Should you share your DNA data?

Background

Much has changed in the two years since we last discussed the topic of sharing DNA data. New, tighter controls on storing and sharing personal data have been implemented following the EU's General Data Protection Regulation, resulting in companies bombarding their customers with requests for consent. The Cambridge Analytica scandal has shaken public confidence in the companies that profit from their data and opened our eyes to the risks of secret and aggregated data collection. We now live in an age when an unprecedented amount of data is collected about our lives, and this data is not a finite resource – there are infinite ways to analyse and extrapolate our potential future actions and desires from it.

At the same time, the full DNA (or genome) of thousands of volunteers is being collected for research and the NHS is preparing for the rollout of medical services that use this data to diagnose, treat and prevent serious illnesses of willing patients. This data has huge potential for improving the health of everyone using the NHS, but genomic data is the most personal type of information, and it requires special care and protection.

This round of Future Debates will consider the benefits and risks of a medical system that uses genomic data. How should it be stored and who should have access? Should everyone's genome be analysed at birth? And is there a duty to donate your genome, like we do with blood and organs?

What is DNA and the genome?

Your DNA contains all the instructions needed to make you, repair you and run you. Like all living things, you have a copy in almost every cell in your body. Your genome is the whole set of your DNA and contains about 20,000 genes – sets of specific recipes or instructions. But genes are only 2% of your genome. People thought the remaining DNA was junk, but it's now known to have an important role in controlling and regulating genes. That's why there's so much interest now in looking at the whole genome. The genome's instructions are written in an alphabet that uses only 4 characters, A, G, T & C but there are 3 billion of them. They can be 'read' one by one using a technique called sequencing.

Decoding a single reference genome – the Human Genome Project - took 13 years and cost almost $3 billion (Hayden, 2014). A genome can now be sequenced for about £700 in less than a day.

How is genome data interpreted and stored?

Your genome data is recorded as a sequence of As, Ts, Gs and Cs. You'd use most of your laptop's memory storing just your own, which is why collections of genomes have to be stored on huge computer mainframes. About 99.8% of human genomes are identical but it's the remaining 0.2% that makes us each unique. But even 0. 2% is around 3 million letters. Some of these changes or 'variants' are just that, variations that have no health significance. To diagnose the cause of illness, a genome is compared to the reference

Human Genome Project one, using online genome browsers, which allow scientists to perform analysis (bioinformatics). If there are changes, researchers need to know details of a person's health to help them decide whether variations they see are healthy ones or pathogenic (cause disease). Genome data isn't held in medical records although the results of analysis are.

What genomes can tell us

Rare disease

Genomic data is a powerful tool for diagnosing rare diseases. While they are individually rare, taken together the amount of people with a 'rare disease' is estimated as 1 in 17 (Burn 2013). About 80% of rare diseases are genetic.

Common disease

The causes of common disease are often complex and are the result of interactions between behavioural, environmental and genetic factors, so genomics isn't as helpful in deciphering these although it may help 'stratify' or indicate which treatment approaches might work best.

Pharmacogenomics

Genomic data can also provide information about how an individual will respond to different drugs. This is of great value in ensuring effective medicines with minimal side effects

Cancer

Cancer cells contain genomes which have many changes in them – up to a million in some. Seeing which changes are most common helps decide which treatments might work best. It might also help us understand why cancers occur.

Key issues in data sharing

Some changes in the genome may be perfectly normal, so it's only when large numbers of individuals with the same condition are compared that you can begin to understand which changes might be important in, for example determining the

severity of disease. So, progress in genomic research depends on the willingness of large numbers of individuals to contribute their genome data to research studies. Researchers have identified the sharing of data as an absolute necessity for accurate interpretation (Wright, Hurles and Firth 2015). However scientific progress must be balanced with concerns about privacy interests and the autonomy of patients and study participants.

A recent report, Wellcome (2016) identified four key principles that the public thought important regarding sharing health information. These were:

Why?

Social benefit vs. commercial profit. Research shows that the public are more willing to support access to health care data if there is a public good and a benefit to wider society. The public are very sceptical and concerned about data being used for profit or commercial reasons. If this is the only reason that data will be accessed by an institution, the public do not want their data shared.

