Curr Cardiovasc Risk Rep (2015) 9: 47
DOI 10.1007/s12170-015-0475-0

DIABETES + INSULIN RESISTANCE (M RUTTER, SECTION EDITOR)

The Hunt for Low-Frequency Alleles Predisposing to Type 2 Diabetes and Related Cardiovascular Risk Factors

Lorena Boquete Vilariño[1] · **Timothy M. Frayling**[1]

Published online: 14 September 2015

Abstract Research into the genetic basis of cardiovascular-related diseases is moving at an extremely fast pace. Developments in technology such as whole-genome sequencing and massive resources of DNA collected from hundreds of thousands of people mean scientists have an unprecedented ability to discover the genetic variation that predisposes to disease. Before 2007, very little was known about the variation in the human DNA sequence and its influence on common diseases. We now know of hundreds of common variants that influence LDL cholesterol levels, type 2 diabetes, hypertension and heart disease to name a few. Attention has now turned to the discovery of the genetic variants that occur in between 1 in 20 and 1 in 1000 individuals. These variants are unlikely to cause disease in the same way that mutations in some genes cause a monogenic disorder with a particular pattern of inheritance. But variants in this frequency range will shed light on biological mechanisms of disease. In this review, we focus on these variants and discuss how a range of study designs have identified low-frequency genetic variants with stronger predisposing effects on type 2 diabetes and related traits than common genetic variants.

Keywords Type 2 diabetes · Genetics · Genetic variants · Single nucleotide polymorphism

This article is part of the Topical Collection on *Diabetes + Insulin Resistance*

✉ Timothy M. Frayling
T.M.Frayling@exeter.ac.uk

[1] University of Exeter Medical School, University of Exeter, RILD building, Barrack Road, Exeter EX2 5DW, UK

Introduction

Diabetes is one of the major risk factors for cardiovascular disease. The obesogenic environment, resulting in a much higher prevalence of obesity today than in previous generations, is the primary cause of the higher prevalence of type 2 diabetes today. However, two individuals with the same BMI can have very different risks of type 2 diabetes [1], and genetic factors affecting insulin secretion, insulin resistance and body fat distribution play a role. Since 2007, genome-wide association studies (GWAS) have identified more than 100 common genetic variants associated with type 2 diabetes and related glycaemic traits [2, 3]. Some of these common variant associations have provided first steps towards an increased biological understanding of diabetes and examples have been discussed in several previous reviews [4–8]. These common variants tend to have relatively small effect sizes on individual risk and cumulatively only explain a small fraction of the heritable component to type 2 diabetes. Here, we review the most recent findings, where studies using whole-genome sequencing, whole exome sequencing, exome-based micro arrays and other approaches have started to uncover variants with lower frequencies but larger effects on diabetes and its intermediate traits such as insulin resistance and insulin secretion (Table 1).

Whole-Genome Sequencing Combined with Imputation into Very Large Studies Identifies New Loci Associated with Type 2 Diabetes

Whole-genome sequencing potentially allows geneticists to identify and analyse the majority of human genetic variants, not just those captured on microarrays. However, whilst the cost of sequencing has fallen dramatically, it is still only

Table 1 Low-frequency or population-specific variants of larger effect on type 2 diabetes risk or with larger effects on diabetes related intermediate traits

