Should we be making use of genetic genealogy to assist in solving crime?

A report on the feasibility of such methods in the UK

September 2020

The Biometrics and Forensics Ethics Group

The BFEG is an advisory non-departmental public body, sponsored by the Home Office. The group provides advice on ethical issues in the use of biometric and forensic identification techniques such as DNA, fingerprints, and facial recognition technology. The BFEG also advises on ethical considerations in the use of large and complex data sets and projects using explainable data-driven technology.

Contents

Executive Summary

The use of genetic genealogy databases by law enforcement came to prominence in the search for a suspect in the 'Golden State killer' in California, USA. Over 8,000 potential suspects had already been considered when police arranged for a preserved sample of DNA from one of the crime scenes to be analysed and uploaded to a genetic genealogy database. Following several thousand hours of genealogical research, and police work, a suspect was identified and charged with 13 murders.

The Biometrics and Forensics Ethics Group (BFEG) was asked to consider the feasibility of the use of genetic genealogy resources for the identification of suspects in criminal cases in the UK.

This report is aimed at those involved in the criminal justice system in the UK, with an interest in understanding how to take advantage of new technologies while maintaining an ethical approach. The report explains how genetic genealogy works, and how it has been used to identify suspects in criminal cases outside the UK. Some of the challenges of the method are described, and its use is compared with traditional familial DNA searching. As most use of genetic genealogy methods by law enforcement has so far occurred in the USA, this report addresses the feasibility, and necessity of using such methods in the UK.

What is genetic genealogy?
Genetic genealogy is the application of DNA analysis and traditional genealogy to infer relationships between individuals. Individuals who are closely related share segments of DNA, and the more distant the relationship, the less DNA they share. We humans share 50% of our DNA with a parent but when we reach a third cousin (common great-great grandparents) we share, on average, only around 0.8%. The amount of sharing is established by finding the sequence of hundreds of thousands of variable points scattered along the human genome, the complete set of DNA we carry. Genetic genealogy has become popular and by early 2019 around 26 million people had provided their DNA to various testing companies, the main ones being 23andMe, AncestryDNA, MyHeritage and FamilyTreeDNA. Most individuals who have provided DNA are of mainly Western European heritage and many are North American residents. Individuals who have been tested by different companies can be compared provided they submit their DNA data to a third-party online database, known as GEDmatch.

Access to genetic genealogy databases for law enforcement use

This process of uploading DNA from a crime scene in the 'Golden State killer' case to GEDmatch violated the terms and conditions of use. These terms stated that the person submitting the DNA had to declare that; it was their own DNA; or they were the legal guardian of the DNA donor; or they were otherwise authorised.

23andMe, AncestryDNA and MyHeritage do not allow law enforcement use of their databases without a warrant. FamilyTreeDNA offers an 'opt-out from law enforcement matching' possibility, and all European users are automatically opted out in line with the EU General Data Protection Regulation (GDPR). Contributors to GEDmatch, which allows law enforcement use of 'public' profiles with permission in serious cases, must actively opt in to law enforcement matching.

The number of profiles available to law enforcement on genealogy databases will affect the chance of successfully identifying potential suspects.

Technical and economic challenges to law enforcement use

Genetic genealogy analysis requires a quantity of high-quality DNA to be recovered from the crime scene. Effectively, this limits its use to a sample of undegraded semen, saliva or blood.

Once a potential link to an individual has been found, confirmation of this will require comparison with a standard short tandem repeat (STR)-based DNA profile, which entails obtaining a reference sample from that person.

If the perpetrator of the crime is not Western European in origin, then the chance of success is likely to be limited because of the under-representation of individuals from other parts of the world in the GEDmatch database.

In the USA, Parabon carried out an assessment of 200 cases for the suitability of a genealogical approach. Around 35% were not suitable, reflecting the fact that relatives may not be in the database or may not be identifiable because they are not genetic relatives (sharing DNA). An experiment at a UK forensic science provider, Eurofins UK, identified four out of ten anonymised UK volunteers using genetic genealogy, suggesting that the applicability of the method to UK cases may be similar to that demonstrated in the USA.

The genealogical research effort is considerable and could be very expensive; on average Parabon reports cases take around 24 hours and cost around $5000 (£4,000).

