OPEN ACCESS

Emerging roles of rare and low-frequency genetic variants in type 1 diabetes mellitus

Haipeng Pang, Ying Xia, Shuoming Luo, Gan Huang, Xia Li, Zhiguo Xie, Zhiguang Zhou

National Clinical Research Center for Metabolic Diseases, Key Laboratory of Diabetes Immunology (Central South University), Ministry of Education, and Department of Metabolism and Endocrinology, The Second Xiangya Hospital of Central South University, Changsha, Hunan, China

Correspondence to
Professor Zhiguang Zhou, Department of Metabolism and Endocrinology, Second Xiangya Hospital, Changsha 410011, China; zhouzhiguang@csu.edu.cn and Professor Zhiguo Xie, Department of Metabolism and Endocrinology, The Second Xiangya Hospital of Central South University, Changsha, China; xiezhiguo@csu.edu.cn

Received 18 July 2020
Revised 6 January 2021
Accepted 10 January 2021
Published Online First 22 March 2021

To cite: Pang H, Xia Y, Luo S, *et al*. *J Med Genet* 2021;**58**:289–296.

ABSTRACT

Type 1 diabetes mellitus (T1DM) is defined as an autoimmune disorder and has enormous complexity and heterogeneity. Although its precise pathogenic mechanisms are obscure, this disease is widely acknowledged to be precipitated by environmental factors in individuals with genetic susceptibility. To date, the known susceptibility loci, which have mostly been identified by genome-wide association studies, can explain 80%–85% of the heritability of T1DM. Researchers believe that at least a part of its missing genetic component is caused by undetected rare and low-frequency variants. Most common variants have only small to modest effect sizes, which increases the difficulty of dissecting their functions and restricts their potential clinical application. Intriguingly, many studies have indicated that rare and low-frequency variants have larger effect sizes and play more significant roles in susceptibility to common diseases, including T1DM, than common variants do. Therefore, better recognition of rare and low-frequency variants is beneficial for revealing the genetic architecture of T1DM and for providing new and potent therapeutic targets for this disease. Here, we will discuss existing challenges as well as the great significance of this field and review current knowledge of the contributions of rare and low-frequency variants to T1DM.

INTRODUCTION

Currently, type 1 diabetes mellitus (T1DM) is defined as an autoimmune disorder classically characterised by pancreatic islet beta-cell destruction triggered by autoreactive T cells, resulting in subsequent severe insulin deficiency and lifelong reliance on exogenous insulin.[1 2] This autoimmune diabetes accounts for 5%–19% of diabetes and represents the main form of diabetes in children and adolescents.[3] Its incidence is increasing worldwide at a rate of 2%–5% per year.[4] This rising incidence and multiple severe diabetic complications lead to increased mortality and morbidity and aggravate the economic burden of the disease. It is accepted that the interplay between genetic factors and environmental precipitators, including ancestry and geographic location, viral and bacterial infections, vitamin D, hygiene and microbiota, leads to specific tissue inflammation, namely, insulitis, insulin-producing cell death and consequent clinical disease.[5–9]

The genetic component of T1DM can be demonstrated by the fact that siblings and offspring of patients with T1DM have a higher risk than the general population, and disease concordance in identical twins is higher than that in dizygotic twins.[10 11] Over the past few years, genome-wide association study (GWAS), which measures and analyses a million or more DNA sequence variations in known linkage regions in unrelated individuals, have identified at least 58 susceptible loci combined with linkage analysis and candidate gene studies (figure 1).[12–14] Most of the identified variants are common (minor allele frequency (MAF) >5%) and have modest effects (OR <1.5), although the effects of susceptibility genes such as human leucocyte antigen (*HLA*), insulin (*INS*) and protein tyrosine phosphatase, non-receptor type 22 (*PTPN22*) are stronger (figure 1).[13] The *HLA* region (OR >6), located on human chromosome 6p21 and identified by linkage analysis, accounts for the largest proportion of T1DM heritability and explains approximately 50% of genetic T1DM risk.[15] In addition to *HLA*, variants within the *INS* and *PTPN22* loci, which were first identified by candidate gene studies, have larger effect sizes (OR >2) than other variants.[13] The *INS* gene on human chromosome 11p15.5 offers the next strongest genetic risk association with T1DM after *HLA* and accounts for approximately 10% of genetic susceptibility to T1DM.[16] It is believed that 'missing heritability' can be at least partially elucidated by rare and low-frequency variants (rare variants defined as variants with MAF ≤1% and low-frequency variants defined as variants with MAF=1%–5%), and some findings have indicated that rare variants have larger effect sizes than common variants.[17–19] From an evolutionary standpoint, risk variants with higher penetrance are more likely to be rare due to negative selection. Taking an extreme example, monogenic/Mendelian disorders such as autoimmune polyendocrinopathy syndrome type I are caused by rare variants with large effect sizes and high penetrance. Intriguingly, recent and previous studies focusing on the identification of rare and low-frequency variants involved in T1DM have found a handful of such variants, and some of them do have large effect sizes.[13 20–23]

However, some studies suggest that most rare variants have only small or modest effects.[24] Therefore, it remains to be seen whether the tendency of rare and low-frequency variants to have large effects is a universal phenomenon. Even though its practical value in clinical medicine may be restricted if the hypothesis that most rare variants have only a small effect is true, there is still intrinsic value in this field. Such studies can lead to the discovery of new candidate genes implicated in disorders or human phenotypes[25] and determine causal genes

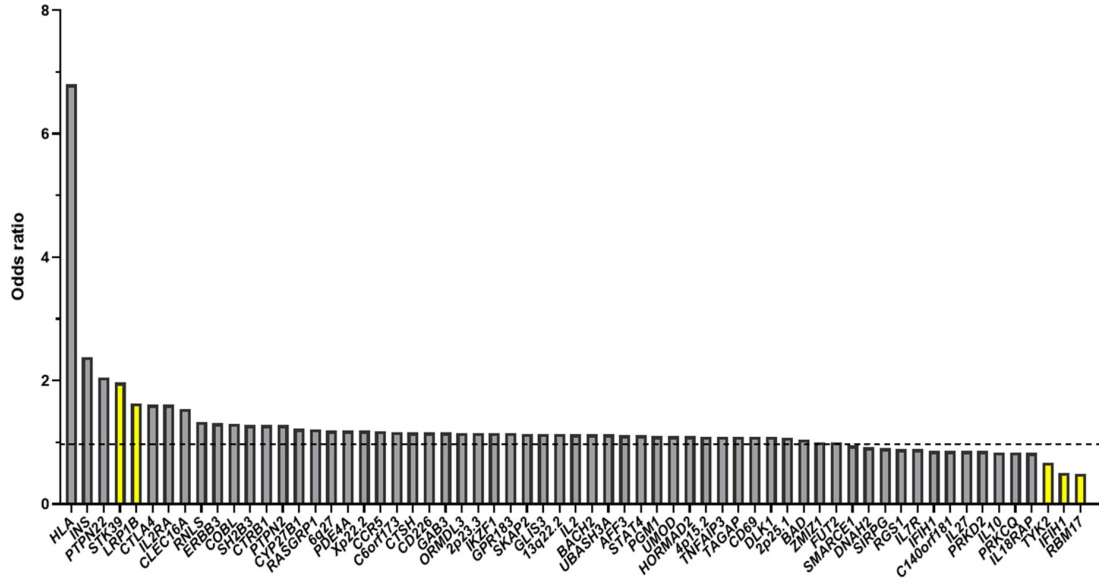

Figure 1 Candidate genes or loci of type 1 diabetes mellitus (T1DM) and their ORs (the yellow bars represent the rare and low-frequency genetic variants of T1DM).[76–79]

in candidate regions identified by GWAS. Other than understanding better its pathophysiology, new loci could lead to the identification of new biomarkers or represent drug targets for T1DM.