Who?

Public vs Private companies. The public are most open to academic researchers, charities or partnerships between these and the public sector having access to their data. Patients who had been seen by clinical genetics in particular welcome the sharing of data for research purposes. This could reflect the fact the importance of data sharing for diagnosis of rare disease does not represent a new concept to these patients.

The public are open to sharing their data with companies who will make a profit, such as pharmaceutical companies. However, the public shows a desire for this to be regulated, for there to be limited profits and to be sure that if sharing their data, there will be a net gain for society. The government is generally not trusted to use data responsibly. A study for the Royal Statistical Society (Royal Society 2014) found that only 13% (n=2019) of people trust the government to use their data appropriately. One study that explored the views of families who been affected by a genetic condition found that the government in the UK was less trusted with data than private companies (Genetic Alliance 2015). There is variation in public opinion on the significance of commercial organisation being involved. In a recent survey, 61% of people would rather see commercial access to data than lose out on benefits from the research. However, a significant minority (25%) would still rather research did not happen if commercial organisations had to have access to their health data. The same survey showed that there was a core segment (17%) of people who would not allow commercial organisations to have access to their health data under any circumstances (Wellcome 2016).

What?

The type of data that is shared is considered in different ways by members of the public. Aggregated or anonymised data is considered less risky on

an individual level, however some studies have shown that members of the public are concerned about the potential for this data to be risky to society as it could lead to groups being discriminated against. The public often thought that sharing non-aggregated data represented a threat to personal security. Additionally, the public demonstrate particular concern about commercial companies having access to genetic data. In one study this was considered a ‘red line’ due to uncertainty about how this might be shared. This attitude is potentiality ameliorated by increased knowledge, as health professionals, who have a more nuanced understating of how data is shared, have a more relaxed attitude to sharing data with commercial companies.

How?

When health data is collected, there is strong public opinion that safeguards should be in place. However, the public rarely has a deep understanding of what these should be. Beliefs about how safe data will be in research is often linked to the public in the institution collecting the data. In the UK for example, involvement of the NHS acts a reassurance that data would be kept safe and secure.

When considering access to health data (including genomics) the public considered it unacceptable if: There is no public benefit, if profit was the only motive, or if the information would allow for identification.

Attitudes towards genetic testing

The public demonstrates a wide range and often some ambivalent opinions regarding the acceptability of genome sequencing. Regarding their personal genome, the public are most comfortable with information about their genome that provides them with knowledge about a medical condition that is either preventable or for which there is clear treatment (Middleton 2016). They are most uncomfortable with adult onset conditions for which there is no, or limited, treatment, such as Huntington's Disease (Bacon, Harris et al. 2015).

There is also emerging evidence from a number of studies that show the public have a growing desire to have as much information about their own genome as possible. For example, a recent study found that 70% of 4,961 participants from the general public wanted information about their genome, even if these had little relevance to their health (Middleton 2016). This finding has been supported by a recent study that found that participants support the return of raw sequence data in genomic research (Middleton 2015). It is important to highlight the variability of the findings on this subject. For example, contrasting with this study above, a recent survey from Holland found that only 24% of 961 participants supported genetic testing for all kinds of disease (Henneman, Vermeulen et al. 2013).

Broader concerns

There is some concern among members of the public that the proliferation of genomic sequencing represents the expansion of the 'surveillance society.' (Wellcome 2016). Concerns have also been voiced from pressure groups and

academics writing from disability rights perspective (Parens and Asch 2003) and from those with concern about race (Duster 2004). From the disability rights perspective, concerns with routine genome sequencing at birth represent a worry that societal judgments about what is 'normal' will become medicalised (Shakespeare 1998). From a race perspective, there is concern that collection of DNA for forensics and criminal cases could represent current racial prejudice and discriminatory attitudes (Duster 2004).

Access to genetic counselling

Patients strongly value the input of genetic counselling when having genomic testing. Recent research by Genetic Alliance UK (Genetic Alliance 2015) found that while patients wanted as much information about their health from genome sequencing as possible, they wanted these findings returned by trained professionals (such as a genetic counsellor). The public thinks it's important to have access to genetic counselling both before and after testing. (Genetic Alliance 2015)