Human Genetics of Obesity and Type 2 Diabetes: Past, Present, and Future

Erik Ingelsson, MD, PhD[1,2]; Mark I. McCarthy, MD[3,4,5]

[1] Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA;

[2] Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305, USA;

[3] Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK;

[4] Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford OX3 7BN, UK;

[5] Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford OX3 7BN, UK.

Address for Correspondence:
Erik Ingelsson, MD, PhD, FAHA
300 Pasteur Dr, mail code: 5773; Stanford, CA 94305; USA
Phone: +1-650-656-0089; E-mail: eriking@stanford.edu

Abstract

Type 2 diabetes (T2D) and obesity already represent two of the most prominent risk factors for cardiovascular disease, and are destined to increase in importance given global changes in lifestyle. Ten years have passed since the first round of GWAS for T2D and obesity. Over this decade, we have witnessed remarkable developments in human genetics. We have graduated from the despair of candidate gene-based studies that generated few consistently-replicated genotype-phenotype associations, to the excitement of an exponential harvest of loci robustly associated with medical outcomes through ever larger genome-wide association study (GWAS) meta-analyses. As well as discovering hundreds of loci, GWAS have provided transformative insights into the genetic architecture of T2D and other complex traits, highlighting the extent of polygenicity and the tiny effect sizes of many common risk alleles. GWAS has also provided a critical starting point for discovering new biology relevant to these traits. Expectations are high that these discoveries will foster development of more effective strategies for intervention, through optimization of precision medicine approaches. In this paper, we review current knowledge, and provide suggestions for the next steps in genetic research for T2D and obesity. We focus on four areas relevant to precision medicine: genetic architecture, pharmacogenetics and other gene-environment interactions, mechanistic inference and drug development. As we describe, the genetic architecture of complex traits has major implications for the prospects of precision medicine, rendering some anticipated approaches decidedly unrealistic. We highlight obstacles to the translation of human genetic findings into mechanism inference but are optimistic that, as these are overcome, there is untapped potential for novel drugs and more effective strategies for treating and preventing T2D and obesity.

Genomic Medicine: Hope and Hype

The pace of discovery in complex trait genetics over the past decade has been breathtaking. From the paltry returns of the candidate variant studies in a few hundred individuals that were typical until 10-15 years ago, we have graduated to genome-wide association studies in sample sizes a thousand times larger than that. At the time of writing, over 300 loci have been robustly associated with type 2 diabetes (T2D)[1, 2] and over 500 with BMI, waist-hip ratio and other obesity traits [3-8]. These discoveries provide the substrate for a systematic assessment of the pathophysiological basis of these conditions, one that embraces both environmental and genetic contributors to predisposition. This has led to increasing hope that these discoveries will pave the way for new approaches to diagnosing, stratifying, monitoring, preventing and treating disease.

Such improvements in prediction, diagnosis and prognosis have long been, and remain, one of the principal goals of genomic medicine. Once the Human Genome Project was completed[9] and low-cost commodity genotyping brought the first successes from genome-wide association analyses, there was a hope that genomics would have a dramatic impact on risk stratification and therapeutic decisions regarding individual patients.[10, 11] This expectation was predicated on the assumption that the most important common risk-variants would have at least moderate effect sizes, and could therefore substantially move the needle in clinical decision-making for individual patients. However, ten years after the landmark study by the Wellcome Trust Consortium[12] which represents the start of a truly remarkable period in biomedical research, there remain relatively few examples of results from GWAS, or human genetics generally, that have impact on clinical decision-making for complex, multifactorial traits such as obesity and type 2 diabetes (T2D).

The chief reason for this is that biology has turned out to be more complicated than many believed – or perhaps hoped – 15 years ago. As discussed elsewhere in this review, rather than being dominated by a small number of common variants with modest to large phenotypic effects, the genetic architecture of complex traits is now known to be characterized by a very large number of causal variants, most of them with only a tiny additive effect on risk.[13-15] This extreme polygenicity – or, as has been recently suggested, omnigenicity[16] – of complex traits does not necessarily mean that GWAS findings cannot be translated into paradigm-shifting new therapies, but it substantially increases the effort required to establish and characterize causal variants, genes and pathways.

When it comes to individual prediction of complex traits, the limited variance in disease risk attributable to the subset of common risk variants that reach genome-wide significance translates into low discriminative capacity. Although some early studies indicated a potential role for common variant risk scores as tools for predicting T2D risk,[17, 18] subsequent studies demonstrated limited increment in performance over clinical models that can be generated from more readily accessible risk factors (such as age, BMI and ethnicity).[19-21] The recent wave of very large GWAS has reinvigorated interest in the potential of genetic risk-profiling, but the potential for clinical utility remains contested.[2, 22-24] In a future where genome sequence forms part of every individual's health record, such information could prove useful for a host of clinical indications, but the benefits are not yet seen to be sufficient to motivate the dedicated collection of such data for any single specific medical purpose (e.g. to obtain slightly more precise estimates of future T2D risk). Further, even though there are reports that knowledge of personal

genetic risk hypothetically could have some impact on individual motivation to modify behavior so as to reduce diabetes risk,[25, 26] randomized clinical trials have failed to show an effect of genetic risk counseling on self-reported motivation or prevention program adherence for individuals at risk of diabetes.[27, 28] Hence, at the moment, there is no evidence to suggest that knowledge about genetic risk for T2D can appreciably change risk assessment or improve motivation and adherence to lifestyle intervention.

The explosion of interest in the potential for precision medicine often leads to a tendency to see that potential purely in terms of genetic- and genomic-profiling. However, although genomics is an important and integral part of precision medicine, these terms are not equivalent. Indeed, integrative approaches using several layers of biological and clinical information are likely to be required if we are to achieve accurate diagnostic and prognostic precision for complex, multifactorial traits. For reasons discussed in detail in this review, genomics may not yet be ripe for widespread clinical implementation in enabling precision treatments for obesity and T2D. But are there other ways in which the insights from human genetics could provide a path towards precision medicine? In this review, we aim to summarize current knowledge and provide some thoughts about the next steps in this respect, with a focus on four areas that are relevant to the future implementation of precision medicine: genetic architecture, pharmacogenetics and other gene-environment interactions, mechanistic inference and drug development.

Genetic Architecture and Why It Matters

There has been a longstanding debate (dating back more than a century) as to whether the genetic differences that influence individual predisposition to common traits such as height or T2D are mostly driven by common alleles widely shared within and across populations, or whether they are more often attributable to rare, or even unique, events, specific to an individual and their family. This question matters because the answer has profound implications for the ways in which genetic data can be used to support personalised medicine. The advent of GWAS, a decade ago, reignited this debate. The first wave of GWAS studies (for T2D, body mass index [BMI] and many other traits) demonstrated the power of this approach to deliver robust discovery of novel association signals. These discoveries featured common variants (by design, of course, given the content of the early commodity genotyping arrays), but it was soon clear that most of these had, at best, modest effects on disease risk.[29] The apparent disparity between the observed heritability of these conditions (derived from twin studies for example) and the desultory performance of the GWAS signals that were being discovered to explain those observations (even when considered in aggregate), spawned endless discussions about the basis for this "missing heritability"; and, in particular, the extent to which this gap might be filled by rare variants of large effect that were largely invisible to GWAS studies.[29-31]

