWatch UK⁹⁵) have uncovered numerous examples where false or misleading claims have been made about genetic risk, in some cases accompanied by incorrect health advice or attempts to market products (usually supplements). These problems often arise due to the inclusion of SNPs that are not actually related to the risk of the disease (usually due to the large number of false statistical associations in the published scientific literature). Secondly, there is no definitive method to interpret an individual’s genetic risk from pieces of information about the risk associated with different SNPs in different studies: these risks may depend on the context (both environmental and biological) and may combine in complex ways, which are not yet fully understood and may not always be predictable.⁹⁶ These two problems mean that different companies may give very different interpretations of a person’s risk based on the same DNA.⁹⁷ Thirdly, there is growing evidence that the predictive value of genetic information for most diseases in most people is (and will remain) rather poor (even when more research is done), meaning that many genetic tests do not provide useful information for a person’s care

(and the usefulness is often exaggerated in marketing materials).⁹⁸ This is also the case for many (but not all) genetic tests which aim to predict drug response (pharmacogenetic tests).⁹⁹ Finally, because rare mutations can sometimes have unexpected serious consequences (even though most tests have poor predictive value) there is the potential for nasty surprises which people may not be prepared for unless they have pre-test counselling to explain the pros and cons.¹⁰⁰ These problems are compounded by weak regulation of genetic tests and other predictive health information.

Commercial companies have repeatedly made misleading claims about genetic test results, including those sold direct to consumer (DTC) online and via private doctors. The EU's new IVD (In-Vitro Diagnostics) draft Regulation is supposed to regulate predictive genetic tests and software but is effectively meaningless as it provides no regulatory check of the companies' claims.¹⁰¹ If adopted by the European Parliament, the new Regulation is likely to be used by commercial lobbyists in US-EU Free-trade negotiations to undermine attempts by the FDA to regulate genetic tests in the United States, including those sold by Google's gene test company 23andMe (which are currently suspended due to the failure of the company to supply evidence for the claims they make to the FDA).^{102,103,104,105,106}

Importantly, there are inherent limitations to the predictive value of genetic tests (and indeed, of any risk predictions) due to the complexity of natural systems, including interactions between environment and biology and the roles of choice and chance. Links between genetic factors and common diseases can provide useful clues about biology and how diseases develop.¹⁰⁷ But most genetic factors seem to change a person's risk of common diseases only very slightly. Rather than a single gene predisposing someone to disease, it now seems likely that everyone possesses hundreds, perhaps thousands, of genetic variants some of which slightly increase their risk, whilst others slightly decrease it.^{108,109} In reality, genetic variants do not have a property called risk, they act through their effects on complex biological pathways, and the risk of a particular genetic variant depends on the rest of biology and on the environment.^{110,111} This means that the idea that an individual's genetic risk of common diseases can be predicted has become increasingly controversial amongst scientists.^{112,113}

In general, common genetic differences are not more but less predictive than most other types of test, and no common genetic variants exist – either singly or in combination - that meet medical screening criteria for the general population.¹¹⁴

Many geneticists are puzzled at the lack of success in finding the expected inherited component of common diseases.^{115,116} However, whilst many believe that more genes will be discovered which explain this 'missing heritability', others have long criticised the calculations (usually made from twin studies), and claim that the assumptions used inevitably exaggerate and oversimplify the role of genes.¹¹⁷ This means that some or all of the so-called "missing heritability" that future research is supposed to find may not actually exist.^{118,119,120}

In reality, much research suggests that however much research is done and even if all genetic variants are identified, they will still have poor predictive value for most diseases in most people and limited clinical utility (i.e. little prospect of bringing any benefit to health).^{121,122,123,124} Inclusion of any gene-gene and gene-environment interactions (assuming they could be identified and even if an exact model of all interactions could be developed) will not improve this situation.¹²⁵

The major differences in people's health and life expectancy observed in Britain and throughout the world have little to do with individual differences in biology.^{126,127} Although some enthusiasts have tried to argue that the complexity of biology was unexpected, the

poor predictive value of genetic tests for common diseases is not really a surprise.^{128,129,130,131,132,133}

Nevertheless, there are many different types of genetic testing which can be useful in specific circumstances. Currently, clinical use of genetic testing in the NHS is restricted to tests of specific genes in specific circumstances, which include: diagnosis of genetic disorders (often in babies and children); carrier testing (identifying rare mutations which must be present in both parents before a child develops a particular disease) for specific diseases within screening programmes or specific families with affected members; testing for predisposition to the relatively rare familial forms of some disorders (particularly breast cancer) in members of high risk families (an example is the mutation in the BRCA1 gene recently identified in actor Angelina Jolie); cascade screening of family members already diagnosed with a genetic condition or predisposition; and pharmacogenetic testing (genetic tests to predict drug response) for a few specific drugs and conditions. Pre-natal testing and screening is also available for a small number of conditions. Cancer patients may also be given somatic (non-inherited) pharmacogenetic tests designed to identify specific genetic mutations or gene expression patterns in cancer tumours. A few cancer drugs are available which are given to specific groups of patients based on these test results.

The Wellcome Trust plan is based on a “disruptive business model” in which specific genetic and genomic tests are not conducted as and when they are necessary for a person’s care: instead costs are sunk up-front in whole genome sequencing for everyone. The purpose of this approach is to change the business model: by first sinking the costs (with the public sector paying for much of the infrastructure and for collecting and storing people’s electronic health records and DNA) and then feeding back predictions made about risk of disease or drug response. This will lead to a massive increase in the market for whole genome sequencing and data storage and analysis services, making a few venture capital investors (subsidised by R&D tax credits) very rich, whether or not any benefit to health is actually delivered. This is problematic because these investments are not accountable to customers (the people who have their DNA sequenced), because they have no choice, or to taxpayers (the people subsidising the public-private partnerships), because the plan has not been democratically debated and decided. Thus, the proposal fits neither a free-market nor a democratic-socialist economic model for investment and fails to be accountable to either the market or the general public. Further, seemingly independent institutions such as Oxford University benefit financially via their spin-out companies (such as Oxford Nanopore), in which they are investors, and from capital investments (such as the public-private infrastructure funding for Oxford University’s Li Ka Shing Centre for Health Information). Thus claims of future benefit (used to justify the costs and other downsides such as loss of privacy) originate almost entirely from those with vested interests in the plan.

