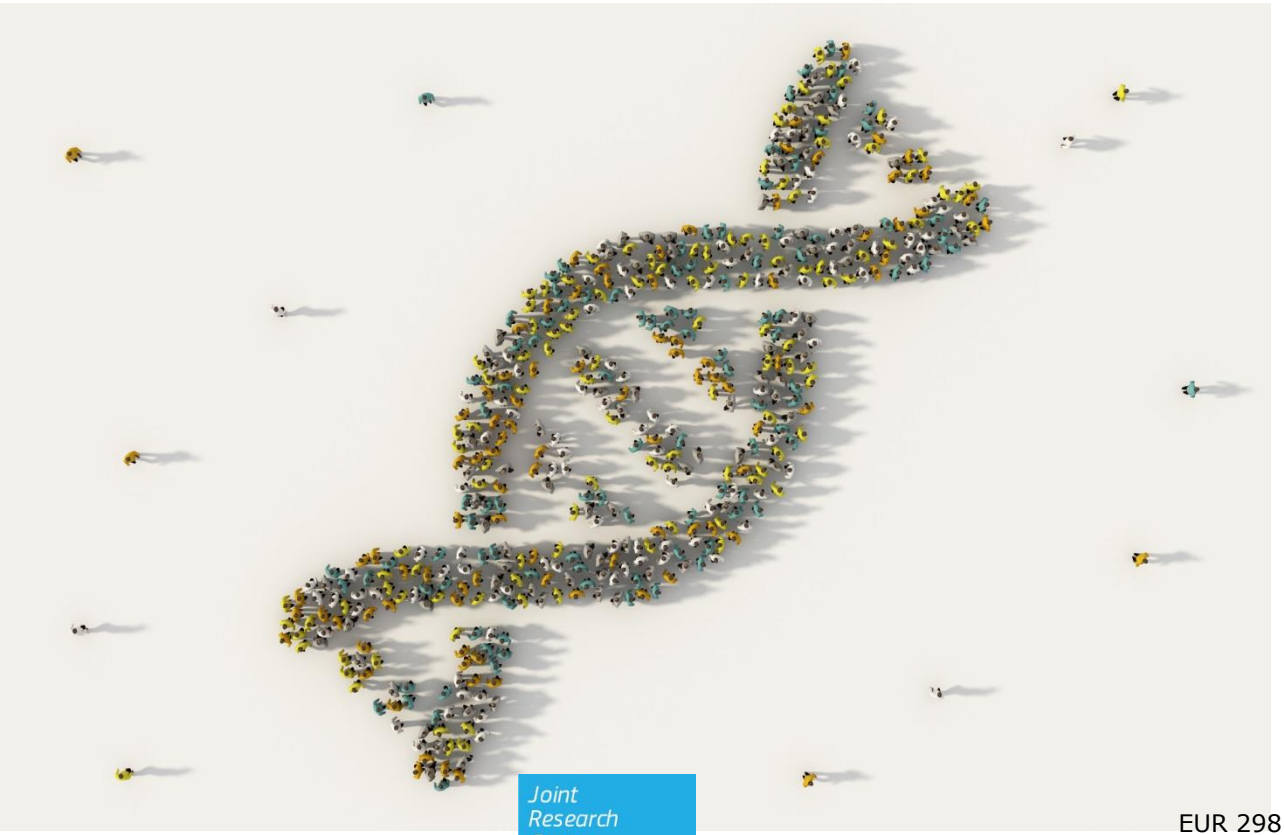
European Commission

JRC TECHNICAL REPORTS

Genome-wide association studies, polygenic scores and social science genetics: overview and policy implications

JRC F7 - Knowledge Health and Consumer Safety

2019



Joint Research Centre

EUR 29815 EN

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information
Name: Guy Van den Eede
Address: Retieseweg 111 B-2440 Geel/Belgium
Email: guy.van-den-eede@ec.europa.eu
Tel.: +32 14 57 14 81

JRC Science Hub
https://ec.europa.eu/jrc

JRC117414

EUR 29815 EN

PDF ISBN 978-92-76-09000-7 ISSN 1831-9424 doi:10.2760/948992

Luxembourg: Publications Office of the European Union, 2019

How to cite this report: JRC F7 - Knowledge Health and Consumer Safety, Genome-wide association studies, polygenic scores and social science genetics: overview and policy implications, JRC Technical Report, Luxembourg: European Commission, EUR 29815 EN, ISBN 978-92-76-09000-7, doi:10.2760/948992, PUBSY No. JRC 117414

Title: Genome-wide association studies, polygenic scores and social science genetics: overview and policy implications.

Contents

Foreword

It is since long very well established that changes in the genetic code may have an effect on the organism concerned. Scientists describe this as a genotype – phenotype relationship in which the genotype refers to the part in the DNA that is responsible for a trait and the term phenotype has been proposed as the composite of the organism's observable characteristics or traits. In some cases the effect may be considered as positive. An example is a small change in the DNA of certain humans that makes them immune to infection by lethal viruses such as the HIV virus. In other cases, the effect may be adverse and for example the majority of rare diseases is caused by such changes in the DNA. Sometimes there is a one-to-one relationship between a change in the DNA and the phenotype, but sometimes a phenotypic change can only be observed when different changes in the DNA occur simultaneously.

When the Human Genome Project was declared complete on April 14, 2003, there were high hopes that the elucidation of the (nearly) complete human genome would allow to understand diseases and to design appropriate treatment. In order to do so, a whole new discipline emerged that studies the whole genome with all possible genetic variants in different individuals to see if any variant is associated with a trait. The hope is that this study, called Genome Wide Association Study (or GWAS), would allow elucidating the genetic background of more complex traits such as depression, intelligence, behavior etc.

It is thus understandable that GWAS has a relevant positive potential, for example on healthcare, but that it should be used and interpreted cautiously and ethically correctly. Therefore, it is justified that the policies on the one hand follow the results obtained from GWAS as they may have an important societal relevance, including a socio-economic importance, but also that they are well aware of the ethical implications.

As the JRC is holding the responsibility for the knowledge management of health-related scientific information for policy, we present and discuss a comprehensive overview with the possible policy implications of recent developments in genetics, with a specific focus on Genome Wide Association Studies (GWAS), polygenic scores (PGS) and social science genetics, an interdisciplinary research field that studies if and how human behaviour and socio-economic outcomes are influenced by genetic factors.

Authors

Contributors (in alphabetical order)

Alexandre Angers, European Commission Directorate General Joint Research Centre, Directorate F – Health, Consumers and Reference Materials, Knowledge for Health and Consumer Safety Unit

Dafni-Maria Kagkli, European Commission Directorate General Joint Research Centre, Directorate F – Health, Consumers and Reference Materials, Knowledge for Health and Consumer Safety Unit

Prof. Dr. Philipp Koellinger, Vrije Universiteit Amsterdam, School of Business and Economics, Department of Economics

Mauro Petrillo, European Commission Directorate General Joint Research Centre, Directorate F – Health, Consumers and Reference Materials, Knowledge for Health and Consumer Safety Unit

Maddalena Querci, European Commission Directorate General Joint Research Centre, Directorate F – Health, Consumers and Reference Materials, Knowledge for Health and Consumer Safety Unit

Barbara Raffael, European Commission Directorate General Joint Research Centre, Directorate F – Health, Consumers and Reference Materials, Knowledge for Health and Consumer Safety Unit

Guy Van den Eede, European Commission Directorate General Joint Research Centre, Directorate F – Health, Consumers and Reference Materials, Knowledge for Health and Consumer Safety Unit

Acknowledgments

A special acknowledgment to Alexandre Binder and to Andres Metspalu for the valuable revision and competent comments.

Abstract

Recent developments in genetics are opening unprecedented possibilities to understand the physical and mental health of individuals. Moreover, with the advent of genome editing techniques it has become possible to introduce subtle changes in the DNA and thus to intervene in the combined sets of traits of an individual, including the curing of certain diseases.

In this report we focus on the first part, namely on the understanding and interpretation of the genetic information and its connection with the functionality in the host organism, and we will zoom in on the great wealth of information the reading, analysing and comprehension of the complete DNA sequence across organisms, including humans may provide. However, since this technology is not based on agreed standard methodologies including validated algorithms, a warning is raised with respect to the possibility of misinterpretation and this in turn may have scientific, ethical and thus policy implications.

This report analyses how these developments are impacting not only on health through personalised medicine, but also on social aspects of individuals and of human well-being. In particular, among these new applications, it highlights the possible policy implications of genome wide association studies (GWAS), polygenic scores (PGS) and social science genetics, an interdisciplinary research field that studies if and how human behaviour and socio-economic outcomes are influenced by genetic factors.

Due to the fact that these developments have the potential to affect many areas of public interest including public health, privacy rights, data security, threats of discrimination, new technologies for forensics, the emergence of new industries and the functioning of markets, schooling, and even direct changes to the human gene pool that can be passed on to future generations, they need to be observed, evaluated and scrutinised. They also raise ethical questions that touch the core of what type of society we want to live in.

The purpose of the current report is not to deliver final, definite answers as to which consequences genetics and genomics might have on the society, but to provide unbiased knowledge on the possible policy implications of recent developments in these fields, with a specific focus on GWAS, PGS and social science genetics. Rather, the purpose is:

- to provide an accessible entry point for policy makers and the wider public to understand the goals and tools of these developments,
- to track the progress of the field until now,
- to identify areas that are potentially relevant from a public policy perspective,
- to open a dialogue between scientists, policy makers, and the general public about how to move forward.

It is also not intended with this report to extensively discuss in detail the possible *applications* of the technology.

In the medical field, there is still some resistance to act on the basis of the results of genome analysis through the polygenic risk score (PRS), even in cases of "high risk" patients, and there is an ongoing debate with respect to the level of evidence required prior to a routine introduction of the approach for personalised care. Here, we will exclude these discussions from the scope of the report.

This report links, more generally, to the ongoing activities within the European Commission Joint Research Centre (JRC) on eHealth and Big Data, both related to the implementation of the EC Digital Single Market strategy. It is complementary to the 2018 JRC Science for Policy Report *Overview of EU National Legislation on Genomics*[1], a

[1] JRC F7 - Knowledge Health and Consumer Safety, *Overview of EU National Legislation on Genomics*, JRC Science for Policy Report, Luxembourg: European Commission, EUR 29404 EN, ISBN 978-92-79-96740-5, doi:10.2760/04463, PUBSY No. JRC113479

mapping of existing national legislations linked to genomics aimed to be used as a baseline for the analyses of possible consequences for EU policies already in place, and to forecast policy gaps and eventual interventions.

1 Introduction

1.1 What the availability of "large number of large data of large samples" (LaN.LaD.LaS) triggers

The successful sequencing of the first human genome in 2001 (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001) started a new era for science, medicine, and society (Lander, 2011). Since then, rapid scientific and technological progress has enabled the fast and cheap collection of genetic data from millions of humans using non-invasive sampling techniques such as spit-kits (Looi et al., 2012). In 2001, an astronomical 100 million USD had to be spent to sequence just one human genome. Now, in early 2019, high-quality sequencing of the entire ≈3 billion base pairs of DNA in the human genome costs less than 1,000 USD and "low-pass" sequencing (i.e. reading each base of the genome less than 10 times) can be performed for less than 100 USD. Furthermore, sequencing the entire genome is not even necessary for many research purposes. Instead, it is often sufficient to analyse only those locations of the genome that are known to vary among humans. The easiest way to do that is to focus on just roughly 1 million well-selected so-called single nucleotide polymorphisms (SNPs) that capture the vast majority of common genetic variation among people (Auton et al., 2015; McCarthy et al., 2016). This type of SNP-data can be collected very quickly and for as little as 40 USD or even less[2].

These developments have led to a rapid increase (large number) in the availability of genetic data (large data) and corresponding increases in the sample sizes (large samples) of scientific studies that use or analyse human DNA; this fact has led to an enormous acceleration of scientific progress across various fields including medicine, anthropology, and even the social sciences (Conley and Fletcher, 2017; Mills and Rahal, 2019; Reich, 2018; Visscher et al., 2017).

Many datasets that contain genetic information have been collected to study specific health outcomes (e.g. case-control cohorts). Other researchers took a more holistic approach and decided to study a wide range of outcomes that were measured in clinical assessment centres, via self-reports, links with digital health records or even registry-based information on a wide range of outcomes, such as conscription records or tax declarations (e.g. in Sweden, Norway, Finland, Denmark, and increasingly also in the US and the Netherlands). Access to many genetic datasets is restricted and typically limited to a small number of scientists and employees at specific institutions. However, a growing number of datasets have protocols in place that enable any bona-fide researcher to obtain access to individual-level data under certain conditions (e.g. dbGaP, Metadac, UK Biobank). This enables a growing number of scientists to work with genetic data, without the requirement to engage in time-consuming fund-raising and data collection activities themselves. Open access to genetic data also fostered statistically well-powered study designs that have dramatically improved research quality and the replicability of genetic association studies (Chabris et al., 2015; Ioannidis, 2005; Visscher et al. 2017).

Thanks to availability of LaN.LaD.LaS, an important recent development is the emergence of consumer genetic companies (e.g. 23andMe, Myheritage.com, FamilyTreeDNA, Ancestry.com) that offer genotyping services and information that are based on the DNA of their customers. These companies operate internationally and have substantially made it possible for people to gain insights into their own DNA. Some of the companies are actively participating in research activities (e.g. 23andMe Research) and have made important contributions to science thanks to their unique ability to collect and to analyse data from millions of people; the customers of these companies do not only voluntarily provide their DNA, they also pay the company for the genotyping of their spit samples, and they often also provide additional information about themselves (e.g. via web-surveys that can be cheaply and quickly administered to large numbers of customers). Often, these companies also enable their customers to download their own

[2] The exact prices depend on sample size and genotyping technology. See, for example, https://gencove.com/.

genetic data and to share it with other companies (e.g. DNA.land), medical professionals, family and friends and to compare their genomes with that of other people (e.g. to find relatives in the database of customers of the company that collected the genetic data). The business models of these companies rest on generating value from the data they collect. Furthermore, many companies started offering consulting services based on genetic data for a variety of purposes including health-related advice, but also tests for non-health related traits such as being a "good salesman"[3] often based on weak scientific evidence.

These developments launched us into a new era in which genetic data becomes ubiquitous, and the barriers to collect and to access genetic data are constantly decreasing – for scientists, health practitioners, but also for private citizens, companies, and possibly even governmental institutions. In this new era, genetic data is not only used to advance medical research, but also to study and potentially influence all aspects of individual differences, including behavioural traits. The possibilities to influence individual characteristics are constantly growing and involve strategies such as targeted environmental interventions (Plomin, 2018; Plomin and Stumm, 2018), *in-vitro* fertilization that selects specific fertilised eggs based on genetic insights (Mastenbroek et al., 2007), or even the potential to manipulate human DNA directly with recently developed technologies such as CRISPR-Cas9 (Grunwald et al., 2019). It is in this area of "genetic data ubiquity" that social science genetics is rooted.

1.2 Social Science Genetics

Social science genetics aims to integrate genetic insights and data into social sciences, in an attempt to deliver richer, more precise answers to old questions in psychology, sociology, economics, and related fields. Furthermore, genetically informed study designs in the social sciences also have the potential to investigate new questions and to obtain new answers that would not have been feasible before.

The roots of social science can be traced back long before the possibility to measure DNA sequence variation directly: starting in the second half of the 20th century, social scientists began to use family-based study designs that exploited known genetic similarities between people (e.g. twins, studies of adoptees, or comparisons of children of twins) to estimate heritability, to separate genetic and environmental mechanisms that transfer inequalities across generations, to study the co-occurrence patterns of genes and environments, and to control for potential genetic confounds in studies that aimed at identifying the causal effects of specific behaviours or environmental conditions.

1.2.1 Heritability and narrow-sense heritability

The starting point of social science genetics is the realization that all human behavioral traits are partly heritable. Thus, the reasons why people differ from each other in how they behave can partly be traced back to genetic factors. This stylised fact is also known as the "first law of behavioural genetics" (Turkheimer, 2000) and is supported by thousands of peer-reviewed scientific articles that employed many different datasets and methods (Polderman et al., 2015).

Heritability is a somewhat opaque concept that is unfortunately often misunderstood or abused, for example to justify discrimination and atrocities ranging from forced sterilization to genocide (Bliss, 2018; Zimmer, 2018)). Thus, any discussion about the role of genetics on individual differences between people has to start with an understanding of what heritability means and, importantly, what it does not mean.

Technically, heritability is defined as the proportion of observed differences in a trait among individuals of a population that is due to genetic differences. Thus, heritability is a population parameter that informs on the origins of observed differences among people

[3] http://www.braincompass.com/en/

of a particular population, but is does not necessarily inform on why an outcome is observed in a specific individual, or why groups differ from each other.