Gene name	GWAS signal?	Strategy	Phenotype	Coding	Effect size	Allele frequency (%)	Type of study
CCND2	No	Whole-genome sequencing	T2D, FG, IS	No	OR=0.55	1.47	Case–control (11,114 cases and 267,140 controls)
MTNR1B	Yes	Exon resequencing	T2D	Yes A42P, L60R, P95L, Y308S	OR=3.88 (four variants)	0.22	Case–control (2186 cases and 5446 controls)
SLC30A8	Yes	Exon resequencing	T2D, non-fasting G	Yes (p.Arg138*)	OR=0.53	0.19	Multi-ethnic meta-analysis of case–control (~150,000 individuals)
PPARG	Yes	Exon resequencing	T2D	Yes (12 variants)	OR=7.22	Approx. 0.08	Multi-ethnic meta-analysis of case–control (9070 cases and 10,682 controls)
HNF1A	No, MODY3 gene	Whole exome	T2D (Latino population)	Yes (p.E508K)	OR=5.48	1.0	Case–control (1794 cases and 1962 controls)
PAM/PPIP5K2	No	Exome chip	IS	Yes (p.Asp563Gly in PAM, p.Ser1228Gly in PPIP5K2)	−0.21±0.04 s.d.	5.3	8229 non-diabetic individuals
GLP1R	No	Exome chip	FG	Yes (A316T)	−0.09±0.01 mmol l^{-1} [27] −0.073±0.015 mmol l^{-1} [25]	1 [27] 1.5 [25]	60,564 [27] or 33,231 non-diabetic individuals [25]
SGSM2	Yes	Exome chip	Fasting PI	Yes (p.Val996Ile)	0.41±0.07 s.d.	1.4	8229 non-diabetic individuals
MADD	Yes	Exome chip	Fasting PI	Yes (p.Arg766X)	−0.32±0.04 s.d.	3.7	8229 non-diabetic individuals
G6PC2	Yes	Exome chip	FG	Yes (e.g., p.His177Tyr)	−0.12±0.02 mmol l^{-1} and −0.034±0.005 mmol l^{-1}	0.8 (p.His177Tyr)	33,231 non-diabetic individuals
HNF1A	No, MODY3 gene		T2D	Yes (p.Gly319Ser)	OR=2	20.9 (in Oji Cree)	Isolated population (Canadian Oji Cree)
TBC1D4	No	Metabochip and imputation	T2D	Yes (p.Arg684Ter)	OR=10.3 (homozygous)	17.0 (in Greenland)	Isolated population (Greenland)

financially feasible to sequence the whole human genome in hundreds of individuals. To get round this problem, geneticists have used the approach of imputation to capture more genetic variation in sample sets too large for whole-genome sequencing. Imputation essentially uses the correlation between variants caused by linkage disequilibrium (the co inheritance of closely linked genetic variants), as measured in a subset of sequenced individuals, to infer the missing genotypes in individuals not sequenced. This approach was recently employed by the Decode Icelandic study to identify a low-frequency non-coding allele in the intron of *CCND2* as associated with type 2 diabetes [9]. The low-frequency allele (approximate minor allele frequency [MAF] 2 %) has the strongest reported effect of any genetic polymorphism on type 2 diabetes. Carriers of one copy of the variant have half the risk of type 2 diabetes compared to non-carriers, a risk confirmed in a replication study [10] and so not biased by the regression to the mean problem often referred to as "winner's curse" in genetic studies. This effect is more than double that of the *TCF7L2* variant [11] that represents the strongest common genetic effect (where common is defined as alleles with a frequency greater than 5 %). The allele protective of diabetes operates through a favourable effect on insulin secretion—individuals carrying the allele are better able to respond to a glucose challenge with increased insulin secretion compared to non-carriers. Intriguingly, the carriers of the protective allele are also taller raising the possibility that the fundamental mechanism of action is cell division and growth of more beta cells—an hypothesis supported by the known role of the CCND2 protein product in beta cell replication [12] and observations that the CCND2 knockout mouse has reduced islet size and glucose intolerance [13]. The *CCND2* story is even more interesting because common variants in and near the gene are association with type 2 diabetes, and these effects appear independent of the low-frequency allele [2].

Exon Resequencing of GWAS Signals Finds Rare Coding Variants in Loci Previously Associated with T2D and Glycaemic Traits

For the majority of GWAS signals, we do not know how the variant affects type 2 diabetes risk or glycaemic traits, but they do provide a short list of potential genes involved in those traits and potential therapeutic targets. DNA variation in genes, especially that which changes the amino acid sequence of the encoded protein (non-synonymous) provides stronger evidence for the role of a particular gene if those coding variants are also associated with disease. However, non-synonymous variants are rarer on average than non-coding variants and are therefore less likely to be captured in GWASs. Recent studies have performed extensive sequencing of strong candidate genes within regions highlighted by GWAS. These studies include those that have identified rare non-synonymous variants in *MTNR1B* [14], *SLC30A8* [15] and *PPARG* [16]. Resequencing of *MTNR1B* found that rare loss-of-function variants in this locus, encoding a melatonin receptor, associate with increased T2D [14]. Melatonin is a circulating hormone secreted mainly from the pineal gland, and is best known for its role in regulating circadian rhythms. Insulin secretion follows a circadian cycle, with secretion being higher during the day [17], and this cycle seems to be regulated by melatonin through the receptor encoded by MTNR1B, expressed in beta cells [18]. Rare protein-truncating variants in *SLC30A8* lower the risk of T2D and associate with lower non-fasting glucose levels in heterozygotes [15]. SLC30A8 encodes a transporter responsible for the influx of zinc into insulin secretory granules [19] and is essential for insulin crystallization [20]. *PPARG* encodes a protein involved in adipocyte differentiation, and nonsynonymous variants in this gene which cause loss-of-function as evaluated by an in vitro assay are associated with a sevenfold increase in risk of T2D [16]. For PPARG, there are now variants across the full frequency spectrum that influence diabetes and insulin resistance, from very rare, protein-truncating mutations that cause severe insulin resistance [21] to low frequency (MAF <0.5 %) [16] to a common variant (Pro12Ala) [22].