Over the past 2-3 years, the ability to directly assay the contribution of lower frequency and rare alleles in complex trait predisposition has advanced to the point where this vexatious question can now be answered with confidence. Large GWAS, conducted on more informative arrays, bolstered by imputation from large, dense reference panels have extended the scope of GWAS to include variants

with minor allele frequencies below 1%.[32] Bespoke genotyping panels (such as exome array) and direct sequencing (of exomes predominantly) have provided access for association analysis to even rarer variants.[33] In parallel, methodological advances (such as Genome-wide Complex Trait Analysis, GCTA) have facilitated assessments of "heritability explained" that go beyond just the genome-wide significant hits and consider the contribution of variants genome-wide to observed variation in trait levels or disease predisposition.[34] These diverse approaches have produced a clear and consistent answer: for T2D and obesity, as well as for many other complex traits, genetic risk is driven predominantly by common risk variants that are shared both within and between populations.[33, 35]

Collectively, these variants (most of which have small effects and lie far below genome-wide significance despite GWAS studies involving many hundreds of thousands of individuals) can explain most of the variation in disease predisposition that is attributable to genetics.[2, 34] It is not widely-appreciated that, whilst there are far more sites of rare variation than there are of common, most of the genetic differences between two individuals reside at the latter. Thus, the allele frequency spectrum of T2D disease-risk variants broadly mirrors the distribution of genetic variation at large. This should not come as a great surprise, particularly for traits of recent origin (such as T2D) which remain largely post-reproductive in terms of their impact on health. These factors will have limited the extent to which T2D risk alleles were exposed to adverse selective pressure, and allowed them to be maintained in populations at relatively high frequency.

Crucially, this is not to deny any role for rare risk alleles in predisposition to these traits. Rare alleles that contribute to T2D risk and variation in BMI are being increasingly revealed as the sample sizes exposed to dense imputed genotyping, exome or whole-genome sequencing grow.[2, 7, 8, 33] Since detectable rare variant associations signals will necessarily be dominated by rare alleles with relatively large effects, their identification can provide valuable and rapid biological insights, particularly when they map to well-annotated sequence (such as coding regions).[36-38] However, the evidence from T2D indicates that, with the possible exception of isolated or endogamous populations, rare variant signals are neither numerous enough nor of sufficient collective effective size to explain much of the variation in phenotype.

These discoveries regarding the genetic architecture of traits such as T2D are hugely relevant to the perspectives we should be adopting with regard to patient stratification. It is increasingly evident that personal risk of T2D or obesity is driven by an individual's pattern of genotypes across many hundreds (or, more likely, thousands) of, mostly common, variant sites. These act in concert with environmental exposures that, judging by the global nature of these conditions, are likely to be pervasive. Thus, whilst each individual's predisposition barcode (across variants and exposures) is likely to be unique, at the same time, these profiles are substantially overlapping across individuals.

In contrast to many Mendelian diseases, this makes it inherently unlikely that it will be possible to define robust phenotypic clusters that map onto distinct aetiological pathways, in a manner consistent with a "traditional" view of precision medicine. One recent study indicating the potential to stratify individuals with newly-diagnosed diabetes on the basis of clinical and biochemical characteristics suggests that this assessment may be overly pessimistic, but it remains unclear how the "group-level"

structure uncovered in this study maps onto clinical response, and the extent to which this provides a platform for individual decision-making.[39]

In addition, longitudinal studies are required to ensure that any clustering detected in cross-sectional studies is truly reflective of aetiological heterogeneity, and does not simply reflect different stages of disease evolution, or other temporal variation (**Figure**). This is particularly true for biological readouts such as novel proteomic or metabolomics biomarkers as well as traditional risk factors for T2D, which, in contrast to DNA variation, change over time and may provide valuable opportunities for dynamic tracking of disease progression.

The above argues for a shift to a more nuanced, probabilistic, molecular taxonomy of disease which recognises that what we consider "disease" is, for complex traits at least, usually the consequence of a confluence of risk across multiple processes (in the case of T2D, these would include insulin secretion, insulin sensitivity, islet autoimmunity, fat distribution, etc.)[40] This "palette" model allows for a redefined perspective on precision medicine, one focused on understanding the mix of abnormalities contributing to disease risk and progression for a given individual at a given point in time, and in developing the tools (e.g. biomarkers for each component process) and interventions (capable of reversing as many of those defects as possible) that allow for a more targeted response. We discuss some of the clinical implications of this model in the next section.

One last question needs to be posed. What can we expect to learn about disease mechanisms when genetics points to the involvement of thousands of regulatory elements and genes influencing many disparate pathways? This is an important and thought-provoking question to which we only have partial answers.[16] First, it is worth emphasising that "thousands of genes" and "many pathways" is not the same as all genes and all pathways, and as we define more of the processes through which disease risk is mediated, we will better understand whether the mechanisms involved in disease-specific biology continue to proliferate, or whether they start to converge on a limited set of key pathways. Second, it is important to consider the importance of temporal- and tissue-specificity, which after all is what defines many diseases in the first place: the dimensionality of disease mechanisms is not merely measured in genes and pathways, but in their impact across cell- and tissue-types and across the life course.

Pharmacogenetics and Other Gene-Environment Interactions

Pharmacogenomic testing to personalize diabetes treatment has been an area of intense research in the past ten years.[41] Given the large inter-individual variations in efficacy and side effects of different treatment regimens, more precise ways to predict individual treatment response would be of huge importance to clinical care. The hope is that genetic information can be used to define subgroups who share more similar responses to the various preventative and therapeutic treatments available. Such developments would epitomize precision medicine, which seeks to deliver the right treatment to the right person at the right time.

Needless to say, there is also a fair bit of hype surrounding these concepts. One frequent expectation arising from the evident heterogeneity in the presentation, course and treatment response

in T2D, is that it will become possible to decompose the T2D monolith into some number of etiologically-distinct, and more phenotypically homogeneous subtypes, each of which will be exquisitely responsive to interventions targeted to reversing the specific lesion responsible. However, as we described above, such a model is at variance with what we know of the genetic architecture of T2D.

In the palette model described earlier,[40] there may well be opportunities to match phenotypic presentation and disease course to specific genetic profiles, but this is likely to be most effective in those (possibly exceptional) individuals in whom the phenotype is dominated by a single phenotypic defect (i.e. those with monogenic disease, or something approaching it). In such circumstances, it is likely that genetic information will allow us to stratify individuals who may optimally respond to different treatments. However, we have some way to go before that such strategies are clinically validated, and the effects are likely to be more nuanced than the more extreme expectations referred to above. It is possible that genetic information may be more useful for target engagement (i.e. selecting a specific diabetes therapy) than for prediction of disease risk because the genetic effects on drug response are likely to be stronger than those seen for associations with disease. This is in part due to the closer biological connection between a genetic variant in or near a gene encoding a drug target and the response to the corresponding drug, and in part because there may be less selective pressure against variants of large effect for novel drug exposures.

Many of those with "typical" T2D will have several processes contributing in parallel to their metabolic state (elevated BMI, unfavorable fat distribution, suboptimal beta-cell function, etc.) and neither their genetic profiles nor their clinical features will necessarily provide a distinctive pattern that can be matched to available therapies. Much will depend on the extent to which it becomes possible, through the integration of information on genetic predisposition, measures of external and internal environmental exposures, clinical phenotype, and biomarker readouts of relevant pathophysiological processes, to characterize the dominant mechanisms contributing to ill-health in a given individual at a given point in time; and to define specific interventions and treatments that demonstrably have differential merit for returning that individual towards health.

