Nuffield Foundation

Nuffield Council on Bioethics

April 2025

# Ethics, genomics and education

## Findings from a workshop exploring the ethical landscape of using polygenic indices in education

# Acknowledgements

Thank you to all of our workshop participants for their generous sharing of time and expertise. Particular thanks go to Dr Emma Meaburn for her role in chairing and co-designing the workshop, and reviewing drafts of this report.

## Key messages

1. The potential use of polygenic indices (PGIs) in education policy and practice raises a range of ethical questions, including around education **values, systems** and **futures**; **consent, agency** and **privacy**; **equity** and **fairness**; **genetic exceptionalism**; and **managing expectations**.
2. The use of PGIs in education raises new ethical issues distinct from uses in other contexts, including around **genetic determinism** and the efficacy of PGI-informed interventions.
3. The existing uses of data in education raise ethical challenges that need to be mitigated, such as those around **privacy** and **consent**.
4. PGIs are poor predictive tools of individual outcomes. This limits their utility and raises important ethical questions about if, when and how they could be applied in education.
5. Misunderstandings arising from the complexity of the science, its limitations and how these are communicated could contribute to an appetite to prematurely translate PGIs into education.

# Introduction

## The NCOB and the Nuffield Foundation have been working in partnership to explore emerging directions in genomic research and their relevance to education.

Our first output as part of this work was a scoping report, published in February 2025, providing an overview of scientific developments in genomic research relevant to education, with a particular focus on advances in the development of polygenic indices (PGIs). The report describes the current research landscape and identifies barriers to filling existing gaps in knowledge. It also highlights some scientific, practical and ethical challenges arising within this research.

However, we felt that further exploration was required of the ethical questions arising from educational genomic research – specifically, those arising in research about PGIs and their potential translation into policy and practice. There is comparatively little discussion in academic and grey literature identifying and exploring these ethical issues. There has also been an increase in access to genomic testing (via direct-to-consumer genetic testing companies (DTCs)); pressure on schools to provide appropriate special educational needs and disability (SEND) support; and political appetite for data-driven innovation. Therefore, using the scoping report as a foundation, we ran an interdisciplinary 'deep-dive' workshop to begin to unpick some of the ethical concerns in more detail with a range of experts.

This is a summary report from that workshop. It sets out key ethical themes and practical considerations that were identified and explored by participants, and some suggestions for further work that is needed to inform the potential translation of genomic data into education policy and practice. This report is therefore based on what participants discussed and our analysis of their views; it is not a definitive account of all the ethical issues around the research and its potential applications in education contexts. The suggestions we make towards the end of this report – of what researchers and decision-makers may wish to explore further – are based on our analysis of participant views.

We expect this publication will be of use and interest to parents and carers, educators and educationalists, geneticists, ethicists and social science researchers; policymakers; and school leaders. For readers less familiar with genomics, we have provided a glossary of key terms and suggest reading the executive summary of our previous **scoping report** for useful further background.

# Background and aims

The workshop aim was to **explore participant views on the ethical issues arising from using polygenic indices as predictive tools in education contexts.**

We provide a summary of polygenic indices and their relevance to education here, but for a more detailed breakdown and comprehensive explanation of their basis and how they work, we ask readers to refer to our previous **scoping report**.

## What are polygenic indices?

Polygenic indices ('PGIs'; sometimes also referred to as 'polygenic scores' or 'polygenic risk scores') are numerical values that sum up the effects of thousands of genetic variants across the genome, in order to estimate how likely a person is to develop or exhibit a particular phenotype. The term 'phenotype' in this report includes the range of traits, behaviours, pathologies and attributes that genetic variants can influence.

## How are they used?

PGIs are increasingly used in research to contribute to the understanding of what might underpin individual differences in a given population. More recently, their potential for translation into healthcare has begun to be assessed, most notably aimed at improving clinical practice and population health management (such as risk-stratified population screening). They are also increasingly marketed to the public for a range of phenotypes via direct-to-consumer genetic testing companies.

## What's the relevance to education?

PGIs are now available that account for a proportion of measured differences between adults in educationally-relevant measures and phenotypes, including years spent in education, mental health and neurodevelopmental conditions, and cognitive abilities. PGIs have been shown to perform as well as more traditional demographic measures of potential achievement – such asparental years spent in education – but with the potential benefit of being theoretically available from conception, specific to the individual, and not requiring additional family data. However, PGIs for other phenotypes related to education (such dyslexia and ADHD) currently fall short of traditional, non-genomic measures in identifying individual risk.