Furthermore, it is crucial to realise that heritability puts no upper limit on the potential relevance of the environment (Goldberger, 1979). In fact, a trait can be perfectly heritable in a particular population because of environmental reasons only.

A simple, hypothetical thought experiment illustrates this point: Imagine a totalitarian society that forces all boys to join the army instead of going to school, and all girls to go to school instead of joining the army. The heritability of being in the army or having a school degree in this hypothetical society would be 100%. In fact, the difference between being in school or in the army could be traced back to whether someone carries one X chromosome or two. And yet, there is no biological reason that would restrict boys from being able to go to school or girls from joining the army. Changing the environment would be expected to change how boys and girls behave and would affect the heritability of the two outcomes. A different environment may even break the link between the number of X chromosomes, schooling, and occupations entirely. **Thus, the heritability of a trait is not a natural constant**. Rather, the heritability of a trait may vary across space and time, depending on environmental conditions.

The heritability of a trait also does not imply a strict form of bio-determinism or the absence of choice and human agency. Rather, the extent to which even highly heritable outcomes can be altered or influenced is often a function of scientific and technological progress. For example, body height is highly heritable. Yet, wearing high heels is a highly effective and frequently practiced way to temporarily extend body height. As another example, consider a rare, monogenetic disorder called Phenylketonuria (PKU), which leads to severe intellectual disability, seizures, and mental disorders. PKU is caused by a mutation in the PAH gene that results in decreased metabolism of the amino acid phenylalanine. Before the causes of the disease were understood, it was impossible to treat the condition or even to diagnose it correctly. However, once the underlying mechanism of PKU had been unravelled, it became clear that a dietary change to foods that are low in phenylalanine is a highly effective treatment that can prevent the outbreak of symptoms or even reverse the progression of the disease. Thus, even if the cause of PKU is 100% genetic, this does not imply that behavioural or environmental interventions cannot be effective (Goldberger, 1979; Zimmer, 2018).

It is also important to realise that the heritability of many behavioural traits works via environmental pathways that can reinforce and amplify tiny genetic differences between people, for example via self-selection into specific environments, and reinforcements from the environment (e.g. being tall leads to advantages in playing basketball, which can be reinforced by success, praise, additional motivation, training, access to superior training facilities and coaches, as well as athletic scholarships, all of which can dramatically amplify small genetic differences between individuals at the outset) (Jencks, 1980; Zimmer, 2018). Although the original source of the observed differences among people can be traced back to genetic differences, the differences in outcomes may strongly depend on environmental conditions and feedback loops that can be changed. Yet, heritability estimates are silent about the mechanism and attribute such effects to the genetic differences between people.

Thus, heritability tells us nothing about the nature of differences between groups or over time and even traits with very high heritability (e.g. height, performance on cognitive tests) can vary substantially between groups or over time, due to persistent environmental differences (e.g. access to sufficient nutrition, health care, and high-quality education) (Dickens and Flynn, 2001).

With these caveats in mind, we can now proceed to the empirical evidence on heritability, the so-called narrow-sense heritability of a trait (i.e. the share of phenotypic variation in the trait that can be attributed to linear genetic effects in a given population). Twin studies are very popular to study narrow-sense heritability, as they use the experiment of monozygotic and dizygotic twins. Monozygotic (MZ) twins are (almost) genetically

identical, while dizygotic (DZ) twins share on average 50% of their genetic make-up (i.e. they are genetically not more similar to each other than all full siblings that had the same parents). Under some assumptions (e.g. no assortative mating among parents, the only difference between MZ and DZ twin pairs is their genetic similarity, all genetic effects are additive and linear), it is possible to interpret stronger phenotypic similarity among MZ twin pairs compared to DZ twin pairs as evidence for a genetic influence on the trait. Annex 1 provides an insight on it.

The availability of molecular genetic data in large samples since around 2008 enabled scientists to investigate the narrow-sense heritability of traits from genetic data directly rather than through twins. Many different methods have since been developed for this purpose (Bulik-Sullivan et al., 2015; Diprete et al., 2018; Evans et al., 2018; Yang et al., 2010), relying on some identifying assumptions. These different approaches to estimate the heritability of traits has now resulted in converging evidence using many different datasets and assumptions, all suggesting that genetic influences on individual differences across virtually all domains are pervasive and robust. However, narrow-sense heritability estimates from molecular genetic data tend to be lower than narrow-sense heritability estimates from twins studies. The reasons for this difference are still a topic of scientific discourse, but the existing evidence suggests that a large part of the gap is due to unobserved genetic variants that are not strongly correlated to the observed genetic variants on which the estimates are currently based (Wainschtein et al., 2019; Yang et al., 2015).

Thus, the existing empirical evidence strongly suggests that differences between people in their behaviour and their achievements in life are all affected somehow by differences in their genomes, including traits such as personality (Borkenau et al., 2001; Floderus-Myrhed et al., 1980), income (Taubman 1976b, a), political orientation (Benjamin et al., 2012; Kandler et al., 2012; Oniszczenko and Jakubowska, 2005), number of children (Rodgers et al., 2001; Tropf et al., 2015), educational attainment (Baker et al., 1996; Heath et al., 1985; Lee et al., 2018; Okbay et al., 2016; Rietveld et al., 2013), risk tolerance (Cesarini et al., 2009; Karlsson Linnér et al., 2019), well-being (Bartels and Boomsma, 2009; Harris et al., 2007; Nes et al., 2006; Okbay et al., 2016), professional choice (van der Loos et al., 2013; Nicolaou et al., 2008; Nicolaou and Shane, 2010), financial decision making (Cesarini et al., 2010), physical and psychiatric health (Polderman et al., 2015), and longevity (Herskind et al., 1996; Ljungquist et al., 1998; Pilling et al., 2017).

1.2.2. Estimates of heritability

The possibility to measure specific differences in DNA among people opened up new possibilities to pursue those goals and to identify which specific genetic variants are linked with behavioural and socio-economic outcomes. In practice, however, whether and when genetic association studies can be successful depends to a large extent on how genetically complex the trait is: if the heritability of the trait is due to just one or a few genes with large effects, it is possible to isolate the relevant genes in relatively small samples. If, however, a trait is genetically complex in the sense that hundreds or even thousands of different genetic variants all contribute a little bit to the observed heritability, isolating the relevant genetic variants is much more challenging: the effects of each genetic variation will tend to be very small (i.e. close to zero), and differentiating between very small and null effects is statistically much more difficult, requiring a combination of very large sample sizes that allow to estimate effect sizes with very small estimation errors as well as more sophisticated strategies to avoid finding false positive and spurious associations (Benjamin et al., 2012; Chabris et al., 2013; McCarthy et al., 2008; Ioannidis et al., 2001).

The first generation of studies that attempted identifying genetic associations with human behaviour and socio-economic outcomes worked under the implicit assumption that a few genes with large effect sizes exist that are driving heritability. They also limited virtually all attempts to identify genetic associations to a few dozen of so-called candidate genes that were believed to have well-understood biological functions. A

typical candidate gene study would look at only one genetic variant at a time, without explicitly considering the role of all other genetic variants. However, it soon became clear that the implicit assumptions of these candidate gene studies are highly problematic: humans have more than 20,000 genes (International Human Genome Sequencing Consortium 2001, Venter et al., 2001), and the majority of them is expressed in the brain (Ardlie et al., 2015; Ramsköld et al., 2009; Stranger et al., 2017). The function of most of these genes is unknown or at least not well-understood, which makes it impossible to rule them out on theoretical grounds, thereby limiting the possibility to restrict the number of plausible hypotheses based on theory. It soon became clear that virtually all human behavioural traits are genetically complex, meaning that their heritability is due to hundreds or even thousands of genes with miniscule effect sizes that were orders of magnitude smaller than the first generation of candidate gene studies implicitly assumed (Chabris et al., 2015; Visscher et al., 2012). Furthermore, publication bias, file-drawer effects, and inadequate statistical methods (e.g. not accounting for multiple hypothesis testing, underpowered study designs and insufficient controls for environmental confounds) plagued the first generation of candidate gene studies in the social science. As a result, the vast majority of these studies failed to replicate, and the few studies that did replicate typically had vastly inflated estimates of the true effect sizes of the identified genetic variants (Beauchamp et al., 2011; Benjamin et al., 2012; Chabris et al., 2012; Chabris et al.,2013; Freese, 2018; Hewitt, 2012; Ioannidis, 2005).

The solution to the replication crisis of the first generation of candidate gene studies were so-called genome-wide association studies (GWAS) and are carried out in very large samples, typically involving a two-stage study design that consists of a discovery and a replication stage carried out.

2 Genome-wide association studies and polygenic scores.

The general idea of GWAS is to scan all measured single nucleotide polymorphisms (SNPs) in a sample for associations with an outcome, using stringent controls for possible environmental confounds and multiple testing (McCarthy et al., 2008). In addition to directly measured SNPs, GWAS typically also scan "imputed" SNPs. The imputation of unmeasured SNPs allows researchers to pool evidence from various genotyping platforms that have non-perfectly overlapping sets of SNPs that were measured. Furthermore, imputation allows researchers to increase statistical power and to gain a more high-resolution insight into the associated genetic variants than directly genotyped SNPs alone (Marchini and Howie, 2010). Imputation "borrows" information about the correlation patterns between all DNA variations that derive from large reference panels of individuals whose DNA was entirely sequenced at high accuracy (Das et al., 2016; McCarthy et al., 2016; The 1000 Genomes Project Consortium et al., 2015). Imputation accuracy increases with sample size of the reference panel and the most recent reference panels allow imputing even rare statistical variants with relatively high accuracy. In practice, most modern GWAS include much more imputed than directly genotyped SNPs (Auton et al., 2015; McCarthy et al., 2016; Walter et al., 2015).

The rapid increase in data availability and computational power led to an equally rapid increase in the number of published GWAS and genetic associations in the past decade (Mills and Rahal, 2019). The increase in data availability also enabled even larger sample sizes for GWAS studies, leading to statistically well-powered analyses that are able to detect increasing numbers of associated genetic variants. Larger samples also led to more precise estimates of the true effect sizes of individual genetic variants. For the vast majority of genetically complex traits (including all behavioural and socio-economic outcomes that have been studied so far), any particular SNP accounts for less than 0.03% variation of the trait in the population (Chabris, 2015; Karlsson Linnér et al., 2019; Okbay et al., 2016; Okbay et al., 2016).

2.1 Public access to GWAS results

GWAS results are typically published in peer-reviewed scientific journals, some of which require to store and share GWAS summary statistics in a way that make these data accessible to other researcher. In fact, it has become best-practice to share GWAS summary statistics with the wider research community when the article that reports the results of the study has been published in a peer-reviewed journal. Some scientists share their summary statistics even earlier. However, there are scientists and research groups who still do not share their results with the wider academic community.

GWAS summary statistics that are useful for follow-up research should include a unique identifier for each genetic variant, the *p*-value, the estimated coefficient and its standard error (or, alternatively, the test statistic), the sample size for each genetic variant, as well as information about which "allele" (one of the variant forms of a given gene) was used as a reference and the minor allele frequency (MAF) of that allele. To protect subject confidentiality, many GWAS summary statistics do not report sample MAFs. Instead, they report MAFs from a reference population panel and they restrict the number of reported digits of estimated coefficients to a necessary minimum to protect the identity of the individuals who contributed to the study (Homer et al., 2008; Visscher and Hill, 2009).

Scientists typically share such GWAS summary statistics on websites (e.g. see the consortium websites of the SSGAC, PGC, or GIANT) or repositories such as dbGaP. The US National Human Genome Research Institute created a catalogue of all published GWAS association results (maintained by the European Bioinformatics Institute (EMBL-EBI): https://www.ebi.ac.uk/gwas/). These openly shared resources are of very high value for the research community and have led to the development of interactive atlases comparing the genetic architecture of thousands of traits (e.g. http://atlas.ctglab.nl/ and https://biobankengine.stanford.edu/).

2.2 Assumptions and challenges

GWAS use standard linear or logistic regression to estimate associations between SNPs and some trait of interest. As a result, all standard assumptions of these regression methods are also valid for GWAS. It is useful to think about the assumptions in GWAS from this general perspective (see Annex 1, Statistical Regression section).

SNPs are not the only type of genetic variation that exist and that differentiate people from each other. In addition to SNPs, the so-called structural variations exist (e.g. insertions, deletions, inversions, copy number variants, translocations) and may influence outcomes of interest. However, in practice, due to the currently prevailing genotyping technologies, the vast majority of GWAS focus on SNPs or include only those structural variants that are directly measured by genotyping arrays. Structural variants can be correlated with SNPs and thus, even if a GWAS focused on SNPs entirely, it may still pick up some signal from unobserved but correlated, structural variants (Auton et al., 2015; Linnér et al., 2019; Sudmant et al., 2015). The unobserved, non-perfectly correlated genetic variants in GWAS contribute to the so-called "missing heritability" problem that we will return to later.

Insufficient controls for population structure can lead to completely misleading GWAS results. For example, a GWAS on educational attainment that does not properly control for genetic ancestry would find an association with the lactose (LCT) gene that codes for the enzyme lactase (Cornelius A. Rietveld et al. 2014). Lactose intolerance is unrelated to cognitive ability or personality, but is much more frequent in south-eastern parts of Europe than in north-western parts. Because the prevalence of genes varies across geography (Novembre et al., 2008), environmentally-driven differences between geographical regions can thus induce a spurious association with the LCT gene.

Unfortunately, there is no guarantee that this approach or even more sophisticated methods will succeed in eliminating all forms of environmental confounds.

A clean and robust way to address this challenge would be to conduct GWAS on samples of dizygotic, non-identical (DZ) twins or siblings. This would exploit the fact that genetic variations between DZ twins or siblings are entirely random and therefore independent from family-specific environmental effects. Unfortunately, the currently available sample sizes for DZ twins and siblings are still too small to allow conducting well-powered GWAS analysis on most genetically complex traits. This may change in the future, thanks to growing data availability and recent efforts of researchers to form consortia of genotyped twin datasets. In the meantime, however, the above mentioned controls in population samples are being used and they can be combined with various follow-up analyses that allow quantifying the extent to which population structure is present in GWAS summary statistics (Bulik-Sullivan et al., 2015; Okbay et al., 2016).

Finally, the GWAS approach of scanning the entire genome for associations with an outcome implies that a very large number of independent hypotheses are being tested. This is a challenge for statistical inference. To illustrate the problem, consider the hypothetical experiment of a researcher who conducts one million independent statistical tests for which the Null hypothesis of no association between x and y is always true. If a p-value threshold of 0.05 is used to evaluate the tests, it implies that the researcher is expected to reject the Null hypothesis 5% of the times even if the Null hypothesis is true. Thus, ≈5% of the 1,000,000 independent tests will be erroneously rejected, implying that the researcher will end up with 50,000 expected false positive results. This is, of course, unacceptable and requires a correction in the testing procedure that adjusts for the massive amount of independent hypotheses that were tested. Empirically, a GWAS tests indeed roughly 1,000,000 independent hypotheses.[4] The actual number of independent tests varies from study to study depending on how many rare genetic variants are included and which population is being studied (Auton et al., 2015). To correct for this multiple testing burden, GWAS impose a very stringent p-value threshold of 5×10^{-8}, and only SNPs that have p-values lower than this are considered to be "genome-wide significant" [5]. Many published GWAS effectively test less than 500,000 independent hypotheses because they are limited to samples of Europeans and common SNPs with minor allele frequencies of >2% - but they still use the genome-wide significance level of $p < 5\times10^{-8}$ as a benchmark for statistical inference. However, this is still under discussion among the scientists (Storey at al., 2003).

One side-effect of having to use these very stringent p-value thresholds for statistical inference is that it lowers the statistical power of the test, i.e. the researcher's ability to reject the Null hypothesis if it is actually false. However, statistical power increases with sample size (N) and therefore, successful GWAS efforts for genetically complex traits typically require extremely large sample sizes, often in the range of $N > 100{,}000$ (Benjamin et al., 2012; Chabris et al., 2015; Ioannidis, 2005; Rietveld et al., 2013; Visscher et al., 2017).

Summary: GWAS assumptions and challenges

- GWAS require stringent controls for multiple testing, reflected in the commonly used thresholds for genome-wide significance of 5×10^{-8}
- GWAS require stringent controls for environmental confounds that may be correlated with genetic data

[4] The actual number of SNPs included in a GWAS is often substantially higher than one million, but this does not necessarily translate into a higher number of statistically independent hypotheses that are being tested because SNPs that are physically close to each other on chromosome tend to be correlated with each other, see (Auton et al. 2015).

