The Importance of Context: Uncovering Species- and Tissue-Specific Effects of Genetic Risk Variants for Type 2 Diabetes

Soren K. Thomsen[1], Mark I. McCarthy[1,2,3] and Anna L. Gloyn[1,2,3]*

[1] Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, UK, [2] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK, [3] Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK

Genome-wide association studies (GWAS) have been highly successful in identifying genetic variation associated with type 2 diabetes (T2D) risk and related quantitative traits (1–3). The vast majority of association signals are located in non-coding regions of the genome, influencing nearby genes through regulation of transcriptional, translational, or splicing activity (4). Due to the highly context-dependent nature of gene expression, the effects of many risk variants are restricted to specific cell types and produce more subtle effects than those observed in organism-wide (or "global") knockouts. In addition, identification of the underlying causal genes and target tissues is often a major challenge, hindering translation into disease mechanisms. Recent studies have shown that the intersection of genetic data and genomic annotations can be used to produce a cellular atlas with which to understand the phenotypes of GWAS signals. Through the generation of directed hypotheses, this integrated framework has the potential to bridge the gap between association signals and disease biology.

THE BASIS FOR TISSUE SPECIFICITY

Across the human population, differences in complex traits, such as height and disease susceptibility, are influenced by the presence of single-nucleotide polymorphisms (SNPs). Some of these genetic variants modulate binding of transcription factors (TFs), which in turn drive differences in gene expression (5, 6). TF binding is also influenced by co-factors and chromatin state, which are highly dependent on cell type and developmental stage. To establish and maintain cellular identity, cell-type-specific TFs tend to bind in clusters, often referred to as cis-regulatory modules (7–9). Intriguingly, association signals for T2D have been found to show a significant overlap with islet-selective enhancer clusters (9, 10). Although other tissues have also been implicated in T2D susceptibility, this is consistent with physiological studies establishing islet dysfunction as a central mechanism of disease-associated variants (1, 11–13).

The overlap between T2D signals and enhancer clusters suggests that the effect of risk variants could be subject to the same tissue specificity observed at the level of regulatory activity (11). In other words, a motif-altering allele would only be expected to produce a molecular phenotype in those contexts (cell type or developmental stage) where the binding site has the potential to be occupied. In support of this notion, T2D risk variants were found to be enriched for nearby binding sites of the pioneer TF FOXA2 in islet and liver (14). T2D-association signals also show a significant overlap with SNPs affecting islet expression of regional transcripts, so-called cis-expression quantitative trait loci (cis-eQTL), most of which are not found to be eQTLs in other tissues (15). Together, these general observations outline some of the context-dependent effects that T2D risk variants can be subject to.

CONTEXT SPECIFICITY OF CAUSAL MECHANISMS FOR T2D SUSCEPTIBILITY

Recent studies have provided more specific evidence to support the notion that context specificity is a key aspect of GWAS causal mechanisms. The following cases collectively provide examples of mechanisms where studying the right tissue, species, and developmental stage proved critical to uncovering the relevant phenotypes.

At the *MTNR1B* locus, a convergence of evidence has pointed to effects of a non-coding T2D-association signal on the pancreatic β cell. Physiological studies have revealed a phenotype indicative of β-cell dysfunction in risk-variant carriers, with some evidence for additional effects on insulin action (12, 16, 17). Fine-mapping efforts identified a single likely causal variant that overlaps active islet and liver enhancers, and a cis-eQTL for *MTNR1B* in islets (14, 15, 18, 19). The risk allele, which increases islet *MTNR1B* expression, was predicted to create a NEUROD1 binding site and shown to selectively bind this key TF in human β cells. These results establish a likely causal mechanism for the non-coding risk allele, and illustrate how motif-altering alleles can generate highly tissue-specific effects. Surprisingly, exon re-sequencing of the *MTNR1B* gene has also shown coding loss-of-function (LOF) mutations to be associated with increased risk of T2D (20). The reason for the opposite directions of effect observed for coding and non-coding risk variants is unclear but may reflect differences between global and islet-specific roles of *MTNR1B*.