Once all the data has been collected and stored, its use by individuals (with or without medical professionals) is then low cost (involving an electronic test of the person’s existing stored variant file, containing their whole genome minus the reference genome). This will allow this data to be mined for marketing, even when it is of no relevance to health, and also overcome the lack of interest of customers in purchasing gene tests from companies like 23andMe online. However, this plan requires enormous resources to be sunk in collecting and storing data which is likely to be of limited value to most people’s health. This is the opposite of the Future Forum’s recommendation that the NHS should be: “*Moving from a focus on collecting data (often too much data) to a focus on using data to generate intelligence to inform action*”.¹³⁴ Clinically useful data is likely to be swamped with clinically useless data which requires significant financial and energy resources to collect and store. Private sector investment will expect a high rate of return, whilst the return for taxpayers is only the claimed (but highly speculative) future benefit to health.

Rejecting the Wellcome Trust's proposal would not mean rejecting medical research or the use of genetic testing altogether. The alternative is to continue to recruit people to research studies which are separate from their care, with their fully informed consent, and to implement specific tests in healthcare as and when their benefits outweigh the risks. In this scenario, new tests will be introduced more gradually for specific groups of people, mainly those who are already ill or who have a strong family history of a particular disease. Whole genome sequencing would remain relatively rare and used mainly for children with undiagnosed genetic disorders and perhaps also for cancer tumours, if and when trials show that this can lead to better treatment. A more measured approach to the introduction of genetic and genomic tests would affect much smaller numbers of people and provide much less of a challenge to existing systems for assessing risks and benefits and for obtaining informed consent and protecting privacy.

At the same time, traditional public health approaches should be used to reduce the incidence of common diseases by tackling health inequalities, poor diets, smoking and pollution.

Consultation question 4:

What are the opportunities for, and the impacts of, use of linked biomedical data in research?

This is a problematic question because it is designed to illicit an answer from public and charitable sector scientists about what they would do if they could get hold of limitless amounts of data for free, without the chore of filling in too much paperwork. It does not consider who is actually going to use the data (e.g. Google, GE Healthcare, Bupa, BGI, the NSA), how this is going to be funded, who decides what is going to be funded, or what outcomes might be expected (see comments on commercial interests above).

GeneWatch UK disagrees with the claim that a focus on individual biological variation is a good priority for health research. Billions of pounds and dollars have already been wasted on poor research in this area. The Wellcome Trust has acted as lobbyists for investment in the area for decades and also promoted much of what is politely termed "genohype" in the media. Few scientists working in the area now seriously believe the claims that individual genomic risk predictions will transform medicine because their predictive value is so poor. It is not surprising that the idea is still being promoted since so many companies and venture capitalists expect to make so much money out of it (see above). But this does not mean that this approach is in the public interest.

This does not mean all such research should be abandoned (see above) but a return to a more targeted hypothesis driven approach (i.e. the opposite of Big Data) is much more likely to be cost-effective. It is also worth remembering that the major public health challenges (such as tackling obesity by improving diets and exercise) do not need more epidemiology, although more research on what works and what doesn't, and more political will to tackle food industry interests would not go amiss. It is worth remembering that the food industry's enthusiasm for genetically targeted diets (based on marketing functional foods to the "genetically susceptible") is undermined by the failure to find meaningful tests for genetic susceptibility to hypertension^{135,136,137,138,139} type 2 diabetes^{140,141, 142} or obesity^{143,144} (except in extremely rare cases). The role of the food industry (ILSI) in this approach is a classic example of vested interests driving the research agenda and diverting resources from more effective approaches, just as the tobacco industry did.

Consultation question 5:

What are the opportunities for, and the impacts of, data linking in medical practice?

Feeding back individual risk predictions will overall be bad for health as explained above, because there are no genetic tests that meet screening criteria for use in the general population and Numbers Needed to Treat will always be high.

The way people metabolise some drugs can sometimes be deduced from analysing their DNA: these tests are known as pharmacogenetic tests. However, these limited, specific applications also do not appear to justify a roll-out of whole genome sequencing to the whole population. It is notable that the “poster child” of pharmacogenetic testing for warfarin prescribing has recently failed to improve health outcomes in clinical trials.¹⁴⁵ Many other drugs do not have such a narrow therapeutic index and variability in response may also be only partly predictable from genetic tests, rendering them largely useless in the clinic. There is of course a possibility that some new tests will work, but they should be used, as they are now before prescribing a specific drug, if and when evidence from trials indicates improved health outcomes.^{146,147} Adverse drug reactions have been rising¹⁴⁸ but this is not because of an increase in genes for adverse drug reactions but because of a variety of factors including more older patients, possibly combined with weaker regulation and greater use of over-the-counter and off-label medicines (a trend that will be exacerbated along with over-treatment). For example, if drug use doubles as a result of treating healthy people based on their genetic risk (as some predictions suggest), adverse drug reactions and side effects will also increase significantly, even if pharmacogenetic tests are useful in some circumstances.

Of course there may be benefits to other aspects of data-sharing in medical practice, but these are undermined by the blanket approach which is designed to attract venture capital investment by exploiting vulnerable people.

Consultation question 6:

What are the opportunities for, and the impacts of, using biomedical data outside biomedical research and health care?

The main commercial opportunity expected by a wide range of commercial companies is to use individuals’ data for personalised marketing, as explained above.

It is a mystery why this is regarded as likely to lead to anything that is remotely in the “public interest”.

The main opportunity expected by governments is total surveillance of every individual and their relatives.

Consultation question 7:

What legal and governance mechanisms might support the ethical linking and use of biomedical data?

The most important principle must be to **give the public a say about what happens to their data and also about how taxpayers’ money is spent and the future of the NHS.** The history of the idea of creating a database of everyone’s medical records and DNA in the NHS has been one of untrustworthy institutions led by a small number of people hiding their true intent from members of the public whilst negotiating behind closed doors. Many aspects of the Health and Social Care Act have been criticised for not having been spelt out in the coalition parties’ manifestos. Failure to consult or inform the public about data-sharing plans is yet another failure to be open with the electorate.