## Why does the ethical landscape need to be explored?

While PGIs are increasingly used as predictive tools in educational research contexts to ask genetically-informed questions, they lack accuracy in predicting individual outcomes and have been criticised for being reliant on a limited range of educationally relevant measurements. They are also complex, capturing not only an individual's biology but also, incidentally, elements of an individual's social and family environments. Their predictive power varies across populations and contexts, and it remains unclear when and how PGI-based screening or educational interventions based on them would be effective. Genetic factors are only part of the story, with an individual's environmental influences impacting their likelihood of exhibiting a particular phenotype. Therefore, any interventions in response to genetic risk must be considered alongside mitigation of the impact of those environmental influences.

Increased knowledge of, appetite for, and access to PGIs for educationally-relevant phenotypes – particularly via DTCs – risks their being incorporated into education practice and policy before the ethical implications, including scientific limitations, have been fully considered. This, in turn, could lead to pressure on schools, teachers and policymakers to respond to genomic information without thoroughly weighing up the evidence and associated ethical concerns. Challenges and pressures arising from existing uses of non-genomic data to shape education may further complicate matters, with any new data policy potentially further embedding these challenges. Translation of PGIs from research to practice contexts in healthcare has raised a number of ethical questions, including around agency, equity, privacy, consent and discrimination, and whether a focus on PGIs instead of other, more modifiable contributions to risk is appropriate (see the recent report from the Ada Lovelace Foundation and NCOB – **"Predicting: the future of health?"** for examples). Consideration of these questions, as well as identifying any new ethical questions arising in the education context, is needed to ensure that future policy and practice avoids unintended and potentially harmful consequences. These could includecreating or reinforcing social and educational inequities; adverse impacts on core rights, such as privacy, agency and consent; and discrimination. Ethical considerations relating to translation will need to be explored on an ongoing basis if and when PGIs grow in predictive capability.

# Approach to the workshop

We invited a range of participants from across genomic science and policy, education, ethics and the social sciences to take part in our one-day workshop. Participants were provided with background reading and heard presentations on the day to ensure a shared baseline level of understanding of key concepts.

19 participants took part in small, mixed-disciplinary group discussions to stimulate debate and ensure a spread of perspectives. Three broad topics were explored:

1. Opportunities and challenges arising from current, non-genomic data use in education (for example: assessment and pupil absence data);
2. Identifying ethical challenges arising from translation of PGIs into education, noting when they are either novel to the discipline and context or in alignment with challenges already faced from non-genomic data use and/or PGI translation in healthcare; and
3. Considering how to navigate increased access to,and appetite for, PGI use in education in light of a political climate promoting data-driven innovation, and greater direct-to-consumer access to testing.

# What themes arose?

In the course of discussions, participants identified and explored a range of ethical issues arising from the potential translation of genomic data into education policy and practice. Some of the issues identified were common to current uses of non-genomic data use in education, or to the current and potential uses of PGIs in healthcare, although others were considered specific to genomics and education – particularly some around equity and fairness.

There was a range of views about whether PGIs ought to be translated into education policy and practice at all, and – if so – what foundational principles, systems and evidence base would need to underpin any such translation. Some participants highlighted that the overall current predictive power of PGIs for educationally-relevant phenotypes is low, and that relying on a poor predictor to make educational decisions is, in itself, ethically problematic. Others noted that the predictive power of PGIs is likely to improve as research progresses and datasets become more diverse and comprehensive,albeit with an as-yet unknown timescale for improvement. Therefore, it was felt that the primary ethical challenges might arise from non-scientific barriers to implementation.

We have drawn out several key themes from the discussions, which are set out in more detail below. They focus on:

- education values, systems and futures;
- consent, agency and privacy;
- equity and fairness;
- added dimensions of genetic exceptionalism; and
- managing expectations.

# Education values, systems and futures

A number of issues relating to current and future education systems arose as part of discussions.

## The goal and purpose of education

Participants reflected that consensus between decision-makers and educators on the intended purpose of education in the UK does not exist but would be helpful in considering the potential utility of PGIs. For example, if the purpose is to improve individual attainment, then PGIs may not be useful as they are unreliable predictors at the individual level and are more useful for group-level insights. Participants discussed this in context of claims made by researchers that PGIs may have most utility in helping to guide SEND interventions. They also discussed whether using PGI data for such purposes would be ethical in light of its poor predictive power at the individual level, and the need for SEND interventions to be tailored to the individual.