Whole Exome Sequencing Finds Rare Variants in Loci Not Previously Associated with T2D

Whilst exome resequencing focuses efforts in a genomic region of interest, whole exome sequencing searches for coding variants associated with T2D in the majority of exons in the genome, with no previous assumptions about which genes might be important. Whole exome sequencing targets the 1 % of the coding genome and so is cheaper than the whole genome (although for various technical reasons, including the need to capture the 1 % of DNA with selective capture methods, the cost differences are not 100-fold). A missense rare mutation in exon 8 of *HNF1A*, the gene responsible for a form of monogenic diabetes, is associated with T2D in Latino populations [23]. This variant, E508K, is observed in approximately 1 in 50 individuals of Latino origin and increased the risk of diabetes with an odds ratio of 4.2 (95 % CI 1.8–9.9) in studies independent from the discovery study [23]. Further analysis showed that this variant is extremely rare in non-Latino populations (it was not found in approximately 10,000 whole exome sequenced non-Latinos). This variant illustrates another important point about rare variants—those that occur at 1 to 2 % MAF in one population may be much rarer or even non-existent in other populations even from the same major racial group. These differences between 2 and 0.1 % (for example) have dramatic effects on statistical power to detect rarer variation associated with diabetes and cardiovascular traits. An example was the

reduced significance of the association between the *CCND2* diabetes protective allele in Iran compared to Iceland, where the MAF was only 0.39 % (1 in 130 individuals) compared to 1.48 % (1 in 35 individuals), suggesting a trend in reduced frequency from north western Europe to the middle east [9]. These allele frequency differences between different populations of the same major racial groups highlight a second important point about genetic study design of particular relevance to type 2 diabetes and cardiovascular disease. Because the prevalence of type 2 diabetes and cardiovascular disease varies between populations, genetic studies can be confounded by population stratification unless properly corrected with genome-wide information. Any allele that is more common in an ethnic group than a second ethnic group will be associated with type 2 diabetes and obesity if the first ethnic group has a higher prevalence of obesity and diabetes. This phenomenon can result in false positive results. Fortunately, genetic information from across the genome can be used to correct for population stratification, and the authors of the E508K *HNF1A* study were careful to employ such techniques [23].

A recent study of myocardial infarction in a Norwegian population based study provided evidence that there may be few low-frequency or rare alleles that increase disease risk with odds ratios of 2 or more. Using exome sequencing, the study captured two-thirds of the coding variation in the population but after genotyping in 2906 individuals with myocardial infarction and 6738 controls, no variants were associated with the disease with sufficient statistical confidence, and the authors estimated that they could exclude effects of OR 2.0 [24].

Exome Array Studies Detect Coding SNPs Associated with T2D and Glycaemic Traits

Exome array approaches are similar to GWASs in their design, as an array is used to genotype individuals for chosen polymorphic positions. These polymorphic positions are situated in exons instead of being scattered around the genome, and include all variants detected at least twice in exome sequencing data from 12,000 individuals. A number of exome chip studies for T2D and glycaemic traits have been published recently [25–27]. These studies, using tens of thousands of individuals, have identified rare coding variants in loci known to be associated with T2D and provided evidence of association in previously unknown regions.