Familial DNA searching compared with genetic genealogy

Familial searching has been used in the UK for serious crimes since 2003. The technique uses standard STR-based DNA profiles and ranks the likelihood of a familial relationship between an unknown individual who has left DNA at a crime scene and individuals on the National DNA Database. This technique can only identify parents, children or siblings and the success rate is around 20%.

The apparent high clear-up rate of cold cases in the USA using genetic genealogy masks the USA's backlog of unanalysed DNA from rape cases, and issues in adding DNA profiles from both suspects and convicted individuals to the US DNA database (CODIS). Of note, the brother of the alleged Golden State killer was a convicted felon and if his DNA profile had been present on CODIS and familial searching had been used then the suspect could have been identified earlier.

The report concludes

The UK already has one of the most efficient DNA databases in the world and conventional methods, with appropriately applied familial searches, will identify the bulk of perpetrators.

If a genealogy approach is used, a proportion of the potential relatives will not be UK based and this could add significantly to the genealogical effort and cost.

Although surveys show public support for its use in the USA, there is concern that such surveys could be biased.

The cases for which a genetic genealogy approach may be considered must be clearly defined to enable an ethical and reasoned decision to be made. Permission from the Forensic Information Database Strategy Board should be required.

The initial use of the method in identifying otherwise unidentifiable bodies, would allow its potential to be tested in a UK setting, while avoiding some of the more contentious issues.

At the time of writing, the whole process is unregulated, and ethical, legal, and safeguarding issues must be considered.

Ethical, legal and safeguarding considerations

The legality and necessity of police use of genetic genealogy in the UK would need to be clearly established, with reference to Article 8 of the European Convention of Human Rights (ECHR) and the Human Rights Act 1998.

The approach should only be used if it can be shown to be based on clear evidence, verified by an independent body, that the established methods already in use for these law enforcement purposes are no longer adequate or effective.

Otherwise, the use of any such novel processes would not meet the tests of necessity and proportionality. This would make the legality of using such novel processes highly suspect.

The legality of using informed consent as the sole appropriate legal basis to obtain highly sensitive data is doubtful, in view of the [Data Protection Act 2018 (Part 3)](), the EU Law Enforcement Directive, and well-established Article 8 ECHR case law.

Avoidance of the unnecessary invasion of an individual's privacy and data security should be paramount and so guidance should be provided to limit direct and indirect activity in investigating potential distant relatives.

Processes would be needed to maintain the security of data, including sensitive medical and personal data.

Genealogists must have the necessary skills and knowledge. Ideally there should be a form of accreditation and a professional body to ensure:

- quality of identifications;
- acceptance of appropriate conduct requirements; and
- confidentiality and privacy.

The genetic analysis would need to be done in an accredited analytical environment that meets chain of custody, security, process and confidentiality requirements.

Legislation for the transmission, length of retention, and destruction of the sample, profile and collected genealogical data would be needed.

Introduction and definitions

There has been considerable interest in the use of genetic genealogy worldwide since the 2017 identification of a man suspected of committing a series of rapes and murders in the 1970s and 1980s in the USA (the 'Golden State killer'). Since 2017 interest has grown markedly, due to its use leading to the identification of potential perpetrators in more than 50 unresolved serious criminal cases in the USA.

This report is aimed at readers interested in the criminal justice system in the UK, and how it could develop to take advantage of new technologies while maintaining an ethical approach. It explains how genetic genealogy works and how it has been used to identify suspects in criminal cases outside the UK. The technical and economic challenges are described, and the use of the method is compared with traditional familial searching that is currently in use. Most activity so far has been in the USA, and the report addresses the feasibility, and necessity of using such methods in the UK.

What is genetic genealogy?

Genetic genealogy is the application of DNA analysis, along with traditional genealogy approaches to infer relationships between individuals that were developed particularly to provide information for family historians. Initially, analysis of genetic markers (short tandem repeats – STRs) on the male- specific Y chromosome was the focus, as a virtually identical set of markers is inherited down the paternal line and is potentially associated with a surname, also inherited paternally in many societies. However, since 2007 genetic-testing companies have been offering more powerful tests of DNA in the autosomes (non-sex chromosomes).

Individuals who are closely related share segments of autosomal DNA that they have inherited from their shared ancestors – the more distant the relationship, the less they have in common. We humans share 50% of our DNA with a parent but when we reach a third cousin (common great-great grandparents) we share, on average, only around 0.8%. It is the amount of sharing that is used to suggest a level of familial relationship. Not only do we each have, on average, around 175 third cousins, there is also about a 2% chance that a given pair of third cousins will share no DNA at all. This chance of non- sharing increases as relationships become more distant and makes identifying the approximate 1,500 fourth cousins and 17,500 fifth cousins much more problematic. There are also increasing numbers of 'false positives' – individuals who appear related at a particular distant degree, but only through the general shared ancestry that is observed within most human populations.