Some questioned what a focus on predicting attainment says about UK society, and whether this contributed to greater social value being attributed to academic achievement than other, equally valuable skills. Some were of the view that many of the personalised interventions which currently exist within education are not especially effective in improving attainment of every person but are effective in improving individual wellbeing and happiness within education, and that this was arguably an equally worthwhile outcome.

It was also debated whether PGI data would actually help to personalise education interventions, or whether it would risk a further embedding of assumptions about learners, i.e. whether it would be assumed that a specific intervention would work well someone with a specific PGI.

## Data-driven education: too much or too little?

Questions about the rationale for current data collection within education arose in discussions. This was seen as a relevant point to consider when reflecting on the potential permissibility of using PGIs.

Participants broadly agreed that whilst excessive 'datafication' of education was generally undesirable and burdensome to educators, the large amount of data already collected at both the local and national levels was more often used to measure school performance rather than to identify or implement support for individual needs. Some participants described a circularity of data use, with perceptions that data was being used to excuse or explain difference in attainment or outcomes rather than address it, although this was also seen as a potential result of using genomic data (i.e. where pupils' 'low' PGIs could be used to excuse poor performance). Some participants, therefore, considered that collecting more data that could be used specifically to target interventions, such as mental health and school environment data, should be explored before considering PGIs as a metric for this purpose. On the other hand, others felt that a potential benefit of using genomic data might be to assist in assessing school performance more effectively by acting as a constant against which school environments can be measured (i.e. whether the school is managing to build upon the genetic 'baseline').

It was also queried whether the use of data within education for punitive purposes – for example, absence data being used to sanction parents – might have impacted trust in education systems and, therefore, lead to a lack of support for (or willingness to engage with) genomic data collection, even if evidence suggested that might be useful for targeting interventions.

# Consent, agency and privacy

## Participants raised a number of issues relating to consent, agency and privacy that may arise as a result of translating PGI data into an education context.

Drawing on existing educational data collection practices as a comparator, they questioned whether parents are aware of, or understand, how data is collected about their children at either a local or national level; whether they expect it to be retained in the ways it currently is; and whether they are fully aware of the actual and potential uses of data repositories. It was agreed that whilst parents might consent explicitly to specific data uses at the point of collection – for example, via a survey – they are potentially unaware of repurposed uses of that data. One group identified a recent example of data collection within education which had highlighted issues with consent: a survey of Scottish schoolchildren about sexual relationships, where parents and children were not given full information about questions to be able to provide informed consent to participate, nor were they made aware about the extent to which responses would be shared with third parties, or linked to other datasets[1]. Some considered that a lack of awareness about data use and sharing would be even more ethically problematic if replicated with PGI use, due to the perceived unique status of genomic data (see 'Genetic exceptionalism' section below for more detail).

The potential impact on children's and families' rights also arose – specifically, the right 'not to know' and the 'right to an open future'.

In terms of the right 'not to know', it was noted that PGI data – about a child's likelihood of developing a particular phenotype – extends beyond the child to other genetically-related family members. Collecting and using such data might, therefore, adversely impact those whose own genomic risk is implicated, but who would rather not know about it.

1 BBC Scotland News. School pupil 'sex survey' data offered to researchers. 4 February 2025 [accessed 20 March 2025 https://www.bbc.co.uk/news/articles/c5yvr65dpgzo].

The potential for adverse psychosocial impact of knowing one's PGI data was discussed, with reference to study findings linking knowledge of PGI data to negative self-expectations.[2]

The right to an open future – the established ethical principle of a right to have options kept open until a person is capable of making their own decisions[3] – was also raised with reference to potential impacts on a child's autonomy if decisions about their trajectory through education (and potentially beyond) were made based on their PGI and without their consent. Many participants considered this to be a novel issue in the education context, as the immutability of genomic data meant it was more likely to be perceived as deterministic and relevant throughout the life course, and, therefore, could follow the child beyond education. Arguably, this would be in contrast with much non-genomic data, which is potentially subject to change and therefore being less likely to remain relevant into adulthood. Further and related concerns about genetic determinism are set out in the section below.