[5] Technically, the p-value threshold of 5×10^{-8} can be motivated by a Bonferroni correction for one million independent tests with a family-wide significance level of 0.05.

- GWAS do not inform on which associated genetic variants are causal: these studies highlight correlations but do not conclude on causation
- GWAS on genetically complex traits require very large sample sizes, often with >100,000 participants

2.3 Quality criteria

Following the disappointing replication record of the first generation of candidate gene studies (Beauchamp et al., 2011; Button et al., 2013; Chabris et al., 2012, Chabris et al., 2015; Hewitt, 2012; Ioannidis, 2005; van der Loos et al., 2011; Marchini et al., 2004; Okbay and Rietveld, 2015),, the genetics community has embraced very strict quality criteria for reports of novel genetic associations that follow directly from the discussions above.

Most importantly, the genetics community has mostly realised the need for statistically well-powered study designs (with N in the order of tens thousands for behavioural outcomes) that include sufficient corrections for multiple hypothesis testing such as the commonly accepted threshold of genome-wide significance ($p < 5\times10^{-8}$), which may have to be adjusted further downward for datasets that include many rare genetic variants or non-European samples (Auton et al., 2015; McCarthy et al., 2008).. Importantly, it has become common practice in the GWAS literature to report replication results for novel genetic associations from independent samples in the same article that reports the novel association. In fact, this is typically a requirement for GWAS to get published in leading field journals (Barban et al., 2016; Okbay et al., 2016; Cornelius A. Rietveld et al., 2014; Rietveld et al., 2013; Wood et al., 2014). This commitment of the field to improve the replication record of GWAS has resulted in a dramatic improvement of the quality of the reported empirical evidence, making large-scale GWAS results probably one of the fields in science with the most transparent and best replication records to-date (Marigorta et al., 2016; Marigorta et al., 2018; Rietveld et al., 2014; Visscher et al., 2017).

Furthermore, peer-review processes of GWAS in top journals often require substantial evidence to convince referees and readers that the reported genetic associations are not due to environmental confounds such as population stratification. A standard precaution in this context is to restrict GWAS to a sample of individuals that have similar ancestry and to exclude genetic outliers and individuals who do not belong to the largest ancestry group in the sample. In practice, this often implies that GWAS are restricted to white individuals of European decent, which led to a Euro-centric domination of the GWAS literature, which poses challenges for the generalizability of GWAS findings to non-European populations (Duncan et al., 2018; Martin et al., 2019; Mills and Rahal, 2019; Rosenberg et al., 2019). The development of a statistical method called Linkage Disequilibrium Score Regression (LD-score regression) was a major step forward that now allows scientists to distinguish to which extent GWAS results represent true association signal or confounds due to uncontrolled population structure or non-independent samples (Bulik-Sullivan et al., 2015). The intercept of LD-score regression is a measure for the extent of confounding that is driving the inflation of the test statistics, and this intercept can then be used to correct the test statistics and *p*-values of all SNPs accordingly (Karlsson Linnér et al., 2019; Lee et al., 2018; Okbay et al., 2016; Okbay et al., 2016).

Another important quality criterion is the extent to which genetic data and the derived association results have been screened for quality and plausibility. Genetic data can be imprecise (e.g. due to genotyping error, strand flips, or inaccurate imputation) and especially rare genetic variants are often measured with error. Furthermore, statistical tests with rare genetic variants have less power than tests with more common variants in the same sample, implying an increased risk of false positive association results and inflated effect size estimates for rare variants. To deal with this challenge, many GWAS efforts impose filters on minor allele frequency that exclude rare variants from the

analyses. Also, many mistakes can happen during the process of analysing the data and transfer the results to other research centres (e.g. miscoding of the dependent variable, wrong column headers, copy-paste errors), all of which can lead to wrong association results. Leading centres in complex trait genetics have developed automated processes that carefully check all data and summary statistics for these errors (Marees et al., 2018; Okbay et al., 2016; Okbay et al., 2016; Winkler et al., 2014). The extent to which a particular GWAS makes use of these quality control tools is often an important sign for the overall quality of the study.

Furthermore, good GWAS participate in the open science movement by making analysis protocols, computer code, and GWAS summary statistics publicly available to increase transparency, to enable independent replication efforts, and to foster scientific progress by allowing other researchers to conduct follow-up studies that use GWAS summary statistics as input (Foster and Deardorff, 2017; Nosek et al., 2015). Last but not least, high quality GWAS publications adhere to recent reporting standards for statistical analyses (Cumming, 2011). In particular, they report all details about the analyses that were conducted and include information about sample sizes, reference allele, minor allele frequency, effect size estimates, and confidence intervals for each SNP in addition to *p*-values.

Leading scientific journals that publish GWAS are strictly enforcing these quality standards. Yet, not all researchers groups and journals adhere to the same level of rigor and deviations from any of these best practices could point to potential problems in the respective GWAS article.

Summary: Important quality criteria of GWAS for genetically complex traits

- Sample size (N > 100,000 for behavioural outcomes)
- Accurate adjustment for multiple hypothesis testing (e.g. using genome-wide significance thresholds of 5×10^{-8} or lower)
- Replication in independent samples
- Discovery sample of homogeneous ancestry, without genetic outliers, and controlled for subtle population structure in the association analyses
- GWAS results tested against presence of population stratification (e.g. LD score regression or genetic prediction within families)
- Strict quality control parameters for genetic data and association statistics
- Effect sizes and confidence intervals shall be reported (not just *p*-values)
- Summary statistics and analysis protocols are publicly available to enable independent replication efforts and follow-up research

2.4 Use of GWAS to make predictions: polygenic scores

Is it possible, using insights derived from GWAS, to make some predictions? Large-scale GWAS on behavioural traits have clearly shown that each common genetic variant considered separately captures only tiny amounts of the overall variation of the trait in the population. Yet, aggregating the effects of many SNPs in so-called polygenic scores (PGS) yields a genetic index that can capture substantial parts of the variation in behavioural traits and thus become broadly useful for social science. In polygenic scoring, researchers take results from a GWAS of a specific trait and apply them in a *new* sample, weighting each person's genetic variants by the effect size from the GWAS and summing across the variants. The resulting PGS is therefore a linear index summarizing an individual's overall genetic liability towards a phenotype (Wray et al., 2014). It can be theoretically shown that the accuracy of PGS primarily depends on the heritability of the trait (+), the GWAS sample size (+), the polygenicity of the trait (-), and whether the genetic architecture of the trait varies across different environments (-) (Dudbridge, 2013; de Vlaming et al., 2017). The upper limit of how much a PGS for a trait could

predict is given by the narrow-sense SNP-based heritability of that trait. Empirical results map theoretical expectations well, showing a clear upward trend of PGS accuracy with growing GWAS sample sizes (Visscher et al., 2017).

An example is provided in Annex 2.

There are many different ways how PGS can be constructed in practice that vary in how the methods deal with the correlation patterns between SNPs (which are ignored in GWAS) and how many SNPs are included in the score (Euesden et al., 2015; Lin et al., 2017; Vilhjálmsson et al., 2015)Typically, methods that include a large number of SNPs and account for correlations patterns between SNPs to some extent tend to perform better, but all PGS constructed from large-scale GWAS typically have at least some degree of explanatory power in hold-out samples.

Nevertheless, the extent to which PGS are predictive of an outcome also depends on whether the genetic architecture of the trait of interest is comparable in the GWAS discovery and the prediction sample (de Vlaming et al., 2017).

In addition, several recent studies have clearly illustrated that PGS can currently not be used to make accurate predictions in samples of different ancestry than the GWAS discovery sample Partly, this is for technical reasons (e.g. different ancestry groups have different minor allele frequencies which means that the estimated GWAS effect sizes from one group are not the correct effect sizes for another group) but partly, this is also because the environmental conditions that influence outcomes tend to be very different for individuals with different ancestries Duncan et al., 2018; Martin et al., 2019; Rosenberg et al., 2019; de Vlaming et al., 2017). As a consequence, differences in PGS values between groups cannot be used to draw conclusions about the reasons for observed differences in a trait between these groups if the environmental conditions for the groups are different and / or if the groups have different ancestry background.

Importantly, even though PGS for social scientific outcomes such as educational attainment begin to capture substantial amounts of sample variation, even the best currently available PGS are not useful for individual-level prediction.

Summary: Polygenic scores (PGS)

- PGS are more predictive for genetically complex traits than any specific genetic variant alone
- PGS get more precise as GWAS sample sizes increase
- The upper limit for how much of the variance of a trait could be captured by a PGS is given by the SNP-heritability of that trait
- PGS are becoming increasingly useful tools for social scientists
- PGS have very limited use for individual-level prediction
- PGS have no or only limited predictive accuracy in samples with different ancestry or environmental background
- PGS cannot be used to draw conclusions about the reasons for phenotypic differences between groups or over time

2.5 Interpretation of GWAS results

2.5.1 Interplay of genetic and environmental factors

Section 1 explained that the heritability of a trait depends on the environment in which the trait is being studied. Furthermore, the heritability of a trait puts no upper limits on the relevance of the environment and it **implies no strong form of biological determinism**. In some cases, the outcome of heritability of a trait may even be entirely

influenced by specific environmental conditions. All of these caveats extend directly to GWAS results and their interpretation. It is tempting to think of GWAS results as direct glimpses into biological mechanisms that may influence a trait. And although GWAS results can be helpful for this purpose (Visscher et al., 2017), a correct interpretation of GWAS results does not only require bioinformatics attempts to link SNP associations with possible biological mechanisms, it also requires asking three additional questions:

First, to which extent are specific GWAS results externally valid, i.e. are the same genetic associations found in different environments? One way to address this question empirically is to estimate the genetic correlation of the same outcome using GWAS results from different environments with bivariate LD score regression (Bulik-Sullivan et al., 2015). For example, the latest GWAS on educational attainment found an average genetic correlation of 0.723 (SE = 0.124) across the 48 cohorts that were meta-analysed. Thus, although educational attainment has similar genetic architectures in the cohorts being studied, there is some degree of heterogeneity that may be due to different environmental conditions. Of note, all of the 48 cohorts included in that study were from rich, industrialised countries and based on white, European-decent individuals, most of which were born in the 20th century, but it can be that the genetic correlations between samples with different backgrounds would differ as well.

Second, to which extent and how are the genetic associations with a phenotype mediated by behavioural and environmental channels? This question is much more difficult to answer than the first one because of the sheer endless space of possibilities how this could happen. For example, a gene may be associated with lung cancer because it may influence a person's tendency to smoke tobacco. Similarly, a gene associated with educational attainment may work via environmental feedback loops that provide specific support and positive reinforcement such as praise for hard work or a preference of teachers for children with specific physical attributes that may be entirely independent from cognitive aptitude, personality, or motivation. Thus, even if GWAS manage to avoid bias from population structure entirely, any identified genetic correlation may still point to important behavioural or environmental pathways that influence the outcome of interest. One approach to tackle this challenge is to conduct follow-up studies that have excellent data about environmental exposures and behaviours in addition to genetic information. In such datasets, polygenic scores for the outcome of interest can be constructed and mediation models can be tested. Studies of this type are still rare, but will hopeful become more common in the future, helping us to come to a better understanding of the implications of the GWAS findings.

Third, to which extent are the GWAS results picking up on subtle environmental confounds such as parental environment? As previously discussed, one of the biggest challenges in GWAS is to correct for environmental causes that are unobserved but correlated with genetic data. The methods that are currently used in GWAS seem to do a reasonably good job of correcting for such biases in many cases. However, behavioural and socio-economic outcomes such as educational attainment are particularly challenging in this regard because the rearing environment and family background are of obvious importance for such outcomes. Yet, it is difficult to separate genetic effects from family-specific environments in population samples that do not include a sufficiently large number of members of the same family that would allow controlling for family-specific environments in a rigorous way (Belsky et al., 2018; Koellinger and Harden, 2018; Kong et al., 2018). Importantly, even if GWAS results replicate consistently across samples and standard tests against population stratification do not raise concerns, this still does not rule out that the observed genetic association with the outcome are partly due to the rearing environments that parents and other relatives create due to their genetic make-up (Koellinger and Harden ,2018; Kong et al., 2018).

Thus, even carefully conducted, well-powered GWAS analyses should be thought of as partial correlations that are not necessarily informative about biology or causal mechanisms by themselves. However, robust genetic correlations that are free of population structure provide an excellent starting point for social and

medical scientists to investigate possible causal mechanisms in follow-up analyses that exploit the fact that genes remain unchanged during the lifetime of an individual and that the genetic endowment of each individual is a random mix of the parental genes (Conley and Zhang, 2018, Davey Smith, 2010, Diprete et al., 2018; Okbay et al., 2016).

2.5.2 Limits to the generalizability of GWAS findings

At least three issues potentially limit the generalizability of GWAS findings to populations other than those that were studied in the GWAS:

First, as discussed above, different populations may be exposed to different environmental conditions that can shape the molecular genetic architecture of a trait. This may limit the genetic correlation of the trait between populations and attenuate the predictive accuracy of polygenic scores (de Vlaming et al., 2017).

Second, populations with different ancestry differ in the minor allele frequencies of genetic variants and their correlations with each other. As a result, the SNPs effects on GWAS estimates from one ancestry group are not directly applicable to a different ancestry group. As a rule of thumb, the greater the differences in environmental conditions and ancestry are between two different population, the less informative GWAS results from one population are for the other (Martin et al., 2019; Neale et al., 2018; Rosenberg et al., 2019). Since the vast majority of current GWAS results are based on white individuals of European decent that live in rich, industrialised countries, this imposes serious limitations to extend the insights from existing GWAS to non-European populations or populations from non-industrialised environments (Martin et al., 2019; Mills and Rahal, 2019). This has two important implications: First, to the extent that GWAS results can lead to societal benefits, these benefits are currently limited to European populations that are already quite advantaged to begin with. And second, GWAS results cannot and should not be used to make any sort of inference about the reasons for differences in particular traits across populations or over time.

Third, selection bias into GWAS samples is a violation of assumption 3 (described in Annex 1) that may induce spurious, misleading correlations that do not only threaten the external validity of results, but the internal validity as well (Haworth et al., 2019; Hughes et al., 2019). The extent to which this potential problem matters is currently an active area of investigation.

Summary: Interpretation of GWAS results

- GWAS results from population samples are partial correlations, not causal estimates of genetic effects
- GWAS results depend on environmental conditions and the ancestry of the population being studied
- GWAS results may point to specific biological mechanisms, but this does not imply biological determinism or irrelevance of the environment for the traits being studied
- GWAS results can also point to important behavioural and environmental channels that influence an outcome
- GWAS results are generally not informative about the reasons for differences in traits between different populations or over time
- Sample selection may limit both the external and the internal validity of GWAS results
- Unbiased GWAS results are a good starting point for follow-up studies that investigate Genome-Environment interactions and causal mechanisms; they do not replace such efforts

2.6 Missing heritability

Although GWAS have led to thousands of replicable genetic associations with medical and behavioural outcomes in the last few years, the discovered genetic variants typically account only for a small share of the estimated heritability of these traits. Polygenic score capture greater shares of the variance of these traits, but even polygenic scores typically still fall short of the heritability estimates derived from twin studies. This empirical pattern has become known as "missing heritability" in the genetics literature and much research has been conducted in the past few years to understand this empirical pattern better. In a nutshell, it is necessary to differentiate between two types of heritability estimates: One is the so-called chip heritability which is estimated from currently observed genetic variants, the other is the so-called total heritability, which takes all genetic effects into account. GWAS findings and polygenic scores derived from GWAS findings can at most account as much as the estimated chip heritability of a trait since all these statistical approaches are restricted by how much of all relevant genetic variation among humans is currently measured and observed. The gap between the predictive accuracy of polygenic scores, and the chip heritability gets smaller as GWAS sample sizes increase and empirical trends match this theoretical expectation (Lee et al., 2018; de Vlaming et al., 2017; Witte et al., 2014).

The remaining gap between chip heritability and total heritability could be either due to upward bias of twin studies or due to unobserved genetic variants. At least for body height and body mass index, it appears that high-accuracy imputation of missing genetic variants or whole-genome sequencing can account for the still-missing heritability gap (Wainschtein et al., 2019; Yang et al., 2015; Yang et al., 2017).

Thus, the "missing heritability" pattern is consistent with theoretical expectations for genetically complex traits, given available GWAS sample sizes and genotyping technology.