Genetic testing in maturity onset diabetes of the young (MODY) and neonatal diabetes has already provided the diabetes field with proof-of-principle examples where genetic knowledge guides treatment.[42] However, it will likely be more challenging to develop robust pharmacogenomic tests to aid therapy decisions in T2D than for these monogenic disorders due to the complex web of genetic and environmental factors influencing T2D risk. Indeed, there has been a multitude of small, post-hoc studies addressing the pharmacogenomics of different diabetes drugs such as metformin[43] and sulfonylureas.[41] In a systematic review of 34 studies published in 2014, there was some evidence of pharmacogenetic interactions for metformin, sulfonylureas, repaglinide, thiazolidinediones, and acarbose consistent with their pharmacokinetics and pharmacodynamics, but a lack of large, prospective, high-quality studies.[44] An example of the kind of large, high-quality study that will be crucial for clinical implementation was published by the Metformin Genetics (MetGen) Consortium last year.[45] In this study of 10,577 participants of European ancestry, an allele in the intron of *SLC2A2*, which encodes the facilitated glucose transporter GLUT2, was associated with greater metformin response. Among obese individuals, homozygote carriers had a 0.33% (3.6 mmol/mol) greater absolute HbA1c

reduction than reference allele carriers. Though this effect size is likely too modest to justify genotyping of this variant for this indication alone, this study provides some evidence of pharmacogenomic markers with potential to inform precision medicine in T2D. Crucially, this is another example of the kind of association which could have clinically utility in a world where individual genome sequence (or GWAS) data were widely available. It also illustrates the need to extend large, international collaborations – the single most important success factor for GWAS of complex traits – into the pharmacogenomic arena.

There have been parallel efforts to find genetic variants that modify the response to lifestyle interventions that might prevent obesity and T2D, and which might therefore support optimization of individually-tailored behavioral modifications. Most such gene-environment studies have suffered from the same problems as the candidate gene studies of the 1990s. They have been based on questionable hypotheses: candidates from known biology, or even GWAS loci associated with main effects on the trait of interest, are not the best candidates for gene-environment interaction effects due to the nature of their discovery (which would generally be biased against discovery of signals that are contingent on environmental exposures). Further, most studies have been way too small and underpowered, lacking replication in independent study samples.[46] A recent systematic review reported that none of the eight gene-macronutrient interactions for T2D-risk that had been reported in prior literature could be replicated.[47] The most promising advance in this area has come from analyses that suggest the combination of clinical information and assays of gut microbiome content can predict individual excursions in glucose profile in response to specific dietary components.[48]

Although there are some successful studies of gene-environment interactions of obesity, they have typically combined risk alleles into genetic risk scores,[49-51] an approach that increases power, but does not provide insights regarding specific loci. Also, these studies too have typically been performed in small samples with limited replication, so their value is still uncertain. For gene-environment effects of physical activity on obesity, there is one locus, *FTO*, which shows a well-replicated interaction. The effect of this variant on BMI has been reported, in a study of ~200,000 individuals, to be attenuated by ~30% in physically active individuals compared to inactive individuals.[52] As in the case of the MetGen Consortium, this highlights the importance for investigators interested in gene-environment interactions to be inspired by successes of the GWAS era, and understand the importance of forming alliances to maximize sample sizes and replicate findings. Indeed, there are strong reasons to argue that this is going to be even more important for gene-environment studies than for analyses of main effects.

Unfortunately, this paucity of robust findings for gene-environment interactions in T2D and obesity has not prevented commercial companies from attempts to monetize the field of consumer genetics. Although there are some serious businesses out there, there are all-too-many examples of genetic tests being oversold without much, if any, science to support them. There are companies who claim that their genetic tests can provide personalized diet or exercise advice[53-57]; predicting what wine you like[58], whether you are likely to like Marmite,[59] or indicating that your child is likely to be a successful soccer player.[60]

Lost in Translation: Mechanistic Challenges

One of the key goals of human genetic enquiry for complex traits such as diabetes and obesity is to reveal the fundamental mechanisms which lead to disease predisposition and development, as well as those that influence the risk of subsequent complications. In contrast to the findings from "traditional" approaches to physiological and epidemiological investigation from which it can be difficult to disentangle causal effects from those which reflect reverse causation or confounding, human genetics offers the opportunity to define mechanistic relationships that link inherited variants to consequent changes at the molecular, cellular, tissue and physiological level.[61, 62]

Mechanistic insights can provide the foundation for uncovering novel interventions that better prevent and treat disease, and for the detection of biomarkers that monitor disease progression. Those same insights may also presage a more rationale mechanistic taxonomy of disease which enables a more personalised approach to disease prevention and management.[63] The value of this approach has been reinforced time and again from studies of monogenic and syndromic conditions, including many early-onset familial forms of diabetes and obesity. The demonstration that many cases of neonatal diabetes result from mutations in the *KCNJ11* gene, encoding a component of the beta-cell K_{ATP} channel, suggested that patients with this rare condition might be amenable to treatment with sulphonylureas, rather than insulin, a hypothesis which received rapid empirical confirmation.[64, 65] In similar fashion, the identification, in some rare instances of extreme early obesity and hyperphagia, of causal mutations in the leptin gene, enabled successful treatment with replacement leptin.[66]

The discoveries just described led to rapid clinical advances and some of the clearest examples of personalised medicine in current medical practise. This was in part because the disease-causing mutations had dramatic and unequivocal impacts on the protein sequence of genes that could be very easily positioned within existing understanding of disease pathogenesis. At the same time, the very specific phenotypic presentation in these cases provided confidence that the remedial treatments concerned were likely to be both effective and safe. The situation for a multifactorial disease like T2D is a deal more complex. As described above, the genetic contribution to T2D predisposition is made up from many hundreds (likely thousands) of mostly common variants of individually small effects.[33] Each of us (diabetic or not) carries our own particular configuration of risk and protective alleles across these variants. Because they are common, causal variants at these loci are typically hidden amongst a forest of highly-correlated alleles (due to linkage disequilibrium), frustrating efforts to identify the specific variants involved. And, because most of the signals map to non-coding sequence, the identity of the genes (protein coding or otherwise) through which they operate is often obscure.[67] All of these represent challenges to proceeding from genetic discovery (for example, of a signal detected by GWAS) to enumeration of the mechanisms involved. Indeed, many thousands of GWAS loci have been identified in the past decade, but a compelling mechanistic description has been forthcoming for only a small proportion. Having said that, recent advances are propelling far more complex trait loci towards mechanistic resolution.

One illustrative example from the T2D arena concerns a GWAS signal mapping near *MTNR1B*, encoding one of the receptors for the hormone melatonin.[68-70] The progressive accretion of GWAS data through consortium efforts provided replication of this signal, and the capacity to disentangle the

correlations between adjacent SNPs through fine-mapping, such that the association signal could be resolved to a single variant (rs10830963).[71] Analysis of intermediate metabolic traits in non-diabetic individuals indicated that the T2D-predisposing action of the risk allele was at least in part mediated through a deficit in insulin secretion, highlighting the pancreatic islet as one key tissue of interest.[72] In genome-wide analyses, it had been demonstrated that T2D GWAS signals were enriched for overlap with binding sites (in islet and liver) for the FOXA2 transcription factor, and the *MTNR1B* risk variant was shown to map into an islet enhancer region containing such a site (though rs10830963 appeared to be modifying a binding site in that enhancer for NEUROD1 rather than FOXA2 itself). The T2D risk allele was subsequently shown to increase enhancer activity, and to preferentially bind NEUROD1 in islet-derived cells.[71] These findings were consistent with evidence from human islets that the T2D risk allele was associated with increased *MTNR1B* RNA expression. This, in turn, chimed with the hypothesis that one of the desirable physiological consequences of the circadian (nocturnal) release of melatonin from the pineal gland is suppression of insulin secretion during a period of restricted energy intake.[73, 74]

However, there are reasons to consider this reductionist narrative – focused around identifying a single causal variant, a clearly defined effector gene, and a single plausible physiological mechanism – as overly simplistic. At some loci, at least, the truth may be a deal more complex.