2 Matthews LJ *et al* (2021) Pygmalion in the genes? On the potentially negative impacts of polygenic scores for educational attainment *Social Psychology of Education* 24. 10.1007/s11218-021-09632-z.

3 Davis, D. S. (1997) Genetic Dilemmas and the Child's Right to an Open Future *The Hastings Center Report* 27(2), 7–15. https://doi.org/10.2307/3527620.

# Equity and fairness

Questions and concerns about equity and fairness arose frequently in discussions, relating to both the integrity of data and its potential translation into education policy and practice.

## Research gaps, lack of diversity and learning lessons from the past

The primary equity-related issue posing a barrier to potential translation was the lack of diversity in genomic datasets from which PGIs are generated[4]. This was a key issue identified in our previous report. Most existing data within genomic datasets comes from populations of European genetic ancestral descent, which limits the accuracy and relevance for populations of other genetic ancestral descent. Many considered that there was potential for inappropriate extrapolation of research findings – possibly caused or exacerbated by poor genomic literacy – to wider populations than those captured in existing datasets (described in academic literature as 'the portability problem'), and for the lack of diversity in genomic research findings resulting in research only benefitting those of European genetic ancestral descent. Both these factors could further embed existing inequities in education and, more broadly, within society. Similar concerns were also raised that a perceived 'genomification' of education more generally might lead to problematic thinking in the sector, akin to that seen from the legacy of eugenics on healthcare.

Research gaps and potential issues with equity were also noted in relation to children with special educational needs and disabilities (SEND). Some participants expressed the view that increasing numbers of children with SEND are being educated outside of state education provision and are, therefore, at institutions that are exempt from statutory assessments and mandatory data reporting. This means that, potentially, any national level policy on use of PGI data may either not extend to independent schools (and therefore fail to benefit the children attending them) or that interventions

4 Despite people of European genetic ancestral descent making up only around 16% of the global population, a 2019 study found that over 94% of participants in genome-wide association studies – the large studies from which PGIs are determined – were of such descent. See Mills MC and Rahal CA (2019) A scientometric review of genome-wide association studies *Communications in Biology* 2: 9.

may be ineffective and inconsistent due to the absence of standardised non-genomic data to inform their introduction.

Participants also noted the pressure on SEND provision within the state education system and resource constraints as posing potential problems for equity, with some questioning whether accepting PGI data as evidence for SEND support would unfairly disadvantage children from families without the genetic literacy and/or financial means to undergo genetic testing.

## Labelling and determinism

As mentioned earlier, the immutability of genetic data was seen to contribute to it potentially being seen as having relevance beyond education and throughout the life course. This was also considered to contribute to genetic determinism, that is, the erroneous perception that genetic make-up is the sole, or most important, driver of individual outcomes. Some participants considered that this might lead to children being negatively labelled in ways that they cannot dispel, leading to limits being placed on expectations about their potential, and fewer educational opportunities offered. They noted that this already happens to an extent with non-genomic data, the example given being low parental achievement and free school meal status often resulting in low expectations about a child's academic ability.

There were mixed views on whether translating PGI data into education practice would result in increased labelling more generally. Some thought it may lead to inappropriate pathologisation of children based on normal human social and behavioural variance, and traits being perceived as problems to solve. Others suggested that responding to PGI-based phenotypic difference rather than diagnostic difference could have a destigmatising effect, normalise variance and increase social acceptance of all differences between children. This led to wider discussions about how the provision of support is prioritised in a low-resource system, and how this might be done as equitably as possible. There were similarly mixed views on whether children having both relevant clinical diagnoses and very low or very high PGIs should be prioritised for support, or whether this might lead to those children in equal or greater need for support but lacking a diagnosis for reasons beyond their control (such as long waiting lists, socioeconomic factors or cultural differences) being unfairly disadvantaged.

## Potential for novel inequities?

Two particular issues arose in discussions as representing potentially novel inequities specific to the translation of genomic data in education, which would not apply to such data being translated into healthcare provision.

The first of these was the disparity in models used between genomically-driven healthcare and education. While there are potential similarities between PGI use in

healthcare and education where both systems have limited resources, participants focused on perceived differences. In healthcare, they considered that using PGIs to determine risk and allocate interventions may lead to efficient use of resources targeted to groups in the population most likely to benefit. In education, they thought that while PGIs could help to allocate resources towards those most in need, this might unintentionally divert support from others, potentially resulting in a net loss of benefit.

