The personal and clinical utility of polygenic risk scores

Article *in* Nature Reviews Genetics · May 2018

DOI: 10.1038/s41576-018-0018-x

CITATIONS
93

READS
1,830

3 authors:

Ali Torkamani
The Scripps Research Institute
199 PUBLICATIONS **4,832** CITATIONS
SEE PROFILE

Nathan E Wineinger
Scripps Health
58 PUBLICATIONS **643** CITATIONS
SEE PROFILE

Eric J Topol
The Scripps Research Institute
947 PUBLICATIONS **63,456** CITATIONS
SEE PROFILE

TRANSLATIONAL GENETICS

The personal and clinical utility of polygenic risk scores

Ali Torkamani[1,2]*, *Nathan E. Wineinger*[1,2] *and Eric J. Topol*[1,3]

Abstract | Initial expectations for genome-wide association studies were high, as such studies promised to rapidly transform personalized medicine with individualized disease risk predictions, prevention strategies and treatments. Early findings, however, revealed a more complex genetic architecture than was anticipated for most common diseases — complexity that seemed to limit the immediate utility of these findings. As a result, the practice of utilizing the DNA of an individual to predict disease has been judged to provide little to no useful information. Nevertheless, recent efforts have begun to demonstrate the utility of polygenic risk profiling to identify groups of individuals who could benefit from the knowledge of their probabilistic susceptibility to disease. In this context, we review the evidence supporting the personal and clinical utility of polygenic risk profiling.

Polygenic risk scores
(PRSs). A weighted sum of the number of risk alleles carried by an individual, where the risk alleles and their weights are defined by the loci and their measured effects as detected by genome wide association studies.

Genetic architecture
The underlying genetic basis of a trait or disease. The combination of the number, type, frequency, relationship between and magnitude of effect of genetic variants contributing to a trait.

[1]*The Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, CA, USA.*

[2]*Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA.*

[3]*Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA.*

**e-mail: atorkama@scripps.edu*

https://doi.org/10.1038/s41576-018-0018-x

Estimating the probabilistic susceptibility of an individual to disease — risk prediction — is central to clinical decision-making, especially in the context of early disease detection and prevention of common adult-onset conditions. Moreover, it can be a powerful tool for personal health management when communicated and understood effectively. Today, clinical risk prediction for common adult-onset diseases often relies on basic demographic characteristics, such as age, gender and ethnicity; basic health parameters and lifestyle factors, such as body mass index, smoking status, alcohol consumption and physical exercise habits; measurement of clinical risk factors proximal to overt disease onset, such as blood pressure levels, blood chemistries or biomarkers indicative of ongoing disease processes; ascertainment of environmental exposures, such as air pollution, heavy metals and other environmental toxins; and family history. Routine genetic profiling is conspicuously absent from this list, often relegated to use only when testing clarifies individual-level risks in the context of a known family history for some common adult-onset diseases.

Early disease detection, prevention and intervention are fundamental goals for advancing human health. Meanwhile, genetic risk estimation is, for all intents and purposes, the earliest measurable contributor to common heritable disease risk. Thus, in theory, genetic profiling could be considered a useful component of health management. Indeed, recent studies suggest that, for a subset of diseases, our knowledge of the genetic factors underlying these conditions has improved to a point where polygenic risk profiling on the basis of calculated polygenic risk scores (PRSs) provides personal and clinical utility.

Here, we review the utility of genetic risk profiling for common adult-onset polygenic conditions, focusing on the leading heritable causes of death in the developed world: Alzheimer disease, cancer (breast and prostate), coronary artery disease and type 2 diabetes mellitus. For these conditions, recent studies have linked polygenic risk prediction to actionable outcomes, including the prioritization of preventive interventions and screening[1–3], prediction of age of disease onset[4], benefit from lifestyle modifications[2,5] and modification of familial disease risk leading to changes in clinical decision-making[6–8]. We begin with an overview of the genetic architecture of common adult-onset diseases. We then describe how genetic risk factors can be combined to produce PRSs and review recent studies that have demonstrated the utility of PRSs for disease risk stratification as well as their implications for early disease detection, prevention, therapeutic intervention and/or life planning. We describe some of the limitations of PRSs and the remaining barriers to clinical and personal utility and lay out potential future directions for the enhancement of the predictive capacity, generalizability and utility of PRSs.

Genetic inheritance of common diseases

The basic components of disease risk are usually broken down into genetic susceptibility, environmental exposures and lifestyle factors. The relative contribution of genetic susceptibility to the predisposition to disease in a population can be quantified by the heritability of the disease in that population. Heritability itself can be defined in several ways[9]; from a quantitative genetics perspective — especially as it relates to missing heritability in genome-wide association studies (GWAS)[10] — it is usually

defined as the proportion of phenotypic variation in a population that can be explained by genetic variation. This definition of heritability is often erroneously interpreted to describe how much genetic factors contribute to disease occurrence in any single individual.

Heritability explained in a population versus individual disease risk. Although heritability is related to the theoretical limit of genetic risk stratification at a population level[11], it does not directly relate to the utility of genetic information for an individual[12]. To illustrate this distinction further, consider *BRCA1* and *BRCA2* testing as an accepted example of the utility of genetic risk information. The prevalence of *BRCA1* and *BRCA2* mutations at birth is <<1% in outbred populations. Thus, at a population level, both the heritability explained by and the total incidence of breast cancer attributable to *BRCA1* and *BRCA2* mutations are low, accounting for approximately 5% of all breast cancer cases[13,14]. Yet, pathogenic *BRCA1* and *BRCA2* mutations confer an estimated 65% and 45% absolute lifetime risk of developing breast cancer, respectively, in contrast to the 12% absolute lifetime risk in the general population[15]. In other words, although the total heritability explained by *BRCA1* and *BRCA2* variants is low, *BRCA1* and *BRCA2* testing can identify a subset of individuals whose absolute risk of disease is significantly higher than that of the average individual in the general population. These high-risk individuals could benefit from a tailored health management strategy, which may include intensive screening or more invasive interventions. This distinction between heritability explained in a population and risk conveyed to an individual applies equivalently to large numbers of genetic variants, which cumulatively may not explain a substantial portion of heritability but may convey clinically meaningful risk to those individuals whose genomes are enriched in risk alleles, especially when considered in combination with other clinical risk factors.

For some common adult-onset diseases, the polygenic risk conveyed to a substantial segment (10–20%) of the population whose genomes are enriched in risk alleles is comparable to the risk conveyed by commonly used clinical risk factors (FIG. 1). The relationship between the genetic variants carried by an individual and the absolute risk of disease conveyed to that individual is governed by the underlying genetic architecture of the disease of interest, including the number and frequency of genetic variants influencing disease risk, the magnitude of the influence of each genetic variant on disease predisposition and the prevalence of the disease in the general population[16,17].

Genetic architecture of common diseases. The underlying genetic architecture for most common adult-onset diseases has not been fully characterized, although recent large-scale genetic studies and advanced analytical techniques can indicate the most plausible architectures[16]. Genetic architecture is often categorized as monogenic versus polygenic, meaning that one or many gene perturbations contribute to the occurrence of disease in an individual, respectively[18]. For common adult-onset diseases, this dichotomous classification is a historical artefact derived from the available technology and study designs most suited to detect rare high-risk (monogenic; via family-based linkage analysis) versus common low-risk (polygenic; via GWAS) genetic risk loci.