For a start, the concept of a single causal variant may, for some GWAS signals, be illusory. Not only may there be multiple independent signals in a given region (revealed by conditional analysis for example),[75] but each independent signal may not be resolvable to a single variant. This could arise where multiple highly-correlated alleles on a single haplotype are causal, each contributing to disease predisposition, for example, through combinatorial effects on enhancer activity. An alternative mechanism, relevant to alleles that create or destroy CpG sites, could involve a joint haplotype-specific impact on local methylation status. For obvious reasons, examples of this kind of behaviour have been hard to demonstrate, though the haplotype-specific methylation signal at the *FTO* locus may be one.[76] At some loci, potentially including *PPARG*, predisposition may reflect the joint effects of both regulatory and coding alleles.[32, 77] The potential for such scenarios to confound efforts to obtain functional evidence for <u>the</u> causal variant is obvious, and speaks to the need to consider potential risk variants in relation to their genomic context.

A subsequent challenge lies in connecting the causal variant(s) identified through fine-mapping and genomic integration with the effector transcripts through which they operate. In the absence of a compelling biological candidate in the region (for example a gene known to harbour rare severe mutations that result in a broadly similar, albeit more extreme, phenotype), these connections are often made on the basis of the detection of *cis*-expression signals or evidence of physical interaction with target gene promoters (typically derived from conformational capture methods).[78, 79] There are issues here with ensuring that the GWAS and *cis* eQTL signals are truly colocalized (rather than merely overlapping);[80] and in establishing that the tissues or cells analyzed are relevant to both the disease of interest and the specific locus under consideration. As studies become larger, increasing numbers of GWAS and *cis* eQTL signals are seen to be composed of multiple independent signals, posing challenges to interpretation when only a subset of these co-localise. We still lack unbiased information on the

sensitivity and specificity of these approaches, though the current evidence would suggest that neither is sufficiently high to command robust inference from any single quantum of experimental data.

A third consideration concerns the phenotypic consequences of the variants of interest. It is interesting to note that the first-discovered "landmark" GWAS loci for BMI, T2D and CAD (*FTO*, *TCF7L2* and *CDKN2A*, respectively) have proven particularly refractory to mechanistic resolution despite large effect sizes and the most extended – and most intensive – research efforts.[81-84] Might it be that the very reason that these loci have the largest common variant impacts for their respective traits (prompting their early discovery) is because they influence multiple processes? Recent work has shown that common variants near the *FTO* gene that influence BMI and obesity compromise the thermogenic potential of adipose tissue via altered regulation of *IRX3* and *IRX5*,[85] but that does not necessarily prove that this is the only mechanism through which the variants operate. Other genes in the region (*RPGRIP1L*, and *FTO* itself) are reasonable candidates with strong experimental support, and central as well as peripheral effects on BMI at this locus are plausible.[86, 87]

Also, going beyond *cis* effects, it is likely that some GWAS loci tag genes or other elements acting as "master regulators" of disease. A well-known example of this is *KLF14*, where an imprinted *trans* eQTL has multiple downstream effects on gene expression[88] and metabolic traits, amongst others T2D and high-density lipoprotein (HDL) cholesterol.[62] Distinguishing between such "true" master regulators of gene expression acting in *trans*, and correlations resulting from the downstream consequences of variation in one phenotype impacting upon others will require careful attention and thoughtful experiments as the pathways and mechanisms underlying obesity and T2D are complex and involve many negative and positive feedback loops.

Given difficulties in assigning function to weak regulatory variants in humans, it is natural for researchers to seek to fold in additional ways of confirming the causal role of proposed effector transcripts. The most obvious route is to observe the consequences of direct perturbation of gene function in humans, rodent models or cellular systems. Here again, there are limits to the facile interpretation of the data arising from such studies. Most obviously, the cellular system used in such studies may lack authenticity (a poorly-chosen cell line, or a tissue inappropriate for the disease of interest or the specific locus under consideration), or the rodent model fail to recapitulate the human phenotype.

Less widely appreciated is the likely divergence in the phenotypic expression of coding and regulatory variants even when both impinge on the function and/or expression of the same gene. Most GWAS signals map to tissue-specific enhancers and their phenotypic consequences will be "filtered" by that tissue specificity. In contrast, the impact of a coding variant will be felt in any tissue that expresses that gene (or, more precisely, any tissue that expresses the isoform in which that variant is transcribed), leading in many cases to a broader, or divergent phenotype.[36, 89] This is analogous to the difference between a tissue-specific and a global knockout mouse. This distinction is highly relevant from the perspective of therapeutic intervention, given that most pharmaceuticals will be active across all tissues in which the target is expressed, and the profile of action (and side effects) may be better captured by coding than regulatory variants. This may explain some of the apparent discrepancies between regulatory and coding variants at the *MTNR1B* locus. Whereas the T2D risk allele discovered by GWAS is,

as described earlier, associated with *increased* MTNR1B expression in islets,[90] a study of rare functionally-deficient coding alleles in the same gene demonstrated an aggregate effect which was also supportive of increased T2D-risk.[91] One explanation for this directional divergence could involve the impact of coding variants on melatonin receptor function at sites outside the islet (such as the brain).[74, 92]

In the face of these challenges, it may be tempting to despair of our capacity to make mechanistic headway. Our view is that, whilst some GWAS loci may feature daunting mechanistic complexity (along the lines described), there is no reason to believe that this is universal. The ability to build a mechanistic narrative from diverse lines of functional evidence, and to place findings in their genome-wide, tissue-specific context is allowing researchers to reduce reliance on single data points of uncertain significance. Emerging high-throughput, genome-wide techniques such as HiChIP[93, 94] and ATAC-Seq[95, 96], that can map chromatin interactions and accessibility with higher resolution than previous methods will improve our ability to disentangle GWAS loci; while single cell RNA sequencing[97, 98] and CRISPR-based pooled gene perturbation methods[99-103] provide unprecedented opportunities for studies of how RNA expression patterns differ between cells within tissues, and how those tissues and cells react to perturbation of multiple genes in parallel.

In this way, more and more GWAS loci will start to reveal their secrets: the hope is that "solving" the biological puzzle at a subset of loci will provide clues that allow other, less immediately tractable, loci to gather mechanistic momentum. Although there is no guarantee that the puzzle can ever be fully solved or even that solving the puzzle would lead to important clinical insights, we believe that combining genomic information with characterization of relevant disease pathways in a temporal and tissue-specific manner are more likely to be successful than sequencing of static DNA variation alone.

Can Genetics Increase Efficiency of the T2D Drug Development Pipeline?