In order to identify the amount of DNA sharing, genetic-testing companies do not sequence the whole of the human genome (the DNA complement we carry). Instead they define the sequence of hundreds of thousands of points scattered along the genome, on each autosomal chromosome. These are points that commonly vary between people (single nucleotide polymorphisms – SNPs) and can be used to infer an identical sequence segment of DNA along a chromosome shared between two individuals. Identifying amounts of sharing across all the chromosomes (and the lengths of the shared DNA segments) can then suggest how closely these individuals are related.

By early 2019 around 26 million people had provided their DNA to the various testing companies, which hold the genetic information in proprietary databases for their users for family (and other) research purposes (International Society of Genetic Genealogy, 2019). Most contributors are individuals of mainly northern European heritage and many are North American residents.

Because related individuals may have been tested by different companies – and therefore cannot be directly associated – a third-party online database was set up for individuals to compare information, known as GEDmatch. This particular database offered useful genomic tools for the family researcher, allowing putative relatives within the database to be identified, as well as information about family trees of members, if provided.

How has it been used in criminal justice

The Golden State killer

The person responsible for a series of murders, rapes and multiple burglaries in California between 1974 and 1986, named the 'Golden State killer' in 2013, had been unidentified until police arranged for a preserved sample of DNA from one of the crime scenes to be analysed using an autosomal SNP test, and uploaded it to the GEDmatch database. Over 8,000 suspects had already been considered. An earlier Y-chromosome match to an individual in a public database led police to a man in a nursing home. After police had successfully subpoenaed the genetic-testing company, Family Tree DNA, to obtain his name and required the elderly man to provide a DNA sample, he was shown not to be the perpetrator, highlighting the potential for invasion of an individual's privacy.

This process of uploading the perpetrator's SNP profile to GEDmatch violated the terms and conditions of use that were in place at the time, as when uploading the crime scene DNA sample the police had to declare that:

- it was either their own DNA; or
- they were the legal guardian of the DNA donor; or
- they were otherwise authorised.

Nevertheless, this issue was ignored because of the success of the process – the ends justified the means. Around 20 individuals who were potentially related to the perpetrator as third or fourth cousins were identified. It then took several thousands of hours of genealogical research to identify the likely suspect, James Joseph DeAngelo; DNA collected from a discarded tissue and from the door handle of his car provided the direct link to the crime scene material (using conventional DNA profiling) and he was charged with 13 murders.

Genetic databases for genealogy

The main companies that provide 'direct-to-consumer' genetic testing to the public are 23andMe, AncestryDNA, MyHeritage and FamilyTreeDNA. There are several others, and it is anticipated that the number will increase as the cost of testing goes down and interest in genetic testing by the public grows. The potential use of these companies to provide genetic information for legitimate law enforcement use is limited: 23andMe, AncestryDNA and MyHeritage do not allow it without a warrant.
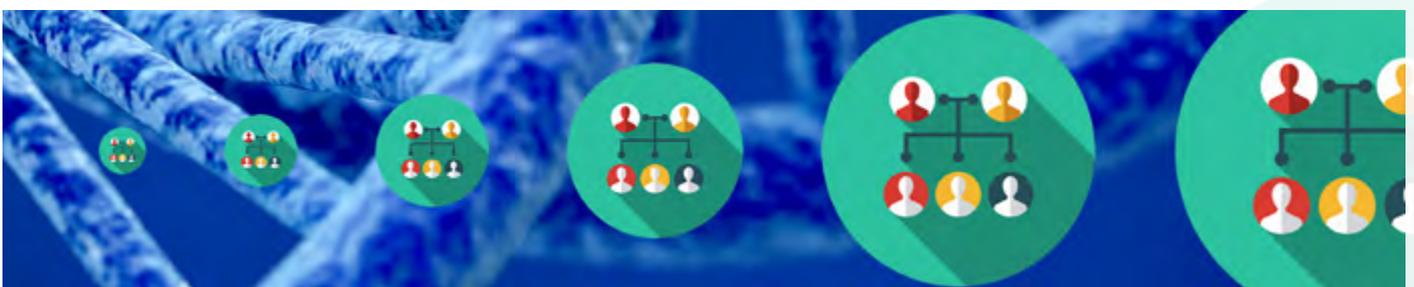
FamilyTreeDNA (with over one million contributors) has been collaborating with the FBI for some time and it faced criticism from customers once this became known. Now it openly engages with the public, appealing to people to have their DNA tested, particularly in areas of the world where coverage is poor. The company offers an 'opt-out from law enforcement matching' possibility, but not many contributors have used this, other than the European users who have all been automatically opted out in line with the EU General Data Protection Regulation (GDPR) and the EU Law Enforcement Directive. European users have to actively consent to law enforcement use and opt in if they wished to contribute. At the time of writing, the use of the FamilyTreeDNA database by law enforcement is limited to the USA.