Coding variants in loci not previously highlighted by GWAS include signals for insulin secretion (two variants occurring in approximately one in ten individuals perfectly correlated due to linkage disequilibrium but in two genes: *PAM* and *PPIP5K2* [9, 26], and fasting glucose (*GLP1R* [25, 27])). *GLP1R* encodes a glucagon-like peptide (GLP1) receptor. This receptor binds the incretin hormone GLP1 and stimulates insulin secretion in beta cells and decreases glucagon release [28]. The association is another example of how DNA variants associated with diseases tend to locate genes encoding proteins of relevance to therapy. Agonists of the GLP1 receptor are used for the treatment of T2D. Other examples where genetic associations locate therapeutic targets include variants near the *PPARG* (thiazolidinediones) and *KCNJ11*/*ABCC8* (sulphonylureas) genes.

Exome array approaches have also located rarer coding variants in loci previously identified by common GWAS associations. These findings usually represent independent variants, but there are examples where the coding variant is correlated with and clearly explains the more common association signal. For example, a variant in the coding region of *TM6SF2* that occurs in one in six to seven people is associated (and robustly replicated) with lower LDL cholesterol levels but higher levels of markers of fatty liver disease and explains an association between a nearby variant in the *NCAN* gene. Subsequent functional studies of the coding variant in *TM6SF2* strongly implicate it as the causal gene [29, 30].

The discovery and replication of variants in coding regions within known GWAS loci implicate causal genes and provide a stronger case for studying a gene in follow-up experiments. Examples include loci harbouring coding variants independent from those GWAS signals (*SGSM2*, *MADD*; associated with fasting proinsulin [26] and *G6CP2*; associated with fasting glucose [25, 27]). *G6PC2* codes for a glucose-6-phosphatase catalytic subunit. Glucose-6-phosphatase is an enzyme that produces glucose from glucose-6-phosphate, so glucose can leave the cells in periods of starvation. Mahajan et al. found three coding variants associated with fasting glucose, with two of them also included in the four signals that Wessel et al. identified in the same gene [25].

Studying Isolated Populations Uncovers New Loci Associated with T2D

A final approach to the identification of genetic variants with large effects is to study isolated populations. In isolated populations, genetic drift may have resulted in alleles reaching relatively high frequencies despite apparently deleterious effects. There are two intriguing examples in diabetes, and more may emerge with sequencing efforts. First, a coding variant in *HNF1A*, G319S, is relatively common in the native Canadian population of the Oji Cree tribe [31]. *HNF1A* encodes HNF-1 α, a protein that among other functions regulates the expression of the main glucose transporter in β cells. The G319S mutation affects the transactivation activity of HNF-1 α [32]. This variant may also alter splicing as well as the amino acid sequence of *HNF1A* [33]. The 319S allele occurs in 33 and 63 % of the Oji Cree population in heterozygous (one copy) and homozygous (two copies) form and increases the risk of

non-insulin-dependent diabetes with odds ratios of 2 and 4, respectively. The allele is virtually non-existent in other populations.

A second more recent example is a coding variant in the *TBC1D4* gene that exists at 17 % frequency in the Greenland population but is virtually absent from all other populations. Homozygous carriers have an odds ratio for diabetes of approximately 10 [34]. TBC1D4 mediates glucose uptake in response to insulin. The variant affects a transcript that is expressed primarily in skeletal muscle, and decreases insulin sensitivity in that tissue by lowering the presence of the glucose transporter GLUT4 [34].

Conclusions

- Different strategies have been adopted to find rarer variants with stronger effects on type 2 diabetes risk and its related traits.
- Rare variants have been found in loci not previously highlighted by GWAS, but also in loci carrying GWAS signals.

Compliance with Ethics Guidelines

Conflict of Interest Lorena Boquete Vilarino reports grants from Diabetes UK, during the conduct of the study. Timothy Frayling reports grants from Diabetes UK, grants from European Research Council, personal fees from Boehringer Ingelheim, during the conduct of the study.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by the author.