The latest Government plan is just the latest in a long history of attempts to build a database of everybody’s DNA within the NHS, which has been promoted by a small group of government advisors, including the Wellcome Trust, since at least 1999.¹⁴⁹ The Department of Trade and Industry (DTI) 1999 report “*Genome Valley*” endorsed claims that genomics

would revolutionise healthcare by allowing predictive profiling, without making any assessment of the likely costs, or of the claimed benefits to health or the economy.¹⁵⁰ The report also highlighted the value of making NHS data available to industry for research as Britain's 'unique selling point' (USP) in the knowledge-based economy. This same plan – changed only by storing the data in the cloud rather than on a single government database - is now being promoted by Prime Minister David Cameron and Conservative ministers, as a result of being lobbied by many of the same vested interests as before. This raises important questions about the role of democracy and its relationship to science and technology. In practice a small circle of advisers is promoting what they claim will be a technical solution to rising healthcare costs and lack of economic growth, in the absence of any public scrutiny or debate about the pros and cons of their vision of the future.

It is unlikely that most people would agree to their medical records and genomes being handed over to Google and the private healthcare industry for commercial exploitation and to the police and security services in the UK and overseas. Thus, the attitude is that the whole thing has to be implemented before people wake up to what is going on, including sinking enormous amounts of taxpayers' money into building infrastructure. GeneWatch UK highlights that:

1. What is in the "public interest" cannot be decided by unaccountable committees who have a vested interest in a particular approach. **Research must have fully informed "opt in" consent**, not a blanket opt-out because there is no other way for people to have a say about who gets access to their data and how it is used. Scientists, ethicists and medical professionals should be careful of promoting a "presumed consent" approach on the basis of how they themselves would like to use the data: they will have no control over who else gets to use the data for what purposes and should not pretend to patients that they will. **Removing people's control over their own data will inevitably lead to commercial exploitation and excessive government surveillance;**
2. Adding genomes or genotypes to shared data is different from sharing medical records because genomes and genotypes are not needed for a person's care and act as biometrics which can be used to track individuals and their relatives. **Indefinite retention of biometrics without consent would breach Article 8 of the European Convention on Human Rights.** The benefits of doing this do not outweigh the harms;
3. Data-sharing without consent with companies like Google will lead to a massive loss of public trust. **It is unacceptable to fail to inform the public about who the "approved researchers" are likely to be and how they might use the data.** Many commercial companies have an interest in biometrics as a means to link multiple data sets for personalised marketing: this will lead to commercial exploitation of vulnerable individuals and end the role of the NHS as a service based on need. Nobody has voted to destroy the NHS in line with Syke's plan but this is exactly what is being implemented;
4. **There is no possibility of preventing total surveillance of every citizen if genome data linked to medical records is stored in the cloud:** the UK police and security services and foreign governments (e.g. the US and China) will have access to this data and it will allow them to track every individual and their relatives. Nobody has voted for this either;
5. **People should not be misled about "anonymisation":** there is a consensus amongst experts that "the computer will know who you are" and privacy will not exist anymore if the plan to share all NHS England medical records and genomes widely goes ahead. Sharing medical records alone (including postcodes, NHS number etc.) is likely to be sufficient to identify most individuals but adding genomes or genotypes would certainly destroy all anonymity. It is wrong to promise people that they will not be identified;

6. People should not be misled about the poor value of genomic risk predictions and the lack of utility of genomic testing for most diseases in most people: this will not be solved by collecting more data because it is inherent in the biological complexity of most diseases (and adverse drug reactions) and the important roles of external/environmental factors, choice and chance. **Misleading propaganda (known as “genohype”) should not form the basis of public policy or communication;**
7. **Individualised medicine (based on individual risk predictions) is likely to be bad for health** overall because it amounts to blanket screening leading to overtreatment for most people for most conditions. The commercial drivers for this approach are personalised marketing; a massive expansion in the drug, healthcare and genomic testing markets; and a (deliberate) distraction from public health approaches that tackle unhealthy products or pollution. Before implementing a major transformation in the NHS it would be a good idea to pay attention to the extensive evidence of poor predictive value of genetic screening.
8. **Collecting and sequencing or genotyping DNA from most people is not necessary or proportionate.** The benefits of genetic and genomic research are limited to specific populations groups and applications, there is no justification for wasting taxpayers money on building the vast infrastructure needed to sequence the whole population;
9. Sequencing or genotyping babies’ DNA may sometimes be necessary or useful for their care (for example to identify an undiagnosed genetic disorder) but **it is not necessary to store this data indefinitely or to share it for other (non-clinical) uses to which the child cannot give or refuse consent.** The principle of data minimisation should apply until the child is old enough to decide whether or not to take part in research or to take any further genetic/genomic tests relevant to adult-onset diseases. Research involving children should require the fully informed consent of their parents for specific purposes.
10. A step by step approach to introducing genomic technologies into healthcare is widely supported by genomic researchers and medical professionals. Rather than sequencing everyone and sharing all data with commercial companies, this would involve funding specific promising areas of research (such as sequencing for undiagnosed genetic disorders, and genotyping of cancer tumours). Instead of being driven by creating the maximum (subsidised) market possible for whole genome sequencing it would be driven by healthcare needs. **In this alternative approach, relatively limited amounts of DNA would be collected and stored and fully informed consent could be retained to recruit people to specific studies, with restricted data-sharing.**
11. Genetic testing remains unregulated and the current draft of the EU’s IVD Regulation will not significantly change this situation because it involves no pre-market regulatory assessment of clinical data. This means that unsubstantiated risk assessments made by computer algorithms will proliferate (and different risk predictions will be made by different companies). **Pre-market assessments by regulators should be required if people are not to be misled about their health and these should cover the clinical validity and clinical utility of tests (i.e. whether or not they improve health outcomes).** It should be noted that this requires specific clinical studies (such as controlled trials) for specific tests plus recommended interventions, not vast databases.
12. There has been no democratic debate about the plan to turn NHS England into a giant database, despite NHS data having been identified as Britain’s “unique selling point” in the knowledge-based economy well over a decade ago and the proposal being backed by successive governments in a (misguided) attempt to secure economic growth. **Enormous amounts of taxpayers’ money have already been wasted and billions more are likely to be spent. The lack of democratic debate and public accountability is totally unacceptable.**