GEDmatch, which has over one million contributors and has been used in most of the reported US cases, originally had three categories for profile donors, of which only the 'public' category was open to law enforcement use with permission in serious cases. When police persuaded GEDmatch to help to identify a teenager who broke into a church and assaulted the elderly organist, this provoked strong criticism and the site's terms and conditions changed (on 18 May, 2019).

The change clarified the scope of crimes for which the database could be used to include robbery and aggravated assault (and unidentified bodies), with authorisation. All contributors were opted out from law enforcement matching and were offered the option to actively opt in. This significantly limited the use of this database for criminal search purposes at the time of implementation; however, by October 2019 about 181,000 GEDmatch users had actively opted back in. GEDmatch was taken over by the forensic genomics company Verogen in December 2019.

Challenges to the use of genetic genealogy approaches

There are considerable technical and economic challenges to using the current approach as it has been practised in the USA.

There is a practical need for a sufficient quantity of high-quality DNA to be recovered from the crime scene; the relevant laboratory analysis uses considerably more DNA than would normally be available and requires it to be of better quality. Effectively, this limits its use to a sample of semen, saliva or blood from a crime scene that has also not become too degraded for the test to work.

Once a potential link to an individual has been found, confirmation of this will require comparison with a standard STR-based DNA profile, which entails obtaining a reference sample from that person. In the USA it is allowable for police to collect discarded items, such as a used paper cup that is likely to bear DNA from their suspect, for this purpose. Whether or not such methods would be acceptable in the UK, when the evidence against an individual is limited to a genealogical link, is unclear.

Two main companies in the USA are assisting police in the genealogical approach: Bode Technology and Parabon. The success of the method depends on a sufficiently large database of genetically- tested individuals, ideally with associated genealogical information. Employing the GEDmatch database before it was restricted, Parabon had assessed about 200 cases of which about 65% were suitable for ongoing genealogical research, implying that the search had identified useful links in these cases. The 35% of unsuitable cases reflects the fact that relatives may not be present in the database or may not be identifiable because they are not genetic relatives (sharing DNA), even though they may be genealogically related.

If the perpetrator of the crime is not Western European in origin, then the chance of success is likely to be limited because of the under-representation of individuals from other parts of the world in the GEDmatch database, and the consequent failure to identify relatives.

The above statistic of potential success is similar to a 2019 experiment conducted by Peter Aldhous, a science journalist at Buzzfeed News (Aldhous, 2019). Dr Aldhous was challenged with the identification of 10 suspects, from amongst 1,500 Buzzfeed employees. He spent around 60 hours undertaking research using GEDmatch, assisted with social information from the employee list (names and ages) and from social media platforms. He identified six individuals – four through genealogy and two from their likely geographic ancestry. He found the process personally uncomfortable because of the personal nature of much of the information he revealed. In a similar experiment conducted by Thomson et al. (2019) at a UK forensic science provider, Eurofins UK, four out of ten anonymised UK volunteers were identified using their matches on GEDmatch followed by classical genealogy research, suggesting that the applicability of the method to UK cases may not be far behind of that of the demonstrated US cases.

The genealogical research effort is obviously considerable and could be very expensive. Identifying a second cousin might only take about a few hours but on average cases take around 24 hours, and costs were around $5,000 (£4,000) per case when undertaken by Parabon.