Identifying rare and low-frequency variants

Recently, advances in next-generation DNA sequencing technologies as well as bioinformatic tools and methods to process and analyse the resulting data have enhanced the ability of researchers to find rare variants, and the decreasing cost of these technologies has made it feasible to apply them to related studies (table 1).[26] The most comprehensive approach is high-depth whole-genome sequencing (WGS) due to its excellent coverage. However, high costs and multiple computational challenges have restricted its application.[21] In addition to WGS with high or low depth, SNP-array genome-wide genotyping and imputation has been used to identify rare variants. Notably, current sequencing depth (especially 30x) of WGS is likely to miss at least some coding variants as compared with whole-exome sequencing (WES, especially >100x).

There are some lower-cost alternatives as well. First, a combination of low-depth WGS and imputation is another choice. Imputation is a statistical method that can determine genotypes that are not directly detected by taking advantage of various previously sequenced reference panels. For instance, Martínez-Bueno and Alarcón-Riquelme identified rare variants that were jointly associated with systemic lupus erythematosus (SLE) within 98 SLE candidate genes by applying genome-wide imputation and other techniques.[27] Notably, some studies have indicated that the newer imputation panels, such as the recent Haplotype Reference Consortium panel and the combined UK10K and 1000 Genomes projects phase III, provide better quality of imputation

Table 1 Technologies and study designs for detecting rare variants

Strategy	Advantages	Disadvantages
High-depth WGS	Cover nearly all rare variants with high confidence.	1. High costs and computational challenges. 2. Miss some coding variants as compared with WES.
Low-depth WGS and imputation	Cost-effective compared with high-depth WGS.	1. Limited accuracy for rare variants. 2. Decreasing accuracy with same number of subjects compared with high-depth WGS.
WES	1. Less expensive. 2. Identify all variants resides in exomic regions. 3. Easily interpreted.	Ignore non-coding regions which account for large proportion of genome.
Targeted sequencing	Cost-effective.	Fail to identify disease-associated rare variants in some studies.
SNP-array genotyping with imputation	Low costs.	Lower accuracy for imputed rare variants.
Extreme phenotype sampling	Boosts power to find rare variants.	1. Requires statistical analysis to remove sampling bias. 2. Difficult to generalise to the wider population. 3. The results may be sensitive to outliers and sampling bias.
Population isolates	1. Lacks phenotypic variability due to cultural and environmental homogeneity. 2. Higher frequency of rare variants resulting from reduced genetic diversity and increased genetic drift.	Risk-conferring variants may be extremely rare and monomorphic due to lack of genetic diversity.
Family studies	1. Detect mutations that underlie Mendelian diseases successfully. 2. Improve statistical power significantly.	Less powerful than case-control designs for common diseases.

WES, whole-exome sequencing; WGS, whole-genome sequencing.

for rare variants compared with early panel, such as the UK10K, which underlines the significance and potential of larger reference panels to impute rare variants.[28 29] Nevertheless, the power of imputation for identifying rare variants is attenuated because its accuracy decreases with decreasing MAF. Additionally, studies have indicated that the utility of population-specific panels leads to improved imputation accuracy of rare variants.[30] Therefore, the utilisation of imputation is relatively limited in non-European populations because of the lack of ethnicity-specific reference cohorts.

Second, using WES finds rare variants within protein-coding regions. Given the reality that only an exceedingly small portion of the human genome is coding sequence and the functions of protein-coding variants are more easily interpreted, WES is considered a cost-effective technique for discovering rare variants. However, an obvious defect is that WES ignores non-coding regions, which account for 98% of the human genome. Moreover, most loci identified by GWAS are located in non-coding regions, and evidence indicates that these regions play critical roles in complex disorders and have significant biological functions.[31 32]

Third, targeted sequencing investigates a specific part of the genome, including candidate genes identified by previous studies and clinically significant genes. For instance, Rivas *et al* identified a protein-truncating variant of the gene *RNF186* that can exert a protective effect against ulcerative colitis via changed localisation and decreased expression by conducting targeted sequencing in regions previously associated with inflammatory bowel disease; they found that this loss-of-function variant was a promising therapeutic target.[33] However, some targeted sequencing studies have failed to detect rare risk variants, indicating the deficiency of this method in discovering rare and low-frequency variants.[24 34]

In addition, burden tests, which collapse information for multiple variants into a single genetic score and analyse the association between the score and disease characteristic, are a common approach in genomics to potentialise identification of rare variants, because aggregating analysis of variants within a gene can improve the power to detect statistical signals between case and control subjects. For example, a study analysed WES data from 393 patients with idiopathic hypogonadotropic hypogonadism (IHH) against 123 136 control subjects from public sequencing database, and identified a significant burden in *TYRO3*, a candidate gene implicated in IHH in mouse models.[35] However, this gene-based burden testing approach will lose power when effects of variants are not in the same direction or the causal variants only account for a small fraction.[36]

Traditional genetic studies have focused mostly on DNA sequences collected from unrelated individuals. However, a variety of new study designs have been applied to finding rare variants with the goal of decreasing sample sizes and costs. The common feature of these designs, including extreme phenotype sampling, population isolates and family studies (table 1), is that they improve the power of rare variant testing by selecting a specific population.[37–39]

Challenges for identifying rare and low-frequency variants

The detection and analysis of rare and low-frequency variants constitute a rising research field, but this field has encountered substantial obstacles and challenges. First, the statistical analysis of rare and low-frequency variants is far more complicated and difficult than the analysis of common variants. For example, because the number of rare variants is greater than the number of common variants, the significance threshold or p value established for GWAS is not appropriate for rare variant association studies.[40] The linkage disequilibrium (LD) r^2 between two rare variants or a common variant and a rare variant cannot be accurately calculated, and as such it is difficult to define if novel rare variants are independent from known rare or common variants.[41 42] A variety of traditional methods used to reduce or eliminate confounding factors and population stratification, such as linear mixed effect models and principal components analysis, are not applicable to the analysis of rare and low-frequency variants because rare variants and the distribution of disease risk are strictly localised. A study indicates that the estimated ancestry scores can be used to control the population stratification if the pool of control is large. Also, off-targeted read might be applied for controlling population stratification in targeted sequencing.[43] Moreover, because these variants are rare, the strategy used to analyse common variants, which is based on analysing a single variant at a time, is underpowered to detect rare variants and can do so only if the effect size or sample size is exceedingly large.[44] Thus, alternative methods have been developed to analyse the aggregate effect of rare variants.[45–47] These methods, such as burden tests, variance component test and exponential combination tests, evaluate association for multiple variants in a gene or a biologically region. Combined analysis of genetic association data with other biological information, such as methylation, gene expression and biological pathways, can also leads to substantial gain In the statistical power of rare variants studies.[48–50]

Second, it still remains challenging to apply genetic information obtained by rare variants association studies to diagnostic and prognostic medicine because some healthy individuals carry deleterious variants. For example, Flannick *et al* found that a large portion of the general population carries low-frequency non-synonymous mutations that can change the length or sequence of coding proteins in maturity-onset diabetes of young genes, and these carriers remain normoglycaemic through middle age.[51] In addition, Bick *et al* discovered that rare variants in sarcomere protein genes could boost the risk of adverse cardiovascular events in Framingham Heart Study participants, and more surprisingly, a large number of non-synonymous variants, including nonsense, missense and splice variants, are present in healthy populations.[52] Therefore, the functional validation of rare and low-frequency genetic variants is necessary to determine the causality in genotype-phenotype analysis.