In reality, the genetic architecture of common adult-onset diseases is likely a continuum of common low-risk to rare high-risk genetic variants that can act cumulatively to drive overall risk in any single individual[19]. Regardless, if we consider the contribution of rare (minor allele frequency (MAF) <0.5%) high-risk genetic variants separately, these variants account for approximately 1–10% of disease incidence and sometimes result in the familial aggregation of disorders, including familial hypercholesterolaemia for coronary artery disease[20], DNA repair deficiencies for cancer[21], amyloid precursor protein processing defects in Alzheimer disease[22] and maturity-onset diabetes of the young in type 2 diabetes mellitus[23]. As in the *BRCA1* and *BRCA2* example, genetic testing for high-risk variants underlying these disorders, often informed by family history, can identify a subset of very high-risk individuals who could benefit from acting on their genetic risk status if relevant interventions are available. For the vast majority of the population without evidence of familial disease (nonfamilial), the presence of a high-risk variant may be masked owing to inaccurate family history, de novo mutation, small family size or sex-specific inheritance; however, often the relevant source of genetic risk is derived from more complex components of genetic architecture.

Polygenic disease susceptibility. The available evidence suggests that the vast majority of the remaining heritability of many common adult-onset diseases is mediated by numerous common (MAF >5%) and low-frequency (MAF >0.5% and <5%) genetic variants that individually contribute small effects, most of which can be captured through genome-wide genotyping and/or imputation[24–31]. It should be noted that statistical modelling of heritability is an intensely and actively debated topic that cannot be fully addressed in this Review; we refer readers to the emerging literature[29,32–35]. Heritability estimates reported in this Review relate to the largest assessed European ancestry populations and should be interpreted with caution. Nevertheless, the results from recent large-scale GWAS (>100,000 individuals) and sequencing efforts for many common adult-onset diseases continue to expand our knowledge of the number of genetic loci associated with disease in a manner that is consistent with heritability models that suggest an infinitesimal[36] or even omnigenic[37] model of inheritance. In this model, the nonfamilial risk of disease is driven by a substantial number of common genetic variants with small additive effect sizes (approximate ranges for effect sizes (odds ratio) are defined as small: 1.0–1.5, moderate: >1.5, and intermediate: >3.0) in combination with a relatively smaller contribution from rare variants of moderate effect size in genes known to be associated with familial disease.

For example, a recent large-scale comprehensive (including imputation) GWAS of coronary artery disease found that the majority of disease heritability is likely explained by common variants with small effect

Heritability
The proportion of total variation between individuals within a population that is due to genetic factors.

Genome-wide association studies
(GWAS). A genetic study designed to rapidly scan for statistical links between a genome-wide set of known genetic variants and a disease or other phenotype of interest.

Alleles
One of two or more alternative forms of a genetic variation.

Absolute risk
Absolute risk is the unqualified probability, or risk, that a certain event will occur; it ranges from 0–100%.

Monogenic
A term used to describe diseases with one contributing gene, that is, familial risk is driven by high-risk variants, which is in contrast to polygenic disease, where several genetic factors contribute to the disease.

Minor allele frequency
(MAF). The frequency at which the second most frequent allele occurs in a population.

Imputation
A technique for the inference of unobserved genotypes based on their statistical relationship with observed genotypes.

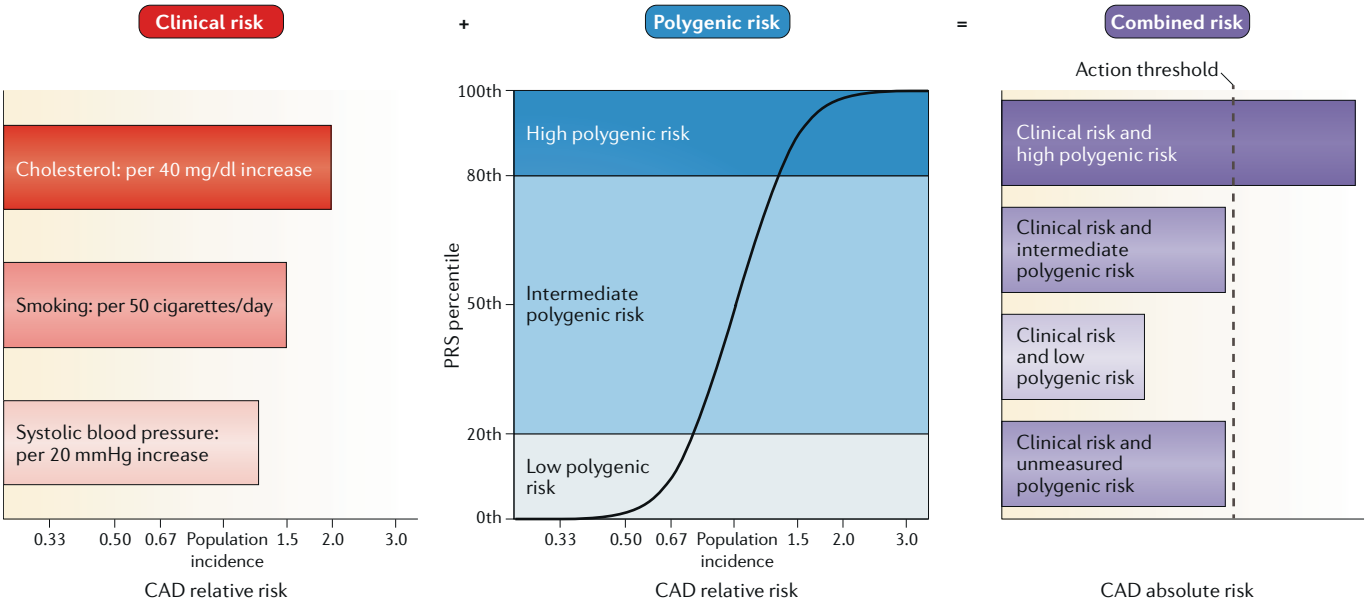

Fig. 1 | **Contrasting and combining clinical risk factors and polygenic risk.** The relative risk conveyed to individuals via commonly measured clinical risk factors (left panel) and polygenic risk estimation (middle panel) for coronary artery disease (CAD) is comparable and, when combined, can lead to different action recommendations (right panel). Relative risks for commonly measured clinical risk factors (left panel) can vary across populations and are approximated here. For polygenic risk (middle panel), the black sigmoidal curve represents the estimated CAD risk relative to average polygenic risk (at population incidence) based on 74 genome-wide association study (GWAS)-significant single nucleotide polymorphisms (SNPs)[38,39]. The bars represent the percentile thresholds typically used to define low (<20th percentile), medium (20th–80th percentile) and high (>80th percentile) polygenic risk. The combination of clinical and polygenic risk estimates (right panel) can lead to combined risk estimates that exceed the appropriate thresholds of risk versus benefit that justify certain medical interventions (action threshold). In this example, an individual with estimated clinical risk near the action threshold in the absence of polygenic risk information (bottom bar) could clarify their total risk with the addition of a polygenic risk estimate to decide against (low polygenic risk) or for (high polygenic risk) taking clinical action. PRS, polygenic risk score.