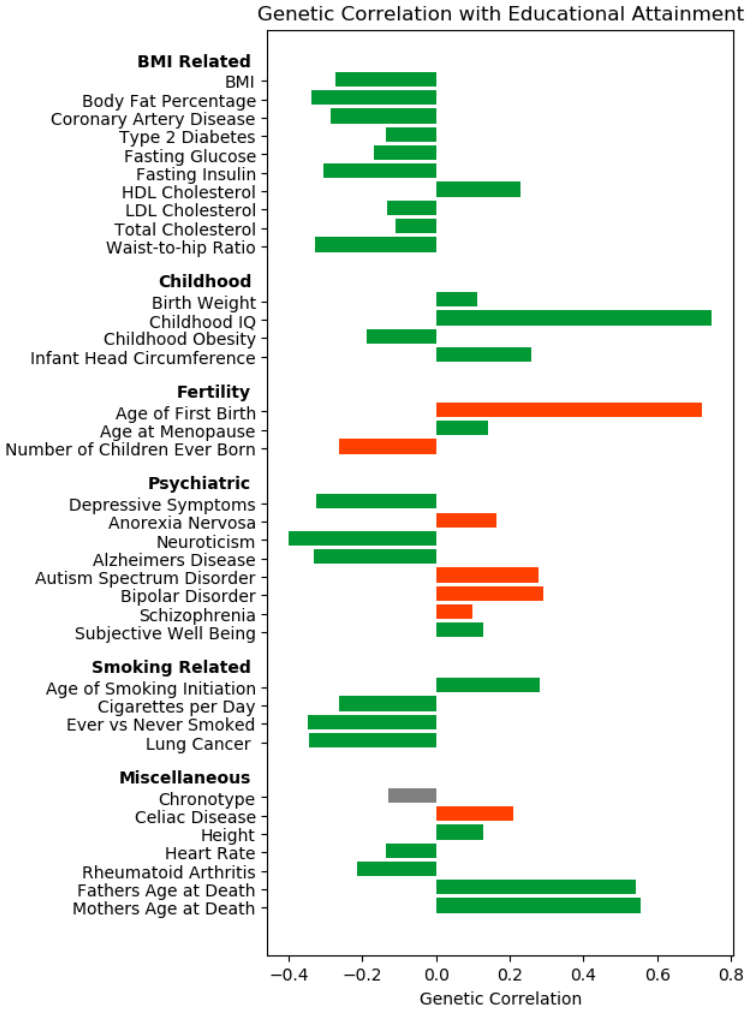
Summary

- The gap between the total heritability of a trait and the variance in the trait captured by GWAS results is closing with increasing GWAS sample sizes and higher resolution genotyping technologies
- The remaining gap may be due to currently unmeasured, but relevant, genetic variants or upward bias in twin studies of heritability.

2.7 GWAS as linker between social science genetics and medical research

Social science genetics is sometime viewed as being independent or isolated from genetic epidemiology and medical research. However, this is a misconception. Instead, there is intensive, frequent, and highly productive collaboration between social science geneticists and genetic epidemiologists that involve sharing of data, computational infrastructure, methods development, training and teaching activities, co-authorship on peer-reviewed publications, as well as joint conferences and workshops. These collaborations arise not only out of a common methodological approach and the widely accepted advantages of open science and data sharing, they also arise out of many links between health outcomes and behavioural phenotypes that both parties are interested in. For example, poverty and economic deprivation are major risk factors for mental and physical diseases (Wilkinson and Marmot, 2003), lower life expectancy (Stringhini et al., 2017), and antisocial behaviour (Piotrowska et al. 2015). Furthermore, socio-economic status is important to health not only for those in poverty, but at all levels of the socio-economic latter (Adler and Al., 1994). Thus, environmental and behavioural influences are of critical importance for many health outcomes, and this provides strong reasons for social and medical scientists to collaborate.

Figure 1: Genetic correlations of educational attainment with traits across the entire lifespan



Source: Figure 1 in Harden and Koellinger (2019).
Note: Genetic correlations of educational attainment (Okbay, Beauchamp, et al. 2016) with 196 traits were computed using bivariate LD-score regression on LDHub (http://ldsc.broadinstitute.org/ldhub/). The Figure only shows a subset of all results that are significant after Bonferroni-correction ($p < 0.05/196 = 2.5\times10^{-4}$, two-sided tests). Green and red color symbolize advantageous and disadvantageous genetic relationships of educational attainment with health, respectively. Grey means no obvious health relevance.

Indeed, social scientific outcomes often resemble medical outcomes to a substantial extent, as reflected by the pervasive and often relatively high genetic correlations between behavioural and medical outcomes (Bansal et al., 2018; Bulik-Sullivan et al., 2015; Karlsson Linnér et al., 2019; Okbay et al., 2016; Turley et al., 2018). For example, Figure 1 shows genetic correlations of educational attainment health-related outcomes throughout the entire lifespan, from birthweight (+), infant head circumference (+) and childhood IQ (+), to BMI-related traits (-), depressive symptoms (-), neuroticism (-), smoking (-) in adulthood, all the way to risk for lung cancer (-), Alzheimer's disease (-), and longevity (+, as implied by parent's age at death).

The genetic correlations between normal range behaviours or socio-economic outcomes on the one side and medical outcomes on the other side imply that social science genetics can help to advance medical research. For example, the findings from social science genetics can be exploited to speed up the genetic discovery for medical traits because it is sometimes possible to study social scientific outcomes in much larger GWAS samples than medical outcomes. That is possible because social scientific outcomes such as educational attainment, risk tolerance, or subjective well-being can be measured easily and cheaply with just one short multiple choice question that is often included in

many genetic datasets. Specific medical outcomes, instead, often require setting up specific case-control cohorts that are expensive and restricted by the rare occurrence of the disease in the population. And even if medical conditions can be studied in population samples, they often require more comprehensive survey batteries (e.g. psychiatric conditions or cognitive performance tests) or expensive procedures (e.g. analysis of blood samples or MRI scans) that limit available GWAS sample sizes to study these outcomes. Several statistical methods have been recently developed and applied that allow leveraging information from genetically correlated traits to boost statistical power (Grotzinger et al., 2019; Cornelius A Rietveld et al., 2014; Turley et al., 2018). The application of these new tools in combination with large-scale GWAS results on social scientific outcomes has already benefitted medical research substantially. For example, the first replicable genetic associations with general cognitive performance have been discovered using educational attainment as a proxy (Rietveld et al., 2014). Similarly, subjective well-being and happiness ratings were used to discover the first replicable genetic associations for depression among Europeans (Okbay et al., 2016). Furthermore, genetic association results for traits such as educational attainment are being used to gain new insights into psychiatric nosology (Bansal et al. 2018) and to study the causes of diseases (Tyrrell et al., 2016).

Thus, it is a misconception to think of genetic epidemiology and social science genetics as isolated fields. Rather, there is intense collaboration and cross-fertilization of ideas between both fields. As a result, social science genetics can and is helping to advance genetic epidemiology in a variety of important ways (and *vice versa*).

Summary

- There is intensive collaboration between social science genetics and genetic epidemiology
- Genetic and phenotypic correlations between social scientific outcomes and health are prevalent and informative
- Social science genetics is helping to advance medical research

3 Important scientific and regulatory considerations

Thanks to the wide-spread availability of genetic data and large-scale GWAS results, the insights into the functions of the human genome and its links with differences among people are rapidly growing. The possible applications of these insights are sheer endless, raising both opportunities and challenges of various kinds for individuals and societies at large. This holds true both for medical and non-medical applications.

Policy makers are advised to actively monitor these developments and to respond to them ideally in a proactive, timely fashion that aims at maximizing welfare at the aggregate level. Given the rapid rate of progress in the field, now is the right time to start having an informed dialogue about how these new technologies could change our lives, what type of society we want to live in in the future, and how policy makers can shape developments into directions that will embrace positive applications and regulate undesired ones. This dialogue should ideally involve policy makers, geneticists, medical researchers, lawyers, technology providers, but also ethicists and stakeholders from society at large. For many of the emerging technologies and applications, there will be a range of possible responses from policy makers, and quite likely there will not be only one correct solution. Rather, different societies may decide to weigh the opportunities and risks of various applications in varying ways, depending on their values, culture, law, their political system and other factors, potentially giving rise to divergent development paths.

This section lists several areas in which social science genetics may have an impact. This list is not supposed to be comprehensive. Quite often, there will be several possible responses of policy makers that would be justifiable and reasonable. In short, the goal of this section is not to provide definite answers, but to start a discussion about the potential implications of social science genetics.

3.1 Recent selection events

Heritable traits can be under natural selection and, as a result, the genetic composition of a species can change over time (Darwin, 1859; Fisher, 1930). Traits that favour survival until reaching a species' reproductive age and traits that are directly influencing reproductive success during lifetime are often targets for natural selection. Contrary to some beliefs, it has recently been shown that the evolution of the human species did not come to an end 40,000 – 50,000 years ago. Rather, our species keeps evolving and with it the composition of our gene pool (Mathieson et al., 2015; Pickrell et al., 2009; Voight et al., 2006). Examples of recent events of natural selection gave rise to lactase persistence (Tishkoff et al., 2007), resistance to malaria (Kwiatkowski 2005), and adaptation to high altitude (Yi et al., 2010).

Recent evidence has shown that genes associated with behavioural outcomes such as educational attainment may be subject to contemporary selection as well. Figure 1 shows that educational attainment has a negative genetic correlation with lifetime number of born children and a strong positive genetic correlation with age at first birth (see also Barban et al., 2016). Consistent with this, Beauchamp (2016) finds in a sample of US-Americans with European ancestry born between 1931 and 1953 evidence for natural selection against genes associated with educational attainment, implying -1.5 months of education per generation. However, this trend has been more than offset by environmental factors that actually led to an increase of 6.2 years in the mean number of years of schooling in the US between 1876 and 1951 and more recent trends that increased the share of Americans holding a College degree from ≈34% in 1939 to ≈55% in 1996 (Acemoglu, 2002). Similarly, Kong et al. (2017) report based on a sample of 129,808 Icelanders born between 1910 and 1990 that the average polygenic score for educational attainment has declined by ≈0.01 standard units per decade, which is substantial on an evolutionary timescale. Furthermore, due to the fact that the polygenic score that was available to Kong et al. (2017) captures only roughly one third of the chip heritability of educational attainment, the actual rate of decline could be two to three

times faster. The negative selection against genes positively associated with educational attainment seems to be partly manifested through delayed reproduction of individuals with higher education. Furthermore, Kong et al. (2017) also show that the polygenic score for educational attainment is negatively associated with lifetime reproductive success even *after* accounting for the actual education level of the individuals. Thus, although part of the mechanisms seems to work via the education-related choices that individuals make, this is not the entire reason for the observed patterns in the data.

According to the extrapolations of Kong et al. (2017), the selection against genes that are positively associated with educational attainment may imply a decline of 0.3 IQ points per decade. However, this implied decline in IQ also seems to be more than compensated by environmental changes that led to an estimated increase in IQ scores by 13.8 points between 1932 and 1978 (equivalent to +3 points per decade) according to a meta-analysis of 27 studies that were conducted in the United States and the United Kingdom. This increase in IQ scores in the 20th century is referred to as the Flynn effect and many scientists (including Flynn himself) have attributed this increase largely to improvements in the environment and technological progress (Courtiol et al., 2016; Dickens and Flynn, 2001).

Should policy makers be concerned about recent selection against genes that are positively associated with educational attainment? The answer is both no and yes. On one hand, the empirical evidence clearly shows that changes and improvements in the environment (e.g. schooling reforms, labour markets that heavily reward high skills, increasing rates of participation of women in higher education and the labour market) have led to large increases both in educational attainment and its associated traits (e.g. performance on cognitive ability tests) in the 20th century (Acemoglu 2002, Trahan et al. 2014). These environmental effects have been overriding and over-compensating the potential effects of genetic selection. Furthermore, the calculations by Beauchamp (2016) and Kong et al. (2017) that suggest potential negative implications of recent selection events on IQ and educational attainment are based on a number of strong assumptions, including (i) that the genetic architecture of IQ and educational attainment is fixed and does not vary across environments or over time, (ii) that GWAS results on educational attainment are immune to population stratification and other environmental confounds, and (iii) that the trends observed in their data are stable and that they will continue. However, the imperfect genetic correlation of educational attainment across samples suggests that the genetic architecture of education is not fixed and that it does depend on specific environmental conditions (Lee et al., 2018). Furthermore, Kong et al. (2018a) demonstrated that a part of the association signal in GWAS on educational attainment can be attributed to family-specific environmental transmission channels rather than biological effects that affect the individual directly. This complicates the interpretation of GWAS findings and the observed changes in the distribution of the polygenic score for educational attainment over time. And third, it is possible that the effects leading to selection for, or against genetic variants associated with educational attainment are cyclical and that the observed trends from the 20th century may <u>not</u> continue (Fisher 1930). Also, to the extent that the assumptions underlying the extrapolations of Beauchamp (2016) and Kong et al. (2017) are justified, it is uncertain if and to which extent environmental improvements would continue to override and dominate the effects of genetic selection in the future.

Finally, the delayed reproductive period of highly educated individuals, which seems to be partly driving the selection patterns reported by Beauchamp (2016) and Kong et al. (2017), are likely pointing to environmental challenges that make it difficult to combine parenthood with higher education and professional careers (Marini, 1984; Mills et al., 2011). This last point is certainly an issue of high policy relevance that also concerns general demographic trends, life quality, and equality.

3.2 Personalised medicine

One of the biggest opportunities and incentive for genetics research is the prospect of personalised medicine – an attempt to identify individual risks and to guide clinical management and decision making that takes genetic assessments into account. Personalised medicine has the potential to affect all stages of health care, starting from baseline risk assessments of susceptibility, preclinical targeted screening, more precise diagnosis and assessment of prognosis, all the way to personalised treatment that may vary in terms of approach, timing, and dosage (Chan and Ginsburg 2011, Ginsburg and Willard 2009).

One major obstacle to realizing these opportunities is the need to identify the molecular genetic architecture of health outcomes and to develop polygenic scores that are predictive and valid. Although a lot of progress has been made in the past few years, progress may not have been as rapid as originally thought due to the unexpectedly complex genetic architecture of most diseases and the challenges that come with identifying many small genetic influences (Joyner and Paneth, 2015; Mills and Rahal, 2019; Rose, 2013; Visscher et al., 2017). Social science genetics can help to speed up such efforts, thanks to the prevalent genetic overlap between health outcomes and behavioural traits as well as the possibility to study social scientific outcomes in GWAS sample sizes that can be much larger than those for specific medical traits. For example, genetic insights from large-scale GWAS on subjective well-being and happiness have led to new insights and improved the predictive accuracy of polygenic scores for depression (Okbay et al. 2016; Turley et al., 2018; Wray et al., 2018).

Furthermore, social science genetics can help to gain new insights into disease subtypes that may improve diagnosis and treatment. For example, results from a GWAS on educational attainment were used to demonstrate that schizophrenia is a genetically heterogeneous diagnosis with at least two sub-types that may require different treatment – one of them being a neurodevelopmental disorder, while the other one is not (Bansal et al., 2018).

However, the vision of personalised medicine is not without challenges (Chan and Ginsburg, 2011; Joyner and Paneth, 2015; Rose 2013) which need to be addressed by the policy makers; these include,

- new needs to educate medical professional about genomics,
- designing clinical decision support systems that provide doctors and caregivers relevant, understandable, and accurate information at the right time,
- new challenges for protecting privacy rights of patients and their relatives,
- new dangers for discrimination based on genetic information,
- new tasks for regulators to assess genetic tests and intervention options, to assess and compare the effectiveness of personalised approaches to medicine, and to update reimbursement policies accordingly,
- new questions arising from consumer's direct access to genetic information that may or may not be health relevant.

3.3 Clinical trials

Clinical trials are an essential tool to assess whether a new therapeutic agent is effective for treating a specific medical condition, which side effects it has and whether potential benefits outweigh risks. Clinical trials are also one of the most expensive parts of new drug development, with costs often in the tens of millions of Euros (Moore et al., 2018). From a statistical perspective, the main goal of a clinical trial is to estimate the effect of the treatment as precisely as possible. In other words, the goal is to estimate the effect size of the treatment with as little estimation error as possible (i.e. to obtain a small standard error on the estimated regression coefficient).

There are two main possibilities to influence the estimation error in a statistical analysis. The first one is to increase sample size to reduce estimation error (Wooldridge, 2002). Whenever this strategy is possible, it is usually desirable from a statistical point of view (Ioannidis, 2005). However, the costs of clinical trials strongly depend on the number of participants (Moore et al., 2018) and increasing the number of trial participants may not be feasible for economic or logistic reasons (e.g. the disease may be very rare in the populations, restricting the number of cases that can be recruited).