Third, many rare and low-frequency variants are geographically localised and population specific, so it is difficult to find suitable replication panels and generate a common population. Nelson *et al* sequenced 202 drug target genes in coding regions in 14 002 people and found that 95% of observed variants are rare and at least 74% are detected in only one or two individuals.[53] Similarly, a study conducted in 2440 individuals of African and European ancestry found that 86% of over 500 000 variants identified are rare, and most are previously unknown.[54] Notably, these studies indicate that the vast majority of rare variant allelic spectra are unique to their sample sets and need to be identified by direct resequencing.

Finally, although some detection studies of rare and low-frequency variants, such as WES and data processing software, are relatively standardised, many aspects of this emerging field, including WES capture technologies and even the definition of rare variants, still do not have uniform standards. Therefore, combining data generated from different groups is problematic.

Table 2 Rare and low-frequency variants associated with T1DM, T2DM and other autoimmune diseases

Disease	Gene/Locus	Method	Reference
T1DM	*IFIH1*	Candidate gene sequencing	23
	PTPN22	Targeted deep sequencing	22
	STK39 LRP1B	Deep imputation of genotyped data	13
	TYK2 RBM17	Fine mapping of T1DM risk loci	20
T2DM	*CCND2 PAM PDX1*	WGS and imputation	80
	Xq23 EHMT2	Reanalysis of data from GWAS	81
	GLP1R G6PC2	HumanExome BeadChip	82
	SGSM2 MADD TBC1D30 KANK1	Analysis of exome array data	83
	PPARG	Targeted gene sequencing	84
RA	*IL2RA IL2RB*	Candidate gene sequencing	85
	TYK2	Immunochip	86
SLE	*BLK*	Genotyping	87
	TREXI	Genotyping	88
IBD	*IL23R*	Candidate gene sequencing	89
	NOD2	Candidate gene sequencing	90 91
	PRDM1	WES	92
	CARD9 RNF186	Targeted gene sequencing	93

GWAS, genome-wide association study; IBD, inflammatory bowel disease; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; WES, whole-exome sequencing; WGS, whole-genome sequencing.

Benefits of identifying rare and low-frequency variants

It has been suggested that rare and low-frequency variants account for a large proportion of the genetic variation in the human genome represented by the 1000 Genomes Project.[55 56] Although a substantial number of SNPs have been identified by GWAS, there is still a so-called 'missing heritability' phenomenon in complex disorders.[57] For instance, GWAS have identified >80 common variants with small effect sizes for T2DM, which can explain only 10% of the total heritability.[58] To address this issue, several hypotheses have been proposed, and great technological advances have provided a better understanding of the genetic architecture of common diseases over the past several years. Rare and low-frequency variants can influence both susceptibility to common complex diseases and their phenotypes (table 2).[59–62] For example, researchers performed WGS in 1038 pulmonary arterial hypertension (PAH, a rare disorder characterised by occlusion of arterioles in the lung) cases and 6385 control subjects and make the total proportion of cases explained by mutations increased to 23.5% from previously established 19.9% by incorporating novel rare variants and genes identified.[63] Also, a study indicated that rare variants of *SLC22A12* gene influence urate reabsorption and the heritability explained by these *SLC22A12* variants exceeds 10%, indicating that rare functional variants make substantial contribution to the 'missing heritability' of serum urate level.[64] In fact, a 'common disease-rare variant model' that assumes rare variants with high penetrance may be involved in increased complex disease risk has been proposed.[59 65] It is obvious that great genetic heterogeneity exists under this model. Intriguingly, in line with this model, some autoimmune diseases, such as T1DM, are extremely heterogeneous.

Besides rare and low-frequency genetic variants, there are some other hypotheses to explain the 'missing heritability'.[59] For example, empirical and theoretical analyses have indicated that multiple genetic variants with small effects are missed because GWAS are underpowered to capture these variants, therefore, taking into account genetic variants with smaller effects that do not reach significance will contribute to disease susceptibility and phenotype variability. Additionally, structural variants, such as CNV, are poorly studied owing to insufficient coverage on SNP chips.[66] The presence of gene-gene (epistasis) and gene-environmental interactions may also contribute to the 'missing heritability'.[67]

In addition, the candidate regions identified by GWAS sometimes harbour several different genes. Identifying rare genetic variants is helpful to pinpoint causal genes within the loci identified by GWAS.[68] Moreover, the identification of rare and low-frequency variants may result in the identification of new candidate genes.[40] For instance, researchers identified a heterozygote truncating mutation within *CLCN1* gene by performing WES in patients with statin-associated myopathy and therefore, determined a novel candidate gene of this disease.[69] Additionally, it has been suggested that rare variants are likely to have appeared more recently than common variants, leading to reduced LD and making them more easily interpretable than common variants.[21]

Moreover, early studies have indicated that rare and low-frequency genetic variants may have larger effects on complex disease phenotypes and susceptibility than common variants.[70] Therefore, it is helpful to reveal the genetic pathways underlying diseases and to provide clinically actionable targets for personalised medicine. As an example, Roth *et al* found that rare and low-frequency genetic variants with large phenotypic effects within the proprotein convertase subtilisin/kexin 9 (*PCSK9*) gene, which encodes products that bind to the low-density lipoprotein (LDL) receptor and increase its degradation, can lower the risk of coronary heart disease (CHD) by reducing the circulating level of LDL cholesterol.[71] Based on this research, a fully human monoclonal antibody targeting PCSK9 has been proven to increase LDL receptor recycling and decrease LDL cholesterol level.[72] These findings provide a new treatment and prevention strategy for hypercholesterolaemia and CHD and offer inspiration for the transformation of genetic discoveries into clinical practice.

Rare and low-frequency variants and T1DM

Focusing on autoimmune diabetes, fully understanding the genetic factors underlying T1DM is beneficial for revealing its pathophysiology, discovering new drug targets and developing predictive and personalised medicine (figure 2). It is especially vital and valuable because T1DM is extremely complex and heterogeneous. The candidate T1DM loci identified by GWAS sometimes contain several distinct genes, and strong LD makes it difficult to pinpoint the precise causative genes in genomic regions. In addition, the fact that many SNPs reside in non-coding regions or do not have obvious functional effects offers few clues to ascertain the causative genes. However, the discovery of rare and low-frequency disease-associated variants is helpful for T1DM candidate gene identification. The T1DM-associated region on human chromosome 2q24 harbours interferon (IFN) induced with helicase C domain 1 (*IFIH1*), *GCA*, *FAP* and part of *KCNH7*. The interaction between *IFIH1* and double-stranded RNA, a byproduct of viral replication, leads to the secretion of IFNs. While *IFIH1* is a plausible susceptibility gene on the basis of its biological function, there is no direct evidence to indicate which of these genes in this locus is responsible for increased T1DM risk. Nejentsev *et al* resequenced the exons and splice sites of 10 candidate genes in pools of DNA from 480 patients and 480 controls and discovered 4 rare or low-frequency variants