sizes[38]. Little to no evidence of low-frequency variants with moderate or larger effects (odds ratio >1.5) was uncovered despite the study being well-powered to detect the majority of such associations — imputation enables the interrogation of the vast majority of (~90%), but not all, low-frequency variants[38,39]. Similarly, a recent large-scale sequencing study of type 2 diabetes mellitus found that the frequency spectrum of detected risk loci was consistent with a common polygenic disease model where the majority of heritability was due to common variants of small effects[23]. Again, little to no evidence for low-frequency risk variants with moderate or larger effects was uncovered despite being well-powered to detect such associations. Meanwhile, enrichment of rare variants with moderate effects on type 2 diabetes mellitus predisposition was observed within genes known to be associated with familial diabetes[23]. Similarly, a recent large-scale comprehensive GWAS for breast cancer found that 41% of familial relative risk of breast cancer can be explained by genetic variants captured by genotyping and imputation, again with no low-frequency variants of moderate effect size detected despite sufficient power to detect such associations[40]. As for type 2 diabetes mellitus, rare variants of intermediate effect size in genes associated with familial breast cancer are known to have an important role in breast cancer predisposition[41]. A similar genetic architecture has been observed for prostate cancer[42–44]. Finally, recent studies indicate that between ~70% and ~90% of late-onset Alzheimer disease heritability may be explained by genetic variants captured by genotyping and imputation in GWAS[22,45], with exome-chip studies finding little evidence for low-frequency risk variants with moderate or larger effects[46]. Again, rare variants with moderate effects have been identified via sequencing studies focused on families enriched with late-onset Alzheimer disease cases[47].

Thus, both projections from statistical modelling and empirical results of comprehensive genomic studies reinforce the conclusion that the genetic architecture for many common adult-onset diseases is composed of a familial form, responsible for 1–10% of disease incidence, linked to highly penetrant rare variants within a small set of genes known to drive familial disease, and a nonfamilial form of disease that is mostly driven by an amalgamation of common variants of small effect distributed throughout the genome, in combination with a smaller contribution from rare variants of moderate effect in genes known to cause familial disease. This observed polygenic architecture is consistent with the hypothesized genetic architecture that motivated the design and pursuit of GWAS — the 'common disease, common variant' hypothesis[48–52] — which posits that the genetic variants responsible for most of the disease risk in the population are shared across members

Relative risk
Relative risk is the probability, or risk, that a certain event will occur in comparison to the event rate in a reference group; often expressed as the ratio of absolute risk between two groups, thus a value of 1.0 means no difference in risk.

of the population. The findings to date simply show that there are many more variants with much smaller effects than were originally anticipated (BOX 1). As any of such individual common genetic variant is incapable of effectively stratifying disease risk, researchers have sought to leverage numerous variants simultaneously for risk prediction.

Development of polygenic risk scores

Before reviewing the evidence supporting the utility of polygenic risk profiling, we briefly describe the development and evaluation of PRSs themselves. For a detailed discussion of the considerations for the development of PRSs, especially the need for appropriate calibration of PRSs for clinical use, see the in-depth review by Chatterjee et al.[53]. In brief, a PRS is most commonly calculated as a weighted sum of the number of risk alleles carried by an individual, where the risk alleles and their weights are defined by the loci and their measured effects as detected by genome wide association studies[53]. In some instances, a lower threshold than genome-wide statistical significance may be used to improve or estimate total predictability, often at the expense of generalizability[17,25,54]. In other instances, models may be recalibrated to account for biases in effect size that are typically inflated in the discovery cohort, to account for multiple linked variants within each disease-associated locus, to re-estimate effect sizes for a sub-phenotype of interest or to adjust for ethnic or demographic factors that may influence the generalizability of models[53,55]. However, the most common approach, which may not be optimized for predictive power, is the simple approach of summing risks across susceptibility loci as reported in the literature. The utility of the PRS is often evaluated by determining whether it, in combination with clinical risk factors, separates the population into categories with sufficiently distinct degrees of absolute risk to drive clinical or personal decision-making (FIG. 1).

Historically, the utility of GWAS-based genetic risk estimates has been assessed, perhaps inappropriately, on the basis of their ability to comprehensively discriminate between diseased and non-diseased individuals — usually quantified by the area under the curve (AUC) of a receiver operating characteristic curve, a plot of the true positive rate (sensitivity) versus false-positive rate (specificity). The AUC is equivalent to the overall probability that the predicted risk of an individual with disease is higher than the predicted risk of an individual without disease[11,12,53]; it provides no information regarding the predicted absolute risk conveyed to any single or subgroup of individuals. In other words, AUC is a population-level metric that is most appropriate for a diagnostic test, the primary purpose of which is the separation of diseased from non-diseased individuals. Yet, the relevant use case for genetic risk information is prognosis, a prediction of the likelihood that a certain outcome, such as the onset of disease, will occur in each individual or subgroup of individuals (FIG. 2).

While the ultimate goal of polygenic risk estimates may be the comprehensive stratification of the entire population through a complete accounting of each individual's genetic susceptibility for disease, the more

Box 1 | A brief history of GWAS risk profiling

The initial results of genome-wide association studies (GWAS) revealed that prediction of disease risk from GWAS findings would not be simple. The history of GWAS findings and the response to those findings have been reviewed extensively[52,89]. Thus, here, we provide a brief summary with a focus on perceptions of the utility of GWAS.

The first large-scale GWAS, published in 2007, with sample sizes on the order of 1,000–5,000 affected individuals, identified a handful (1–3) of associated loci for coronary artery disease, type 2 diabetes mellitus, and prostate, breast and colorectal cancer, explaining <5% of disease heritability[90–94], with the notable exception of Alzheimer disease owing to *APOE*-ε4, which explained a substantial proportion (~5%) of heritability[95]. The next 5 years would see GWAS sample sizes grow to ~10,000s of affected individuals. The number of distinct associated loci per disease grew to the order of tens of associated loci per disease, but with these factors explaining only ~10% of the heritability per disease[52]. The response to these findings ranged from disappointment to more visceral condemnation of the pursuit of GWAS[52].

In the meantime, direct-to-consumer genetic testing companies, 23andMe, Pathway Genomics, deCODE Genetics and Navigenics being the most prominent, were commercializing GWAS findings and returning genetic risk estimates to individuals, perhaps prematurely, despite the lack of studies demonstrating the analytical or clinical validity of the tests. These personal genetic testing services were met with varied interest from the general public but considerable concern from the medical and scientific community regarding the validity, marketing, psychological harm, public health utility and unnecessary health-care utilization impact of genetic risk profiling[96]. These concerns culminated in the US Food and Drug Administration temporarily banning 23andMe from the sale of its Personal Genome Service owing to its violation of federal marketing guidelines for medical devices[97].

Thanks to strategies aimed at improving the statistical power of GWAS, there has been a growing acceptance of their validity, at least as an appropriate experimental design for the discovery of genetic risk factors for common disease[89]. The empirical evidence is simply overwhelming. The latest GWAS meta-analyses include >100,000 individuals and have begun to explain a more appreciable proportion of disease heritability. For example, ~80 loci explain ~20% of coronary artery disease heritability[38,39], ~100 loci explain ~20% of type 2 diabetes heritability[23,98], ~150 loci explain ~20% of the familial relative risk of breast cancer[40], ~100 loci explain ~33% of the familial relative risk of prostate cancer[43] and ~20 loci explain ~30% of Alzheimer disease heritability[22]. Concurrently, public interest in commercial genetic services has grown rapidly. In November 2017, during the first days of the start of the US Christmas shopping season (from 'Black Friday' to 'Cyber Monday'), AncestryDNA reported the sale of 1.5 million genetic testing kits[99]. While exact figures are not available, to date, at least 12 million individuals have been genetically profiled, mostly via AncestryDNA and 23andMe — albeit through genealogical services that do not return common disease risk estimates[100]. Meanwhile, perceptions of the utility of GWAS-based polygenic risk profiling seem to be largely unchanged.