Although there have been indisputable scientific and technological advances in biomedicine over the past 50 years, this has not translated to an improved efficiency of pharmaceutical drug research and development (R&D). On the contrary, we have seen a remarkable decline in R&D efficiency, as the number of drugs approved per amount spent on R&D has halved every nine years since 1950.[104] The consequence of this disturbing and clearly unsustainable trend – described by some as "Eroom's law", to highlight the contrast with Moore's law regarding computing costs – is that every new drug reaching market costs at least $1 billion[104] (and according to some estimates, up to $3-5 billion USD) to develop.[105-107] The reasons for this failure of pharmaceutical R&D have been discussed widely. Even when considering other reasons for drugs not reaching market, such as strategic, commercial or operational aspects, there seems general agreement that the main obstacle lies in lack of efficacy rather than safety issues.[108-111] This may be particularly true for drugs failing late in development (phase III trials), especially problematic from an economical and ethical standpoint.[109]

Many have argued that the main reason for the lack of efficacy is the limited predictive value of the preclinical models of disease that have dominated the discovery and evaluation of therapeutic

targets.[112] The dramatic advances in human genetics over the past decade offer the prospect for improved assessment of drug development efficacy in early pharmaceutical R&D through provision of a stronger evidence base for linking perturbation of a given target or pathway to a desirable phenotypic outcome in humans. In addition, human genetics can offer an opportunity to detect less welcome consequences of perturbation that may presage safety or toxicity concerns. There are already several examples in the public domain demonstrating these concepts, and more are likely to follow given the current reallocation of resources to human genetics at a growing number of pharmaceutical companies.

Retrospective proof-of-principle examples of this concept (where the drug was already in use before the genetic finding was made) include *PPARG*[113] and thiazoladinediones,[114] as well as *HMGCR*[115] and statins.[116] There are also several recent examples where genetic studies have predicted the outcome of an ongoing randomized controlled trial, even though the genetic finding may not have been the basis for initiating drug development. These include the reports of inactivating variants in *NPC1L1*[117] that correctly predicted the clinical effect of ezetimibe on recurrent cardiovascular events,[118]; and of loss-of-function (LoF) variants in *APOC3*[119] that predicted a clinical effect of ISIS 304801, an antisense inhibitor of APOC3 synthesis, on triglyceride levels.[120]

However, there are also several emerging prospective examples where the genetic discovery preceded and underpinned the development of the drug. These include variation in *CFTR* (causing cystic fibrosis)[121] and ivacaftor (a CFTR potentiator);[122]; and, from the cardiovascular area, the well-known example of PCSK9 inhibitors.[123-127] The *CFTR*-ivacaftor gene-drug pair illustrates a time-frame typical for pharmaceutical R&D, with over 20 years from initial scientific discovery to commercial availability of a drug. In contrast, the PCSK9 story highlights how the new era in human genetics can accelerate this process. The first study describing a gain-of-function variant in *PCSK9* causing hypercholesterolemia was published in 2003[125]: it was rapidly followed by landmark studies that described loss-of-function alleles, particularly in African-Americans, that were associated with decreased low-density lipoprotein cholesterol (LDL-C)[124] and, importantly, protection against coronary heart disease (CHD).[123] After a tight head-to-head competition between several pharmaceutical companies, two antibody-based PCSK9 inhibitors - evolocumab and alirocumab - were approved by the Food and Drug Administration (FDA) and European Medicines Agency (EMA) in 2015, barely a decade after the initial genetic discovery. It would be misleading to claim that the *PCSK9* target was discovered using GWAS – on the contrary, it was mapped using a traditional family-based design – but we argue that it is likely that there will be many other examples arising from GWAS that can be developed into successful drugs.

Currently, many hope that large-scale exome sequencing will provide new leads for development of drugs treating T2D and other diseases. Certainly, an advantage of exome sequencing is the focus on exonic variants, which makes the translation easier than with lead variants from GWAS, as discussed in detail in the previous section. An example lies in the study of loss-of-function alleles in the *SLC30A8* gene (encoding an islet-specific zinc transporter): these have been associated with decreased risk of T2D and highlight ZnT8 inhibition as a potential novel strategy for T2D therapy.[128] On the other hand, most exome sequencing studies so far have been too small to provide much additional value beyond previous GWAS. As recently demonstrated for T2D, most variants discovered on exome sequencing were common and/or located in regions already identified by GWAS.[33]

It is important therefore not to overlook the enormous potential in already existing knowledge about loci robustly associated with T2D and other diseases that have been generated through the past decade of GWAS. Presumably, there are many pointers to novel therapeutic regimens among these loci: the challenge is to systematically work through this treasure trove of associations to find them, starting with establishing causal genes in the loci, followed by careful characterization of the downstream mechanisms to understand underlying biology and to identify druggable mechanisms.[89, 129] There are strong indications that GWAS loci are enriched with genes that are successfully targeted with existing drugs,[130, 131] and that a drug development pipeline built on support from human genetics will be more efficient and lead to improvements in clinical development success.[131, 132] For example, the proportion of drug targets with genetic support from GWAS or the Online Mendelian Inheritance in Man (OMIM) database increases significantly across stages in the drug development pipeline; from ~2% in the preclinical stage to ~8% for approved drugs – a proportion that is even higher for drugs treating diabetes and metabolic diseases.[131]

Even if the often-cited *PCSK9* example may represent particularly low-hanging fruit, there is no good reason to believe that human genetics as a drug discovery vehicle should be less efficient for obesity and T2D than for lipid-lowering treatment. Indeed, there are several examples of GWAS loci with genes encoding targets of already established drugs against T2D.[71, 133] In addition to the already mentioned *PPARG*-thiazoladinediones[113, 114] pair; other examples of GWAS hit-drug combinations include *ABCC8*-sulfonylurea[134, 135] and *GLP1R*-GLP-1 agonists.[136, 137] Although these genes and corresponding drugs were established before the GWAS era, a necessity given the lead time in drug development, such re-identification of druggable targets illustrate the promise of GWAS for discovery of new targets that can be entered into the drug development pipeline.

Conclusions

Ten years have passed since the first round of GWAS for T2D and obesity. This decade has been transformative for human genetics, and has already had a major impact on the understanding of these diseases. We have learned much about their genetic architecture, including the high degree of polygenicity and tiny effect sizes of most risk variants, which have major implications for precision medicine. Several mechanistic challenges complicate the translation of novel loci, but overcoming these will expose untapped potential for the development of novel drugs to treat T2D and obesity. In parallel, there is the expectation that we will move to a more informed, molecular taxonomy of disease that provides a more precise framework for the delivery of preventative and therapeutic interventions. There is a clear need for increased focus on translational and mechanistic research to make sense of the hundreds of loci associated with these diseases, but this work can be conducted with the knowledge that, provided we can find ways to better capitalize on them, the rewards will be momentous.

Conflict of Interest Disclosures

Dr. Ingelsson is a scientific advisor for Precision Wellness and Olink Proteomics for work unrelated to the present project. Dr McCarthy is a scientific advisor for Pfizer and NovoNordisk and in receipt of research funding from Pfizer, NovoNordisk, EliLilly, Janssen, Roche, Merck, Takeda, Sanofi-Aventis, Servier, Boehringer Ingelheim and AstraZeneca.

References

1. Scott RA, Scott LJ, Magi R, Marullo L, Gaulton KJ, Kaakinen M, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*. 2017;66:2888-2902. doi: 10.2337/db16-1253.
2. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *bioRxiv*. 2018. doi: 10.1101/245506.
3. Lu Y, Day FR, Gustafsson S, Buchkovich ML, Na J, Bataille V, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat Commun*. 2016;7:10495. doi: 10.1038/ncomms10495.
4. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197-206. doi: 10.1038/nature14177.
5. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518:187-196. doi: 10.1038/nature14132.
6. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, et al. Correction: The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet*. 2016;12:e1006166. doi: 10.1371/journal.pgen.1006166.