-
- ¹ A DNA database in the NHS: Your freedom up for sale? GeneWatch UK Briefing. April 2013. http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/DNAinNHS_GWbriefing_fi_n.pdf
- ² Children could have DNA tested at birth. *The Telegraph*. 8th December 2013. <http://www.telegraph.co.uk/health/healthnews/10501788/Children-could-have-DNA-tested-at-birth.html>
- ³ GeneWatch UK (2004) Bar-coding babies: good for health? Briefing Number 27. <http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/brief27.pdf> .
- ⁴ Human Genetics Commission (2005) Profiling the newborn: a prospective gene technology? March 2005. www.hgc.gov.uk .
- ⁵ Sykes R (2000) *New medicines, the practice of medicine, and public policy*. Nuffield Trust, 2000. London, The Stationary Office.
- ⁶ MRC(2007) The use of personal health information in medical research. MRC/Ipsos MORI. 26 June 2007. <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003810>
- ⁷ Armstrong V, Barnett J, Cooper H, Monkman M, Moran-Ellis J, Shepherd R (2007) Public attitudes to research governance: A qualitative study in a deliberative context. Wellcome Trust. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtx038443.pdf
- ⁸ CfH (2008) Consultation on the wider use of patient information.
- ⁹ HGC (2008) NHS Connecting for Health – Consultation on Public, Patients, and other interested parties views on Additional Uses of Patient Data: Response by the Human Genetics Commission. 15 December 2008. <http://www.hgc.gov.uk/UploadDocs/DocPub/Document/HGC%20response%20to%20CfH%20consultation%20on%20additional%20uses%20of%20patient%20data%2015-12-08.doc>
- ¹⁰ NHS Connecting for Health (2009) Summary of responses to the consultation on additional uses of patient data. 27th November 2009. http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Consultations/Responsestoconsultations/DH_109310
- ¹¹ NHS England (2013) Better information means better care. January 2013. http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Documents/NHS_Door_drop_26-11-13.pdf
- ¹² Dialogue by Design (2007) *Science Horizons, Strands 1-3: Summary Report*. ScienceWise/Department for Universities, Innovation and Skills. September 2007. <http://www.sciencewise-erc.org.uk/cms/assets/Uploads/Project-files/sciencehorizons-summary-report.pdf>
- ¹³ Dialogue by Design (2007) *Science Horizons: Deliberative Panel Report*. ScienceWise/Department for Universities, Innovation and Skills. September 2007. <http://www.sciencewise-erc.org.uk/cms/assets/Uploads/Project-files/sciencehorizons-deliberative-panel-Report.pdf>
- ¹⁴ Dialogue by Design (2007) *Science Horizons: Deliberative Panel Report*. September 2007. http://www.sciencehorizons.org.uk/resources/sciencehorizons_deliberative_panel.pdf
- ¹⁵ World Medical Association (WMA) Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. <http://www.wma.net/en/30publications/10policies/b3/>
- ¹⁶ Dowty T, Korff D (2009) Protecting the Virtual Child: The Law and Children's Consent to Sharing Personal Data. Action on Rights for Children. The Nuffield Foundation. January 2009. <http://medconfidential.org/wp-content/uploads/2013/03/Protecting-the-virtual-child.pdf>
- ¹⁷ Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine Oviedo, 4.IV.1997. <http://conventions.coe.int/Treaty/en/Treaties/Html/164.htm>
- ¹⁸ National DNA database needed for personalised medicine drive. *The Telegraph*. 25th January 2012. <http://www.telegraph.co.uk/health/healthnews/9038712/National-DNA-database-needed-for-personalised-medicine-drive.html>
- ¹⁹ A DNA database in the NHS: Your freedom up for sale? GeneWatch UK Briefing. April 2013. http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/DNAinNHS_GWbriefing_fi_n.pdf
- ²⁰ <http://caldicott2.dh.gov.uk/>

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/210830/ethics_advice_letter_to_CMO.pdf

²² Opinion 3/2012 on developments in biometric technologies. Article 29 Data Protection Working Party. 00720/12/EN WP193. Adopted on 27th April 2012. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf

²³ Data Protection Act 1998. <http://www.legislation.gov.uk/ukpga/1998/29/contents>

²⁴ Case of S. and Marper v. the United Kingdom. Grand Chamber of the European Court of Human Rights. Judgment. Strasbourg, 4 December 2008.

<http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-90051>

²⁵ DNAScan <https://promo.gelifesciences.com/na/K12242/#.UswTJbTFqWE>

²⁶ The Global Alliance for Rapid DNA <http://rapiddna.org/about-the-global-alliance/board-members/>

²⁷ Railroad Agrees to Stop Gene-Testing Workers. *Washington Post*. 19th April 2001.

<http://www.washingtonpost.com/ac2/wp-dyn/A34877-2001Apr18?language=printer>

²⁸ Department of Health (2012) Agreement extended on predictive genetic tests and insurance. 26 June 2012. <http://www.dh.gov.uk/health/2012/06/genetics/>

²⁹ Chabris CF, Hebert BM, Benjamin DJ, et al. (2012) Most reported genetic associations with general intelligence are probably false positives. *Psychol Sci*. **23**(11):1314–1323. doi:10.1177/0956797611435528.

³⁰ De Moor MHM, Costa PT, Terracciano A, et al. (2012) Meta-analysis of genome-wide association studies for personality. *Mol. Psychiatry*. **17**(3):337–349. doi:10.1038/mp.2010.128.

³¹ Michael Gove held talks with 'IQ genes' professor. *The Independent*. 13th October 2013.

<http://www.independent.co.uk/news/education/education-news/michael-gove-held-talks-with-iq-genes-professor-8876500.html>

³² Inside 23andMe founder Anne Wojcicki's \$99 DNA Revolution. *Fast Company*. 14th October 2013.