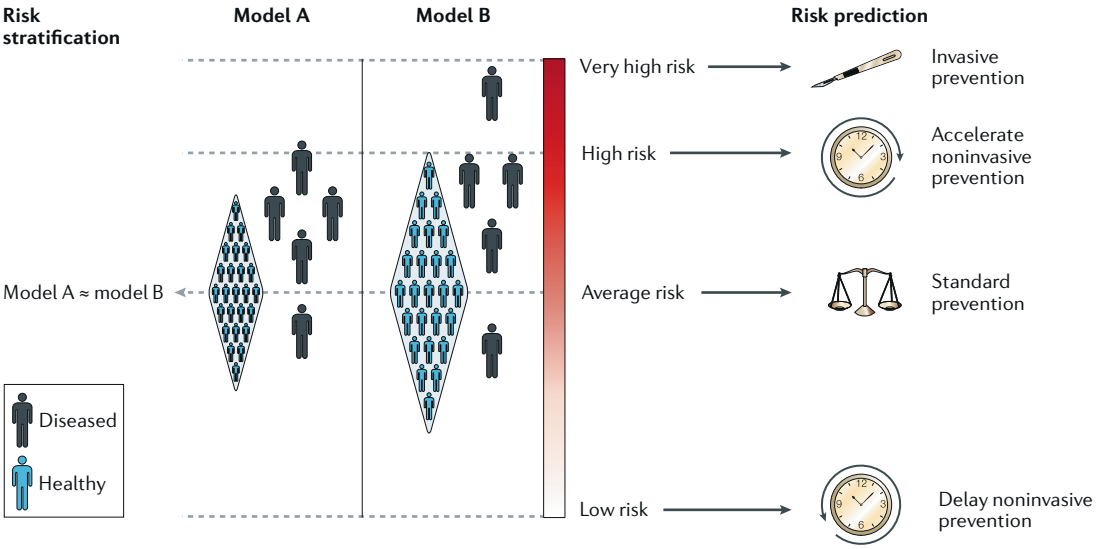

Fig. 2 | **Contrasting risk stratification versus risk prediction.** A visual depiction of the difference between the utility of risk information for population-level risk stratification (ordering individuals by risk) and prediction of individual-level and group-level disease susceptibility (absolute risk of disease). Two hypothetical models (A and B) are depicted in the middle, where the absolute vertical positioning of the human figures corresponds to the probability that an individual has or will acquire a disease. When comparing the two models, the relative distribution of diseased individuals (black) versus healthy individuals (blue) is approximately equivalent, that is, the vertical positioning of the black figure relative to the blue figures in each model is similar, which would result in receiver operator characteristic curves with approximately equivalent area under the curve values. However, the absolute vertical positioning of all figures, especially the highest-risk black figure, differs substantially between the two models, leading to different conclusions in the utility of each model, as depicted on the right-hand side of the figure. Specific high-risk and low-risk disease probability thresholds would differ on a disease-by-disease basis based on the risk of disease versus risk–benefit balance of the intervention. Note that in this hypothetical example, changing disease probabilities for the individuals at the extremes of the distribution does not influence area under the curve measures substantially but results in risk estimates with differing utility implications.

realistic and practical goal is the identification of a subset of individuals at elevated risk of disease on the basis of genetic factors in combination with clinical risk factors. We return to our *BRCA1* and *BRCA2* testing example to illustrate this difference. We do not expect *BRCA1* and *BRCA2* testing to comprehensively stratify all women by their total genetic susceptibility for breast cancer, nor do we expect *BRCA1* and *BRCA2* testing to identify the complete set of women at elevated genetic risk of breast cancer. Rather, we expect *BRCA1* and *BRCA2* testing to identify a subset of women at high risk, knowing that some women that test negative for *BRCA1* or *BRCA2* mutations may be at high genetic risk due to other unmeasured genetic factors. And we assume that, perhaps more rarely, some *BRCA1* or *BRCA2* mutation-positive women are actually at average or low genetic risk as a result of other unmeasured protective genetic factors.

Similarly, the immediate utility of PRSs should be judged on the basis of probabilistic risk conveyed to subgroups of individuals who test 'positive' for polygenic risk. Individual-level PRS values are often used to stratify the population into distinct tiers of risk based on percentile rank cut-off values (that is, top 1%, top 10%, and so on), which results in the assignment of differing levels of probabilistic risk to groups of individuals in each tier. Note that risk is not precisely defined for each individual in each tier, much like risk is not precisely defined for each *BRCA1* or *BRCA2* positive individual. The threshold rank to consider a PRS test positive then depends on the balance between the level of probabilistic disease risk conveyed to individuals at a PRS tier cut-off, often in combination with risk conveyed by other clinical risk factors, versus the risks and benefits associated with a contemplated intervention (FIG. 1). Generally, as described below, the medical community has already defined the appropriate threshold of risk versus benefit that justifies certain medical interventions — the more invasive or risky the intervention, the higher the level of absolute risk that must be mitigated by the intervention to justify its application.

The utility of polygenic risk scores

We can roughly categorize PRS utility based on three major classes of interventions: PRS-informed therapeutic intervention (the part that PRS can play in the selection of interventions to treat or prevent disease); PRS-informed disease screening (the role that PRS can have in the decision to initiate and the interpretation of disease screens); and PRS-informed life planning (the personal utility that PRSs can provide, even in the absence of preventive actions). The utility of PRSs depends heavily on a fairly complex interplay between disease-specific and intervention-specific risks and benefits, making generalization difficult. Thus, we return to our examples of the leading heritable causes of death in the developed world to review the evidence supporting the role of PRSs in each of these intervention categories (FIG. 3).