The second possibility to reduce estimation error in a statistical analysis is to add control variables to the regression that are correlated with the outcome of interest. These control variable "absorb" residual variation in the outcome and thereby increase the precision of the estimate of the treatment of interest (Benjamin et al., 2012; Wooldridge, 2002). In principle, any variable that can be measured among the participants of a clinical trial that is correlated with the health outcome is potentially useful for this purpose. This also includes polygenic scores that are derived from genetic data. Collecting genetic data from the participants of clinical trials is particularly useful because (i) it is relatively cheap to collect genetic data, especially in the context of a clinical trial, (ii) it opens up possibilities to investigate potential gene-treatment interaction effects that may guide personalised medicine efforts, and (iii) once the genetic data has been collected, it is possible to construct polygenic scores for a large number of potentially relevant outcomes from publicly available GWAS summary statistics (Welter et al. 2014). This includes GWAS summary statistics for social scientific outcomes, some of which may have high utility in clinical trials. For example, a clinical trial that evaluates a new drug that is supposed to improve memory performance would benefit from including a polygenic score for educational attainment as a control variable because this score would very likely capture some of the residual variance in memory performance in the sample. Similarly, a clinical trial for a new drug against depression would benefit from polygenic scores for subjective well-being or neuroticism as control variable (Okbayet al., 2016). Alternatively, multivariate statistical techniques could be used to "boost" the accuracy for a polygenic score for clinical depression by "borrowing" information from large-scale GWAS on subjective well-being and neuroticism (Grotzinger et al., 2019; Turley et al., 2018).

Using polygenic scores to boost the statistical power and accuracy of clinical trials has the potential to save costs because it allows reduction of the size of the clinical trial without sacrificing precision of the obtained results (Benjamin et al., 2012; Rietveld et al., 2013). Thus, the results of social science genetics have the potential to make clinical trials more cost effective and / or more accurate, which is highly desirable for all stakeholders. As long as the participants of the clinical trial give informed consent that allows collecting and using their genetic data for the purposes described above, this type of application has substantial benefits for science and society with no risks that go beyond the typical risks of participating in a clinical trial.

3.4 Policy evaluations

A major goal in economics and social sciences is to understand and to evaluate the effects of specific policy reforms or interventions. Following the same line of argument as in section 3.3, polygenic scores for social scientific outcomes can also be very useful in policy evaluations because they have the potential to yield more precise estimates of the effects of the policy or to study the effects of the policy in a smaller sample without loss of accuracy at lower costs (Benjamin et al., 2012; Rietveld et al., 2013). Furthermore, they enable new ways to investigate that responds to policy interventions, and who benefits or loses as a results of a policy or environmental exposure (Barcellos et al., 2018; Harden et al., 2019). Again, as long as study participants provide informed consent that allows collecting and using their genetic data for the purposes described above, this type of application of social science genetics has substantial benefits for science and society with very limited risks for the study participants and no obvious downside for society at large.

3.5 Personalised education

The robust, replicable genetic associations and the availability of polygenic scores for traits such as educational attainment and cognitive performance (Lee et al., 2018; Savage et al., 2018) that begin to capture non-trivial shares of variation among people (R^2 > 10% and R^2 > 5% for education and intelligence in samples of people of European descents) has prompted some scientists to speculate on the possibility to use genetic data to offer children personalised education (Asbury and Plomin, 2013; Plomin, 2018; Plomin and Stumm, 2018; Sokolowski and Ansari, 2018; Wilby, 2014), akin to the vision of personalised medicine described in section 3.2.

Although the goal to provide educational experiences to children in ways that attempt to maximise their potential and success in life is laudable, there are important caveats to the idea to use polygenic scores in this context that must be considered. First, as explained above, GWAS are not necessarily only detecting meaningful biological signals. GWAS can be affected by subtle kinds of population stratification and indirect genetic effects from parents and other close relatives. Furthermore, the environment may induce genetic associations. For example, if teachers discriminate against children with blue eyes, a GWAS on educational attainment would find associations with eye colour genes, but not for generalizable, biological reasons that affect the schooling potential of children. An educational policy that would be based on polygenic score values would therefore run a substantial risk to perpetuate environmentally induced disadvantages of children or to target the wrong cause for individual differences in schooling outcomes (e.g. discrimination in the class room instead of biological limitations of learning abilities).

Second, using polygenic scores to rank, sort, and target children comes with a substantial risk to stigmatise and demotivate those with low score values. In general, very little is known (until now) about how children or adults react to information about their "genetic potential", whether this information is meaningful or not. It could be that good news about one's "potential" from genetic data are either motivating or hampering personal efforts and vice versa. One of the first studies that investigated how people react to learning their genetic risk for disease found that the effects of the perceived genetic risk were often self-fulfilling and sometimes greater than the effects associated with actual genetic risk (Turnwald et al., 2019). This raises important questions about whether it is desirable to inform children, parents, or teachers about "genetic risks" or "potential" of school children or whether such attempts may be counter-productive, even if they are well-intended.

Third, even though polygenic scores with $R^2 \approx 10\%$ may be highly predictive of expected mean phenotypic differences between quantiles of the polygenic score distribution, their predictive accuracy is too low to make useful inference for any specific individual (Lee et al. 2018, SSGAC 2018). Clinicians would not make a diagnosis or prescribe a treatment based on a biomarker that hardly improves the chance to make a correct diagnosis compared to guessing without any information at all. Similarly, it is probably not a good idea to base educational policies, recommendations, tracks, and experiences on a biomarker that carries very little information for any specific individual.

Fourth, even if all of the above issues would be resolved, it is not obvious what a "personalised education" should look like, which specific goals it should try to accomplish, how much it would cost, whether it would be superior to good education systems that do not rely on genetic information, and which degree of autonomy it would leave children and their parents.

For all of the above reasons, it is too early and potentially undesirable to think about policies that use genetic information to design personalised educational experiences.

3.6 Assisted reproductive technologies and "designer babies"

A designer baby is a baby whose genetic makeup has been selected or altered, often to include a particular gene or to remove genes associated with disease.

In late 2018, a company called Genomic Prediction, based in New Jersey (USA), hit the news when it started to offer a technology that allows screening *in-vitro* fertilization (IVF) embryos for risk of having low intelligence (Ball, 2018; Janssens, 2018; Wilson, 2018). Embryo screening in IVF was already possible earlier, for example to identify rare genetic diseases of which the parents are known to be carriers. Some countries permit such technologies, and some even allow using this technology to screen for non-medical traits such as an embryo's sex for "family balancing" purposes (Ball, 2018). But Genomic Prediction is the first company that uses polygenic scores to identify embryos that have a high chance of an intelligence quotient (IQ) 25 points below the average, which would correspond to a mild mental retardation. Currently, Genomic Prediction only offers screening for potentially very low IQ. However, the same technology could in principle also be used to screen for very high IQ embryos or for other genetically complex traits that parents may or may not find desirable. The example of Genomic Prediction shows that companies are eager to offer such technologies, even though the predictive accuracy of polygenic scores is still limited and despite the serious and complicated ethical issues this raises. The example illustrates how fast GWAS may influence the sector and demands the policy makers' and society's response on how to determine the margins and deal with these possibilities.

From a scientific perspective, the potential influence of such technologies is still relatively limited for genetically complex traits such as IQ. Because even the best currently available polygenic scores predict only small shares of the overall variation of the trait, very extreme values of the polygenic score would be needed to identify potential phenotypic outliers with high certainty. However, extreme values of polygenic scores do not occur often by definition, and the number of embryos a couple can choose from during an IVF procedure is limited. Yet, as GWAS sample sizes continue to grow and polygenic scores will improve their accuracy, the detection of potential outliers from extreme polygenic scores will become more and more feasible. In the example of Genomic Prediction, it is unclear how accurate their test for low IQ actually is and how often an embryo would be flagged as potentially "at risk". Furthermore, it is unclear if the company is restricting their tests to parents of European descent, and if not, how (if at all) they are planning to apply polygenic scores from European GWAS results to non-European populations. Aside from the pressing and imminent ethical problems that this new technology raises, there are substantial remaining concerns about how accurate the offered tests are, in which contexts and populations they could potentially work, and *why* they "work" in the first place.

Even if such technologies have a high accuracy to identify potential phenotypic outliers for genetically complex traits based on actual biological reasons, there would still be concerns about what such procedures actually select for.

The biggest remaining scientific challenge is called pleiotropy – i.e. the tendency of most genes to affect more than just one outcome, while not all outcomes that a gene affects are necessarily desirable (Verbanck et al., 2018; Welter et al., 2014). For example, the genes that are positively associated with educational attainment also tend to increase the risk for some psychiatric disorders such as anorexia nervosa, bipolar disorder, and autism. Thus, selection for specific traits is likely to have undesirable side effects as well. Furthermore, if such selection would happen at a grand scale, it would have the potential to change the gene pool of the species in path-dependent ways that may be problematic and difficult to reverse (e.g. by restricting genetic diversity that makes the species potentially more resilient against environmental threats and shocks, or by increasing the prevalence of undesirable side-effects). In addition, the legalization and use of such technologies may foster discrimination and stigmatization of minorities in societies, up to

the point that some people or groups may feel or actually be threatened in their right to exist.

In summary, assisted reproductive technologies that screen embryos for genetic characteristics are in general feasible and to some extent already available on the market. There are important remaining concerns about how precise and effective these technologies are and what side effects they may have. Depending on the values and political systems that countries have, it is likely that the use of these technologies will be regulated or forbidden in some, but tolerated and possibly even actively promoted in others. Also, it is very likely that access to these technologies will in practice often be restricted to the small share of the human population that is educated, wealthy, and of European ancestry. Such developments could exacerbate already existing inequalities in health and wealth further (Martin et al., 2019; Mills and Rahal, 2019).

Assisted reproductive technologies such as sperm banks that enable to choose sperm donors based on characteristics such as occupation, education, health, or height have been in existence for decades (e.g. Fairfax Cryobank - How to choose a sperm donor). Selection based on the observed phenotypes of the parents is potentially much more effective than selection based on current polygenic scores that only capture a part of the overall heritability of a trait. Thus, the possibility to choose sperm donors with specific characteristics comes with some of the same challenges mentioned above, including the potential selection for undesirable side effects and rising inequalities. Yet, these practices are already available in many countries for a long time and they have helped large numbers of couples to have healthy children that would otherwise have difficulties or not possibilities to do so at all. Also, the availability of such sperm banks did probably not contribute much (if at all) to rising inequalities or unwanted side-effects for the children that were born or societies at large.

In summary, assisted reproductive technologies to create "designer babies" already exist and some early forms that did not use genetic data directly have already been used for decades. Genetic data, polygenic scores, and possibilities to edit genes directly keep expanding the scope of what is technologically feasible at a rapid pace, making it urgent for societies and policy makers to examine the technological possibilities, their potential advantages and risks, as well as the complex and serious ethical and political issues they raise. The policy options in response to these considerations range from active support of such technologies, *laissez faire*, all the way to banning and punishing their use.

3.7 Companies offering genetic tests: security issues and legal caveats

Genetic tests can be very valuable for many purposes and are sometimes the most cost-effective, safest, or even the only way to obtain desired information. However, they typically also involve potential risks, ranging from personal anxiety, depression, changes in life plans, privacy concerns or guilt at the individual level to potentials for stigmatization and discrimination at the social level, to name just a few (Shoenbill et al., 2014).

A variety of companies have sprung up that offer genetic tests either for private consumers, health professionals, or companies. The available tests cover a broad range of potential applications from medical diagnostics and risk assessment, genealogy, all the way to assessments of "genetic potential" for non-medical traits such as educational attainment and personality. The quality, reliability, and usefulness of the provided assessments vary widely and range from high-accuracy detection of monogenetic disorders to obscure, unwarranted claims (e.g. being "a good sales person"), raising concerns as these findings are not sufficiently supported by scientific evidence.

This emerging industry raises a number of important issues that policy makers and regulators should monitor closely. This includes questions about whether companies have sufficiently high data security standards, for which purposes they use these data, who will have access to the data, whether their test results are sufficiently reliable, accurate,

and backed up by scientific evidence, whether customers are aware of and provided informed consent for all the ways the company will use the collected data, how accurately and clearly companies communicate test results to their customers, and whether they take sufficient measures to prevent the potential misuse or misinterpretation of their test results. In principle, genetics companies may offer services that are highly valuable to their customers. However, data security breaches, unforeseen or unconsented uses of the data, and wrong or potentially misleading test results could cause substantial harm and damage.

A prominent example of the regulatory challenges is 23andMe, one of the leading consumer genetics companies in the world. In 2013, the company ran a television commercial in the US, in which attractive young people said that for $99 one could learn "hundreds of things about your health," including that you "might have an increased risk of heart disease, arthritis, gallstones, [or] hemochromatosis" (http://www.ispot.tv/ad/7qoF/23-and-me). In response to the commercial, the Food and Drug Administration (FDA) sent 23andMe a warning letter ordering it to "immediately discontinue marketing the PGS [Saliva Collection Kit and Personal Genome Service] until such time as it receives FDA marketing authorization for the device" (Warning Letters - 23andMe, Inc. 11/22/13). 23andMe complied with the FDA's demands and discontinued running the commercial (Annas and Elias, 2014). By issuing the warning letter, the FDA required 23andMe to demonstrate their tests are safe and effective. 23andMe has since then been engaged in a dialogue with the FDA that resulted in formal applications for specific testing services that the company offers, some of which have gained FDA approval in the meantime, including consumer genetic tests for cancer risk and the right dosage of a specific drug (Brown, 2019). While many of the health-related tests that 23andMe originally offered were not available anymore to US costumers after the FDA's warning letter, the company continued to offer these tests to consumers outside the US (including the EU). Of note, 23andMe continued to offer its customers information about non-medical traits (e.g. preference for bitter taste, being a morning or night person) in the meantime since the FDA's regulatory scope only covers medical applications.

Nevertheless, questions about the effectiveness and safety of genetic tests on non-medical outcomes can also be raised. Furthermore, 23andMe customers have the possibility to download their own genetic data to inspect it themselves or to share it further, for example with websites such as dna.land, which calculates a polygenic score value for educational attainment for its customers. It is unknown to which extent customers who accessed such information were able to understand the meaning and the limitations of these findings correctly and to which extent this information may or may not have influenced their behaviour.

Another challenge, for law and policy makers, arises from the fact that individuals who submit their genetic data for testing are unintentionally also releasing genetic information about their relatives, who typically did not consent to the submission and analysis of their data. For example, this led to the identification of a serial murderer and rapist who was active in California in the 1970 and 80 three decades after he committed his last crime. The police analysed DNA from one of the crime scenes and checked it against genetic profiles from genealogical websites that collect DNA samples to help people learn about their family backgrounds, coming up with a match that pointed to a relative of the offender (Gafni, 2018).

In summary, it is advisable for policy makers and regulators to verify if these companies are complying with existing laws and regulations, whether their products are safe and effective, and to determine whether additional legislation may be necessary.

3.8 Insurance markets: moral dilemmas

Genetic data contains information about health risks, expected longevity, and to some extent also about an individual's personality and propensity to engage in risky behaviours (Karlsson Linnér et al., 2019; Pilling et al., 2017; Welter et al., 2014), all of which is

potentially relevant information for insurance companies and customers considering whether to purchase insurance or not.

Insurance markets suffer from market failure due to asymmetric information. In particular, customers who believe their chance to benefit from insurance are more likely to purchase it. If insurances cannot differentiate between high and low risk customers and they offer a policy based on the average risk, only the high risk customers would be expected to buy insurance, inducing losses for the insurance company, and ultimately making the market unsustainable. Furthermore, customers who bought insurance may become less careful or actively try to provoke an insurance payment. This moral hazard problem adds to the adverse selection issue in insurance markets (Rothschild and Stiglitz, 1976; Varian, 2010).

The availability of genetic data amplified these problems in insurance markets: If only costumers have information about their risks, this increases the adverse selection problems into insurances and the associated potential for market failure. If, on the other hand, insurances were allowed to access genetic data from potential customers, this would lead to price discrimination, making it more expensive or even impossible to get insurance for those who need it most because they were born with higher genetic risks, without any personal wrong-doing. This creates a moral dilemma for society – there is a need to balance the feasibility of insurance markets against new forms of discrimination based on genetic data. In the US, the Genetic Information Nondiscrimination Act (GINA) was signed into law in 2008 with the specific purpose to prohibit discrimination on the basis of genetic information with respect to health insurance and employment (Ajunwa, 2016).