<http://www.fastcompany.com/3018598/for-99-this-ceo-can-tell-you-what-might-kill-you-inside-23andme-founder-anne-wojcickis-dna-r>

³³ 'Big data' to revolutionise healthcare. *ENEWSPF*. 3rd May 2013. <http://enewspf.com/latest-news/health-and-fitness/42769-prime-minister-david-cameron-joins-sir-ka-shing-li-for-launch-of-90m-initiative-in-big-data-and-drug-discovery-at-oxford-university.html>

³⁴ Prime Minister joins Sir Ka-shing Li for launch of £90m initiative in big data and drug discovery at Oxford University. University of Oxford and Li Ka Shing Foundation Press Release. 3rd May 2013.

<http://www.hutchison-whampoa.com/en/media/topic/2591/LKSF-Oxford/>

³⁵ Manchester, Peking University Team up on Genomic Medicine Center. *GenomeWeb*. 17th October 2013.

<http://www.genomeweb.com/manchester-peking-university-team-genomic-medicine-center>

³⁶ Chancellor announces Sino-British health partnership. 14th October 2013.

<http://www.manchester.ac.uk/aboutus/news/display?id=10858>

³⁷ Peking University (2013) Health experts from Manchester and PKU join forces on genetic research.

http://english.pku.edu.cn/News_Events/News/Outlook/10373.htm

³⁸ Futurish Gamification. *Global Gaming Business*. 31st March 2011.

<http://ggbmagazine.com/issue/vol-10-no-4-april-2011/article/futurish-gamification>

³⁹ From password-free email accounts to medicine made from DNA: IBM reveals its five predictions for our world in 2018. *Daily Mail*. 18th December 2013. <http://www.dailymail.co.uk/sciencetech/article-2525770/The-death-passwords-rise-high-street-treating-cancers-DNA-level-IBM-reveals-predictions-five-years.html>

⁴⁰ HGP10: Genomic Data Privacy and Risk - George Church. March 2013.

<http://www.youtube.com/watch?v=7wAmQvoFdBo>

⁴¹ Williams R, Johnson P (2004) Circuits of surveillance. *Surveillance & Society* **2**(1): 1-14.

<http://www.surveillance-and-society.org/articles2%281%29/circuits.pdf>

⁴² National DNA Database Strategy Board Annual Report 2012-13. Home Office.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/252885/NDNAD_Annual_Report_2012-13.pdf

⁴³ Kaye J (2006) Police collection and access to DNA samples. *Genomics, Society and Policy*, **2**(1), 16-27. <http://www.lancs.ac.uk/fss/journals/gsp/vol2no1/kayeabstract.htm>

⁴⁴ House of Lords Science and Technology Committee (2008) Memorandum submitted by Academy of Medical Sciences. Examination of Witnesses. 16th July 2008.

<http://www.publications.parliament.uk/pa/ld200708/ldselect/ldscstech/999/stii080716ev8.pdf>

⁴⁵ UK Biobank Ethics and Governance Council. Annual Review 2012.

<http://www.egcukbiobank.org.uk/assets/wtp052058.pdf>

- ⁴⁶ ALSPAC (2013) Sixteen new genetic regions for allergies discovered (30th June 2013). <http://www.bristol.ac.uk/alspac/news/2013/183.html>
- ⁴⁷ Bønnelykke K, Matheson MC, Pers TH, et al. (2013) Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat Genet.* advance online publication. doi:10.1038/ng.2694.
- ⁴⁸ Hinds DA, McMahon G, Kiefer AK, et al. (2013) A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat Genet.* advance online publication. doi:10.1038/ng.2686.
- ⁴⁹ Schmidt H, Callier S. How anonymous is “anonymous”? Some suggestions towards a coherent universal coding system for genetic samples. *J Med Ethics*, **38**(5):304–309.
- ⁵⁰ Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2013) Identifying Personal Genomes by Surname Inference. *Science*, **339**, 321-324.
- ⁵¹ Genetic privacy. *Nature Editorial*. 17th January 2013. <http://www.nature.com/news/genetic-privacy-1.12238>
- ⁵² Sweeney L, Abu A, Winn J (2013) Identifying Participants in the Personal Genome Project by Name. <http://dataprivacylab.org/projects/pgp/>
- ⁵³ <http://www.qof.ic.nhs.uk/>
- ⁵⁴ <http://www.pegasus.nhs.uk/Resources/Core%20Info/core5.11.php>
- ⁵⁵ Guidance concerning Non Paternity issues. NHS Screening Programmes. Sickle Cell and Thalassaemia.
- ⁵⁶ House of Lords Science and Technology Committee. Oral Evidence. Corrected Transcript. 2nd July 2008. <http://www.publications.parliament.uk/pa/ld200708/ldselect/ldsctech/999/stii080702ev6.pdf>
- ⁵⁷ Find your genetic father...online. *ZDNet*. 6th November 2006. <http://www.zdnet.com/blog/emergingtech/find-your-genetic-father-online/70>
- ⁵⁸ GPs concerned over separated parents seeking children's medical records. *The Telegraph*. 4th January 2013. <http://www.telegraph.co.uk/health/healthnews/9779564/GPs-concerned-over-separated-parents-seeking-childrens-medical-records.html>
- ⁵⁹ The Big Opt Out. <http://www.thebigoptout.com/about-nhs-confidentiality/>
- ⁶⁰ Pewen WF (2012) Protecting Our Civil Rights in the Era of Digital Health. *The Atlantic*. 2nd August 2012. <http://www.theatlantic.com/health/archive/2012/08/protecting-our-civil-rights-in-the-era-of-digital-health/260343/>
- ⁶¹ ‘Father’ baffled by DNA results. *Khaleej Times*. 31st January 2013. http://www.khaleejtimes.com/kt-article-display-1.asp?section=courtroundup&xfile=data/courtroundup/2013/january/courtroundup_january79.xml
- ⁶² Ioannidis JPA (2013) Informed consent, big data, and the oxymoron of research that is not research. *Am J Bioeth*. **13**(4):40–42.
- ⁶³ Woolf M (2010) NHS uses babies' blood for secret DNA database. *The Sunday Times*. 23rd May 2010. [Subscription required] http://www.thesundaytimes.co.uk/sto/news/uk_news/Health/article297353.ece
- ⁶⁴ DNA from millions of newborn babies is secretly stored on NHS database. *Daily Mail*. 24th May 2010. <http://www.dailymail.co.uk/news/article-1280891/NHS-creates-secret-database-babies-blood-samples-parental-consent.html>
- ⁶⁵ NHS (2005) Code of Practice for the Retention and Storage of Residual Spots. <http://newbornbloodspot.screening.nhs.uk/cms.php?folder=2547>
- ⁶⁶ Newborn blood spot screening in the UK: Policies and standards April 2005. <http://newbornbloodspot.screening.nhs.uk/standards#fileid11025>
- ⁶⁷ Stein, R. Blood samples raise questions of privacy. *The Washington Post*. 30th June 2009. <http://www.washingtonpost.com/wp-dyn/content/article/2009/06/29/AR2009062903118.html>
- ⁶⁸ Anderson, J (2009) Texas to destroy 5mn blood samples stored without consent. *The Money Times*. 23rd December 2009. <http://www.themoneytimes.com/featured/20091223/texas-destroy-5mn-blood-samples-stored-without-consent-id-1094946.html>
- ⁶⁹ Hospital set to destroy baby blood samples. *The Sunday Times*. 24th January 2010. <http://www.timesonline.co.uk/tol/news/world/ireland/article7000049.ece>
- ⁷⁰ Greer, D. (2010) Class sues over infant blood sampling in B.C. *Courthouse News Service*. 27th May 2010. <http://www.courthousenews.com/2010/05/27/27621.htm>
- ⁷¹ Kaptchuk, T.J. 2003. Effect of interpretative bias on research evidence. *British Medical Journal* **326**: 1453-1455.
- ⁷² Stelfox, H., Chua G., O'Rourke, K., Detsky, A. 1998 Conflict of interest in the debate over calcium-channel antagonists. *New England Journal of Medicine* **2**: 101-106.