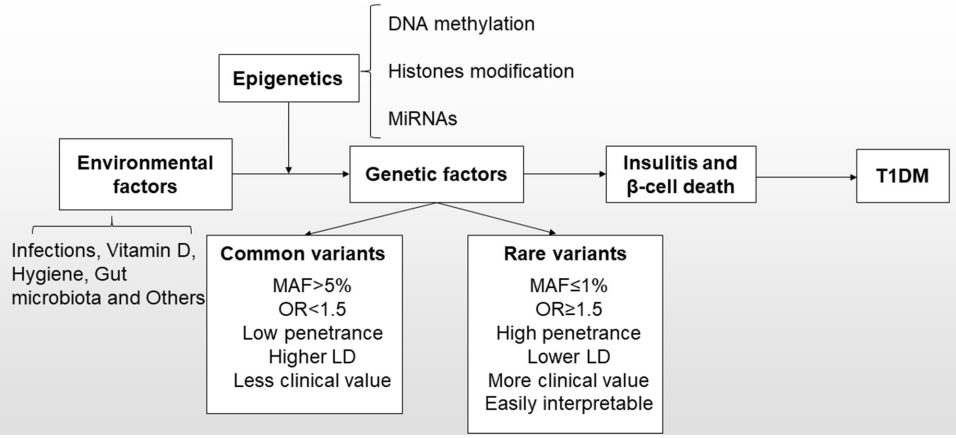

Figure 2 The development of type 1 diabetes mellitus (T1DM). T1DM is caused by interplay between genetic and environmental factors, and epigenetics serves as a bridge between the two. To date, >50 candidate loci have been identified by genome-wide association study. The genetic variants within these risk regions can be divided into common variants, low-frequency variants and rare variants according to their different minor allele frequencies. The rare and low-frequency variants are likely to have more practical value in the treatment of T1DM because their ORs are larger than those of common variants. However, as the study of rare and low-frequency variants is an emerging research field, some hypotheses are still controversial and need further investigation. LD, linkage disequilibrium; MAF. minor allele frequency.

(OR=0.51–0.74, MAF <3%) with low LD within *IFIH1* that could change the structure or expression of its product, melanoma differentiation-associated protein 5 and protect against T1DM.[23] This finding suggests that *IFIH1* is the disease-causing gene. Moreover, Ge *et al* found several rare deleterious variants, including two novel frameshift mutations (ss538819444 and ss37186329) and two missense mutations (rs74163663 and rs56048322) within *PTPN22* by deeply sequencing the protein-coding regions of 301 genes in 49 loci previously identified by GWAS in 70 T1DM cases of European ancestry.[22] This finding further confirmed that *PTPN22* is a T1DM candidate gene on chromosome 1p13.2. Subsequent genotyping in 3609 families with T1DM indicated rs56048322 (MAF=0.87%), which leads to the production of two alternative *PTPN22* transcripts and a novel isoform of its encoding protein, LYP, through affecting splicing of *PTPN22*, was significantly associated with T1DM independent of T1DM-associated common variant rs2476601. Functional analysis showed this isoform of LYP can cause hyporesponsiveness of $CD4^+$ T cell to antigen stimulation in patients with T1DM.

Additionally, as mentioned above, most variants that confer T1DM risk are common and have modest effects, limiting the clinical application of their discovery. However, some research has suggested that rare and low-frequency variants might have larger effect sizes than common variants. Theoretically, if a disorder affects reproduction, such as an autoimmune disease with early onset, genetic variants with strong effects will be maintained at a relatively low frequency through negative selection.[21] Forgetta *et al* applied deep imputation of genotyped data in 9358 patients with T1DM and 15 705 controls from European cohorts to identify novel rare and low-frequency variants with large effect sizes on T1DM risk.[13] Three novel rare and low-frequency variants, including rs192324744 in LDL receptor-related protein 1B (*LRP1B*, MAF=1.3%, OR=1.63), rs60587303 in serine threonine kinase 39 (*STK39*, MAF=0.5%, OR=1.97) and the intergenic variant rs2128344 (MAF=0.55%, OR=2.12), were found and validated by subsequent de novo genotyping.[13] Notably, the effects of these SNPs (ORs ≥1.5) are comparable to those of the lead variants in *INS* and *PTPN22*. In vitro experiments indicated that *STK39* is involved in T cell activation and effector functions and that inhibition of Stk39 can augment the inflammatory response by enhancing interleukin (IL)-2 signalling; therefore, *STK39* may be a promising clinical intervention target.[13]

Besides, previous study through fine mapping of known T1DM susceptible loci has identified a low-frequency variant rs34536443 (MAF=4%, OR=0.67) within tyrosine kinase 2 (*TYK2*) and a rare variant rs41295121 (MAF=1%, OR=0.49) within RNA binding motif protein 17 (*RBM17*, in the same locus as *IL2RA*).[20] *TYK2*, belonging to Janus kinase (JAK) family, is associated with regulation of type I IFN signalling pathway. Some studies have demonstrated that rs34530443 plays protective roles in multiple autoimmune disorders and the underlying mechanisms might lie in the diminishment of IL-12, IL-23 and type I IFN signalling.[73] The specific function of rs41295121 in context of autoimmunity and T1DM needs further investigation.

As for some practical issues such as sample sizes and high costs, a study indicated that a well-powered rare variant association study should include discovery sets with at least 25 000 cases and a substantial replication set.[44] There are some alternative methods to decrease the sample sizes or costs in the context of T1DM. For example, combined analysis of rare variants within a T1DM-associated gene or region can lead to substantial reduction of required sample sizes. In addition, preferential selection of individuals with extreme phenotype on the basis of known risk factors, including age of disease onset, family history of diabetes and diabetic auto-antibodies, can also improve the association power because rare variants might be enriched among them.[74]

Overall, among the identified T1DM loci, the candidate genes with rare or low-frequency variants include *TYK2*, *IFIH1*, *RBM17*, *PTPN22*, *STK39* and *LRP1B*.[13 20 22 23] Many unidentified variants may remain to be dissected, because studies focused on other diseases suggest that rare and low-frequency variants account for the majority of all variants.[27 75]

CONCLUSION

Driven by advancements in sequencing technologies, there has been great improvement in the identification of rare and low-frequency variants that cause complex human diseases, such as T1DM. The benefits of this field can be stated as follows: (1) characterisation of rare and low-frequency variants may lead to

a full understanding of the genetic component of this disorder; (2) detection of rare and low-frequency variants can pinpoint the genes that are actually responsible for increased T1DM risk within the loci identified by GWAS; (3) some new candidate genes for T1DM can be found due to enhanced power to discover rare variants; (4) rare and low-frequency variants are expected to make a significant contribution to human phenotypes and disease susceptibility because some studies indicate the majority of protein-coding variants tend to be evolutionarily recent and rare[54]; (5) accumulated evidence indicates that rare and low-frequency variants have larger phenotypic effects than common variants, suggesting that they will offer more actionable clinical targets and hold tremendous promise in predictive and personalised medicine.

However, some issues remain to be addressed. First, controversy persists about the importance of rare and low-frequency variants in common diseases. Encouragingly, recent studies have found that some such variants, such as rs60587303 in *STK39*, indeed have larger effect sizes than common variants in the pathogenesis of T1DM. Second, the candidate genes for T1DM that have rare or low-frequency variants included only *TYK2*, *RBM17*, *IFIH1*, *PTPN22*, *STK39* and *LRP1B*, which means there may still be many unidentified variants. Moreover, most studies in this field have examined European populations. However, rare and low-frequency variants are geographically localised and population specific. In particular, the heritable background of T1DM varies among different ethnic groups. These facts will limit the practical application of rare and low-frequency variants.

In conclusion, the identification of rare and low-frequency genetic variants will provide new insights into the pathophysiology of T1DM and offer new potential drug targets in the post-GWAS era, despite the many challenges and uncertainties remaining in this field.

Contributors HP searched references, wrote the first draft of the paper and revised the text. YX, SL, GH and XL critically revised the text and provided substantial scientific contributions. ZX and ZZ proposed the project and revised the manuscript. All the authors approved the final version of the manuscript.