In the case of *PTF1A,* studying the right tissue proved necessary but not sufficient to elucidating the underlying mechanism for non-coding mutations in the region. Previous work had identified a group of patients suffering from unexplained isolated pancreatic agenesis, which includes neonatal diabetes as a clinical feature (21). To filter causal mutations from incidental variation, one study used pancreatic endoderm to define regulatory regions that are active during pancreatic development (22). Their strategy identified a distal enhancer that harbors mutations abolishing enhancer activity toward *PTF1A* (23). Coding LOF variants in *PTF1A* had previously implicated the gene in syndromic pancreatic agenesis, characterized by severe neurological features in affected individuals. The observation that the identified enhancer region is not active in any cell type other than pancreatic endoderm provides a plausible explanation for the absence of any cerebral defects (22). Remarkably, even adult pancreatic tissue did not show active chromatin marks in the region, highlighting that studying the right developmental stage was critical to the success of the approach.

The mechanisms underlying GWAS signals are sometimes studied using individuals that carry LOF mutations in positional candidate genes. The observed phenotypes will be a function of global effects across all the tissues where the gene is expressed, which may confound or mask the more context-dependent actions of regulatory risk alleles. At the *CDKN2A* locus, non-coding T2D signals have been robustly associated with measures of islet dysfunction, and a number of studies have established effects of *CDKN2A* on insulin secretion and cellular senescence in β-cells (12, 24, 25). By contrast, coding LOF mutations in *CDKN2A*, which are a cause of familial melanoma, were recently

shown to result in a metabolic phenotype consistent with effects on both liver and β-cells (26). This discrepancy was proposed to arise from islet-specific TF binding of the enhancer region containing the T2D signals.

In animal studies, context-dependent knockouts can provide improved spatial and temporal resolution for targeting candidate causal genes. Even so, the disease relevance of the observed phenotypes is determined by the confidence with which the target tissue of the risk allele is known. One example is provided by an intronic T2D signal at the *TCF7L2* locus, which has been the focus of conflicting observations. Tissue-specific knockout studies have demonstrated primary roles of *TCF7L2* in a number of different tissues, including liver and islets, whereas the non-coding GWAS signal has been consistently associated with a relatively narrow insulin secretion defect. Genomic annotations provide a clue as to the underlying reason, with the risk variants being located in an islet-specific region of open chromatin (8, 10, 27). Furthermore, the region has chromatin marks indicative of regulatory activity in islets, but not in a wide range of other tissues (10). The annotations can, thus, be applied as a filter to exclude non-disease-relevant tissues, and guide efforts to study the effect of the risk allele in the most appropriate context.

Coding GWAS variants can also produce context-dependent mechanisms through the restricted expression of gene isoforms. A striking example of this is provided by a coding variant identified in the *TBC1D4* gene in a small founder population of Greenlandic Inuit (28). The risk allele, which produces a truncated transcript that results in nonsense-mediated decay, is positioned in an exon excluded from the short isoform of the transcript. Unlike the widely expressed short isoform, the long form is predominantly expressed in skeletal muscle (29). The decreased insulin sensitivity resulting from reduced expression of *TBC1D4* is, therefore, selectively imposed on muscle tissue. As a result, the risk variant has a different effect on fasting glucose from that observed in individuals carrying LOF mutations affecting both isoforms (28).

Similar to alternative spliceforms, gene homologs can contribute to concealing the primary effect of a GWAS signal in a specific context. At the *ADCY5* locus, T2D risk alleles have been linked to both decreased islet expression of the *ADCY5* gene and β-cell dysfunction, though the underlying molecular mechanism remains unclear (12, 15). Expression studies in rodents have shown *Adcy5* to be nearly undetectable compared with the closely related homolog *Adcy6* (30, 31). By contrast, human islets show roughly equal expression of these orthologous genes, hinting at a non-conserved function of *ADCY5* between the species (30, 32). It also highlights an underlying species-specificity that makes rodents less well suited as models for mechanistic studies.

LIMITATIONS OF TRADITIONAL APPROACHES

The examples above demonstrate that the molecular phenotypes of GWAS signals can be modulated by a multitude of context-dependent factors. In the case of non-coding variants, tissue- or developmentally restricted activity of the surrounding chromatin can limit effects on gene expression. For coding

variants, alternative splicing or expression of homologs can mask a broader phenotype to produce context-dependent effects. These insights have important implications for how we design studies to translate genetic signals into molecular mechanisms.