- ⁷³ Bhandari, M, Busse, J.W., Jackowski, D, et al., 2004. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. *Canadian Medical Association Journal* **170**(4):477-80.
- ⁷⁴ Friedman, L., Richter, E. 2004. Relationship between conflicts of interest and research results. *Journal of General Internal Medicine* **19**: 51–56.
- ⁷⁵ Lexchin, J., Bero, L.A., Djulbegovic, B., Clark, O. 2003. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *British Medical Journal* **326**: 1167-1170.
- ⁷⁶ Hartmann, M., Knoth, H., Schulz, D., Knoth, S. 2003. Industry sponsored studies in oncology vs. studies sponsored by nonprofit organisations. *British Journal of Cancer* **89**: 1405-1408.
- ⁷⁷ Katan, M.B. 2007. Does Industry Sponsorship Undermine the Integrity of Nutrition Research? *PLoS Medicine* **4**(1): e6 doi:10.1371/journal.pmed.0040006
- ⁷⁸ GeneWatch UK (2010) History of the Human Genome. June 2010. http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/HGPhistory_2.pdf
- ⁷⁹ Wallace HM (2009) Big Tobacco and the human genome: driving the scientific bandwagon? *Genomics, Society and Policy*, **5**(1), 1-54. <http://www.hss.ed.ac.uk/genomics/V5N1/documents/Wallace.pdf>
- ⁸⁰ Chen J, Qian F, Yan W, Shen B (2013) Translational biomedical informatics in the cloud: present and future. *Biomed Res Int*. 2013:658925. doi:10.1155/2013/658925.
- ⁸¹ Buchanan AV, Sholtis S, Richtsmeier J, Weiss KM (2009) What are genes “for” or where are traits “from”? What is the question? *Bioessays*, **31**(2):198–208.
- ⁸² Mewes HW (2013) Perspectives of a systems biology of the brain: the big data conundrum understanding psychiatric diseases. *Pharmacopsychiatry*, **46** Suppl 1:S2–9.
- ⁸³ All Watched Over by Loving Machines of Grace. *BBC Two*. May 2011. <http://www.bbc.co.uk/programmes/b011k45f>
- ⁸⁴ Behind AIG's Fall, Risk Models Failed to Pass Real-World Test. *Wall Street Journal*. 31st October 2008. <http://online.wsj.com/article/SB122538449722784635.html>
- ⁸⁵ Concato J, Feinstein AR, Holford TR (1993) The risk of determining risk with multivariable models. *Annals of Internal Medicine*, **118**, 201-210.
- ⁸⁶ Terwilliger JD, Weiss KM (2003) Confounding, ascertainment bias, and the blind quest for a genetic 'fountain of youth'. *Ann Med*, **35**,532-544
- ⁸⁷ I Had My DNA Picture Taken, With Varying Results. *New York Times*. 30th December 2013. http://www.nytimes.com/2013/12/31/science/i-had-my-dna-picture-taken-with-varying-results.html?pagewanted=1&_r=3&hp&_hpid=hp-science%3Agenetics%3Ahomepage%2Ft
- ⁸⁸ Wallace HM (2005) The development of UK Biobank: Excluding scientific controversy from ethical debate. *Critical Public Health* **15**(4): 323–333.
- ⁸⁹ Burke W, Tarini B, Press NA, Evans JP (2011) Genetic screening. *Epidemiol Rev.* **33**(1):148–164.
- ⁹⁰ Gilham, I., Rowland, T. (2001) Predictive medicine: Potential benefits from the integration of diagnostics and pharmaceuticals. *International Journal of Medical Marketing* **2**, 18-22.
- ⁹¹ Clarke AE, Mamo L, Fishman JR, Shim JK, Fosket JR (2003) Biomedicalization: Technoscientific Transformations of Health, Illness, and U.S. Biomedicine. *American Sociological Review*, **68**(2):161.
- ⁹² US Government Accountability Office (2006) Nutrigenetic testing: tests from four websites mislead consumers. July 2006. <http://www.gao.gov/products/GAO-06-977T>
- ⁹³ US Government Accountability Office (2010) Direct-To-Consumer Genetic Tests: Misleading Test Results Are Further Complicated by Deceptive Marketing and Other Questionable Practices July 2010. <http://www.gao.gov/products/gao-10-847t>
- ⁹⁴ Janssens ACJW, Gwinn M, Bradley LA, Oostra BA, van Duijn CM, Khoury MJ (2008) A Critical Appraisal of the Scientific Basis of Commercial Genomic Profiles Used to Assess Health Risks and Personalize Health Interventions. *The American Journal of Human Genetics* **82**, 593–599. [http://www.ajhg.org/AJHG/fulltext/S0002-9297\(08\)00145-6](http://www.ajhg.org/AJHG/fulltext/S0002-9297(08)00145-6) .
- ⁹⁵ GeneWatch UK investigations into Sciona: <http://www.genewatch.org/sub-425647> ; Nicotest: <http://www.genewatch.org/sub-406396> ; Respiragene: <http://www.genewatch.org/sub-566073> ; Genovations: <http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/Genovations.rtf> and Genetic Health: <http://www.genewatch.org/sub-558225> .
- ⁹⁶ Buchanan AV, Sholtis S, Richtsmeier J, Weiss KM (2009) What are genes “for” or where are traits “from”? What is the question? *Bioessays*, **31**(2):198–208. <http://europemc.org/articles/PMC2807122/>
- ⁹⁷ Rival genetic tests leave buyers confused. *The Sunday Times*. 7th September 2008. http://www.thesundaytimes.co.uk/sto/news/uk_news/article234529.ece