The classic solution to resolve the market failure in insurance markets is to make insurance obligatory for everyone, to require insurances to accept any potential customers, and potentially also to limit or forbid price discrimination among different customer types. In such a market set-up, there would be limited or no insurance-relevant information in genetic data and the solidarity principle of the insurance would primarily benefit those with the highest risks. However, obligatory insurance schemes limit the freedom of individuals to decide if they want insurance or not. Obligatory insurance schemes can also impose substantial financial burden on people that may exceed their personal willingness or ability to pay. Therefore, obligatory insurance schemes typically require substantial transfer of wealth in society that may be undesirable or politically infeasible.

The practical relevance of genetic data and polygenic scores for insurance markets hinges upon whether this new source of information has predictive value that goes beyond what insurances can learn from the information that is already available to them and whether genetic information would change the behaviour of customers in the insurance market. These are interesting questions that should be fully addressed

The availability of genetic data creates new challenges for the insurance markets that need to be monitored and requires policy intervention. In particular, it is important for EU lawmakers to probe whether an anti-discrimination law similar to GINA in the US is necessary and desirable in the EU Member States.

3.9 Labour markets: danger of discrimination

Genetic data could potentially also be of interest to employers, e.g. to screen potential job candidates and to manage career trajectories. Some companies already offer genetic tests to companies for such purposes (e.g. http://www.braincompass.com/en/). However, these potential applications are highly problematic for a number of reasons. First, they substantially increase the dangers of discrimination and stigmatization against specific individuals or groups and may reinforce glass ceiling effects. Again, GINA was passed in the US with the specific purpose to prevent discrimination against employees based on genetic information. And despite GINA being passed into law in 2008, there is a trend towards increased occurrences of genetic discrimination in the US, suggesting that

genetic data is becoming an increasingly serious risk for discrimination that needs to be confronted by policy makers (Ajunwa, 2016). Second, given the dramatic failure of candidate gene studies and the relatively low predictive accuracy of polygenic scores for social scientific outcomes, genetic data currently has very little (if any) predictive accuracy at the individual level. Of note, even completely worthless genetic tests or beliefs can be enough to create grounds for discrimination. Third, genetic data contains information about health risk and many other sensitive topics that many employees would not feel comfortable sharing with their employer. Thus, the potential availability of genetic information to employers is a new threat to the right for privacy of employees that needs to be protected. Fourth, similar to the prospect of "personalised education", the claimed benefits of using genetic data with the purpose to help employees to reach their full potential are vague, not supported by scientific evidence, and may be potentially counterproductive despite good intentions.

Thus, the application of genetic data in labour markets is potentially highly problematic and it is difficult to imagine future developments that would tilt the risk-benefit ratio into a more favourable direction. Therefore, this is one potential area that policy makers should monitor closely and potentially intervene to prevent potential harm.

3.10 Data security and privacy

The widespread collection and availability of genetic data has called data security and privacy concerns into focus (Church et al., 2009; Erlich and Narayanan, 2014; Nyholt et al., 2009). The emerging debate highlights the trade-off between privacy and data protection on one hand and other important social goals such as scientific progress, greater security, better public health, or free enterprise on the other (Gafni, 2018; Kaye, 2008; Shoenbill et al., 2014). Recent inquiries have demonstrated the difficulty (or stronger even, the impossibility) to completely guarantee data security and privacy of individuals who agreed to provide genetic information for desirable research purposes (Braun et al., 2009; Couzin, 2008; Erlich and Narayanan, 2014; Greenbaum et al., 2008; Homer et al., 2008; Lunshof et al., 2008; Visscher and Hill, 2009). These challenges are even larger when individuals give companies or specific websites access to their genetic data (Kaye, 2008).

Currently, access to genetic data appears to be more tightly regulated for *bona fide* researchers than for companies, even though the work of *bona fide* researchers is much more likely to contribute to charitable goals. For example, researchers at universities and public research institutes typically have to obtain approval for their intended use of genetic data from ethical committees or institutional review boards that do not exist in the corporate world. Furthermore, researchers at public institutions are often held to much more strict data security standards than private companies. In addition, there are typically strict limitations for the scope of potential uses for publicly financed data collections that contain genetics. Data collected by genetic testing companies or other entities outside of the public sector typically have no restrictions for their use of genetic data other than the laws of the country in which they are active (Kaye 2008). This implies new responsibilities for law makers to ensure that legal frameworks take the specific challenges of genetic data into account and that good, socially acceptable compromises are found between data protection and privacy concerns on one hand and other socially desirable goals on the other.

In particular, it may be desirable to make malicious de-identification practices of genetic data and their unconsented use illegal and to specify and enforce penalties for violations.

Furthermore, it would be beneficial to raise public awareness about

- new challenges that the availability of genetic data raises for the privacy of individuals who were genotyped, as well as their relatives;
- the rights that individuals have to determine for which purposes their (genetic) data can be used;

- the potential consequences that arise from violations of the consented uses of genetic data;
- the trade-offs between data protection and privacy on the one hand and other socially desirable goals on the other (e.g. better public health, greater security).

3.11 Unequal access and participation: inequality

Currently, GWAS research is largely based on samples from rich, industrialised countries and samples of white individuals with European decent (Martin et al. 2019, Mills and Rahal 2019). As discussed above, this limits the usefulness of GWAS results for other populations. To the extent that GWAS discoveries and polygenic scores for medical as well as non-medical outcomes have valuable applications that could improve health, quality of life, and well-being, these applications are currently mostly restricted to white, European decent populations. This could increase already existing differences in health and life quality further (Martin et al., 2019).

Furthermore, genetic tests and access to genetic data is in practice often restricted to wealthy, educated individuals who actively seek this information and who have the financial means to afford it. This, too, could potentially increase existing inequalities further.

Finally, an agreed definition of what population means should be provided in order to avoid confusion and generate misunderstanding. What is the definition of "population" in a GWAS-specific context? The "white European" are often mentioned, but opposed to what? Non-white European? White non-European? Are all white European the same "population"? Population is a term that can mean many different things for many different people (including policy makers vs. the scientists that work on GWAS) and a correct definition should be discussed and properly understood to correctly frame a regulatory action.

3.12 Fairness considerations

The genetic endowment of each person is the outcome of a random draw of the genotypes of the parents. Social science genetics demonstrates clearly and irrefutably that the results of this external event influences who we are and what we do to some degree, including many traits or outcomes that are often seen as meritorious such as performing well at school or getting a desirable job. Social science genetics also shows the strong links between behavioural or socio-economic traits, health outcomes, and longevity, and how all of them can be to some extent traced back to random genetic endowments.

This fact challenges positions that argue that individuals deserve to take full advantage of all the benefits that come with higher education, higher socio-economic status, and higher income. Instead, the realization that success in life partly depends on a random draw from the genetic lottery can strengthens arguments in favour of solidarity and redistribution (Harden, 2018b).

To be clear, the results from GWAS or social science genetics do not replace the need to take a philosophical and political stand about what is fair, just, desirable, or meritorious, and what type of society we want to live in. But it is important to realise that the insights gained from social science genetics are fully compatible with agendas that aim to combat inequalities and that embrace diversity. The results from social science genetics can also be used to develop and test policies that help achieve such goals.

4 Discussion/Conclusion

We are currently in an era in which genetic data are cheap to generate, ubiquitous, and accessible not only to scientists and health practitioners, but also to citizens, private companies, and governmental institutions. In this "new era" of diagnostics, genetic data are not only used to advance medical research, but also to study all aspects of individual differences. The discussions about the potential use of these new "genomic applications" insights constantly gain momentum among scientists, but also on social media platforms, TV, radio, and printed media. (Harden, 2018a; Harden, 2018b; Plomin, 2018; Zimmer, 2018). The applications that are being discussed range from uses in research settings to better understand environmental effects or the causes and potential cures of diseases, to much more problematic applications such as genetic tests in the labour market or for health insurance. The discussion did not even stop at "designer babies": instead, the technology for gene editing and for genetic selection of embryos is already available and will rapidly be getting better. These developments raise a host of difficult and important ethical questions. They also raise entirely new challenges for policy makers and regulators who will be asked to play an important role in how these new technologies will shape our future.

In order to determine the best course of action, it will be necessary to consult and involve experts with relevant expertise from various backgrounds including geneticists, physicians, social scientists, legal scholars, economists, computer scientists, moral philosophers, and bioethicists. Furthermore, it is necessary to involve in the debate stake holders from the general public in the debate, including (but not limited to) minorities who are threatened by exclusion or discrimination, couples seeking help from reproductive technologies to get children, socially disadvantaged groups, patients facing or battling with genetically caused diseases, as well as teachers and educators.

Given the rapid technological and scientific progress in the field of genetics, now is the right time to start having an informed dialogue on how these new technologies could change citizens lives, on what type of society we want to live in in the future, and on how policy makers can shape developments into directions that will embrace positive applications and regulate undesired ones,; otherwise it will be challenging for policy makers, regulators, and the general public to keep up and to make the right decisions.

This report has emphasised the many links between medical genetics and social science genetics in terms of scientific methods, data, possible applications, the many researchers who are making active contributions across disciplinary boundaries, and the countless ways how behaviour and health are influencing each other, as well as the traces these relationships leave in genetic association studies. Thus, although social science genetics is sometimes viewed as a novel field of scientific inquiry that is clearly separated from medical genetics (Bliss, 2018; Freese, 2018) – potentially living within the intellectual silos of clearly defined departmental structures within the social science (e.g. "genoeconomics", Benjamin et al., 2012) – this report has taken the perspective that the conception of such borderlines is artificial and potentially counterproductive. Rather, medical genetics and social science genetics often share similar goals and have influenced each other with ideas and methods and struggle with similar scientific and ethical challenges. Thus, many of the policy implications of medical genetics also apply to social science genetics and vice versa.

The current priorities for policy makers are summarised as follows:

- The existing anti-discrimination laws in EU member countries should offer sufficient protection against discrimination based on genetic information, in particular in labour and insurance markets. In the US, GINA took effect in 2009 to achieve this purpose. However, GINA currently does not include a disparate impact case of action, which would not only cover intentional acts of discrimination, but also unintentional acts. It has been argued that including disparate impact in GINA would be a necessary addition to make the law effective in attempts to combat genetic discrimination (Ajunwa, 2016). This debate also

needs to take place in Europe. In particular, policy makers should consider the possibility to forbid companies to obtain genetic data from applicants, employees, and individuals who act as suppliers to companies.

- The availability of consumer genetics companies and polygenic scores for many outcomes create new challenges for insurance markets, in particular for health and life insurance. Policy action will be needed not only to prevent new forms of discrimination based on genetic profiles (see above), but also to keep these markets viable. One possible solution to these challenges would be mandatory (health) insurance schemes, but those may not be politically viable.
- Policy makers and regulators should oversee companies in the emerging gene testing industry, e.g. to verify if these companies are complying with existing laws and regulations, whether their products are reliable and trustable, and to determine whether additional legislation may be necessary. The Additional Protocol to the Convention on Human Rights and Biomedicine concerning Genetic Testing for Health Purposes (CETS No. 203) of the European Council is an important step into this direction. However, additional discussions and decisions seem necessary to determine how genetic tests for behavioural and socio-economic outcomes (e.g. educational attainment, personality, preferences) should be treated. Given that gene testing companies often operate internationally and via the Internet, there is additional urgency to harmonise potential regulatory responses throughout the EU and abroad.
- Ethically highly controversial applications of genetics are emerging trends to test and select embryos not only for monogenetic disorders, but also for genetic complex traits including cognitive performance. Although the existing technologies still have numerous limitations, it is very likely that there will be a demand for such offers. It is urgent for society at large to debate whether and under which circumstances technologies to select or genetically modify human stem cells and embryos will be legal and who will bear the financial costs. This debate should be held in the context of other assisted reproductive technologies (e.g. selecting sperm donors in sperm banks) that are already frequently practiced in many countries. Given the rapid development of science and technology, policy makers need to address this issue urgently.
- The widespread availability of genetic data implies new responsibilities for law makers to ensure that legal frameworks take the specific challenges of genetic data into account and that good, socially acceptable compromises are found between data protection and privacy concerns on the one hand and other socially desirable goals (e.g. scientific progress, better health care, lower crime rates) on the other.
- There is a need to raise public awareness about the new opportunities and challenges that genetics (including social science genetics) brings to society and to start a society-wide debate about the emerging ethical questions.
- Unequal access to genetic technologies and the limited scope of the majority of current genetic research on European populations may lead to growing inequalities in health and wealth within European countries as well as between Europe and the rest of the world. Policy makers could play an important role in combating such rises in inequality.
- The insights generated by social science genetics can be interpreted and used in many different ways. However, it is important to realise that these insights are fully compatible with political agendas that aim to combat inequalities and that embrace diversity in society.

5 References

Acemoglu D (2002) Technical change, inequality, and the labor market. *J. Econ. Lit.* 40(1):7–72.

Adler NE, Al. E (1994) Socioeconomic status and health. The challenge of the gradient. *Am. Psychol.* 49(1):15–24.

Administration FD Press Announcements - FDA authorizes first direct-to-consumer test for detecting genetic variants that may be associated with medication metabolism.

Ajunwa I (2016) Genetic data and civil rights. *Harv. Civ. Rights-Civil Lib. Law Rev.* 51.

Annas GJ, Elias S (2014) 23andMe and the FDA. *N. Engl. J. Med.* 370(11):985–988.

Anon Fairfax Cryobank - How to choose a sperm donor. Retrieved (April 10, 2019), https://fairfaxcryobank.com/how-to-choose-a-sperm-donor.

Anon Warning Letters - 23andMe, Inc. 11/22/13.

Ardlie KG, Deluca DS, Segre A V., Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, et al. (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (80-.).* 348(6235):648–660.

Asbury K, Plomin R (2013) *G is for Genes: The Impact of Genetics on Education and Achievement.* (Wiley-Blackwell).

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.

Baker L, Treloar S, Reynolds C, Heath A, Martin N (1996) Genetics of educational attainment in Australian twins: Sex differences and secular changes. *Behav. Genet.* 26(2):89–102.

Ball P (2018) Super-smart designer babies could be on offer soon. But is that ethical? *The Guardian* (November 19) https://www.theguardian.com/commentisfree/2018/nov/19/designer-babies-ethical-genetic-selection-intelligence.

Bansal V, Mitjans M, Burik CAP, Karlsson Linner R, Okbay A, Rietveld CA, Begemann M, et al. (2018) Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. *Nat. Commun.* 9(3078).

Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, Shen X, et al. (2016) Genome-wide analysis identifies 12 loci influencing human reproductive behaviour. *Nat. Genet.* 48(12):1462–1472.

Barcellos SH, Carvalho LS, Turley P (2018) Education can reduce health disparities related to genetic risk of obesity: Evidence from a British reform. *Proc. Natl. Acad. Sci.*

Bartels M, Boomsma DI (2009) Born to be happy? The etiology of subjective well-being. *Behav. Genet.* 39(6):605–615.

Beauchamp JP (2016) Genetic evidence for natural selection in humans in the contemporary United States. *Proc. Natl. Acad. Sci.* 113(28):7774–7779.

Beauchamp JP, Cesarini D, van der Loos MJHM, Koellinger PD, Groenen PJF, Fowler JH, Rosenquist JN, Thurik AR, Christakis NA (2011) Molecular Genetics and Economics. *J. Econ. Perspect.* 25(4):57–82.

Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A, Conley D, et al. (2018) Genetic analysis of social-class mobility in five longitudinal studies. *Proc. Natl. Acad. Sci. U. S. A.* 115(31):E7275–E7284.

Belsky DW, Harden KP (2019) Phenotypic annotation: Using polygenic scores to translate discoveries from genome-wide association studies from the top down. *Curr. Dir.*

Psychol. Sci. 28(1):82–90.

Benjamin DJ, Cesarini D, Chabris CF, Glaeser EL, Laibson DI, Gudnason V, Harris TB, et al. (2012) The promises and pitfalls of genoeconomics. *Annu. Rev. Econom.* 4(1):627–662.

Benjamin DJ, Cesarini D, van der Loos MJHM, Dawes CT, Koellinger PD, Magnusson PKE, Chabris CF, et al. (2012) The genetic architecture of economic and political preferences. *Proc. Natl. Acad. Sci.*

Bliss C (2018) *Social by Nature: The Promise and Peril of Sociogenomics* (Stanford University Press, Stanford, CA).

Borkenau P, Riemann R, Angleitner A, Spinath FM (2001) Genetic and environmental influences on observed personality: Evidence from the German Observational Study of Adult Twins. *J. Pers. Soc. Psychol.* 80(4):655–668.

Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* 44(2):512–25.

Braun R, Rowe W, Schaefer C, Zhang J, Buetow K (2009) Needles in the haystack: Identifying individuals present in pooled genomic data Gibson G, ed. *PLoS Genet.* 5(10):e1000668.

Brown K V. (2019) 23andMe just got FDA approval for a DNA cancer test. *Time* (January 22) http://time.com/5510009/23andme-fda-approval-cancer-test/.