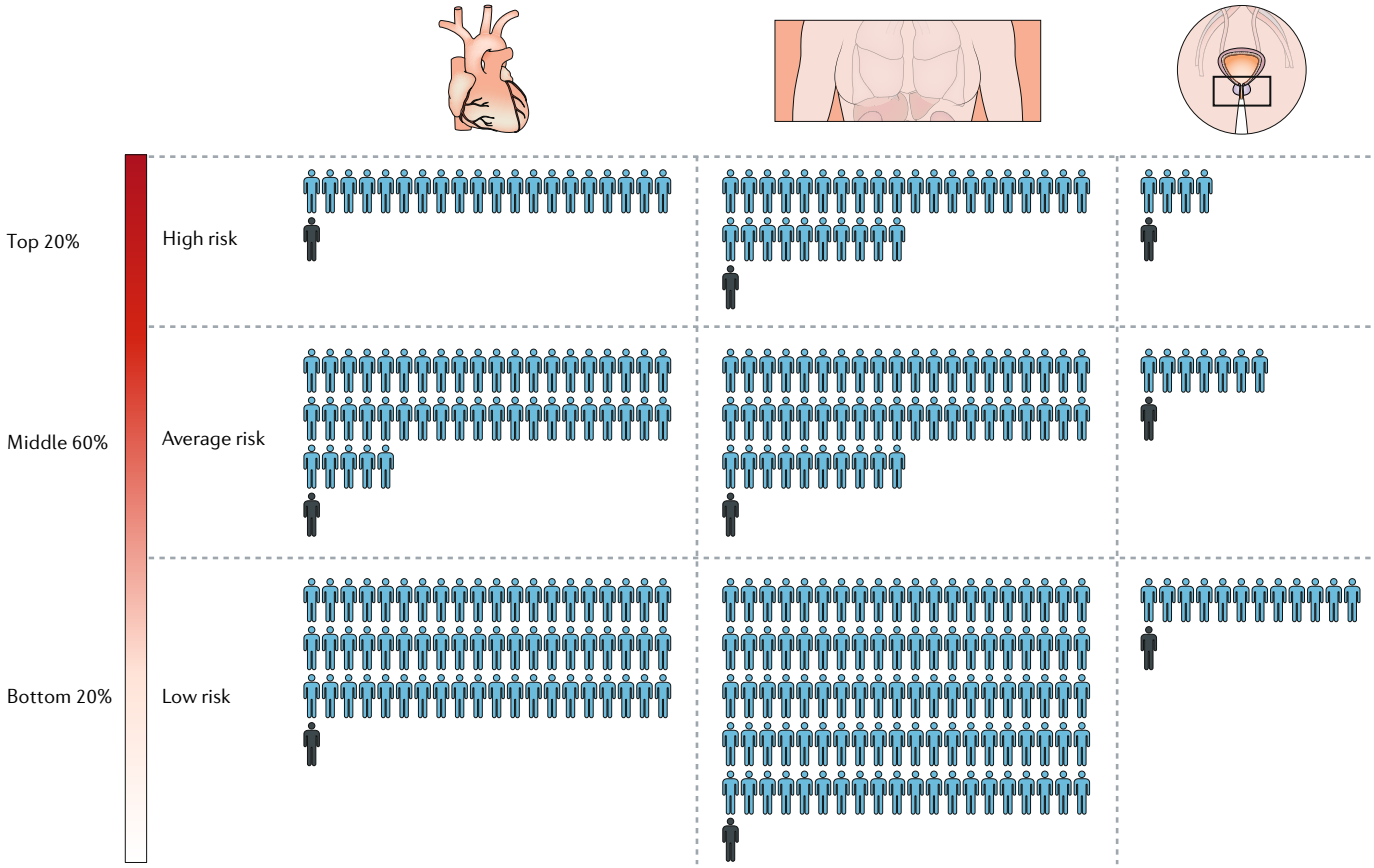

Fig. 3 | **Risks and benefits by polygenic risk score tier.** The number of individuals treated or screened relative to the number of individuals receiving a benefit from the intervention is broken down by polygenic risk score (PRS) tier (top quintile, middle three quintiles (from the 20th to the 80th percentile) and bottom quintile of genetic risk). The underlying data are derived from REF.[64] for coronary artery disease (left — number needed to treat with statins to prevent a heart attack), REF.[2] for breast cancer (middle — number of women screened to detect incident breast cancer) and REF.[3] for prostate cancer (right — positive predictive value of prostate-specific antigen (PSA) testing). Blue figures are healthy individuals. Black figures are unhealthy individuals.

PRS-informed therapeutic intervention. Individualized management of disease is central to the philosophy of precision medicine, with genetic factors often invoked for this strategy to personalize health care[56]. The potential utility of PRSs in prioritizing therapeutic interventions is exemplified by recent studies relating coronary artery disease PRSs to the prophylactic use of cholesterol-lowering therapies. According to the US Preventive Services Task Force, current guidelines for the prevention of a first heart attack (primary prevention) recommend the initiation of low-dose to moderate-dose statins in individuals between the ages of 40 and 75 years of age who have at least one risk factor for an adverse cardiovascular event (obesity, diabetes mellitus, high blood pressure or smoking) and a >10% 10-year absolute risk of an adverse cardiac event, or a 7.5–10% 10-year absolute risk of an adverse cardiac event in which case the ultimate decision is made by considering the potential harms versus benefits of statin therapy[57]. However, the initiation of statins is not without controversy. For primary prevention of heart attack, fewer than 2 out of 100 individuals taking statins for 5 years avoid a heart attack or stroke, whereas 1 in 100 develops diabetes mellitus as a result of the therapy[58,59]. Other professional societies have recommended alternative guidelines[60,61], all of which incorporate some form of clinical risk calculation, which is known to overestimate absolute risk[62,63]. This uncertainty necessitates individualization of the choice to initiate a lifetime of statin therapy — uncertainty that has been shown to be partially addressed by the use of a coronary artery disease PRS.

Although coronary artery disease PRS does not substantially improve the overall stratification of heart attack risk across the entire general population, when combined with clinical risk estimates, a PRS may modify the estimated risk of some individuals so that their combined risk is at or above the level of risk recommended for the initiation of statin therapy (FIG. 1). To this end, numerous studies have shown that coronary artery disease PRSs are useful, independent of family history, for the identification of some high-risk individuals who receive greater benefit from the initiation of statin therapy[1,64–68]. Mega et al.[64] and Natarajan et al.[1] showed that patients within the highest quintile of genetic risk of coronary artery disease are at an ~30% increased risk (hazard ratio) of an adverse coronary event; upon initiation of statin therapy for primary prevention, these individuals achieve an ~45% relative risk reduction of the 10-year

risk of a heart attack or coronary artery disease related death. The outcome differs for individuals of intermediate polygenic risk (60% of individuals within the second to fourth quintiles of risk) who achieve ~25% relative risk reduction, whereas individuals in the lowest quintile of genetic risk show no or little statistical benefit[1,64]. Similarly, Tikkanen et al.[67] and Ripatti et al.[68] showed that a coronary artery disease PRS led to the reclassification of ~12% of individuals from an overall intermediate risk category into a high-risk category, translating into a stronger statin use recommendation. Moreover, when presented with their genetic risk of coronary artery disease, individuals with higher genetic risk were more likely to initiate and adhere to statin therapy[69,70]. Finally, the emerging literature suggests a genome-wide PRS identifies the top 2.5% of individuals who are at a 400% increased risk, equivalent to the adverse coronary event risk associated with familial hypercholesterolaemia, which would lead to a recommendation of aggressive cholesterol-lowering therapy[71].

PRS-informed disease screening. Findings have also demonstrated the utility of PRSs in the decision to initiate and the interpretation of disease screening in cancer. Current guidelines, as stated by the US Preventive Services Task Force, recommend the initiation of biennial screening mammography for women at 50 years of age, with consideration of individual risk factors for the decision to start screening mammography between the ages of 40 and 49 years[72]. This age-based criterion was defined by the balance between the average risk of breast cancer at various age thresholds and the risk of harms due to false-positive mammography results. Based on this risk-to-benefit threshold, a breast cancer PRS, in conjunction with known clinical risk factors, was shown to identify 16% of the population who could make an informed decision to start screening at 40 years of age, given that their risk exceeded that of an average 50-year-old[2,73]. Alternatively, 32% of the population could delay screening, as their risk at 50 years of age was lower than that of an average 40-year-old[2]. Similar results have been shown in colorectal cancer, in which a PRS would lead to a recommendation to initiate colonoscopy screening at 42 years for individuals in the highest PRS decile versus 52 years for individuals in the lowest PRS decile[74].