Funding This work was supported by the National Natural Science Foundation of China (grant numbers 82070813, 81873634, 81400783), the National Key R&D Program of China (grant numbers 2016YFC1305000, 2016YFC1305001, 2018YFC1315603), the Science and Technology Major Project of Hunan Province (grant number 2017SK1020), Hunan Province Natural Science Foundation of China (grant numbers 2018JJ2573, 2020JJ2053).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Zhiguang Zhou http://orcid.org/0000-0002-4451-4906

REFERENCES

1 DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. *Lancet* 2018;391:2449–62.
2 Sun X, Pang H, Li J, Luo S, Huang G, Li X, Xie Z, Zhou Z. The NLRP3 inflammasome and its role in T1DM. *Front Immunol* 2020;11:1595.
3 American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2009;32 Suppl 1:S62–7.
4 Tuomilehto J. The emerging global epidemic of type 1 diabetes. *Curr Diab Rep* 2013;13:795–804.
5 Giwa AM, Ahmed R, Omidian Z, Majety N, Karakus KE, Omer SM, Donner T, Hamad ARA. Current understandings of the pathogenesis of type 1 diabetes: genetics to environment. *World J Diabetes* 2020;11:13–25.
6 Tremblay J, Hamet P. Environmental and genetic contributions to diabetes. *Metabolism* 2019;100S.
7 Zheng P, Li Z, Zhou Z. Gut microbiome in type 1 diabetes: a comprehensive review. *Diabetes Metab Res Rev* 2018;34:e3043.
8 Xie Z, Chang C, Zhou Z. Molecular mechanisms in autoimmune type 1 diabetes: a critical review. *Clin Rev Allergy Immunol* 2014;47:174–92.
9 Wang Z, Xie Z, Lu Q, Chang C, Zhou Z. Beyond genetics: what causes type 1 diabetes. *Clin Rev Allergy Immunol* 2017;52:273–86.
10 Lee HS, Hwang JS. Genetic aspects of type 1 diabetes. *Ann Pediatr Endocrinol Metab* 2019;24:143–8.
11 Jerram ST, Leslie RD. The genetic architecture of type 1 diabetes. *Genes* 2017;8. doi:10.3390/genes8080209. [Epub ahead of print: 22 Aug 2017].
12 Størling J, Pociot F. Type 1 diabetes candidate genes linked to pancreatic islet cell inflammation and beta-cell apoptosis. *Genes* 2017;8. doi:10.3390/genes8020072. [Epub ahead of print: 16 Feb 2017].
13 Forgetta V, Manousaki D, Istomine R, Ross S, Tessier M-C, Marchand L, Li M, Qu H-Q, Bradfield JP, Grant SFA, Hakonarson H, Paterson AD, Piccirillo C, Polychronakos C, Richards JB, DCCT/EDIC Research Group. Rare genetic variants of large effect influence risk of type 1 diabetes. *Diabetes* 2020;69:784–95.
14 Pang H, Luo S, Huang G, Xia Y, Xie Z, Zhou Z. Advances in knowledge of candidate genes acting at the beta-cell level in the pathogenesis of T1DM. *Front Endocrinol* 2020;11:119.
15 Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen W-M, Howson JMM, Todd JA, de Bakker PIW, Rich SS, Raychaudhuri S. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet* 2015;47:898–905.
16 Steck AK, Rewers MJ. Genetics of type 1 diabetes. *Clin Chem* 2011;57:176–85.
17 Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 2005;37:161–5.
18 Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004;305:869–72.
19 Groop L, Pociot F. Genetics of diabetes--are we missing the genes or the disease? *Mol Cell Endocrinol* 2014;382:726–39.
20 Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, Achuthan P, Guo H, Fortune MD, Stevens H, Walker NM, Ward LD, Kundaje A, Kellis M, Daly MJ, Barrett JC, Cooper JD, Deloukas P. Type 1 diabetes genetics C, Todd JA, Wallace C, Concannon P, rich SS. fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* 2015;47:381–6.
21 Sazonovs A, Barrett JC. Rare-Variant studies to complement genome-wide association studies. *Annu Rev Genomics Hum Genet* 2018;19:97–112.
22 Ge Y, Onengut-Gumuscu S, Quinlan AR, Mackey AJ, Wright JA, Buckner JH, Habib T, Rich SS, Concannon P. Targeted deep sequencing in Multiple-Affected sibships of European ancestry identifies rare deleterious variants in PTPN22 that confer risk for type 1 diabetes. *Diabetes* 2016;65:794–802.
23 Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009;324:387–9.
24 Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, Barrett JC, Blackburn H, Brand O, Burren O, Capon F, Compston A, Gough SCL, Jostins L, Kong Y, Lee JC, Lek M, MacArthur DG, Mansfield JC, Mathew CG, Mein CA, Mirza M, Nutland S, Onengut-Gumuscu S, Papouli E, Parkes M, Rich SS, Sawcer S, Satsangi J, Simmonds MJ, Trembath RC, Walker NM, Wozniak E, Todd JA, Simpson MA, Plagnol V, van Heel DA. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* 2013;498:232–5.
25 Auer PL, Teumer A, Schick U, O'Shaughnessy A, Lo KS, Chami N, Carlson C, de Denus S, Dubé M-P, Haessler J, Jackson RD, Kooperberg C, Perreault L-PL, Nauck M, Peters U, Rioux JD, Schmidt F, Turcot V, Völker U, Völzke H, Greinacher A, Hsu L, Tardif J-C, Diaz GA, Reiner AP, Lettre G. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nat Genet* 2014;46:629–34.
26 Dolled-Filhart MP, Lee M, Ou-Yang C-wen, Haraksingh RR, Lin JC-H. Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. *ScientificWorldJournal* 2013;2013:1–10.
27 Martínez-Bueno M, Alarcón-Riquelme ME. Exploring impact of rare variation in systemic lupus erythematosus by a genome wide imputation approach. *Front Immunol* 2019;10:258.
28 Chou W-C, Zheng H-F, Cheng C-H, Yan H, Wang L, Han F, Richards JB, Karasik D, Kiel DP, Hsu Y-H. A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples. *Sci Rep* 2016;6:39313.
29 Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A, Evenden L, Gibson J, Amador C, Howard DM, Navarro P, Morris A, Deary IJ, Hocking LJ, Padmanabhan S, Smith BH, Joshi P, Wilson JF, Hastie ND, Wright AF, McIntosh AM, Porteous DJ, Haley CS, Vitart V, Hayward C. Exploration of haplotype research Consortium imputation