References

1. Ntuk UE, Gill JM, Mackay DF, Sattar N, Pell JP. Ethnic-specific obesity cutoffs for diabetes risk: cross-sectional study of 490,288 UK biobank participants. Diabetes Care. 2014;37(9):2500–7.
2. 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(9):981–90.
3. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44(9):991–1005.
4. McCarthy MI. Dorothy Hodgkin lecture 2010. From hype to hope? A journey through the genetics of type 2 diabetes. Diabet Med. 2011;28(2):132–40.
5. McCarthy MI. Genomic medicine at the heart of diabetes management. Diabetologia. 2015.
6. Pal A, McCarthy MI. The genetics of type 2 diabetes and its clinical relevance. Clin Genet. 2013;83(4):297–306.
7. Yaghootkar H, Frayling TM. Recent progress in the use of genetics to understand links between type 2 diabetes and related metabolic traits. Genome Biol. 2013;14(3):203.
8. Bonnefond A, Froguel P. Rare and common genetic events in type 2 diabetes: what should biologists know? Cell Metab. 2015;21(3):357–68.
9. 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(3):294–8.
10. Yaghootkar H, Stancakova A, Freathy RM, Vangipurapu J, Weedon MN, Xie W, et al. Association analysis of 29,956 individuals confirms that a low frequency variant at CCND2 halves the risk of type 2 diabetes by enhancing insulin secretion. Diabetes. 2015.
11. 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(3):320–3.
12. Georgia S, Bhushan A. Beta cell replication is the primary mechanism for maintaining postnatal beta cell mass. J Clin Invest. 2004;114(7):963–8.
13. Kushner JA, Ciemerych MA, Sicinska E, Wartschow LM, Teta M, Long SY, et al. Cyclins D2 and D1 are essential for postnatal pancreatic beta-cell growth. Mol Cell Biol. 2005;25(9):3752–62.
14. 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(3):297–301.
15. 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(4):357–63.
16. Majithia AR, Flannick J, Shahinian P, Guo M, Bray MA, Fontanillas P, et al. 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(36):13127–32.
17. Boden G, Chen X, Urbain JL. Evidence for a circadian rhythm of insulin sensitivity in patients with NIDDM caused by cyclic changes in hepatic glucose production. Diabetes. 1996;45(8):1044–50.
18. 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(1):89–94.
19. Chimienti F, Devergnas S, Pattou F, Schuit F, Garcia-Cuenca R, Vandewalle B, et al. In vivo expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion. J Cell Sci. 2006;119(Pt 20):4199–206.
20. Lemaire K, Ravier MA, Schraenen A, Creemers JW, Van de Plas R, Granvik M, et al. Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice. Proc Natl Acad Sci U S A. 2009;106(35):14872–7.
21. Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, et al. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. Nature. 1999;402(6764):880–3.
22. 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(1):76–80.
23. Consortium STD, Estrada K, Aukrust I, Bjorkhaug L, Burtt NP, Mercader JM, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA. 2014;311(22):2305–14.
24. Holmen OL, Zhang H, Zhou W, Schmidt E, Hovelson DH, Langhammer A, et al. No large-effect low-frequency coding

variation found for myocardial infarction. Hum Mol Genet. 2014;23(17):4721–8.
25. Mahajan A, Sim X, Ng HJ, Manning A, Rivas MA, Highland HM, et al. Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genet. 2015;11(1):e1004876.
26. Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stancakova A, Stringham HM, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat Genet. 2013;45(2):197–201.
27. Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun. 2015;6:5897.
28. Nauck MA, Heimesaat MM, Orskov C, Holst JJ, Ebert R, Creutzfeldt W. Preserved incretin activity of glucagon-like peptide 1 [7-36 amide] but not of synthetic human gastric inhibitory polypeptide in patients with type-2 diabetes mellitus. J Clin Invest. 1993;91(1):301–7.
29. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. Nat Genet. 2014;46(4):345–51.
30. Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjaerg-Hansen A, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2014;46(4):352–6.
31. Hegele RA, Cao H, Harris SB, Hanley AJ, Zinman B. The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree. J Clin Endocrinol Metab. 1999;84(3):1077–82.
32. Triggs-Raine BL, Kirkpatrick RD, Kelly SL, Norquay LD, Cattini PA, Yamagata K, et al. HNF-1alpha G319S, a transactivation-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community. Proc Natl Acad Sci U S A. 2002;99(7):4614–9.
33. Harries LW, Sloman MJ, Sellers EA, Hattersley AT, Ellard S. Diabetes susceptibility in the Canadian Oji-Cree population is moderated by abnormal mRNA processing of HNF1A G319S transcripts. Diabetes. 2008;57(7):1978–82.
34. 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(7513):190–3.