Familial searching in the UK

Familial searching in the UK has been used since 2003 in serious crimes where there is a DNA STR profile attributable to an offender, but the offender's DNA profile is not in the National DNA Database (NDNAD). The technique ranks the genetic likelihood of a familial relationship between the crime scene STR profile and individuals on the NDNAD. Other case-relevant information (such as the locality of the crime) is also used to prioritise the investigation and sometimes Y chromosome tests may also be employed to compare the identified relative with the crime scene material for further confirmation, although since the implementation of the Protection of Freedoms Act (PoFA) 2012 this requires the named individuals to have new samples taken.

Because of the very small number of DNA regions surveyed by an STR-based DNA profile (16 in the UK, compared to hundreds of thousands of SNPs in the genealogical tests), this technique can at best only identify parents, children or siblings on the database. Since 2012, 120 cases have been authorised for familial searches, of which 9 have been resolved through this familial approach (and a further 14 by other means). Generally, however, it is reported that the success rate is around 20%. It has been suggested that genetic genealogy could assist with the remainder.

The apparent high clear-up rate of cold cases in the USA using these techniques seems impressive. However, it masks the inefficiencies of the US legal system, which has enormous backlogs of unanalysed rape cases, and significantly fails to collect DNA profiles from both suspects and convicted individuals to be placed on their CODIS STR-based reference database. Indeed, the brother of the alleged Golden State killer, James Joseph DeAngelo, is a convicted felon and if his DNA profile had been present in the database and familial searching had been used then the former would have been identified many years prior to 2017.

Findings

Potential for use in the UK

There are many questions that need answering and factors to consider before a genealogical approach could be used in the UK,

The UK already has one of the most efficient DNA databases in the world and conventional methods, with appropriately applied familial searches, will identify the bulk of perpetrators.
If a genealogy approach is used, a proportion of the potential relatives will not be UK based and this could add significantly to the genealogical effort. The US-centric nature of the GEDmatch database may limit the power of identification and the question of cost-benefit will need to be considered. The recent acquisition of GEDmatch by the company Verogen means that its content and its terms and conditions of use may change. Other companies are also working within this field of activity, and therefore the genetic genealogical approach needs to be kept under review.
Although there appears to be significant public support for its use in the US-based surveys undertaken to date, there is concern that such surveys could be biased (Guerrini et al., 2018).
Identification of cases where genetic genealogy may be appropriate must be carefully defined to enable an ethical and reasoned decision to be made (avoiding historic issues such as the identification and prosecution of women who abandoned their newborn babies decades ago, based on analysis of the deceased baby's DNA followed by a forensic genealogical approach).
At the time of writing, familial searching is done only in the most serious of unsolved crimes and permission to conduct these requires approval of the Forensic Information Database Services (FINDS) Strategy Board. Only 17 were authorised in 2018/19. A similar restriction could appropriately be applied before a genealogy search is considered.
It is worth noting that the initial use of the method in the identification of otherwise unidentifiable bodies (similar to the DNA Doe project in the USA, which so far aided the identification of 11 deceased individuals via GEDmatch), would allow the potential of this method to be tested in a UK setting, while avoiding some of the more contentious issues.
At the time of writing this report the whole process is unregulated, and ethical, legal, and safeguarding issues must be considered.

Ethical, legal and safeguarding considerations

The genealogical approach should only be used once traditional methods have been exhausted and must be authorised by the appropriate body so that its use is proportionate.
The issue of whether an individual's consent can be judged to be 'freely given' for collection, retention and use of genealogical data in the UK, and whether this would meet the requirements of consent under the Data Protection Act 2018 and the EU General Data Protection Regulation (GDPR), while maintaining necessary legal safeguards needs to be addressed. Consent may not be 'freely given' because the data obtained may reveal sensitive social and health information about many other individuals, for example, other family members, current and in the future.
The legality and necessity of police use of genetic genealogy (and associated interference with privacy) would need to be clearly established, with reference to Article 8 of the European Convention of Human Rights (ECHR). Of note, in the USA, where genetic genealogy has been most used, the ECHR does not apply. This is in contrast to the UK, which is subject to the ECHR under the Human Rights Act 1998 and the additional requirements and safeguards under Part 3 of the Data Protection Act 2018.
Genetic genealogy should only be used as a policing tool if it can be shown to be based on clear evidence, verified by an independent body, and that the established methods already in use for these law enforcement purposes are no longer adequate or effective. Otherwise, the use of any such novel processes would not meet the tests of necessity and proportionality. This would make the legality of using such novel processes highly suspect.
The legality of using informed consent as the sole appropriate legal basis to obtain the above highly sensitive data is doubtful, in line with the Data Protection Act 2018 (Part 3), the EU Law Enforcement Directive, and well-established Article 8 ECHR case law.
If it could be shown that the use of genetic genealogy was clearly needed in light of evidence that current processes were no longer adequate, a binding legal framework would have to be enacted that explicitly permits the collection and use of such genetic data, accompanied by relevant legal safeguards. In particular, legislation for transmission, length of retention, and destruction of sample, profile and collected genealogical data would be needed.
Avoidance of the unnecessary invasion of an individual's privacy and data security should be paramount, and so guidance should be provided to limit direct and indirect activity in investigating potential distant relatives.
Processes would be needed to maintain the security of data. In addition to genealogical information, genetic information would be revealed in the analysis, including sensitive medical and personal data, that may also be misinterpreted if revealed.
Genealogists must have the necessary skills and knowledge. Ideally there should be a form of accreditation and a professional body to ensure the quality of identifications, and acceptance of appropriate conduct requirements to ensure confidentiality and privacy.
The genetic analysis would need to be done in an accredited analytical environment that meets chain of custody, security, process, and confidentiality requirements.