Traditionally, particular genes have been selected for follow-up studies based on a combination of known candidate-gene biology and proximity to the GWAS signal (**Figure 1**, *left*). For whole-organism gene knockouts, this could involve engineering
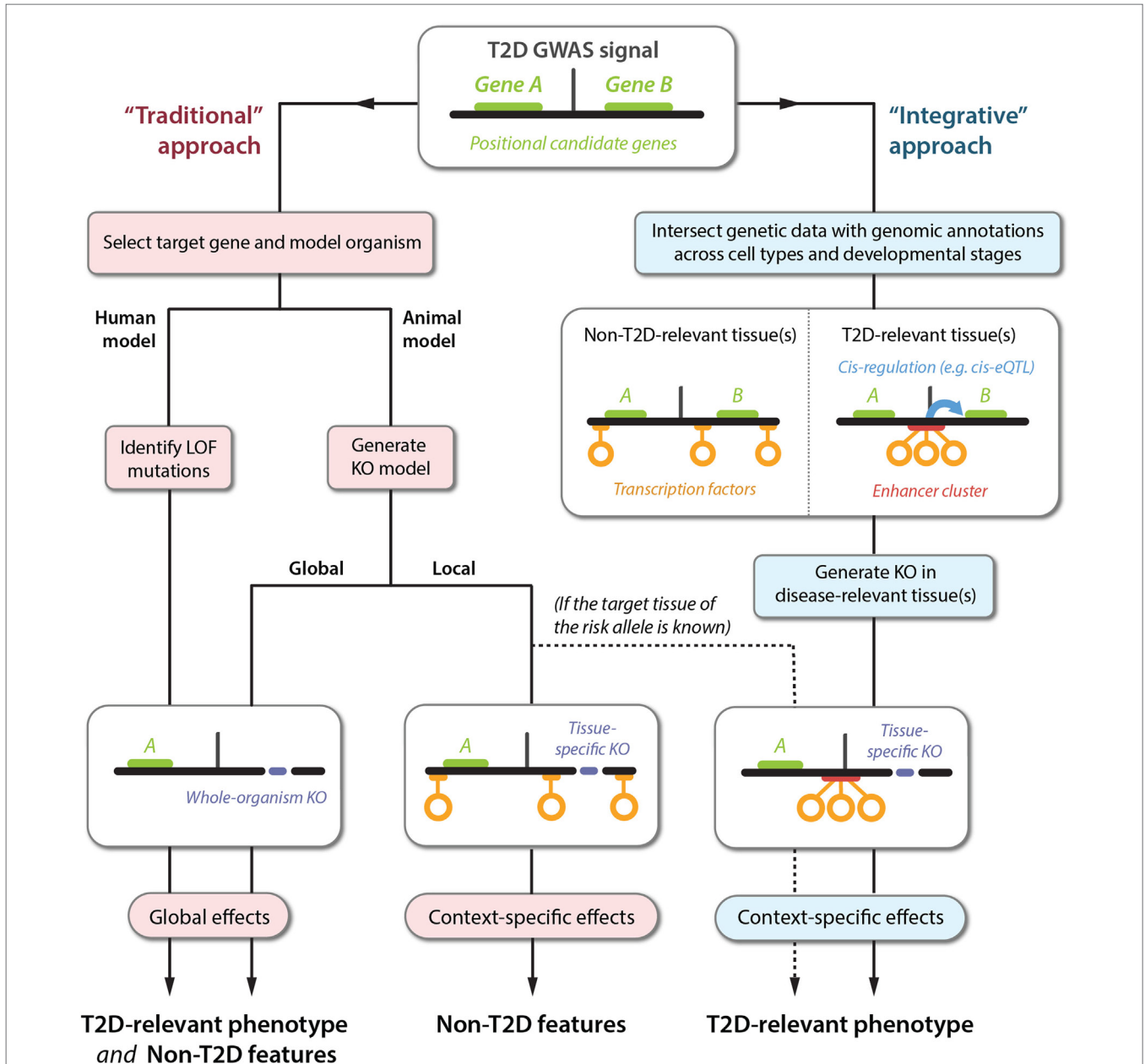