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Consortium R, Genomics Consortium P, et al. (2015) An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47(11):1236–1241.

Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, et al. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47(3):291–295.

Burgess S, Butterworth A, Thompson SG (2013) Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* 37(7):658–665.

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14(5):365–376.

Cesarini D, Dawes CT, Johannesson M, Lichtenstein P, Wallace B (2009) Genetic variation in preferences for giving and risk taking. *Q. J. Econ.* 124(2):809–842.

Cesarini D, Johannesson M, Lichtenstein P, Sandewall Ör, Wallace B (2010) Genetic variation in financial decision-making. *J. Finance* 65(5):1725–1754.

Chabris CF, Hebert BM, Benjamin DJ, Beauchamp JP, Cesarini D, van der Loos M, Johannesson M, et al. (2012) Most reported genetic associations with general intelligence are probably false positives. *Psychol. Sci.* 23(11):1314–1323.

Chabris CF, Lee JJ, Benjamin DJ, Beauchamp JP, Glaeser EL, Borst G, Pinker S, Laibson DI (2013) Why it is hard to find genes associated with social science traits: theoretical and empirical considerations. *Am. J. Public Health* 103 Suppl:S152-66.

Chabris CF, Lee JJ, Cesarini D, Benjamin DJ, Laibson DI (2015) The fourth law of behaviour genetics. *Curr. Dir. Psychol. Sci.* 24(4):304–312.

Chan IS, Ginsburg GS (2011) Personalized medicine: Progress and promise. *Annu. Rev. Genomics Hum. Genet.* 12(1):217–244.

Church G, Heeney C, Hawkins N, de Vries J, Boddington P, Kaye J, Bobrow M, Weir B, Weir B (2009) Public access to genome-wide data: Five views on balancing research with privacy and protection Gibson G, ed. *PLoS Genet.* 5(10):e1000665.

Conley D, Fletcher J (2017) *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future* (Princeton University Press, Princeton, N.J.).

Conley D, Zhang S (2018) The promise of genes for understanding cause and effect. *Proc. Natl. Acad. Sci.*:201805585.

Courtiol A, Tropf FC, Mills MC (2016) When genes and environment disagree: Making sense of trends in recent human evolution. *Proc. Natl. Acad. Sci. U. S. A.* 113(28):7693–5.

Couzin J (2008) Genetic privacy. Whole-genome data not anonymous, challenging assumptions. *Science (80-.).* 321(5894):1278.

Cumming G (2011) *Understanding the New Statistics* First. (Taylor & Amp; Francis Ltd).

Darwin C (1859) *On the Origin of Species by Means of Natural Selection* (John Murray, London).

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, et al. (2016) Next-generation genotype imputation service and methods. *Nat. Genet.* 48(August).

Davey Smith G (2010) Mendelian randomization for strengthening causal inference in observational studies: Application to gene x environment interactions. *Perspect. Psychol. Sci.* 5(5):527–545.

Davey Smith G, Hemani G (2014) Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23(R1):R89-98.

Dickens WT, Flynn JR (2001) Heritability estimates versus large environmental effects: the IQ paradox resolved. *Psychol. Rev.* 108(2):346–69.

Diprete TA, Burik CAP, Koellinger PD (2018) Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data. *Proc. Natl. Acad. Sci.* 115(22):E4970–E4979.

Domingue BW, Belsky DW, Fletcher JM, Conley DC, Boardman JD, Harris KM (2018) The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. *Proc. Natl. Acad. Sci.*:201711803.

Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 9(3):e1003348.

Duncan L, Shen H, Gelaye B, Ressler K, Feldman M, Peterson R, Domingue B (2018) *Analysis of polygenic score usage and performance in diverse human populations* (Cold Spring Harbor Laboratory).

Erlich Y, Narayanan A (2014) Routes for breaching and protecting genetic privacy. *Nat. Rev. Genet.* 15(6):409–421.

Euesden J, Lewis CMCM, O'Reilly PF, O'Reilly PF (2015) PRSice: polygenic risk software. *Bioinformatics* 31(9):1466–1468.

Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, Bjelland DW, et al. (2018) Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. Genet.* 50(5):737–745.

Fisher RA (1930) *The Genetical Theory of Natural Selection* (The Clarendon Press, Oxford).

Floderus-Myrhed B, Pedersen N, Rasmuson I (1980) Assessment of heritability for personality, based on a short-form of the Eysenck personality inventory: A study of 12,898 twin pairs. *Behav. Genet.* 10(2):153–162.

Foster ED, Deardorff A (2017) Open Science Framework (OSF). *J. Med. Libr. Assoc.* 105(2):203.

Freese J (2018) The arrival of social science genomics. *Contemp. Sociol. A J. Rev.* 47(5):524–536.

Gafni M (2018) GEDmatch DNA website helped crack Golden State Killer case. *The Mercury News* (April 26) https://www.mercurynews.com/2018/04/26/ancestry-23andme-deny-assisting-law-enforcement-in-east-area-rapist-case/.

Ginsburg GS, Willard HF (2009) Genomic and personalized medicine: foundations and applications. *Transl. Res.* 154(6):277–287.

Goldberger AS (1979) Heritability. *Economica* 46(184):327–347.

Greenbaum D, Du J, Gerstein M (2008) Genomic anonymity: Have we already lost it? *Am. J. Bioeth.* 8(10):71–74.

Grotzinger AD, Rhemtulla M, Vlaming R de, Ritchie SJ, Mallard TT, Hill WD, Ip HF, et al. (2019) Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.*

Grunwald HA, Gantz VM, Poplawski G, Xu XRS, Bier E, Cooper KL (2019) Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* 566(7742):105–109.

Harden KP (2018a) Heredity is only half the story. *Spectator* https://spectator.us/heredity-dna-robert-plomin/.

Harden KP (2018b) Why progressives should embrace the genetics of education. *The New York Times* https://www.nytimes.com/2018/07/24/opinion/dna-nature-genetics-education.html.

Harden KP, Domingue BW, Belsky DW, Boardman JD, Crosnoe R, Malanchini M, Nivard M, Tucker-Drob EM, Harris KM (2019) *Genetic associations with mathematics tracking and persistence in secondary school* (Cold Spring Harbor Laboratory).

Harden KP, Koellinger PD (2019) *Social science genetics, invited contribution under consideration at Nature Human Behaviour*

Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson P V, Sigurdsson G, Thorgeirsson G, et al. (2007) Age, gene/environment susceptibility–reykjavik study: Multidisciplinary applied phenomics. *Am. J. Epidemiol.* 165(9):1076–1087.

Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, Carslake D, et al. (2019) Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.* 10(1):333.

Heath AC, Berg K, Eaves LJ, Solaas MH, Corey LA, Sundet J, Magnus P, Nance WE (1985) Education policy and the heritability of educational attainment. *Nature* 314(6013):734–736.

Herskind AM, McGue M, Iachine IA, Holm N, Sørensen TIA, Harvald B, Vaupel JW (1996) Untangling genetic influences on smoking, body mass index and longevity: a multivariate study of 2464 Danish twins followed for 28 years. *Hum. Genet.* 98(4):467–475.

Hewitt JK (2012) Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behav. Genet.* 42(1):1–2.

Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson J V., Stephan DA, Nelson SF, Craig DW (2008) Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays Visscher PM, ed. *PLoS Genet.* 4(8):e1000167.

Hughes RA, Davies NM, Davey Smith G, Tilling K (2019) Selection bias when estimating average treatment effects using one-sample instrumental variable analysis. *Epidemiology*:1.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921.

Ioannidis JPA (2005) Why most published research findings are false. *PLoS Med.* 2(8):e124.

Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. *Nat Genet.* 29(3):306-9.

Janssens ACJW (2018) Why you shouldn't worry too much about designer babies. *Discover* (December 10) http://blogs.discovermagazine.com/crux/2018/12/10/why-you-shouldnt-worry-too-much-about-designer-babies/#.XKxaD5gzaUk.

Jencks C (1980) Heredity, environment, and public policy reconsidered. *Am. Sociol. Rev.* 45(5):723–736.

Joyner MJ, Paneth N (2015) Seven questions for personalized medicine. *JAMA* 314(10):999.

Kandler C, Bleidorn W, Riemann R (2012) Left or right? Sources of political orientation: the roles of genetic factors, cultural transmission, assortative mating, and personality. *J. Pers. Soc. Psychol.* 102(3):633–45.

Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, Lebreton M, et al. (2019) Genome-wide association analyses of risk tolerance and risky behaviours in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* 51(2):245–257.

Kaye J (2008) The regulation of direct-to-consumer genetic tests. *Hum. Mol. Genet.* 17(R2):R180–R183.

van Kippersluis H, Rietveld CA (2017) Pleiotropy-robust Mendelian randomization. *Int. J. Epidemiol.*

Koellinger PD, Harden KP (2018) Using nature to understand nurture. *Science* 359(6374):386–387.

Kong A, Frigge ML, Thorleifsson G, Stefansson H, Young AI, Zink F, Jonsdottir GA, et al. (2017) Selection against variants in the genome associated with educational attainment. *Proc. Natl. Acad. Sci. U. S. A.* 114(5):E727–E732.

Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, Benonisdottir S, et al. (2018) The nature of nurture: Effects of parental genotypes. *Science* 359(6374):424–428.

Kwiatkowski DP (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.* 77(2):171–192.

Lander ES (2011) Initial impact of the sequencing of the human genome. *Nature* 470(7333):187–97.

Lee J, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, et al. (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* 50:1112–1121.

Lin X, Lim IY, Wu Y, Teh AL, Chen L, Aris IM, Soh SE, et al. (2017) Developmental pathways to adiposity begin before birth and are influenced by genotype, prenatal environment and epigenome. *BMC Med.* 15(1):50.

Linnér RK, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, Lebreton M, et al. (2019) Genome-wide association analyses of risk tolerance and risky behaviours in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.*:261081.

Ljungquist B, Berg S, Lanke J, McClearn GE, Pedersen NL (1998) The effect of genetic factors for longevity: A comparison of identical and fraternal twins in the Swedish

Twin Registry. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* 53A(6):M441–M446.

Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL (2018) Mixed-model association for biobank-scale datasets. *Nat. Genet.* 50:906–908.

Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, et al. (2015) Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* 47(3):284–290.

Looi ML, Zakaria H, Osman J, Jamal R (2012) Quantity and quality assessment of DNA extracted from saliva and blood. *Clin. Lab.* 58(3–4):307–12.

van der Loos MJHM, Koellinger P, Groenen PJ, Rietveld CA, Rivadeneira F, van Rooij F van, Uitterlinden A, Hofman A, Thurik R (2011) Candidate gene studies and the quest for the entrepreneurial gene. *Small Bus. Econ.* 37(3):269–275.

van der Loos MJHM, Rietveld CA, Eklund N, Koellinger PD, Rivadeneira F, Abecasis GR, Ankra-Badu GA, et al. (2013) The molecular genetic architecture of self-employment. Cherny S, ed. *PLoS One* 8(4):e60542.

Lunshof JE, Chadwick R, Vorhaus DB, Church GM (2008) From genetic privacy to open consent. *Nat. Rev. Genet.* 9(5):406–411.

Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11(7):499–511.

Marchini JL, Cardon LR, Phillips MS, Donnelly P (2004) The effects of human population structure on large genetic association studies. *Nat. Genet.* 36(5):512–7.

Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM (2018) A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 27(2):e1608.

Marigorta UM, Rodríguez JA, Gibson G, Navarro A (2018) Replicability and prediction: Lessons and challenges from GWAS. *Trends Genet.* 34(7):504–517.

Marigorta UM, Rodriguez JA, Navarro A (2016) GWAS replicability across time and space. Appasani K, ed. *Genome-Wide Assoc. Stud. From Polymorph. to Pers. Med.* (Cambridge University Press, Cambridge, UK), 53–66.

Marini MM (1984) Women's educational attainment and the timing of entry into parenthood. *Am. Sociol. Rev.* 49(4):491.

Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2019) *Current clinical use of polygenic scores will risk exacerbating health disparities* (Cold Spring Harbor Laboratory).

Mastenbroek S, Twisk M, van Echten-Arends J, Sikkema-Raddatz B, Korevaar JC, Verhoeve HR, Vogel NEA, et al. (2007) In vitro fertilization with preimplantation genetic screening. *N. Engl. J. Med.* 357(1):9–17.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, et al. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528(7583):499–503.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN (2008) Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9(5):356–369.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, et al. (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48(10):1279–1283.

Middeldorp CM, de Moor MHM, McGrath LM, Gordon SD, Blackwood DH, Costa PT, Terracciano A, et al. (2011) The genetic association between personality and major depression or bipolar disorder: A polygenic score analysis using genome-wide

association data. *Transl. Psychiatry* 1:e50.

Mills M, Rindfuss RR, McDonald P, te Velde E (2011) Why do people postpone parenthood? Reasons and social policy incentives. *Hum. Reprod. Update* 17(6):848–860.

Mills MC, Rahal C (2019) A scientometric review of genome-wide association studies. *Commun. Biol.* 2(1):9.

Moore TJ, Zhang H, Anderson G, Alexander GC (2018) Estimated costs of pivotal trials for novel therapeutic agents approved by the US Food and Drug Administration, 2015-2016. *JAMA Intern. Med.* 178(11):1451.

Neale B, Abbot L, Bryant S, Churchhouse C, Ganna A, Howrigan DP, Palmer DS, Walters RK, Carey C, Team TH (2018) GWAS of 4,203 traits in the UK Biobank. Retrieved (August 20, 2018), http://www.nealelab.is/uk-biobank/.

Nes RB, Røysamb E, Tambs K, Harris JR, Reichborn-Kjennerud T (2006) Subjective well-being: genetic and environmental contributions to stability and change. *Psychol. Med.* 36(7):1033–1042.

Nicolaou N, Shane S (2010) Entrepreneurship and occupational choice: Genetic and environmental influences. *J. Econ. Behav. Organ.* 76(1):3–14.

Nicolaou N, Shane S, Cherkas L, Hunkin J, Spector TD (2008) Is the tendency to engage in entrepreneurship genetic? *Manage. Sci.* 54(1):167–179.

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, et al. (2015) Promoting an open research culture. *Science (80-.).* 6(6242):1422–5.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, et al. (2008) Genes mirror geography within Europe. *Nature* 456(7218):98–101.

Nyholt DR, Yu CE, Visscher PM (2009) On Jim Watson's APOE status: genetic information is hard to hide. *Eur. J. Hum. Genet.* 17(2):147–149.

Okbay A, Baselmans BML, De Neve JE, Turley P, Nivard MG, Fontana MA, Meddens SFW, et al. (2016) Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* 48(6):624–633.

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, et al. (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533:539–542.

Okbay A, Rietveld CA (2015) On improving the credibility of candidate gene studies: A review of candidate gene studies published in Emotion. *Emotion* 15(4):531–537.

Oniszczenko W, Jakubowska U (2005) Genetic determinants and personality correlates of sociopolitical attitudes in a Polish sample. *Twin Res. Hum. Genet.* 8(1):47–52.

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, et al. (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 19(5):826–37.

Pilling LC, Kuo CL, Sicinski K, Tamosauskaite J, Kuchel GA, Harries LW, Herd P, Wallace R, Ferrucci L, Melzer D (2017) Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany. NY).* 9(12):2504–2520.

Piotrowska PJ, Stride CB, Croft SE, Rowe R (2015) Socioeconomic status and antisocial behaviour among children and adolescents: A systematic review and meta-analysis. *Clin. Psychol. Rev.* 35:47–55.

Plomin R (2018) *Blueprint: How DNA Makes Us Who We Are* (Allen Lane, Pinguin Random House, London et al.).

Plomin R, Stumm S von (2018) The new genetics of intelligence. *Nat. Rev. Genet.*

19:148–159.

Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 47(7):702–709.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38(8):904–9.

Ramsköld D, Wang ET, Burge CB, Sandberg R (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* 5(12):e1000598.

Reich D (2018) *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past* (Oxford University Press, Oxford, UK).

Rietveld CA, Conley D, Eriksson N, Esko T, Medland SE, Vinkhuyzen AAE, Yang J, et al. (2014) Replicability and Robustness of Genome-Wide-Association Studies for Behavioural Traits. *Psychol. Sci.* 25(11):1975–1986.

Rietveld CA, Esko T, Davies G, Pers TH, Turley P, Benyamin B, Chabris CF, et al. (2014) Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. U. S. A.* 111(38):13790–13794.

Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NG, Westra HJ, et al. (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340(6139):1467–1471.

Rodgers JL, Kohler HP, Kyvik KO, Christensen K (2001) Behaviour genetic modeling of human fertility: Findings from a contemporary Danish twin study. *Demography* 38(1):29–42.