- ⁹⁸ Khoury MJ, Janssens ACJW, Ransohoff DF (2013) How can polygenic inheritance be used in population screening for common diseases? *Genet. Med.* Published online 14 February 2013. <http://www.nature.com/gim/journal/vaop/ncurrent/full/gim2012182a.html>
- ⁹⁹ Kalf RRJ, Bakker R, Janssens ACJW (2013) Predictive ability of direct-to-consumer pharmacogenetic testing: when is lack of evidence really lack of evidence? *Pharmacogenomics*, **14**(4):341–344.
- ¹⁰⁰ Harvard Mapping My DNA Turns Scary as Threatening Gene Emerges. *Bloomberg*. 15th February 2012. <http://www.bloomberg.com/news/2012-02-15/harvard-mapping-my-dna-turns-scary-as-threatening-gene-emerges.html>
- ¹⁰¹ IVD Regulation: problems still not solved. GeneWatch UK. May 2013. http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/IVD_Regulation_brief.pdf
- ¹⁰² 23andMe Wants FDA Approval for Personal DNA Testing. What Can It Reveal? *Time*. 2nd August 2012. <http://healthland.time.com/2012/08/02/23andme-wants-fda-approval-for-personal-dna-testing-what-can-it-reveal/>
- ¹⁰³ FDA panel advises caution on personal genetic testing. *Los Angeles Times*. 9th March 2011. <http://articles.latimes.com/2011/mar/09/nation/la-na-genetic-testing-20110309>
- ¹⁰⁴ Hayes WS, Allingham-Hawkins D (2013) Genetic testing must have more oversight. *Managed Healthcare Executive*. 1st April 2013. <http://managedhealthcareexecutive.modernmedicine.com/managed-healthcare-executive/news/user-defined-tags/oncology/genetic-testing-must-have-more-oversigh>
- ¹⁰⁵ FDA (2013) 23andMe, Inc. 22nd November 2013. <http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2013/ucm376296.htm>
- ¹⁰⁶ Gene testing firm 23andMe halts marketing. *BBC*. 3rd December 2013. <http://www.bbc.co.uk/news/science-environment-25209956>
- ¹⁰⁷ Hirschorn JN (2009) Genome Wide Association studies – illuminating biological pathways. *New England Journal of Medicine*, **360**, 1699-1701.
- ¹⁰⁸ Munafò MR (2009) The clinical utility of genetic tests. *Addiction*, **104**, 127-128.
- ¹⁰⁹ Gartner CE, Barendregt JJ, Hall WD (2009) Multiple genetic tests for susceptibility to smoking do not outperform simple family history. *Addiction*, **104**, 118-126.
- ¹¹⁰ Porteous JW (2004) A rational treatment of Mendelian genetics. *Theor Biol Med Model.*, **1**:6. DOI: 10.1186/1742-4682-1-6
- ¹¹¹ Tryggvadottir L, Sigvaldason H, Olafsdottir GH, Jonasson JG, Jonsson T, Tulinius H, Eyfjord JE (2006) Population-based study of changing breast cancer risk in Icelandic BRCA2 mutation carriers, 1920–2000. *J Natl Cancer Inst*, **98**(2):116-122.
- ¹¹² Janssens ACJW, van Duijn CM (2008) Genome-based prediction of common diseases: advances and prospects. *Human Molecular Genetics*, **17**(R2), R166-R173
- ¹¹³ Wilkie A (2006) Polygenic inheritance and genetic susceptibility screening. *Encyclopedia of Life Sciences*. DOI: 10.1002/9780470015902.a0005638. <http://mrw.interscience.wiley.com/emrw/9780470015902/els/article/a0005638/current/abstract?hd=All,9780470015902.a0005638>
- Article Online Posting Date: September 15, 2006
- ¹¹⁴ Jakobsdottir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE (2009) Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genetics*, **5**(2), e1000337.
- ¹¹⁵ Maher B (2008) The case of the missing heritability. *Nature*, **456** (6), 18-21.
- ¹¹⁶ Wade N (2009) Genes show limited value in predicting diseases. *New York Times*. 15th April 2009. http://www.nytimes.com/2009/04/16/health/research/16gene.html?_r=2&ref=global-home
- ¹¹⁷ Rose SPR (2006) Heritability estimates – long past their sell-by date. *International Journal of Epidemiology*, **35**, 525-527. <http://ije.oxfordjournals.org/cgi/reprint/35/3/525>
- ¹¹⁸ Wallace HM (2006) A model of gene-gene and gene-environment interactions and its implications for targeting environmental interventions by genotype. *Theoretical Biology and Medical Modelling*, **3** (35), doi:10.1186/1742-4682-3-35. <http://www.tbiomed.com/content/3/1/35>
- ¹¹⁹ Zuk et al. (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *PNAS Early Edition*. On: <http://www.pnas.org/content/early/2012/01/04/1119675109.full.pdf>
- ¹²⁰ Joseph, J (2012) The “Missing Heritability” of Psychiatric Disorders: Elusive Genes or Non-Existent Genes? *Applied Developmental Science*, **16**(2), 65–83, 2012
- ¹²¹ Wilkie A (2006) Polygenic inheritance and genetic susceptibility screening. *Encyclopedia of Life Sciences*. DOI: 10.1002/9780470015902.a0005638.