PRSs can also aid in the interpretation of screening tests with high false-positive rates, with prostate cancer screening being the most infamous. The US Preventive Services Task Force recommends against prostate-specific antigen (PSA) screening as the harms owing to false positives and overtreatment of benign disease outweigh the benefits[75]. This recommendation is complicated by the fact that prostate cancer is fairly common, occurring in one in nine men during their lifetime. However, death from aggressive disease is uncommon, occurring in ~1% of men diagnosed with prostate cancer. Despite the fact that most prostate cancer is benign, prostate cancer remains the second highest cause of cancer death in men, and even initially detected low-risk disease often evolves to require treatment[76]. A prostate cancer PRS has been shown to help identify men at significantly elevated risk of disease who attain much greater risk-to-benefit balance from PSA testing[3,77]. Specifically, the positive predictive value for the detection of aggressive prostate cancer via PSA testing is ~25% for individuals in the top 5% of genetic risk versus ~12.5% in the general population. Individuals with PRS scores in the top 50% account for 76% of aggressive prostate cancer, with the top 20% of individuals accounting for 42% of all cases of aggressive prostate cancer[3]. Importantly, the prostate cancer PRS is not specific to aggressive disease, and thus its use to prioritize PSA screening addresses only the problem of false-positive PSA tests and not the overtreatment of benign disease[78]. Active surveillance programmes using additional clinical risk factors and balancing against competing health risks of the individual may help determine whether treatment is necessary[79]. Thus, a PRS could potentially prioritize screening for a subgroup of men at high risk of prostate cancer, with appropriate counselling regarding the action taken following a positive test.

PRS-informed life planning. Finally, PRSs may have utility even in the absence of, or personal desire to avoid, preventive screening and therapeutic interventions. For coronary artery disease, individuals in the top quintile of genetic risk have the ability to offset much of this risk by maintaining optimal lifestyle habits, which reduces their overall risk of disease by nearly half[5]. For breast cancer, if healthy lifestyle choices were preferentially targeted to and employed by women in the top decile of genetic risk, an estimated ~20% of all preventable breast cancer cases would be avoided[2]. Theoretically, clarifying a high-risk individual's perception of their susceptibility to disease and quantifying the benefits of healthy behaviours would be one among many effective tools for inducing and maintaining behaviour change[80]. Alzheimer disease provides a more interesting use case, as there is some debate about whether lifestyle choices can mitigate Alzheimer disease onset. A PRS for Alzheimer disease was recently shown to be able to dramatically stratify individuals by average age of disease onset[4]. Those individuals with a PRS in the top quartile had an average age of disease onset of 75 years versus 95 years for those in the lowest quartile. Alzheimer disease is influenced by one strong genetic risk factor, *APOE* status. Yet, even when the effect of this factor was removed by limiting the analysis to individuals homozygous for the *APOE-ε3* allele, the difference in average age of disease onset was still 10 years in the top versus bottom decile of genetic risk[4]. Although the adoption of healthy behaviours might not influence Alzheimer disease risk, this information could inform financial, legal and care planning.

Perspectives

Barriers and caveats to PRS utility. PRS-based genetic risk estimates are beginning to show promise in their ability to identify subgroups of individuals who may benefit from the prioritization of preventive actions. However, there remain a number of scientific, clinical and social hurdles to bring PRSs into practice[81]. While we have compared the utility of high-risk variants to cumulative polygenic risk estimation, it is likely that there is greater

uncertainty in the risk estimates at the individual level among high-risk individuals identified by a polygenic versus familial disease genetic test. A major component of this uncertainty is a result of the fact that interrogation of high-risk variants typically involves identification of a directly causal variant, whereas polygenic risk estimates may incorporate variants that are not perfectly correlated with the causal genetic factor or factors. This results in some uncertainty in the estimation of the effect size associated with each individual variant integrated into a PRS and reduces the generalizability of PRS risk estimates in populations beyond the population studied. This issue is most pronounced in the transferability of risk estimates from European ancestry populations, the population in which most GWAS have been executed, to African ancestry populations[82]. Thus, inequities in access to useful genetic risk estimates are of major concern.

It is tempting to invoke missing heritability, the unknown component of genetic risk, as another source of uncertainty specific to polygenic risk estimates. However, this argument relies on what seems to be a false dichotomy between deterministic familial disease variants and probabilistic polygenic risk. In fact, recent evidence suggests that familial risk is both probabilistic and modified by polygenic risk[19]. For example, absolute risk of breast cancer in *BRCA1* or *BRCA2* mutation carriers is markedly influenced by known polygenic risk factors, with effect sizes that are consistent with but slightly smaller in *BRCA1* or *BRCA2* mutation carriers and noncarriers[7]. The cumulative influence of known polygenic risk factors sufficiently modifies total absolute risk levels in *BRCA1* and *BRCA2* mutation carriers so as to change mammography screening recommendations. For example, *BRCA1* carriers in the lowest versus highest PRS decile for breast cancer risk have been shown to have a 21% versus 39% absolute risk of developing breast cancer by age 50 years and a 56% versus 75% risk of developing breast cancer by age 80 years[7]. Similarly, prostate cancer risk by 80 years of age varies from 7% to 26% for *BRCA1* carriers and from 19% to 61% for *BRCA2* carriers in the 5th and 95th percentile of a prostate cancer PRS, respectively[8]. Moreover, there are known polygenic factors not associated with polygenic risk of disease that specifically modify risk in *BRCA1* and *BRCA2* mutation carriers[83,84]. Similarly, polygenic risk strongly modifies the risk of heart attack in individuals with familial hypercholesterlaemia[6]. Thus, the level of uncertainty in estimated risk owing to unmeasured genetic factors may be no greater for high-risk individuals identified via measured high-risk variants or polygenic factors. The total absolute risk conferred to high-risk individuals identified by high-risk variants may be substantially greater than that conferred by polygenic factors; however, this simply translates into different recommendations that can be made to these individuals based on the balance of risk versus benefit of the available interventions. Nevertheless, it should be noted that there may be differences in disease presentation, severity and available therapeutic interventions for disease for familial versus polygenic genetic susceptibility[7,85].

Other barriers to PRS utility include physician and public education regarding the interpretation of polygenic risk, especially in the understanding of various and dynamic risk metrics. A recent trend of unsupported genetic tests for athletic ability, dietary recommendations and others may be difficult for non-experts to differentiate from the validated approaches presented above. For those tests with utility, physician and public opinion may be negatively biased owing to early commercialization efforts. This lack of trust could bias the manner in which PRSs are utilized. For example, to improve both the efficacy and efficiency of clinical practice via the use of PRSs, individuals of low polygenic and clinical risk and their physicians would potentially need to delay screening or decide against therapeutic intervention, which is likely a more challenging recommendation to adopt than the decision to accelerate screening and therapeutic intervention owing to elevated overall risk. Moreover, strategies to effectively communicate risk information to physicians and individuals must be developed and perhaps customized to different target audiences, especially when the intention is to drive the uptake of preventive behaviours. Unwanted psychosocial impact of PRS information in the form of anxiety, fatalistic thinking or the adoption of a false sense of security must be considered as part of this communication strategy. These concerns have been demonstrated to be minor[86–88], although PRS results may become more alarming as models continue to improve to the point where they are able to convey substantially greater levels of risk. As PRS estimates continue to improve, stronger protections against genetic discrimination are needed to encourage the adoption of this approach. Finally, large-scale prospective studies examining the clinical utility of PRSs should be conducted.