for genome-wide association studies in 20,032 generation Scotland participants. *Genome Med* 2017;9:23.
30 Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, Ripatti S, Morris AP, Metspalu A, Esko T, Mägi R, Palta P. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet* 2017;25:869–76.
31 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
32 Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F, Parkinson H. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005–12.
33 Rivas MA, Graham D, Sulem P, Stevens C, Desch AN, Goyette P, Gudbjartsson D, Jonsdottir I, Thorsteinsdottir U, Degenhardt F, Mucha S, Kurki MI, Li D, D'Amato M, Annese V, Vermeire S, Weersma RK, Halfvarson J, Paavola-Sakki P, Lappalainen M, Lek M, Cummings B, Tukiainen T, Haritunians T, Halme L, Koskinen LLE, Ananthakrishnan AN, Luo Y, Heap GA, Visschedijk MC, MacArthur DG, Neale BM, Ahmad T, Anderson CA, Brant SR, Duerr RH, Silverberg MS, Cho JH, Palotie A, Saavalainen P, Kontula K, Färkkilä M, McGovern DPB, Franke A, Stefansson K, Rioux JD, Xavier RJ, Daly MJ, Barrett J, de Lane K, Edwards C, Hart A, Hawkey C, Jostins L, Kennedy N, Lamb C, Lee J, Lees C, Mansfield J, Mathew C, Mowatt C, Newman B, Nimmo E, Parkes M, Pollard M, Prescott N, Randall J, Rice D, Satsangi J, Simmons A, Tremelling M, Uhlig H, Wilson D, Abraham C, Achkar JP, Bitton A, Boucher G, Croitoru K, Fleshner P, Glas J, Kugathasan S, Limbergen JV, Milgrom R, Proctor D, Regueiro M, Schumm PL, Sharma Y, Stempak JM, Targan SR, Wang MH. A protein-truncating R179X variant in RNF186 confers protection against ulcerative colitis. *Nat Commun* 2016;7:12342.
34 Tang H, Jin X, Li Y, Jiang H, Tang X, Yang X, Cheng H, Qiu Y, Chen G, Mei J, Zhou F, Wu R, Zuo X, Zhang Y, Zheng X, Cai Q, Yin X, Quan C, Shao H, Cui Y, Tian F, Zhao X, Liu H, Xiao F, Xu F, Han J, Shi D, Zhang A, Zhou C, Li Q, Fan X, Lin L, Tian H, Wang Z, Fu H, Wang F, Yang B, Huang S, Liang B, Xie X, Ren Y, Gu Q, Wen G, Sun Y, Wu X, Dang L, Xia M, Shan J, Li T, Yang L, Zhang X, Li Y, He C, Xu A, Wei L, Zhao X, Gao X, Xu J, Zhang F, Zhang J, Li Y, Sun L, Liu J, Chen R, Yang S, Wang J, Zhang X. A large-scale screen for coding variants predisposing to psoriasis. *Nat Genet* 2014;46:45–50.
35 Guo MH, Plummer L, Chan Y-M, Hirschhorn JN, Lippincott MF. Burden testing of rare variants identified through exome sequencing via publicly available control data. *Am J Hum Genet* 2018;103:522–34.
36 Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014;95:5–23.
37 Locke AE, Steinberg KM, Chiang CWK, Service SK, Havulinna AS, Stell L, Pirinen M, Abel HJ, Chiang CC, Fulton RS, Jackson AU, Kang CJ, Kanchi KL, Koboldt DC, Larson DE, Nelson J, Nicholas TJ, Pietilä A, Ramensky V, Ray D, Scott LJ, Stringham HM, Vangipurapu J, Welch R, Yajnik P, Yin X, Eriksson JG, Ala-Korpela M, Järvelin M-R, Männikkö M, Laivuori H, Dutcher SK, Stitziel NO, Wilson RK, Hall IM, Sabatti C, Palotie A, Salomaa V, Laakso M, Ripatti S, Boehnke M, Freimer NB, FinnGen Project. Exome sequencing of Finnish isolates enhances rare-variant association power. *Nature* 2019;572:323–8.
38 Bureau A, Begum F, Taub MA, Hetmanski JB, Parker MM, Albacha-Hejazi H, Scott AF, Murray JC, Marazita ML, Bailey-Wilson JE, Beaty TH, Ruczinski I. Inferring disease risk genes from sequencing data in multiplex pedigrees through sharing of rare variants. *Genet Epidemiol* 2019;43:37–49.
39 Peloso GM, Rader DJ, Gabriel S, Kathiresan S, Daly MJ, Neale BM. Phenotypic extremes in rare variant study designs. *Eur J Hum Genet* 2016;24:924–30.
40 Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. *Genome Med* 2015;7:16.
41 Marian AJ. Molecular genetic studies of complex phenotypes. *Transl Res* 2012;159:64–79.
42 Li B, Liu DJ, Leal SM. Identifying rare variants associated with complex traits via sequencing. *Curr Protoc Hum Genet* 2013;Chapter 1.
43 Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, Branham KE, Heckenlively J, Fulton R, Wilson RK, Mardis ER, Lin X, Swaroop A, Zöllner S, Abecasis GR, FUSION Study. Ancestry estimation and control of population stratification for sequence-based association studies. *Nat Genet* 2014;46:409–15.
44 Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* 2014;111:E455–64.
45 Ho Y-Y, Guan W, O'Connell M, Basu S. Powerful association test combining rare variant and gene expression using family data from genetic analysis workshop 19. *BMC Proc* 2016;10:251–5.
46 Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet* 2013;92:841–53.
47 Santorico SA, Hendricks AE. Progress in methods for rare variant association. *BMC Genet* 2016;17 Suppl 2:6.
48 Ho Y-Y, Baechler EC, Ortmann W, Behrens TW, Graham RR, Bhangale TR, Pan W. Using gene expression to improve the power of genome-wide association analysis. *Hum Hered* 2014;78:94–103.
49 Kao PYP, Leung KH, Chan LWC, Yip SP, Yap MKH. Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions. *Biochim Biophys Acta Gen Subj* 2017;1861:335–53.
50 Richardson TG, Timpson NJ, Campbell C, Gaunt TR. A pathway-centric approach to rare variant association analysis. *European Journal of Human Genetics* 2017;25:123–9.
51 Flannick J, Beer NL, Bick AG, Agarwala V, Molnes J, Gupta N, Burtt NP, Florez JC, Meigs JB, Taylor H, Lyssenko V, Irgens H, Fox E, Burslem F, Johansson S, Brosnan MJ, Trimmer JK, Newton-Cheh C, Tuomi T, Molven A, Wilson JG, O'Donnell CJ, Kathiresan S, Hirschhorn JN, Njølstad PR, Rolph T, Seidman JG, Gabriel S, Cox DR, Seidman CE, Groop L, Altshuler D. Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat Genet* 2013;45:1380–5.
52 Bick AG, Flannick J, Ito K, Cheng S, Vasan RS, Parfenov MG, Herman DS, DePalma SR, Gupta N, Gabriel SB, Funke BH, Rehm HL, Benjamin EJ, Aragam J, Taylor HA, Fox ER, Newton-Cheh C, Kathiresan S, O'Donnell CJ, Wilson JG, Altshuler DM, Hirschhorn JN, Seidman JG, Seidman C. Burden of rare sarcomere gene variants in the Framingham and Jackson heart study cohorts. *Am J Hum Genet* 2012;91:513–9.
53 Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu S-A, Fraser D, Warren L, Aponte J, Zawistowski M, Liu X, Zhang H, Zhang Y, Li J, Li Y, Li L, Woollard P, Topp S, Hall MD, Nangle K, Wang J, Abecasis G, Cardon LR, Zöllner S, Whittaker JC, Chissoe SL, Novembre J, Mooser V. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 2012;337:100–4.
54 Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 2012;337:64–9.
55 Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
56 Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
57 Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;11:446–50.
58 Stančáková A, Laakso M. Genetics of type 2 diabetes. *Endocr Dev* 2016;31:203–20.
59 Génin E. Missing heritability of complex diseases: case solved? *Hum Genet* 2020;139:103–13.
60 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010;42:565–9.
61 Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 2012;109:1193–8.
62 Wang X-J, Xu X-Q, Sun K, Liu K-Q, Li S-Q, Jiang X, Zhao Q-H, Wang L, Peng F-H, Ye J, Wu Y, Jiang R, Zhang J, Huang W, Wei W-B, Yan Y, Li J-H, Liu Q-Q, Li S, Wang Y, Zhang S-Y, Zhang X, Jing Z-C. Association of rare PTGIS variants with susceptibility and pulmonary vascular response in patients with idiopathic pulmonary arterial hypertension. *JAMA Cardiol* 2020;5.
63 Gräf S, Haimel M, Bleda M, Hadinnapola C, Southgate L, Li W, Hodgson J, Liu B, Salmon RM, Southwood M, Machado RD, Martin JM, Treacy CM, Yates K, Daugherty LC, Shamardina O, Whitehorn D, Holden S, Aldred M, Bogaard HJ, Church C, Coghlan G, Condliffe R, Corris PA, Danesino C, Eyries M, Gall H, Ghio S, Ghofrani H-A, Gibbs JSR, Girerd B, Houweling AC, Howard L, Humbert M, Kiely DG, Kovacs G, MacKenzie Ross RV, Moledina S, Montani D, Newnham M, Olschewski A, Olschewski H, Peacock AJ, Pepke-Zaba J, Prokopenko I, Rhodes CJ, Scelsi L, Seeger W, Soubrier F, Stein DF, Suntharalingam J, Swietlik EM, Toshner MR, van Heel DA, Vonk Noordegraaf A, Waisfisz Q, Wharton J, Wort SJ, Ouwehand WH, Soranzo N, Lawrie A, Upton PD, Wilkins MR, Trembath RC, Morrell NW. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat Commun* 2018;9:1416.
64 Misawa K, Hasegawa T, Mishima E, Jutabha P, Ouchi M, Kojima K, Kawai Y, Matsuo M, Anzai N, Nagasaki M. Contribution of Rare Variants of the *SLC22A12* Gene to the Missing Heritability of Serum Urate Levels. *Genetics* 2020;214:1079–90.
65 Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 2009;19:212–9.
66 Nakatochi M, Kushima I, Ozaki N. Implications of germline copy-number variations in psychiatric disorders: review of large-scale genetic studies. *J Hum Genet* 2021;66:25–37.
67 Ebbert MTW, Ridge PG, Kauwe JSK. Bridging the gap between statistical and biological epistasis in Alzheimer's disease. *Biomed Res Int* 2015;2015:1–7.
68 Lettre G. Rare and low-frequency variants in human common diseases and other complex traits. *J Med Genet* 2014;51:705–14.
69 Neřoldová M, Stránecký V, Hodaňová K, Hartmannová H, Piherová L, Přistoupilová A, Mrázová L, Vrablík M, Adámková V, Hubáček JA, Jirsa M, Kmoch S. Rare variants in known and novel candidate genes predisposing to statin-associated myopathy. *Pharmacogenomics* 2016;17:1405–14.