The second potential inequity novel to education identified by participants centred on the efficacy of interventions. In a healthcare context, they considered that it is generally known which preventative measures will be of benefit in mitigating risk, but in education, it is far less clear which interventions would be useful in mitigating risk. For example, there is a good evidence base to show that the risk of developing cardiovascular disease can be mitigated by diet modifications and exercise, but there is currently a much weaker evidence base to support specific interventions to mitigate the risk of 'poor maths ability' or similar.

# Genetic exceptionalism

Genetic exceptionalism is the notion that genetic data are uniquely personal and therefore meriting special protection, and is well-established in ethico-legal discourse. This was the focus of some debate when considering whether, and how, PGIs might be translated into policy and practice.

Participants identified that there was potential for a dimension of genetic exceptionalism to arise in the education context that may not arise within healthcare. This related to the status of PGI data in comparison to other education data. Although other personal data is held in education contexts, PGIs would be uniquely sensitive in context as they would likely be the only type of education data that a person might not want to know about themselves. This is in contrast with the healthcare context, where there might feasibly be quite a lot of health risk data stored about a person that they would rather not know. PGI data would therefore have an added layer of contextual genetic exceptionalism when applied to education policy and practice.

This led to discussions about whether education is equipped to handle such sensitive data. There were mixed views on this. Some participants highlighted that quite a lot of sensitive personal data are already processed safely by individual schools and national repositories, but there was broad agreement that the level of genomic literacy within schools was not as high as it would need to be to ensure appropriate PGI data handling.

There were further discussions around the origins of genetic exceptionalism. Some participants thought that the sensitivity of genetic data partly stems from the fact that it can reflect both biological and environmental influences due to confounding. This blurring of causes can make it difficult to determine whether interventions should target genetic or environmental factors – or whether any intervention will lead to the outcome sought. PGIs can, therefore, be uniquely sensitive because of both the range of factors they capture, and the resulting absence of clear, evidence-based actions in response to them. Some groups also discussed a possible further layer of genetic exceptionalism arising from the potential for inconsistent use of PGI-based decision-making, which could result in people experiencing the same challenges receiving inequitable treatment. A comparison with healthcare was used, where poor eyesight could be equitably compensated with the provision of spectacles to all experiencing it, but in education, decision-making based on PGIs could potentially result in differential treatment between children who have similar challenges to each other.

# Managing expectations

Participants were mindful of increasing access to PGI data via DTC companies, particularly in the context of political interest in data-driven innovation. They therefore considered how pushes to translate PGI data into education might be ethically navigated.

Discussions centred around the need to ensure that the predictive limitations of genomic data were well-understood among educators, decision makers and the wider public, as well as realism about what the resource constraints within education can feasibly accommodate.

## Use of language

Participants noted a number of issues with the use of language in relation to genomic prediction that they thought could cause issues with managing expectations.

The differences in language use between genomics and education were highlighted as a particular issue. An example raised was of the word 'attainment', used in education and in common parlance as synonymous with assessment performance/achievement, but in behavioural genomics it has become synonymous with 'years spent in education'. Similarly, the use of the word 'explain' in genomics (in statements such as "PGIs explain X% of variance in educational attainment") was also seen as potentially misleading, as it is not used to imply causation or justification, but would likely be understood as such in everyday language.

A lack of clarity in language used in genomic research was seen to risk creating or adding to hype, and fortifying problematic perceptions of genetic determinism. Overplaying the significance of PGIs and their potential impact was also seen to risk interventions being implemented without strong evidence bases to underpin them, or without a realistic understanding of their limitations. Funding structures in genomic research were seen to exacerbate this potential for hype, with the 'race to translate' possibly fuelling overstatement and misleading language, which then risks becoming ubiquitous when captured and repeated by media and the commercial sector.

Language issues were noted with how the benefits of using PGIs were sometimes described, with the implication that they would result in interventions tailored specifically to an individual's needs. Whilst this would likely be appealing to families

and decision-makers alike and guide their expectations about what PGIs could do for a child, there were doubts on whether there was the evidence base to support this (see below). Participants also questioned whether individualised interventions would be practically possible in a resource-constrained education system. The expectations of both parents and decision-makers about the benefits of PGI use may therefore not be realisable, and result in situations that are difficult for educators to manage. An example was raised of schools and educators being put under pressure if children did not live up to their PGI-predicted potential, potentially resulting in a sense of what some termed 'genomically-driven entitlement'.