Future directions. The utility of PRS risk estimation is currently limited by its simplicity. A number of research directions can be considered for future development to improve their comprehensiveness, to reduce or communicate uncertainty and to improve generalizability. For risk communication, it would be useful to develop methods to convey the uncertainty associated with genetic risk estimates at an individual level, especially a method that incorporates uncertainty resulting from measured as well as unmeasured factors. For example, statistical models[17] could be used to project the expected allele frequency and effect size of yet to be identified GWAS loci to estimate the distribution of the future genetic risk estimate of an individual via PRSs that include these currently unmeasured factors. Importantly, we expect the risk of reclassification owing to future genetic findings to be modest, especially given the paucity of low-frequency risk variants with moderate or larger effects. For individual-level decision-making, quantifying this uncertainty would be an important component to consider. To improve comprehensiveness and reduce uncertainty, risk estimates that integrate familial risk and polygenic risk estimates are needed. This would help alleviate some, but not all, concerns associated with the false sense of security that may be conveyed by a negative familial risk or polygenic test and would be an improvement to both familial risk and polygenic genetic testing modalities.

Comprehensiveness could also be improved by using whole-genome prediction models rather than those limited to GWAS-significant variants, although this would lead to further concerns regarding generalizability. Ultimately, more dynamic methods to estimate effects associated with individual genetic variants given the genetic, demographic and clinical risk factor background of the individual should be developed. One solution could be the development of multiple models per disease to fit various use cases—even when measuring risk of the same event in different circumstances. For example, the balance in the prognostic value of demographic, lifestyle and clinical risk factors shifts throughout the lifespan and should be accounted for. Finally, given the advances in machine learning and artificial intelligence, extensions to PRS models via these technologies may be a more practical solution to addressing the shifting influence of genetic factors and their interaction with other prognostic factors, although the use of these technologies for genetic risk prediction is associated with its own issues, including the size of the large-scale data sets required to effectively train these models and the difficulties in interpreting black box solutions. Ultimately, we believe that both our accumulated knowledge and the explosive growth in public interest have brought us to a tipping point where large-scale studies demonstrating the utility of polygenic risk estimation should be pursued.

Published online: 22 May 2018

1. Natarajan, P. et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation* **135**, 2091–2101 (2017).
2. Maas, P. et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol.* **2**, 1295–1302 (2016).
This study clearly lays out the utility of a breast cancer PRS for risk-based rather than age-based recommendations for breast cancer screening mammography.
3. Seibert, T. M. et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ* **360**, j5757 (2018).
4. Desikan, R. S. et al. Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. *PLoS Med.* **14**, e1002258 (2017).
5. Khera, A. V. et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N. Engl. J. Med.* **375**, 2349–2358 (2016).
6. Paquette, M. et al. Polygenic risk score predicts prevalence of cardiovascular disease in patients with familial hypercholesterolemia. *J. Clin. Lipidol* **11**, 725–732.e5 (2017).
7. Kuchenbaecker, K. B. et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J. Natl Cancer Inst.* **109**, djw302 (2017).
This study clearly lays out the case for the combined testing of monogenic and polygenic disease risk factors.
8. Lecarpentier, J. et al. Prediction of breast and prostate cancer risks in male BRCA1 and BRCA2 mutation carriers using polygenic risk scores. *J. Clin. Oncol.* **35**, 2240–2250 (2017).
9. Witte, J. S., Visscher, P. M. & Wray, N. R. The contribution of genetic variants to disease depends on the ruler. *Nat. Rev. Genet.* **15**, 765–776 (2014).
This reference provides a detailed breakdown of various measures and interpretations of heritability.
10. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
11. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* **6**, e1000864 (2010).
12. Cook, N. R. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* **115**, 928–935 (2007).
13. Anglian Breast Cancer Study Group. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. *Br. J. Cancer* **83**, 1301–1308 (2000).
14. Peto, J. et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J. Natl Cancer Inst.* **91**, 943–949 (1999).
15. Antoniou, A. et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.* **72**, 1117–1130 (2003).
16. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* **19**, 110–124 (2018).
17. Chatterjee, N. et al. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat. Genet.* **45**, 400–405 (2013).
18. Badano, J. L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat. Rev. Genet.* **3**, 779–789 (2002).
19. Katsanis, N. The continuum of causality in human genetic disorders. *Genome Biol.* **17**, 233 (2016).
20. Hartiala, J. et al. The genetic architecture of coronary artery disease: current knowledge and future opportunities. *Curr. Atheroscler Rep.* **19**, 6 (2017).
21. Amos, C. I. et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol. Biomarkers Prev.* **26**, 126–135 (2017).
22. Ridge, P. G. et al. Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol. Aging* **41**, 200 e13–200.e20 (2016).
23. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
A very large-scale, comprehensive GWAS for type 2 diabetes mellitus that finds no evidence for low-frequency variants of moderate effect size despite being powered to detect such associations.
24. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).
25. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
26. Zhu, Z. et al. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* **96**, 377–385 (2015).
27. Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits and implications for the future. Preprint at *bioRxiv*, 175406 (2017).
28. Speed, D. et al. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).
29. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).
30. Evans, L. et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Preprint at *bioRxiv*, 115527 (2017).
31. Stahl, E. A. et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* **44**, 483–489 (2012).
32. Browning, S. R. & Browning, B. L. Population structure can inflate SNP-based heritability estimates. *Am. J. Hum. Genet.* **89**, 191–193 (2011).
33. Krishna Kumar, S., Feldman, M. W., Rehkopf, D. H. & Tuljapurkar, S. Limitations of GCTA as a solution to the missing heritability problem. *Proc. Natl Acad. Sci. USA* **113**, E61–E70 (2016).
34. Yang, J., Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *Proc. Natl Acad. Sci. USA* **113**, E4579–E4580 (2016).
35. Bhatia, G. et al. Subtle stratification confounds estimates of heritability from rare variants. Preprint at *bioRxiv*, 048181 (2016).
36. Barton, N. H., Etheridge, A. M. & Veber, A. The infinitesimal model: definition, derivation, and implications. *Theor. Popul. Biol.* **118**, 50–73 (2017).
37. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
This reference lays out the theoretical basis for the omnigenic model of inheritance.
38. Nikpay, M. et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
A very large-scale, comprehensive GWAS for coronary artery disease that finds no evidence for low-frequency variants of moderate effect size, despite being powered to detect such associations.
39. Howson, J. M. M. et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat. Genet.* **49**, 1113–1119 (2017).
40. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
41. Easton, D. F. et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N. Engl. J. Med.* **372**, 2243–2257 (2015).
42. Mancuso, N. et al. The contribution of rare variation to prostate cancer heritability. *Nat. Genet.* **48**, 30–35 (2016).
43. Al Olama, A. A. et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat. Genet.* **46**, 1103–1109 (2014).
44. Fletcher, O. & Houlston, R. S. Architecture of inherited susceptibility to common cancer. *Nat. Rev. Cancer* **10**, 353–361 (2010).
45. Gatz, M. et al. Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).
46. Sims, R. et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* **49**, 1373–1384 (2017).
47. Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet. Med.* **18**, 421–430 (2016).
48. Lander, E. S. The new genomics: global views of biology. *Science* **274**, 536–539 (1996).
49. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17**, 502–510 (2001).
50. Chakravarti, A. Population genetics — making sense out of sequence. *Nat. Genet.* **21**, 56–60 (1999).
51. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517 (1996).
52. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).