<http://mrw.interscience.wiley.com/emrw/9780470015902/els/article/a0005638/current/abstract?hd=All.9780470015902.a0005638>

Article Online Posting Date: September 15, 2006

¹²² Clayton, DG (2009) Prediction and Interaction in Complex Disease Genetics: Experience in Type 1 Diabetes. *PLoS Genetics*, 5(7): e1000540. On:

<http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1000540>

¹²³ Millikan RC (2006) Commentary: The Human Genome: philosopher's stone or magic wand? *International Journal of Epidemiology*, **35**, 578-580.

¹²⁴ Roberts NJ, Vogelstein, JT, Parmigiani G, Kinzler, KW, Vogelstein B, Velculescu VE (2012) The predictive capacity of personal genome sequencing. *Science Translational Medicine* 3003380. Published ahead of print 2 April 2012.

¹²⁵ Aschard H, Chen J, Cornelis MC, Chibnik LB, Karlson EW, Kraft P (2012) Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. *The American Journal of Human Genetics*, **90**, 962-972.

¹²⁶ WHO(2008) Closing the gap in a generation. Commission on Social Determinants of Health. http://whqlibdoc.who.int/publications/2008/9789241563703_eng.pdf

¹²⁷ <http://www.who.int/mediacentre/news/releases/2008/pr29/en/index.html>

¹²⁸ Strohmman RC (1997) The coming Kuhnian revolution in biology. *Nature Biotechnology*, **15**:194-200

¹²⁹ Holtzman NA, Marteau TM (2000) Will genetics revolutionize medicine? *New Engl J Med*, **343**, 141-144.

¹³⁰ Vineis P, Schulte P, McMichael AJ (2001) Misconceptions about the use of genetic tests in populations. *Lancet*, **357**: 709-712.

¹³¹ Baird P (2001) The Human Genome Project, genetics and health. *Community Genetics*, **4**, 77-80.

¹³² Cooper RS, Psaty BM (2003) Genetics and medicine: distraction, incremental progress, or the dawn of a new age? *Ann Intern Med*, **138**, 576-580.

¹³³ Vineis P, Ahsan H, Parker M (2004) Genetic screening and occupational and environmental exposures. *Occup Environ Med*, **62**, 657-662.

¹³⁴ Information: A report from the NHS Future Forum. 10th January 2012.

http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_132086.p df

¹³⁵ Liu W, Zhao W, Chase GA (2004) Genome scan meta-analysis for hypertension. *American Journal of Hypertension*, **17**, 1100-6.

¹³⁶ Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, et al. (2009) Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics*, **41**(6), 666-676.

¹³⁷ Liu W, Zhao W, Chase GA (2004) Genome scan meta-analysis for hypertension. *American Journal of Hypertension*, **17**, 1100-6.

¹³⁸ Ehret GB (2010) Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension. *Curr Hypertens Rep*. 12(1):17–25.

¹³⁹ Larsson E, Wahlstrand B, Hedblad B, et al. (2013) Hypertension and genetic variation in endothelial-specific genes. *PLoS ONE*, 8(4):e62035.

¹⁴⁰ Meigs JB, Shrader P, Sullivan LM, McAtteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PWF, D'Agostino RB, Cupples LA (2008) Genotype score in addition to common risk factors for prediction of Type 2 Diabetes. *New England Journal of Medicine*, **359**, 2208-2219.

¹⁴¹ Mihaescu R, van Hoek M, Sijbrands EJG, et al. (2009) Evaluation of risk prediction updates from commercial genome-wide scans. *Genetics in Medicine* **11**(8):588–594.

¹⁴² Mihaescu R, Meigs J, Sijbrands E, Janssens AC (2011) Genetic risk profiling for prediction of type 2 diabetes. *PLoS Curr*. 2011;3. doi:10.1371/currents.RRN1208.

¹⁴³ Hofker M, Wijmenga C (2009) A supersized list of obesity genes. *Nature Genetics*, **41**, 139-140.

¹⁴⁴ Sandholt CH, Hansen T, Pedersen O (2012) Beyond the fourth wave of genome-wide obesity association studies. *Nutr Diabetes*. **2**(7):e37. doi:10.1038/nutd.2012.9.

¹⁴⁵ Genetic Testing Falls Flat in Large Patient Drug Trials. *Technology Review*. 19th November 2013. <http://www.technologyreview.com/view/521861/genetic-testing-falls-flat-in-large-patient-drug-trials/>

¹⁴⁶ Gardiner SJ, Begg EJ(2006) Pharmacogenetics, drug-metabolizing enzymes, and clinical practice. *Pharmacological Reviews*, 58(3), 521-590.

¹⁴⁷ Lenzini, P. et al. (2011) Integration of genetic, clinical, and INR data to refine warfarin dosing. *Clin Pharmacol Ther*. 2010 May ; **87**(5): 572–578. doi:10.1038/clpt.2010.13.

¹⁴⁸ Wu T-Y, Jen M-H, Bottle A, et al. (2010) Ten-year trends in hospital admissions for adverse drug reactions in England 1999-2009. *J R Soc Med*. **103**(6):239–250. doi:10.1258/jrsm.2010.100113.

¹⁴⁹ GeneWatch UK (2009) Bioscience for Life? Appendix A. The history of UK Biobank, electronic medical records in the NHS, and the proposal for data-sharing without consent. January 2009. http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/UK_Biobank_fin_1.pdf

¹⁵⁰ Genome Valley: the economic potential and strategic importance of biotechnology in the UK. December 1999. <http://www.berr.gov.uk/files/file28709.pdf>