## Limitations of the science

The current limitations of PGIs as predictive tools also came up as an important issue to manage in response to increased appetite for genomic data use. Participants noted the 'portability problem' (see above); PGIs' weakness as individual predictors; and, in particular, the confounding factors that affect them. The presence of confounding factors means that PGIs may pick up on non-genomic data, such as social or environmental factors, and therefore be falsely perceived as more deterministic than they are. Confounding can arise from the phenomenon of gene-environment correlation, where an individual's genes influence the environments that they are exposed to or seek out. For example, a child with a genetic predisposition for struggling with reading may be less likely to be asked to read aloud in class, or they may demonstrate avoidant behaviours towards reading, with both resulting in fewer opportunities to read. Therefore, when that child experiences reading difficulties, it is difficult to identify disentangle the role of genes from the environment.

The presence of such confounding, and the resulting difficulties in disentangling cause and effect, was considered by most of our participants to be an ethical issue when translating PGI data into education contexts, with the potential for interventions being erroneously driven by correlative effects rather than causative ones (which can also occur with interventions based on non-genomic data). It was noted that, although PGIs' predictive power may improve with further research and expanded datasets, the timeframe for this was uncertain. Some participants questioned whether using PGIs in policy or practice contexts could ever be ethical where non-genomic data was an equally or more powerful predictor – as is currently the case for the majority of measures captured in education.

Even in a scenario where PGIs' predictive power improved, such that it rivalled or exceeded that of other predictive data, questions remained about the pathway from PGI to desired educational outcome. Participants noted that the mechanisms behind PGIs are not yet fully understood and, therefore, little is known about what an effective PGI-based intervention might look like.

## Responsibility and ownership

Participants also touched on accountability in a PGI-informed education system and what would need to be clearly defined in terms of responsibility and ownership to ensure ethical data use. This included identifying the appropriate data controller(s) and repository, and whether this would sit within healthcare or education provision, and clarity around who is responsible for responding to reported data. There was general support for the view that the latter would need to be centralised to avoid excessive burden on individual schools and educators.

# Conclusions and next steps

Based on the evidence we have gathered to date from our research, expert and stakeholder engagement and this workshop it is clear that extreme caution is required when considering whether real-world translation of PGIs into education policy and practice is justifiable, both in terms of their scientific limitations and the ethical preparedness of the education system to navigate translation appropriately.

The workshop highlighted a number of pre-existing issues with how data is already used in education, which, without action to mitigate, could be compounded by the addition of genomic data. It also suggested that although some of the challenges facing translation of genomic data into policy and practice are common to other contexts, others could be unique to, or exacerbated by, the education context and so require specific safeguards to ensure that core ethical principles are adhered to, and children's rights are upheld. Further research is needed before PGIs can responsibly be considered in education, not just to improve their predictive power but also to gain a better understanding of what effective interventions might look like.

Based on our work in this area to date, we make the following suggestions for next steps:

1. Policymakers, including the DfE, should engage in expert-informed dialogue around current and potential ethical issues arising from data-driven education, and discuss ethically-informed options for how any future translation of PGI data would need to be structured, including on how to improve genomic literacy among educators and the wider public;
2. Researchers should work with parents, carers, children and young people to better understand the appetite for PGI-informed education if/when a greater evidence base for effective implementation has been established;
3. Researchers and research funders should support further work exploring PGI mechanisms; their impact on differing environments; the pathways from PGI to chosen educational outcome; and the development of evidence bases to support educational interventions in response to PGI data; and
4. Policymakers, including the DfE, should engage with educators to better understand how current interventions are identified, implemented and evaluated, and gather views on classroom-level barriers to using genomic data to identify support needs if these do develop sufficiently to be used.

# Glossary of terms

**Complex** (multifactorial) **phenotypes**: These arise as a result of a multitude of both genetic and environmental influences, and their interplay across the life course. This means that while the DNA sequence we inherit might influence how we think, feel and act, DNA alone does not determine who we are and how we develop.