Rose N (2013) Personalized medicine: Promises, problems and perils of a new paradigm for healthcare. *Procedia - Soc. Behav. Sci.* 77:341–352.

Rosenberg NA, Edge MD, Pritchard JK, Feldman MW (2019) Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evol. Med. Public Heal.* 2019(1):26–34.

Rothschild M, Stiglitz J (1976) Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Q. J. Econ.* 90(4):629.

Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, Nagel M, et al. (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* 50(7):912–919.

Shoenbill K, Fost N, Tachinardi U, Mendonca EA (2014) Genetic data and electronic health records: a discussion of ethical, logistical and technological considerations. *J. Am. Med. Informatics Assoc.* 21(1):171–180.

Sokolowski HM, Ansari D (2018) Understanding the effects of education through the lens of biology. *npj Sci. Learn.* 3(1):17.

SSGAC (2018) Frequently Asked Questions about "Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment." Retrieved https://www.thessgac.org/faqs.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 100(16):9440-9445.

Stranger BE, Brigham LE, Hasz R, Hunter M, Johns C, Johnson M, Kopen G, et al. (2017) Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat. Genet.* 49(12):1664–1670.

Stringhini S, Carmeli C, Jokela M, Avendaño M, Muennig P, Guida F, Ricceri F, et al.

(2017) Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1·7 million men and women. *Lancet* 389(10075):1229–1237.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, et al. (2015) An integrated map of structural variation in 2,504 human genomes. *Nature* 526(7571):75–81.

Taubman P (1976a) Earnings, education, genetics, and environment. *J. Hum. Resour.* 11(4):447–461.

Taubman P (1976b) The determinants of earnings: Genetics, family, and other environments: A study of white male twins. *Am. Econ. Rev.* 66(5):858–870.

The 1000 Genomes Project Consortium, Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39(1):31–40.

Trahan LH, Stuebing KK, Fletcher JM, Hiscock M (2014) The Flynn effect: A meta-analysis. *Psychol. Bull.* 140(5):1332–1360.

Tropf FC, Stulp G, Barban N, Visscher PM, Yang J, Snieder H, Mills MC (2015) Human fertility, molecular genetics, and natural selection in modern societies Perry JR., ed. *PLoS One* 10(6):e0126821.

Turkheimer E (2000) Three laws of behaviour genetics and what they mean. *Curr. Dir. Psychol. Sci.* 9(5):160–164.

Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, et al. (2018) Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.*

Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, et al. (2018) Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 50(2):229–237.

Turnwald BP, Goyer JP, Boles DZ, Silder A, Delp SL, Crum AJ (2019) Learning one's genetic risk changes physiology independent of actual genetic risk. *Nat. Hum. Behav.* 3(1):48–56.

Tyrrell J, Jones SE, Beaumont R, Astley CM, Lovell R, Yaghootkar H, Tuke M, et al. (2016) Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *BMJ* 352:i582.

Varian HR (2010) Asymmetric information. *Intermed. Microeconomics - A Mod. Approach*. (W.W. Norton & Company, New York & London), 718–739.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GC, Smith HO, Yandell M, Al E (2001) The sequence of the human genome. *Science (80-.).* 291(5507):1304–1351.

Verbanck M, Chen CY, Neale B, Do R (2018) Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50(5):693–698.

Vilhjálmsson BJ, Jian Yang HKF, Alexander Gusev SL, Stephan Ripke GG, Po-Ru Loh, Gaurav Bhatia R Do, Tristan Hayeck HHW, Consortium SWG of the PG, et al. (2015) Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* 97(4):576–592.

Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. *Am. J. Hum. Genet.* 90(1):7–24.

Visscher PM, Hill WG (2009) The limits of individual identification from sample allele frequencies: Theory and statistical analysis. Gibson G, ed. *PLoS Genet.* 5(10):e1000628.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* 101(1):5–22.

de Vlaming R, Okbay A, Rietveld CA, Johannesson M, Magnusson PKE, Uitterlinden AG, van Rooij FJA, et al. (2017) Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies Marchini J, ed. *PLoS Genet.* 13(1):e1006495.

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome Hurst L, ed. *PLoS Biol.* 4(3):e72.

Wainschtein P, Jain DP, Yengo L, Zheng Z, TOPMed Anthropometry Working Group TO for PMC, Cupples LA, Shadyab AH, et al. (2019) *Recovery of trait heritability from whole genome sequence data* (Cold Spring Harbor Laboratory).

Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JRB, et al. (2015) The UK10K project identifies rare variants in health and disease. *Nature* 526(7571):82–90.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42(Database issue):D1001-1006.

Wilby P (2014) Psychologist on a mission to give every child a Learning Chip. *The Guardian* (February 18) https://www.theguardian.com/education/2014/feb/18/psychologist-robert-plomin-says-genes-crucial-education.

Wilkinson RG, Marmot M (2003) *Social Determinants of Health: The Solid Facts* (World Health Organization).

Wilson C (2018) A new test can predict IVF embryos' risk of having a low IQ. *New Scientist* (November 15) https://www.newscientist.com/article/mg24032041-900-exclusive-a-new-test-can-predict-ivf-embryos-risk-of-having-a-low-iq/.

Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, Ferreira T, et al. (2014) Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* 9(5):1192–1212.

Witte JS, Visscher PM, Wray NR (2014) The contribution of genetic variants to disease depends on the ruler. *Nat. Rev. Genet.* 15(11):765–776.

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46(11):1173–1186.

Wooldridge JM (2002) *Econometric Analysis of Cross Section and Panel Data* (Massachusetts Institute of Technology, Cambridge, MA).

Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM (2014) Research Review: Polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry Allied Discip.* 55(10):1068–1087.

Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, et al. (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* 50(5):668–681.

Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, Robinson MR, et al. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* 47(10):1114–20.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42(7):565–569.

Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46(2):100–6.

Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM (2017) Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* 49(9):1304–1310.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, et al. (2010) Sequencing of fifty human exomes reveals adaptation to high altitude. *Science (80-.).* 329(5987):75–78.

Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, Robinson MR, et al. (2018) Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* 9(224).

Zimmer C (2018) *She Has Her Mother's Laugh: The Powers, Perversions, and Potential of Heredity* (Penguin Random House, New York, NY).

List of abbreviations and definitions

DNA	DeoxyriboNucleic Acid
DZ	Dizygotic
FDA	Food and Drug Administration
GINA	Genetic Information Nondiscrimination Act
GWAS	Genome Wide Association Studies
JRC	European Commission Joint Research Centre
IQ	Intelligence Quotient
IVF	In-Vitro Fertilization
LCT	Lactose
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MZ	Monozygotic
PGS	Polygenic Scores
PKU	Phenylketonuria
SNP	Single-nucleotide polymorphism

Annexes

Annex 1. Deepening on narrow-sense heritability of a trait

Twin studies use the naturally occurring experiment of monozygotic and dizygotic twins. Monozygotic (MZ) twins are (almost) genetically identical, while dizygotic (DZ) twins share on average 50% of their genetic make-up (i.e. they are genetically not more similar to each other than all full siblings that had the same parents). Under some assumptions (e.g. no assortative mating among parents, the only difference between MZ and DZ twin pairs is their genetic similarity, all genetic effects are additive and linear), it is possible to interpret stronger phenotypic similarity among MZ twin pairs compared to DZ twin pairs as evidence for a genetic influence on the trait. In fact, with this type of data, it is possible to decompose the observed differences among MZ and DZ twins into three variance components: genetic effects (h^2_{twin} – referred to as common, narrow-sense heritability that captures only linear genetic effects on the phenotype), shared environmental effects (c^2_{twin} e.g. family-specific environments), and individual environmental effects (e^2_{twin}). The latter is usually simply the residual that cannot be attributed towards h^2_{twin} and c^2_{twin}. In contrast to the narrow-sense heritability of a trait, estimates of the broad-sense heritability would also include non-linear effects of genes (e.g. dominance or gene-gene interactions). Estimating broad-sense heritability is more difficult in practice and requires additional sources of information. However, in practice, the observed empirical patterns are typically close to what is expected if the simple linear model underlying narrow-sense heritability estimates would be true (Polderman et al. 2015).

A large meta-analysis of all published twin studies between 1958 and 2012 rejected the null hypothesis of no genetic influences for all traits they examined, including behavioural and environmental traits, cognitive performance, social values and social interactions (Polderman et al. 2015). The average heritability across all traits was found to be 49%, and some behavioural outcomes were found to have a heritability that was higher than the average (e.g. intellectual functions $h^2_{twin} \approx 0.87$).

In particular, GWAS estimate the following structural model:

Equation 1: $y_i = \mu + \sum_{j=1}^{J} \beta_j x_{ij} + \varepsilon_i$

where y_i is the outcome of interest, μ is an intercept, β_j is the effect of SNP j, x_{ij} are all relevant nucleotide polymorphisms, ε_i is the effect of exogenous residual factors, $j \in [1, \dots, J]$, J > 1,000,000, and $i \in [1, \dots, N]$.

The main regressors of interest in GWAS are, of course, the genetic variants (x_{ij}), which have specific properties: Humans inherit two copies of each of 23 chromosomes from their parents -- one from their mother and one from their father. Because the base-pairs on the two strands of the DNA double helix are complementary to each other such that **A**denine always bonds with **T**hymine and **G**uanine always pairs with **C**ytosine, it is sufficient to observe only the base pairs on one of the strands to infer what the base pair on the opposite strand must look like. Thus, the values of all SNPs x_{ij} in GWAS are simply recorded as 0, 1, or 2, reflecting whether individual *i* carry 2 copies of the reference allele (x_{ij} = 0), 2 copies of the alternative allele (x_{ij} = 2), or 1 copy of each allele (x_{ij} = 1).

If the true value of β_j in equation 1 would be known, it would tell us the treatment effect from a hypothetical thought experiment in which SNP j (and nothing else) is changed at conception of individual *i.* In other words, if $\beta_j \neq 0$, we would know that SNP j has a causal effect on the outcome y_i, even if the exact pathway through which this effect works is unknown. In practice, however, the true effect of β_j is unknown. The best we can hope to achieve is to obtain a precise estimate of β_j (referred to as $\widehat{\beta_j}$), where precision refers to the standard error of the estimated coefficient $\widehat{\beta_j}$.

Standard linear regression yields unbiased estimates of $\widehat{\beta_j}$ under the following first five assumptions:

1. Sufficient degrees of freedom: $N > J$.
2. Linearity: Effect of SNP x_{ij} is linear in the number of minor alleles.
3. Random sample: We have a random sample of size N, following the population model (equation 1).
4. Non-collinearity: In the sample, none of the x_{ij} is constant, and there are no exact linear relationships among the x_{ij}.
5. Exogeneous error: The error ε_i has an expected value of zero given any value of x_{ij}, $E(\varepsilon_i|x_{ij}) = 0$.
6. Homoscedasitic: The error has the same variance given any values of x_{ij}, $Var(\varepsilon_i|x_{ij}) = \sigma^2$

If in addition to the first five assumptions the sixth one is also satisfied, the estimated standard error of $\widehat{\beta_j}$ will also be unbiased (Wooldridge 2002). The assumptions of logistic regression are very similar and will not be explicitly discussed here. In the context of GWAS, there are a number of issues that arise from the assumptions above.

First, the error term ε_i is often thought of as the environmental effect in equation 1. However, this is misleading because often the effect of x_{ij} on y_i may operate through environmental channels such as the self-selection of individuals with specific genotypes into specific environments (e.g. basketball practice) that affect the outcome (e.g. basketball performance). Thus, it is better to think of ε_i as component of environmental factors that are not endogenous to genetic endowment.

Assumption 1 (sufficient degrees of freedom) will almost always be violated in a GWAS: High-density scans of genomes provide millions of SNPs per individual, which typically implies that more SNPs than individuals are observed in any given dataset (i.e. $J > N$). As a result, the rank condition of OLS fails and β_j cannot be estimated. GWAS deal with this problem in a naïve way that simply estimates one regression for each SNP j separately. Because J is very large, special software packages (e.g. PLINK) are used to automate this process. However, SNPs that are physically located close to each other are often correlated. Thus, the naïve way of regressing one SNP at a time also means that assumption 5 is violated if SNP j is non-causal but correlated with a causal SNP j. As a results, GWAS results do not tell us which SNP is causal, and which SNPs are only associated with the trait because they are correlated with another (possibly unobserved) causal genetic variant.

It may also be possible that assumption 2 (linearity) is violated. For example, this would be the case if one allele has a dominant effect or if the causal allele is recessive. If this is the case, linear regression would still identify the relevant causal x_{ij}, which would be a useful starting point for investigating genetic dominance or recessiveness. Furthermore, the genotype x_{ij} may interact with factors in the unobserved error term ε_i in affecting the outcome. Examples would be gene×environment interactions (G×E) or gene×gene interactions (G×G). In these cases, linear regression will estimate the weighted average effect of x_{ij}. Thus, if GWAS point to a particular SNP (i.e. $\widehat{\beta_j} \neq 0$), this is still a useful starting point for investigating G×E and G×G effects of this SNP.

Assumption 3 (random sampling) is also often violated in GWAS studies. For example, many datasets that contain genetics have been created to study specific disease outcomes and have a case-control design or the collected data are not representative for the entire population (e.g. large genotyped datasets such as the UK Biobank or 23andMe over represent individuals of high socio-economic status). In such cases, the estimated $\widehat{\beta_j}$ does not generalise to the overall population, even though it may be an unbiased estimate for the parts of the population being studied. For case-control cohorts, a particular challenge is that an association between case-control status and y_i can lead to biased estimates of $\widehat{\beta_j}$. To address this concern, it is possible to analyze cases and control separately and then meta-analyze their GWAS summary statistics or to focus the analysis on the controls only.

Another issue related to non-random sampling (assumption 3) arises if the participants in a particular study are family members or distant relatives. In this case,

linear regression will still estimated $\hat{\beta}_j$ are still unbiased, but the standard errors of $\hat{\beta}_j$ will be attenuated, which would increase the number of false positive discoveries. The standard solution to this challenge is to estimate the degrees of genetic relatedness among the individuals in a dataset directly from the genetic data and to adjust the estimated standard errors in the regression accordingly, for example using clustered standard errors or linear mixed models (Loh et al. 2015, 2018, Yang et al. 2014).

Assumption 4 (non-collinearity) is typically not a problem in GWAS since a separate regression is estimated for each SNP.

However, violations of assumption (5 – exogenous ε_i) are a very serious threat to GWAS and a major area of methodological development in statistical genetics. The basic problem is that if SNP x_{ij} is correlated with unobserved environmental factors, the estimated β_j will be biased such that an environmental cause will be erroneously attributed to genetics. The technical term for this problem that is most commonly used in the literature is population stratification: Population stratification refers to differences in minor allele frequencies (MAF) and correlations between SNPs for populations with different ancestry backgrounds (e.g. Asian vs. European). Because differences in ancestry background often also correlate with geography, culture, economic development, and many other environmental factors that vary across regions, it is important to control for differences in ancestry in GWAS to avoid finding spurious, environmentally induced correlations between SNPs and outcomes. In practice, GWAS typically deal with this challenge by (i) restricting their analysis to one ancestry group (typically individuals of European decent) and (ii) adding the first few principal components from the genetic data as control variables to the GWAS (Price et al. 2006). This latter approach tends to capture major differences in ancestry in a particular sample pretty well. Furthermore, even more sophisticated statistical approaches such as linear mixed models have recently been developed to tackle this challenge more comprehensively (Loh et al. 2015, Yang et al. 2014).

Annex 2. Out-of-sample predictive accuracy of a PGS

Let's take as an example the out-of-sample predictive accuracy of a PGS for educational attainment as a function of sample size. While the results of the first well-powered GWAS on educational attainment from 2013 ($N \approx 100{,}000$) yielded a PGS that captured roughly 2% of the variation in years of schooling, the most recent results from 2018 ($N \approx 1{,}000{,}000$) yield PGS that capture up to 13% of variance. The R^2 of this most recent generation of GWAS on educational attainment mirrors the effect size seen for traditional social science variables, such as the relationship between family income and educational attainment. Furthermore, this PGS remains predictive even after regressions control for traditional social science variables (Lee et al. 2018) as well as in studies that compare educational outcomes of siblings and DZ twins (Domingue et al. 2018). This suggests that current PGS for EA capture at least to some extent biological effects that contribute to different schooling outcomes and this new type of information is not just redundant with already known factors. As a result, PGS are becoming an increasingly useful tool for social scientists (Bansal et al. 2018, Belsky and Harden 2019, Diprete et al. 2018).

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: http://europea.eu/contact

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: http://europa.eu/contact

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU publications
You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact).

KJ-NA-29815-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/948992

ISBN 978-92-76-09000-7