References

Aldhous, P. (2019) 'We Tried to Find 10 BuzzFeed Employees Just Like Cops Did for the Golden State Killer', *BuzzFeed News*. Available at: https://www.buzzfeednews.com/article/peteraldhous/golden-state-killer-dna-experiment-genetic-genealogy [Accessed 21 June 2020].

Guerrini, C. J., Robinson, J. O., Petersen, D. and McGuire, A. L. (2018) 'Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique' *PLoS Biol*., 16 [online]. Available at: http://dx.doi.org/10.1371/journal.pbio.2006906 [Accessed 9 December 2019].

International Society of Genetic Genealogy (2019) *Autosomal DNA testing comparison chart* [online]. Available at: https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart [Accessed 9 December 2019].

Thomson, J., Clayton, T., Cleary, J., Gleeson, M., Kennett, D., Leonard, M. and Rutherford, D. (2019) 'The effectiveness of forensic genealogy techniques in the United Kingdom – an experimental assessment', *Forensic Science International, Genetics*. [online]. Available at: https://doi.org/10.1016/j.fsigss.2019.10.169 [Accessed 9 December 2019].

Data Protection Act (2018, Part 3) [online]. Available at: http://www.legislation.gov.uk/ukpga/2018/12/part/3/enacted [Accessed 23 February 2020].

Privy Council ruling (2015) Available at: https://www.jcpc.uk/cases/jcpc-2015-0079.html [Accessed 24 February, 2020].

Authors

This report was authored by Mark Jobling and Denise Syndercombe Court, with additional contributions from Nóra Ni Loideain, Charles Raab, and Jennifer Temkin, and has been ratified by all members of the Biometrics and Forensic Ethics Group (see Appendix 1 for a list of members).

Appendix 1:
Membership of the Biometrics and Forensics Ethics Group

Chair

Professor Mark Watson-Gandy, a practising barrister at Three Stone chambers and Visiting Professor at the University of Westminster.

Group members

Dr Adil Akram, Consultant Psychiatrist, South West London and St George's Mental Health NHS Trust

Professor Louise Amoore, Professor of Human Geography, Durham University

Professor Liz Campbell, Chair in Criminal Jurisprudence, Monash Law, Australia

Professor Simon Caney, Professor in Political Theory, University of Warwick

Dr Richard Guest, Reader in Biometrics Systems Engineering and Deputy Head of the School of Engineering and Digital Arts, University of Kent

Professor Nina Hallowell, Associate Professor, Nuffield Department of Population Health, University of Oxford

Dr Julian Huppert, Director and Fellow at the Intellectual Forum, Jesus College Cambridge

Professor Mark Jobling, Professor of Genetics, University of Leicester

Isabel Nisbet, has an academic background in moral philosophy, with additional knowledge of medical law

Dr Nóra Ni Loideain, Director of the Information Law and Policy Centre, Institute of Advanced Legal Studies, University of London

Professor Charles Raab, The University of Edinburgh and Turing Fellow, Alan Turing Institute

Professor Tom Sorell, Professor of Politics and Philosophy, University of Warwick

Professor Denise Syndercombe-Court, Professor of Forensic Genetics, King's College London

Professor Jennifer Temkin, Professor of Law, The City Law School (City University of London)

Dr Peter Waggett, Director of Research at IBM

OGL