**Confounding:** A confounder is a third variable, often unmeasured, that influences both the outcome and the risk factor, generating a spurious association between the two. This can lead to incorrect conclusions about cause and effect. **Gene–environment correlation** can be a source of confounding in observational (non-experimental) research. For example, reading ability is genetically influenced, but offspring can both receive the genetic variants associated with reading ability and be influenced by a home reading environment. This is a form of passive gene–environment correlation, and it can confound observational associations between parental and offspring reading characteristics. Active and evocative gene–environment correlation can also create confounding if the environment that an individual experiences is influenced by their genotype.

**Environment:** In genomic research, 'environment' is taken to mean anything other than DNA sequence. Environmental factors examined in genetically informed educational research include family, school and neighbourhood-level factors. Environments that make two individuals in a family, such as biologically related siblings, similar are called 'shared', while those that make them dissimilar are called 'non-shared'.

**Gene:** Genes are arranged along chromosomes and consist of a sequence of DNA that is transcribed to produce a protein or other functional product. These products carry out biological functions inside or outside the cell or regulate the transcription of other genes. There are approximately 20,000 protein-encoding genes in the human genome.

**Genetic ancestry:** Ancestry is directly inferred from genomic variation data and not self-defined. It is used in genomics research to group genomes by how similar they are in patterns of genomic variation. Individuals are assigned to a genetic ancestry group as part of a GWAS analysis to avoid biased results owing to **population stratification**. The ancestry groupings are often given continental-level descriptors such as 'European', 'Asian' or 'African'. However, this approach overlooks the fact that ancestry is a continuum and not well captured by these population-level descriptors. Further, use of population-level descriptors that are linked to biology and can overlap with, or be conflated with, race and ethnicity can perpetuate harmful thinking and is both ethically and scientifically fraught. In response to these issues, the National Academies of Sciences, Engineering and Medicine (NASEM) recently released a report detailing a set of 13 recommendations on the use of population descriptors in genetics and genomics research.

**Genetics:** The study of genes and how they function, as well as how they are inherited.

**Genetic determinism:** The belief that a person's genes are the main or sole cause of their phenotypes – such as intelligence or personality traits – and are therefore fixed and unchangeable. This belief overlooks the importance of social and environmental factors, and does not account for the complex interplay between genes and environments. Deterministic interpretations of genetic research have been associated with harmful ideologies and movements, such as eugenics.

**Genetic exceptionalism:** The notion that genetic data is uniquely personal and sensitive, and therefore requires special consideration and protection beyond that of other personal data.

**Genome:** The complete set of DNA instructions present in a cell.

**Genome-wide association study (GWAS):** The main research method used to identify genetic variants associated with heritable phenotypes. It involves comparing DNA variation data from a very large number of individuals that differ for the phenotype of interest (e.g. individuals with varying academic performance) to identify alleles that correlate with phenotype variation. GWASs can use a case-control study design when the phenotype of interest is dichotomous (e.g. cases with ADHD and controls without ADHD), or a quantitative approach when the phenotype is quantitative (e.g. intelligence). GWASs do not necessarily identify the causal allele but rather a region of the genome that is correlated with the phenotype. These regions are followed up through further experiments to understand how the variation impacts biology.

**Genomics:** The study of the genomes of individuals and organisms that examines both the coding and non-coding regions. This term is also used when talking about related laboratory and bioinformatic techniques. The study of genomics in humans focuses on areas of the genome associated with health and disease.

**Phenotype:** Any measurable characteristic of an individual, for example a physical phenotype such as weight or height, cognitive phenotype such as intelligence, or behavioural phenotype such as aggression. Phenotypes may be classified as a diagnosable disorder (an individual either has conduct disorder or does not), or measured on a scale (individuals show different levels of aggressive behaviours). The term might be used interchangeably with 'trait'.

**Polygenic index (PGI):** A cumulative measure of an individual's genetic propensity for a specific phenotype based on the weighted sum of many thousands of DNA variants distributed throughout the genome. PGIs for diagnosable medical conditions, such as cancer, tend to be referred to as polygenic risk scores (PRS) or polygenic scores (PGS).

**Polygenicity** (adjective: **polygenic**): The contribution of many DNA variants (thousands or tens of thousands) to the variation in a phenotype.

Nuffield Foundation
100 St John Street
London
EC1M 4EH

www.nuffieldfoundation.org
info@nuffieldfoundation.org
Nuffield Foundation

Nuffield Council on Bioethics
100 St John Street
London
EC1M 4EH

www.nuffieldbioethics.org
bioethics@nuffieldbioethics.org
Nuffield Council on Bioethics