53. Chatterjee, N., Shi, J. & Garcia-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
This reference provides a detailed overview of recommended approaches to developing PRS models and translating them to clinically useful measures of risk.
54. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
55. Vilhjalmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
56. Collins, F. S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
57. US Preventive Services Task Force. Statin use for the primary prevention of cardiovascular disease in adults: US Preventive Services Task Force recommendation statement. *JAMA* **316**, 1997–2007 (2016).
58. Macedo, A. F. et al. Unintended effects of statins from observational studies in the general population: systematic review and meta-analysis. *BMC Med.* **12**, 51 (2014).
59. Sattar, N. et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* **375**, 735–742 (2010).
60. Redberg, R. F. & Katz, M. H. Statins for primary prevention: the debate is intense, but the data are weak. *JAMA* **316**, 1979–1981 (2016).
61. Greenland, P. & Bonow, R. O. Interpretation and use of another statin guideline. *JAMA* **316**, 1977–1979 (2016).
62. Cook, N. R. & Ridker, P. M. Calibration of the pooled cohort equations for atherosclerotic cardiovascular disease: an update. *Ann. Intern. Med.* **165**, 786–794 (2016).
63. Rana, J. S. et al. Accuracy of the atherosclerotic cardiovascular risk equation in a large contemporary, multiethnic population. *J. Am. Coll. Cardiol.* **67**, 2118–2130 (2016).
64. Mega, J. L. et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet* **385**, 2264–2271 (2015).
A landmark study demonstrating the utility of PRSs for the prioritization of statin therapy.
65. Abraham, G. et al. Genomic prediction of coronary heart disease. *Eur. Heart J.* **37**, 3267–3278 (2016).
66. Tada, H. et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur. Heart J.* **37**, 561–567 (2016).
67. Tikkanen, E., Havulinna, A. S., Palotie, A., Salomaa, V. & Ripatti, S. Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. *Arterioscler. Thromb. Vasc. Biol.* **33**, 2261–2266 (2013).
68. Ripatti, S. et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* **376**, 1393–1400 (2010).
69. Kullo, I. J. et al. Incorporating a genetic risk score into coronary heart disease risk estimates: effect on low-density lipoprotein cholesterol levels (the MI-GENES clinical trial). *Circulation* **133**, 1181–1188 (2016).
70. Umans-Eckenhausen, M. A., Defesche, J. C., van Dam, M. J. & Kastelein, J. J. Long-term compliance with lipid-lowering medication after genetic screening for familial hypercholesterolemia. *Arch. Intern. Med.* **163**, 65–68 (2003).
71. Khera, A. V. et al. Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. Preprint at *bioRxiv*, 218388 (2017).
72. Siu, A. L. & US Preventive Services Task Force. Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Ann. Intern. Med.* **164**, 279–296 (2016).
73. Mavaddat, N. et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J. Natl Cancer Inst.* **107**, djv036 (2015).
74. Hsu, L. et al. A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology* **148**, 1330–1339.e14 (2015).
75. Bibbins-Domingo, K., Grossman, D. C. & Curry, S. J. The US Preventive Services Task Force 2017 draft recommendation statement on screening for prostate cancer: an invitation to review and comment. *JAMA* **317**, 1949–1950 (2017).
76. Hamdy, F. C. et al. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N. Engl. J. Med.* **375**, 1415–1424 (2016).
77. Pashayan, N. et al. Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genet. Med.* **17**, 789–795 (2015).
78. Eeles, R. et al. The genetic epidemiology of prostate cancer and its clinical implications. *Nat. Rev. Urol.* **11**, 18–31 (2014).
79. Tosoian, J. J. et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. *J. Clin. Oncol.* **29**, 2185–2190 (2011).
80. Morganstein, J. The Handbook of Health Behavior Change edited by Kristin A. Reikert, Judith K. Ockene and Lori Pbert. *Psychiatry* **79**, 95–96 (2016).
81. Wray, N. R. et al. Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
82. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
This analysis highlights the lack of transferability of PRS populations of dissimilar ancestry.
83. Gaudet, M. M. et al. Identification of a BRCA2-specific modifier locus at 6p24 related to breast cancer risk. *PLoS Genet.* **9**, e1003173 (2013).
84. Couch, F. J. et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet.* **9**, e1003212 (2013).
85. Wang, J. et al. Polygenic versus monogenic causes of hypercholesterolemia ascertained clinically. *Arterioscler. Thromb. Vasc. Biol.* **36**, 2439–2445 (2016).
86. Green, R. C. et al. Disclosure of APOE genotype for risk of Alzheimer's disease. *N. Engl. J. Med.* **361**, 245–254 (2009).
87. Collins, R. E., Wright, A. J. & Marteau, T. M. Impact of communicating personalized genetic risk information on perceived control over the risk: a systematic review. *Genet. Med.* **13**, 273–277 (2011).
88. Bloss, C. S., Schork, N. J. & Topol, E. J. Effect of direct-to-consumer genomewide profiling to assess disease risk. *N. Engl. J. Med.* **364**, 524–534 (2011).
89. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
90. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
91. Easton, D. F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
92. Yeager, M. et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).
93. Gudmundsson, J. et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* **39**, 631–637 (2007).
94. Zanke, B. W. et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **39**, 989–994 (2007).
95. Coon, K. D. et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J. Clin. Psychiatry* **68**, 613–618 (2007).
96. Caulfield, T. & McGuire, A. L. Direct-to-consumer genetic testing: perceptions, problems, and policy responses. *Annu. Rev. Med.* **63**, 23–33 (2012).
97. Gutierrez, A. 23andMe, Inc. 11/22/13. *U.S. Food and Drug Administration* https://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2013/ucm376296.htm (2013).
98. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. et al. Genome-wide *trans*-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
99. Molteni, M. Ancestry's genetic testing kits are heading for your stocking this year. *Wired* https://www.wired.com/story/ancestrys-genetic-testing-kits-are-heading-for-your-stocking-this-year/ (2017).
100. Regalado, A. 2017 was the year consumer DNA testing blew up. *MIT Technol. Rev.* https://www.technologyreview.com/s/610233/2017-was-the-year-consumer-dna-testing-blew-up/ (2018).

Acknowledgements
This work is supported by The Scripps Translational Science, a National Institutes of Health-National Center for Advancing Translational Sciences (NIH-NCATS) Clinical and Translational Science Award (CTSA; 5 UL1 TR001114). Further support is from U54GM114833 and the Foundation Leducq.

Author contributions
The authors contributed equally to all aspects of this article.

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Genetics thanks N. Chatterjee, P. Kraft and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

RELATED LINKS
MyGeneRank: mygenerank.scripps.edu