7. Turcot V, Lu Y, Highland HM, Schurmann C, Justice AE, Fine RS, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet*. 2018;50:26-41. doi: 10.1038/s41588-017-0011-x.

8. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. *bioRxiv*. 2018. doi: 10.1101/304030.

9. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921. doi: 10.1038/35057062.

10. O'Donnell CJ and Nabel EG. Cardiovascular genomics, personalized medicine, and the National Heart, Lung, and Blood Institute: part I: the beginning of an era. *Circ Cardiovasc Genet*. 2008;1:51-7. doi: 10.1161/CIRCGENETICS.108.813337.

11. Collins FS and McKusick VA. Implications of the Human Genome Project for medical science. *JAMA*. 2001;285:540-4. doi, https://www.ncbi.nlm.nih.gov/pubmed/11176855

12. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661-78. doi: 10.1038/nature05911.

13. Zhu Z, Bakshi A, Vinkhuyzen AA, Hemani G, Lee SH, Nolte IM, et al. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am J Hum Genet*. 2015;96:377-85. doi: 10.1016/j.ajhg.2015.01.001.

14. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46:1173-86. doi: 10.1038/ng.3097.

15. Hemani G, Yang J, Vinkhuyzen A, Powell JE, Willemsen G, Hottenga JJ, et al. Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *Am J Hum Genet*. 2013;93:865-75. doi: 10.1016/j.ajhg.2013.10.005.

16. Boyle EA, Li YI and Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017;169:1177-1186. doi: 10.1016/j.cell.2017.05.038.

17. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med*. 2008;359:2220-32. doi: 10.1056/NEJMoa0801869.

18. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med*. 2008;359:2208-19. doi: 10.1056/NEJMoa0804742.

19. Vassy JL, Hivert MF, Porneala B, Dauriz M, Florez JC, Dupuis J, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes*. 2014;63:2172-82. doi: 10.2337/db13-1663.

20. Vassy JL, Durant NH, Kabagambe EK, Carnethon MR, Rasmussen-Torvik LJ, Fornage M, et al. A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. *Diabetologia*. 2012;55:2604-2612. doi: 10.1007/s00125-012-2637-7.

21. Vassy JL and Meigs JB. Is genetic testing useful to predict type 2 diabetes? *Best Pract Res Clin Endocrinol Metab*. 2012;26:189-201. doi: 10.1016/j.beem.2011.09.002.

22. Khera AV, Chaffin M, Aragam K, Emdin CA, Klarin D, Haas M, et al. Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. *bioRxiv*. 2017. doi: 10.1101/218388.

23. Knowles JW and Ashley EA. Cardiovascular disease: The rise of the genetic risk score. *PLoS Med*. 2018;15:e1002546. doi: 10.1371/journal.pmed.1002546.

24. Natarajan P, Young R, Stitziel NO, Padmanabhan S, Baber U, Mehran R, 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*. 2017;135:2091-2101. doi: 10.1161/CIRCULATIONAHA.116.024436.

25. Markowitz SM, Park ER, Delahanty LM, O'Brien KE and Grant RW. Perceived impact of diabetes genetic risk testing among patients at high phenotypic risk for type 2 diabetes. *Diabetes Care*. 2011;34:568-73. doi: 10.2337/dc10-1960.

26. Tarini BA, Herman WH and Lee JM. Effects of hypothetical type 2 diabetes genetic testing on parents' efforts to prevent diabetes in children. *Clin Pediatr (Phila)*. 2013;52:821-8. doi: 10.1177/0009922813488644.

27. Li SX, Ye Z, Whelan K and Truby H. The effect of communicating the genetic risk of cardiometabolic disorders on motivation and actual engagement in preventative lifestyle modification and clinical outcome: a systematic review and meta-analysis of randomised controlled trials. *Br J Nutr*. 2016;116:924-34. doi: 10.1017/S0007114516002488.

28. Grant RW, O'Brien KE, Waxler JL, Vassy JL, Delahanty LM, Bissett LG, et al. Personalized genetic risk counseling to motivate diabetes prevention: a randomized trial. *Diabetes Care*. 2013;36:13-9. doi: 10.2337/dc12-0884.

29. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747-53. doi: 10.1038/nature08494.

30. Lupski JR, Belmont JW, Boerwinkle E and Gibbs RA. Clan genomics and the complex architecture of human disease. *Cell*. 2011;147:32-43. doi: 10.1016/j.cell.2011.09.008.

31. Dickson SP, Wang K, Krantz I, Hakonarson H and Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol*. 2010;8:e1000294. doi: 10.1371/journal.pbio.1000294.

32. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat Genet*. 2018;50:559-571. doi: 10.1038/s41588-018-0084-1.

33. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536:41-47. doi: 10.1038/nature18642.

34. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet*. 2015;47:1114-20. doi: 10.1038/ng.3390.

35. Agarwala V, Flannick J, Sunyaev S, Go TDC and Altshuler D. Evaluating empirical bounds on complex disease genetic architecture. *Nat Genet*. 2013;45:1418-27. doi: 10.1038/ng.2804.

36. Pal A, Potjer TP, Thomsen SK, Ng HJ, Barrett A, Scharfmann R, et al. Loss-of-Function Mutations in the Cell-Cycle Control Gene CDKN2A Impact on Glucose Homeostasis in Humans. *Diabetes*. 2016;65:527-33. doi: 10.2337/db15-0602.

37. Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature*. 2014;512:190-3. doi: 10.1038/nature13425.

38. Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet*. 2014;46:294-8. doi: 10.1038/ng.2882.

39. Ahlqvist E, Storm P, Karajamaki A, Martinell M, Dorkhan M, Carlsson A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*. 2018;6:361-369. doi: 10.1016/S2213-8587(18)30051-2.

40. McCarthy MI. Painting a new picture of personalised medicine for diabetes. *Diabetologia*. 2017;60:793-799. doi: 10.1007/s00125-017-4210-x.

41. Florez JC. Pharmacogenetics in type 2 diabetes: precision medicine or discovery tool? *Diabetologia*. 2017;60:800-807. doi: 10.1007/s00125-017-4227-1.

42. Hattersley AT and Patel KA. Precision diabetes: learning from monogenic diabetes. *Diabetologia*. 2017;60:769-777. doi: 10.1007/s00125-017-4226-2.

43. Florez JC. The pharmacogenetics of metformin. *Diabetologia*. 2017. doi: 10.1007/s00125-017-4335-y.

44. Maruthur NM, Gribble MO, Bennett WL, Bolen S, Wilson LM, Balakrishnan P, et al. The pharmacogenetics of type 2 diabetes: a systematic review. *Diabetes Care*. 2014;37:876-86. doi: 10.2337/dc13-1276.

45. Zhou K, Yee SW, Seiser EL, van Leeuwen N, Tavendale R, Bennett AJ, et al. Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. *Nat Genet*. 2016;48:1055-1059. doi: 10.1038/ng.3632.

46. Franks PW, Pearson E and Florez JC. Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. *Diabetes Care*. 2013;36:1413-21. doi: 10.2337/dc12-2211.

47. Li SX, Imamura F, Ye Z, Schulze MB, Zheng J, Ardanaz E, et al. Interaction between genes and macronutrient intake on the risk of developing type 2 diabetes: systematic review and findings from European Prospective Investigation into Cancer (EPIC)-InterAct. *Am J Clin Nutr*. 2017;106:263-275. doi: 10.3945/ajcn.116.150094.

48. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized Nutrition by Prediction of Glycemic Responses. *Cell*. 2015;163:1079-1094. doi: 10.1016/j.cell.2015.11.001.

49. Wang T, Heianza Y, Sun D, Huang T, Ma W, Rimm EB, et al. Improving adherence to healthy dietary patterns, genetic risk, and long term weight gain: gene-diet interaction analysis in two prospective cohort studies. *BMJ*. 2018;360:j5644. doi: 10.1136/bmj.j5644.

50. Qi Q, Chu AY, Kang JH, Huang J, Rose LM, Jensen MK, et al. Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. *BMJ*. 2014;348:g1610. doi: 10.1136/bmj.g1610.

51. Qi Q, Chu AY, Kang JH, Jensen MK, Curhan GC, Pasquale LR, et al. Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med*. 2012;367:1387-96. doi: 10.1056/NEJMoa1203039.

52. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. *PLoS Genet*. 2017;13:e1006528. doi: 10.1371/journal.pgen.1006528.

53. DNA Fit. https://www.dnafit.com/us/.

54. DNActiv8. https://www.dnactiv8.co.uk/.
55. Vitagene. https://vitagene.com/.
56. Helix. https://www.helix.com/.
57. Orig3n. https://orig3n.com/. Accessed 08/24/2017.
58. Vinome. https://www.vinome.com/.
59. Marmite Gene Project. https://www.marmite.co.uk/geneproject. Accessed 09/14/2017.
60. Soccer Genetics. https://www.soccergenomics.com/home/.
61. Evans DM, Brion MJ, Paternoster L, Kemp JP, McMahon G, Munafo M, et al. Mining the human phenome using allelic scores that index biological intermediates. *PLoS Genet*. 2013;9:e1003919. doi: 10.1371/journal.pgen.1003919.
62. Small KS, Todorcevic M, Civelek M, El-Sayed Moustafa JS, Wang X, Simon MM, et al. Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat Genet*. 2018;50:572-580. doi: 10.1038/s41588-018-0088-x.
63. Ashley EA. Towards precision medicine. *Nat Rev Genet*. 2016;17:507-22. doi: 10.1038/nrg.2016.86.
64. Sagen JV, Raeder H, Hathout E, Shehadeh N, Gudmundsson K, Baevre H, et al. Permanent neonatal diabetes due to mutations in KCNJ11 encoding Kir6.2: patient characteristics and initial response to sulfonylurea therapy. *Diabetes*. 2004;53:2713-8. doi, https://www.ncbi.nlm.nih.gov/pubmed/15448106
65. Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, Slingerland AS, et al. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med*. 2004;350:1838-49. doi: 10.1056/NEJMoa032922.
66. Montague CT, Farooqi IS, Whitehead JP, Soos MA, Rau H, Wareham NJ, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*. 1997;387:903-8. doi: 10.1038/43185.
67. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9:356-69. doi: 10.1038/nrg2344.
68. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, Sparso T, Holmkvist J, Marchand M, et al. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet*. 2009;41:89-94. doi: 10.1038/ng.277.
69. Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, et al. Variants in MTNR1B influence fasting glucose levels. *Nat Genet*. 2009;41:77-81. doi: 10.1038/ng.290.
70. Lyssenko V, Nagorny CL, Erdos MR, Wierup N, Jonsson A, Spegel P, et al. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet*. 2009;41:82-8. doi: 10.1038/ng.288.
71. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Magi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet*. 2015;47:1415-25. doi: 10.1038/ng.3437.

72. Dimas AS, Lagou V, Barker A, Knowles JW, Magi R, Hivert MF, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes*. 2014;63:2158-71. doi: 10.2337/db13-0949.

73. Tuomi T, Nagorny CLF, Singh P, Bennet H, Yu Q, Alenkvist I, et al. Increased Melatonin Signaling Is a Risk Factor for Type 2 Diabetes. *Cell Metab*. 2016;23:1067-1077. doi: 10.1016/j.cmet.2016.04.009.

74. Mulder H. Melatonin signalling and type 2 diabetes risk: too little, too much or just right? *Diabetologia*. 2017;60:826-829. doi: 10.1007/s00125-017-4249-8.

75. DIAbetes Genetics Replication Meta-analysis C, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, South Asian Type 2 Diabetes Consortium, Mexican American Type 2 Diabetes Consortium and Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples Consortium. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet*. 2014;46:234-44. doi: 10.1038/ng.2897.

76. Bell CG, Finer S, Lindgren CM, Wilson GA, Rakyan VK, Teschendorff AE, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS One*. 2010;5:e14040. doi: 10.1371/journal.pone.0014040.

77. Lappalainen T, Montgomery SB, Nica AC and Dermitzakis ET. Epistatic selection between coding and regulatory variation in human evolution and disease. *Am J Hum Genet*. 2011;89:459-63. doi: 10.1016/j.ajhg.2011.08.004.

78. Schmitt AD, Hu M and Ren B. Genome-wide mapping and analysis of chromosome architecture. *Nat Rev Mol Cell Biol*. 2016;17:743-755. doi: 10.1038/nrm.2016.104.

79. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648-60. doi: 10.1126/science.1262110.

80. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet*. 2010;6:e1000895. doi: 10.1371/journal.pgen.1000895.

81. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007;316:889-94. doi: 10.1126/science.1141634.

82. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38:320-3. doi: 10.1038/ng1732.

83. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491-3. doi: 10.1126/science.1142842.

84. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488-91. doi: 10.1126/science.1142447.

85. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med*. 2015;373:895-907. doi: 10.1056/NEJMoa1502214.

86. Stratigopoulos G, Burnett LC, Rausch R, Gill R, Penn DB, Skowronski AA, et al. Hypomorphism of Fto and Rpgrip1l causes obesity in mice. *J Clin Invest*. 2016;126:1897-910. doi: 10.1172/JCI85526.

87. Church C, Moir L, McMurray F, Girard C, Banks GT, Teboul L, et al. Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet*. 2010;42:1086-92. doi: 10.1038/ng.713.

88. Civelek M, Wu Y, Pan C, Raulerson CK, Ko A, He A, et al. Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am J Hum Genet*. 2017;100:428-443. doi: 10.1016/j.ajhg.2017.01.027.

89. Thomsen SK and Gloyn AL. Human genetics as a model for target validation: finding new therapies for diabetes. *Diabetologia*. 2017;60:960-970. doi: 10.1007/s00125-017-4270-y.

90. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc Natl Acad Sci U S A*. 2014;111:13924-9. doi: 10.1073/pnas.1402665111.

91. Bonnefond A, Clement N, Fawcett K, Yengo L, Vaillant E, Guillaume JL, et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat Genet*. 2012;44:297-301. doi: 10.1038/ng.1053.

92. Bonnefond A and Froguel P. The case for too little melatonin signalling in increased diabetes risk. *Diabetologia*. 2017;60:823-825. doi: 10.1007/s00125-017-4255-x.

93. Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho SW, et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet*. 2017;49:1602-1612. doi: 10.1038/ng.3963.

94. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*. 2016;13:919-922. doi: 10.1038/nmeth.3999.

95. Kumasaka N, Knights AJ and Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat Genet*. 2016;48:206-13. doi: 10.1038/ng.3467.

96. Buenrostro JD, Wu B, Chang HY and Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*. 2015;109:21 29 1-9. doi: 10.1002/0471142727.mb2129s109.

97. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015;161:1202-1214. doi: 10.1016/j.cell.2015.05.002.

98. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*. 2014;343:776-9. doi: 10.1126/science.1247651.

99. Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*. 2016;167:1867-1882 e21. doi: 10.1016/j.cell.2016.11.048.

100. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. 2016;167:1853-1866 e17. doi: 10.1016/j.cell.2016.11.038.

101. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*. 2017;14:297-301. doi: 10.1038/nmeth.4177.

102. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell*. 2016;167:1883-1896 e15. doi: 10.1016/j.cell.2016.11.039.

103. Xie S, Duan J, Li B, Zhou P and Hon GC. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol Cell*. 2017;66:285-299 e5. doi: 10.1016/j.molcel.2017.03.007.

104. Scannell JW, Blanckley A, Boldon H and Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov*. 2012;11:191-200. doi: 10.1038/nrd3681.

105. Herper M. How Much Does Pharmaceutical Innovation Cost? A Look At 100 Companies. Forbes 2013; http://www.forbes.com/sites/matthewherper/2013/08/11/the-cost-of-inventing-a-new-drug-98-companies-ranked/.

106. DiMasi J. Tufts CSDD Cost Study. 2015; http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study.

107. DiMasi JA, Grabowski HG and Hansen RW. The cost of drug development. *N Engl J Med*. 2015;372:1972. doi: 10.1056/NEJMc1504317.

108. Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov*. 2014;13:419-31. doi: 10.1038/nrd4309.

109. Arrowsmith J and Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. *Nat Rev Drug Discov*. 2013;12:569. doi: 10.1038/nrd4090.

110. DiMasi JA, Feldman L, Seckler A and Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther*. 2010;87:272-7. doi: 10.1038/clpt.2009.295.

111. Morgan P, Van Der Graaf PH, Arrowsmith J, Feltner DE, Drummond KS, Wegner CD, et al. Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug Discov Today*. 2012;17:419-24. doi: 10.1016/j.drudis.2011.12.020.

112. Plenge RM, Scolnick EM and Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*. 2013;12:581-94. doi: 10.1038/nrd4051.

113. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet*. 2000;26:76-80. doi: 10.1038/79216.

114. Lehmann JM, Moore LB, Smith-Oliver TA, Wilkison WO, Willson TM and Kliewer SA. An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). *J Biol Chem*. 1995;270:12953-6. doi, http://www.ncbi.nlm.nih.gov/pubmed/7768881

115. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature genetics*. 2008;40:189-97. doi: 10.1038/ng.75.

116. Endo A. The discovery and development of HMG-CoA reductase inhibitors. *J Lipid Res*. 1992;33:1569-82. doi, http://www.ncbi.nlm.nih.gov/pubmed/1464741

117. Myocardial Infarction Genetics Consortium I, Stitziel NO, Won HH, Morrison AC, Peloso GM, Do R, et al. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med*. 2014;371:2072-82. doi: 10.1056/NEJMoa1405386.

118. Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, et al. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. *N Engl J Med*. 2015;372:2387-97. doi: 10.1056/NEJMoa1410489.

119. Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, Lange LA, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371:22-31. doi: 10.1056/NEJMoa1307095.

120. Gaudet D, Alexander VJ, Baker BF, Brisson D, Tremblay K, Singleton W, et al. Antisense Inhibition of Apolipoprotein C-III in Patients with Hypertriglyceridemia. *N Engl J Med*. 2015;373:438-47. doi: 10.1056/NEJMoa1400283.

121. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*. 1989;245:1066-73. doi, http://www.ncbi.nlm.nih.gov/pubmed/2475911

122. Wainwright CE, Elborn JS, Ramsey BW, Marigowda G, Huang X, Cipolli M, et al. Lumacaftor-Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR. *N Engl J Med*. 2015;373:220-31. doi: 10.1056/NEJMoa1409547.

123. Cohen JC, Boerwinkle E, Mosley TH, Jr. and Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med*. 2006;354:1264-72. doi: 10.1056/NEJMoa054013.

124. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK and Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet*. 2005;37:161-5. doi: 10.1038/ng1509.

125. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet*. 2003;34:154-6. doi: 10.1038/ng1161.

126. Sabatine MS, Giugliano RP, Wiviott SD, Raal FJ, Blom DJ, Robinson J, et al. Efficacy and Safety of Evolocumab in Reducing Lipids and Cardiovascular Events. *N Engl J Med*. 2015. doi: 10.1056/NEJMoa1500858.

127. Robinson JG, Farnier M, Krempf M, Bergeron J, Luc G, Averna M, et al. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *N Engl J Med*. 2015;372:1489-99. doi: 10.1056/NEJMoa1501031.

128. Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burtt NP, et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet*. 2014;46:357-63. doi: 10.1038/ng.2915.

129. Grotz AK, Gloyn AL and Thomsen SK. Prioritising Causal Genes at Type 2 Diabetes Risk Loci. *Curr Diab Rep*. 2017;17:76. doi: 10.1007/s11892-017-0907-y.

130. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med*. 2017;9. doi: 10.1126/scitranslmed.aag1166.

131. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856-60. doi: 10.1038/ng.3314.

132. Hurle MR, Nelson MR, Agarwal P and Cardon LR. Impact of genetically supported target selection on R&D productivity. *Nat Rev Drug Discov*. 2016. doi: 10.1038/nrd.2016.187.

133. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44:981-90. doi: 10.1038/ng.2383.

134. Flanagan SE, Clauin S, Bellanne-Chantelot C, de Lonlay P, Harries LW, Gloyn AL, et al. Update of mutations in the genes encoding the pancreatic beta-cell K(ATP) channel subunits Kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism. *Hum Mutat*. 2009;30:170-80. doi: 10.1002/humu.20838.

135. Nagashima K, Takahashi A, Ikeda H, Hamasaki A, Kuwamura N, Yamada Y, et al. Sulfonylurea and non-sulfonylurea hypoglycemic agents: pharmachological properties and tissue selectivity. *Diabetes Res Clin Pract*. 2004;66 Suppl 1:S75-8. doi: 10.1016/j.diabres.2003.12.011.

136. Maiorino MI, Chiodini P, Bellastella G, Capuano A, Esposito K and Giugliano D. Insulin and Glucagon-Like Peptide 1 Receptor Agonist Combination Therapy in Type 2 Diabetes: A Systematic Review and Meta-analysis of Randomized Controlled Trials. *Diabetes Care*. 2017;40:614-624. doi: 10.2337/dc16-1957.

137. Potts JE, Gray LJ, Brady EM, Khunti K, Davies MJ and Bodicoat DH. The Effect of Glucagon-Like Peptide 1 Receptor Agonists on Weight Loss in Type 2 Diabetes: A Systematic Review and Mixed Treatment Comparison Meta-Analysis. *PLoS One*. 2015;10:e0126769. doi: 10.1371/journal.pone.0126769.

Figure. Drawing the wrong conclusions from clustering of cross-sectional data. Your alien space craft crashes near a ski resort. In an effort to understand the diversity of human life, you apply your favorite clustering algorithm to the lifeforms visible through the window. You conclude that there are four subtypes of human life with very distinctive and non-overlapping appearances and behaviors. However, as you await rescue, you realize that this clustering was only transient, and an artefact arising from the restriction to a single cross-sectional view. Over time, you observe that almost all lifeforms progress through all four stages at various points, and that membership of a given cluster at a single point in time provides very little insight into long-term behavior.