70 Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008;40:695–701.
71 Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 2006;354:1264–72.
72 Roth EM, McKenney JM, Hanotin C, Asset G, Stein EA. Atorvastatin with or without an antibody to PCSK9 in primary hypercholesterolemia. *N Engl J Med* 2012;367:1891–900.
73 Gorman JA, Hundhausen C, Kinsman M, Arkatkar T, Allenspach EJ, Clough C, West SE, Thomas K, Eken A, Khim S, Hale M, Oukka M, Jackson SW, Cerosaletti K, Buckner JH, Rawlings DJ. The *TYK2*-P1104A Autoimmune Protective Variant Limits Coordinate Signals Required to Generate Specialized T Cell Subsets. *Front Immunol* 2019;10:44.
74 Guey LT, Kravic J, Melander O, Burtt NP, Laramie JM, Lyssenko V, Jonsson A, Lindholm E, Tuomi T, Isomaa B, Nilsson P, Almgren P, Kathiresan S, Groop L, Seymour AB, Altshuler D, Voight BF. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genet Epidemiol* 2011;35:n/a–246.
75 Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, Teslovich TM, Caulkins L, Koesterer R, Barajas-Olmos F, Blackwell TW, Boerwinkle E, Brody JA, Centeno-Cruz F, Chen L, Chen S, Contreras-Cubas C, Cordova E, Correa A, Cortes M, DeFronzo RA, Dolan L, Drews KL, Elliott A, Floyd JS, Gabriel S, Garay-Sevilla ME, Garcia-Ortiz H, Gross M, Han S, Heard-Costa NL, Jackson AU, Jorgensen ME, Kang HM, Kelsey M, Kim BJ, Koistinen HA, Kuusisto J, Leader JB, Linneberg A, Liu CT, Liu J, Lyssenko V, Manning AK, Marcketta A, Malacara-Hernandez JM, Martinez-Hernandez A, Matsuo K, Mayer-Davis E, Mendoza-Caamal E, Mohlke KL, Morrison AC, Ndungu A, MCY N, O'Dushlaine C, Payne AJ, Pihoker C, Broad Genomics P, Post WS, Preuss M, Psaty BM, Vasan RS, Rayner NW, Reiner AP, Revilla-Monsalve C, Robertson NR, Santoro N, Schurmann C, WY S, Soberon X, Stringham HM, Strom TM, CHT T, Thameem F, Tomlinson B, Torres JM, Tracy RP, van Dam RM, Vujkovic M, Wang S, Welch RP, Witte DR, Wong TY, Atzmon G, Barzilai N, Blangero J, Bonnycastle LL, Bowden DW, Chambers JC, Chan E, Cheng CY, Cho YS, Collins FS, de Vries PS, Duggirala R, Glaser B, Gonzalez C, Gonzalez ME, Groop L, Kooner JS, Kwak SH, Laakso M, Lehman DM, Nilsson P, Spector TD, Tai ES, Tuomi T, Tuomilehto J, Wilson JG, Aguilar-Salinas CA, Bottinger E, Burke B, Carey DJ, Chan JCN, Dupuis J, Frossard P, Heckbert SR, Hwang MY, Kim YJ, Kirchner HL, Lee JY, Lee J, Loos RJF, Ma RCW MAD, O'Donnell CJ, Palmer CNA, Pankow J, Park KS, Rasheed A, Saleheen D, Sim X, Small KS, Teo YY, Haiman C, Hanis CL, Henderson BE, Orozco L, Tusie-Luna T, Dewey FE, Baras A, Gieger C, Meitinger T, Strauch K, Lange L, Grarup N, Hansen T, Pedersen O, Zeitler P, Dabelea D, Abecasis G, Bell GI, Cox NJ, Seielstad M, Sladek R, Meigs JB, Rich SS, Rotter JI, Discov E, Charge L. exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* 2019;570:71–6.
76 Johnson MB, Cerosaletti K, Flanagan SE, Buckner JH. Genetic mechanisms highlight shared pathways for the pathogenesis of polygenic type 1 diabetes and monogenic autoimmune diabetes. *Curr Diab Rep* 2019;19:20.
77 Bakay M, Pandey R, Grant SFA, Hakonarson H. The genetic contribution to type 1 diabetes. *Curr Diab Rep* 2019;19:116.
78 Fløyel T, Kaur S, Pociot F. Genes affecting β-cell function in type 1 diabetes. *Curr Diab Rep* 2015;15:97.
79 Pociot F, Akolkar B, Concannon P, Erlich HA, Julier C, Morahan G, Nierras CR, Todd JA, Rich SS, Nerup J. Genetics of type 1 diabetes: what's next? *Diabetes* 2010;59:1561–71.
80 Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, Helgadottir HT, Johannsdottir H, Magnusson OT, Gudjonsson SA, Justesen JM, Harder MN, Jørgensen ME, Christensen C, Brandslund I, Sandbæk A, Lauritzen T, Vestergaard H, Linneberg A, Jørgensen T, Hansen T, Daneshpour MS, Fallah M-S, Hreidarsson AB, Sigurdsson G, Azizi F, Benediktsson R, Masson G, Helgason A, Kong A, Gudbjartsson DF, Pedersen O, Thorsteinsdottir U, Stefansson K. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet* 2014;46:294–8.
81 Bonàs-Guarch S, Guindo-Martínez M, Miguel-Escalada I, Grarup N, Sebastian D, Rodriguez-Fos E, Sánchez F, Planas-Fèlix M, Cortes-Sánchez P, González S, Timshel P, Pers TH, Morgan CC, Moran I, Atla G, González JR, Puiggros M, Martí J, Andersson EA, Díaz C, Badia RM, Udler M, Leong A, Kaur V, Flannick J, Jørgensen T, Linneberg A, Jørgensen ME, Witte DR, Christensen C, Brandslund I, Appel EV, Scott RA, Luan Jian'an, Langenberg C, Wareham NJ, Pedersen O, Zorzano A, Florez JC, Hansen T, Ferrer J, Mercader JM, Torrents D. Re-Analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nat Commun* 2018;9:321.
82 Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA, Dauriz M, Hivert MF, Raghavan S, Lipovich L, Hidalgo B, Fox K, Huffman JE, An P, Lu Y, Rasmussen-Torvik LJ, Grarup N, Ehm MG, Li L, Baldridge AS, Stancakova A, Abrol R, Besse C, Boland A, Bork-Jensen J, Fornage M, Freitag DF, Garcia ME, Guo X, Hara K, Isaacs A, Jakobsdottir J, Lange LA, Layton JC, Li M, Hua Zhao J, Meidtner K, Morrison AC, Nalls MA, Peters MJ, Sabater-Lleal M, Schurmann C, Silveira A, Smith AV, Southam L, Stoiber MH, Strawbridge RJ, Taylor KD, Varga TV, Allin KH, Amin N, Aponte JL, Aung T, Barbieri C, Bihlmeyer NA, Boehnke M, Bombieri C, Bowden DW, Burns SM, Chen Y, Chen YD, Cheng CY, Correa A, Czajkowski J, Dehghan A, Ehret GB, Eiriksdottir G, Escher SA, Farmaki AE, Franberg M, Gambaro G, Giulianini F, Goddard WA. Low-Frequency and rare exome CHIP variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat Commun* 2015;6:5897.
83 Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stančáková A, Stringham HM, Sim X, Yang L, Fuchsberger C, Cederberg H, Chines PS, Teslovich TM, Romm JM, Ling H, McMullen I, Ingersoll R, Pugh EW, Doheny KF, Neale BM, Daly MJ, Kuusisto J, Scott LJ, Kang HM, Collins FS, Abecasis GR, Watanabe RM, Boehnke M, Laakso M, Mohlke KL. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet* 2013;45:197–201.
84 Majithia AR, Flannick J, Shahinian P, Guo M, Bray M-A, Fontanillas P, Gabriel SB. Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proc Natl Acad Sci U S A* 2014;111:13127–32.
85 Diogo D, Kurreeman F, Stahl EA, Liao KP, Gupta N, Greenberg JD, Rivas MA, Hickey B, Flannick J, Thomson B, Guiducci C, Ripke S, Adzhubey I, Barton A, Kremer JM, Alfredsson L. rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. *Am J Hum Genet* 2013;92:15–27.
86 Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, Zhernakova A, Stahl E, Viatte S, McAllister K, Amos CI, Padyukov L, Toes REM, Huizinga TWJ, Wijmenga C, Trynka G, Franke L, Westra H-J, Alfredsson L, Hu X, Sandor C, de Bakker PIW, Davila S, Khor CC, Heng KK, Andrews R, Edkins S, Hunt SE, Langford C, Symmons D, Concannon P, Onengut-Gumuscu S, Rich SS, Deloukas P, Gonzalez-Gay MA, Rodriguez-Rodriguez L, Ärlsetig L, Martin J, Rantapää-Dahlqvist S, Plenge RM, Raychaudhuri S, Klareskog L, Gregersen PK, Worthington J. High-Density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet* 2012;44:1336–40.
87 Delgado-Vega AM, Dozmorov MG, Quirós MB, Wu Y-Y, Martínez-García B, Kozyrev SV, Frostegård J, Truedsson L, de Ramón E, González-Escribano MF, Ortego-Centeno N, Pons-Estel BA, D'Alfonso S, Sebastiani GD, Witte T, Lauwerys BR, Endreffy E, Kovács L, Vasconcelos C, da Silva BM, Wren JD, Martin J, Castillejo-López C, Alarcón-Riquelme ME. Fine mapping and conditional analysis identify a new mutation in the autoimmunity susceptibility gene blk that leads to reduced half-life of the blk protein. *Ann Rheum Dis* 2012;71:1219–26.
88 Namjou B, Kothari PH, Kelly JA, Glenn SB, Ojwang JO, Adler A, Alarcón-Riquelme ME, Gallant CJ, Boackle SA, Criswell LA, Kimberly RP, Brown E, Edberg J, Stevens AM, Jacob CO, Tsao BP, Gilkeson GS, Kamen DL, Merrill JT, Petri M, Goldman RR, Vila LM, Anaya J-M, Niewold TB, Martin J, Pons-Estel BA, Sabio JM, Callejas JL, Vyse TJ, Bae S-C, Perrino FW, Freedman BI, Scofield RH, Moser KL, Gaffney PM, James JA, Langefeld CD, Kaufman KM, Harley JB, Atkinson JP. Evaluation of the TREX1 gene in a large multi-ancestral lupus cohort. *Genes Immun* 2011;12:270–9.
89 Momozawa Y, Mni M, Nakamura K, Coppieters W, Almer S, Amininejad L, Cleynen I, Colombel J-F, de Rijk P, Dewit O, Finkel Y, Gassull MA, Goossens D, Laukens D, Lémann M, Libioulle C, O'Morain C, Reenaers C, Rutgeerts P, Tysk C, Zelenika D, Lathrop M, Del-Favero J, Hugot J-P, de Vos M, Franchimont D, Vermeire S, Louis E, Georges M. Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. *Nat Genet* 2011;43:43–7.
90 Lesage S, Zouali H, Cézard J-P, Colombel J-F, Belaiche J, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, Finkel Y, Modigliani R, Gower-Rousseau C, Macry J, Merlin F, Chamaillard M, Jannot A-S, Thomas G, Hugot J-P, Group E-I, Group E, Group G. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;70:845–57.
91 Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, Boucher G, Ripke S, Ellinghaus D, Burtt N, Fennell T, Kirby A, Latiano A, Goyette P, Green T, Halfvarson J, Haritunians T, Korn JM, Kuruvilla F, Lagace C, Neale B, KS L, Schumm P, Torkvist L. deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011;43:1066–73.
92 Ellinghaus D, Zhang H, Zeissig S, Lipinski S, Till A, Jiang T, Stade B, Bromberg Y, Ellinghaus E, Keller A, Rivas MA, Skieceviciene J, Doncheva NT, Liu X, Liu Q, Jiang F, Forster M, Mayr G, Albrecht M, Häsler R, Boehm BO, Goodall J, Berzuini CR, Lee J, Andersen V, Vogel U, Kupcinskas L, Kayser M, Krawczak M, Nikolaus S, Weersma RK, Ponsioen CY, Sans M, Wijmenga C, Strachan DP, McArdle WL, Vermeire S, Rutgeerts P, Sanderson JD, Mathew CG, Vatn MH, Wang J, Nöthen MM, Duerr RH, Büning C, Brand S, Glas J, Winkelmann J, Illig T, Latiano A, Annese V, Halfvarson J, D'Amato M, Daly MJ, Nothnagel M, Karlsen TH, Subramani S, Rosenstiel P, Schreiber S, Parkes M, Franke A. Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. *Gastroenterology* 2013;145:339–47.
93 Beaudoin M, Goyette P, Boucher G, Lo KS, Rivas MA, Stevens C, Alikashani A, Ladouceur M, Ellinghaus D, Törkvist L, Goel G, Lagacé C, Annese V, Bitton A, Begun J, Brant SR, Bresso F, Cho JH, Duerr RH, Halfvarson J, McGovern DPB, Radford-Smith G, Schreiber S, Schumm PL, Sharma Y, Silverberg MS, Weersma RK. Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLoS Genet* 2013;9:e1003723.