FIGURE 1 | Translating GWAS signals into disease mechanisms using traditional and integrative approaches. The top panel shows a schematic representation of a GWAS locus with a non-coding association signal for T2D risk located near two genes, *A* and *B*. The left-hand side represents the traditional approach followed for elucidating the causal mechanisms leading to a T2D phenotype. This approach can produce a range of phenotypes that are difficult to translate into causal GWAS mechanisms, because the selected genes are often knocked out globally (bottom left panel), or in a tissue-specific context that is irrelevant to understanding the effects of the T2D susceptibility phenotype (bottom center panel). Sequencing approaches can also be used to identify gain-of-function mutations (not shown), which are subject to the same limitations. The right-hand side outlines an alternative approach using integration of emerging datasets to produce directed hypotheses. For optimal resolution, several types of genetic and genomic datasets can be integrated, including TF binding, enhancer regions (defined based on chromatin state), and cis-regulatory relationships (e.g., identified by cis-eQTL or chromatin-conformation capture studies). In this example, the GWAS signal is observed to disrupt an enhancer cluster with cis-regulatory activity toward gene *B*. Importantly, the disruption exclusively affects enhancer activity in disease-relevant tissues. This proposes a follow-up experiment for manipulation of gene *B* in a specific context (bottom right panel), producing a phenotype that is likely to be directly relevant to understanding the molecular basis of T2D susceptibility.

of animal models or using genetic testing to identify individuals carrying LOF variants. As discussed, the relevance of these approaches for delineating disease-relevant mechanisms is limited by the potential for unspecific global phenotypes. Tissue-specific knockouts provide higher spatiotemporal resolution but require the target tissue(s) of the GWAS signal to be known. Human physiological associations can narrow down the list of likely relevant tissues, but these measures are often too crude to pinpoint specific cell types.

In principle, using animal models to target non-coding regions could produce more precise disease models, but this strategy is constrained by the low conservation at the level of regulatory architecture. In a study of the *Cdkn2a* locus in mice, targeted deletion of a 70 kb non-coding interval established enhancer activity toward nearby genes, but the relevance to humans has been questioned by subsequent findings (33, 34). The region encodes a long non-coding RNA that has no clear ortholog in rodents, highlighting the possibility of divergent cis-regulatory mechanisms. More generally, TF binding sites have been shown to diverge even faster than the underlying sequence itself (35). For two key liver TFs, the majority of binding events were shown to be species specific, while only 10–30% of hepatic enhancer clusters have corresponding rodent orthologs (36, 37). Although subsets of conserved clusters may aid in the prioritization of causal variants, these observations suggest that a different approach is required to delineate GWAS mechanisms (37, 38).

Even in those cases where animal models do provide targeted gene manipulation in an appropriate context (whether through tissue-specific knockout or transcriptional dysregulation), the resulting phenotypes may not be directly relevant for understanding human disease. Though rodent models continue to be an important tool for studying type 2 diabetes pathogenesis, it has become increasingly clear that murine pancreatic islets differ in a number of ways from their human counterparts (39–41). Certain monogenic forms of diabetes are, therefore, not well recapitulated in rodents, and molecular mechanisms elucidated in animal models should be interpreted with caution (42).

TOWARD AN INTEGRATED UNDERSTANDING OF GWAS SIGNALS

To successfully study disease-relevant phenotypes, experimental designs can be guided by the integration of genetic association data and genomic annotations (**Figure 1**, *right*). At the core of this framework is the overlaying of a static dataset – a list of variants linked to disease susceptibility and/or physiological traits – with layers of highly dynamic functional information that provide spatial and temporal dimensions. These layers encompass diverse datasets, and include information centered on single variants, such as histone marks and TF binding sites, and higher-order information that signifies relationships between distinct elements, such as chromatin interactions and cis-eQTL data.

Genomic annotations provide a cellular atlas with which to navigate and interpret genetic data in the context of specific cell types and developmental stages. For instance, if a set of likely causal variants has been identified from GWAS or fine-mapping studies, the tissue of action may be inferred from comparing chromatin states across cell types (8–11, 43–46). Conversely, if the target tissue is known from physiological associations, the causal mechanism can be pinpointed by overlaying with relevant functional annotations (14, 22, 47–51). This process generates a plethora of directed hypotheses that can be followed up with specific functional experiments. Increasingly, such studies are likely to be focused on differentiated cells derived from human stem cells, which can provide disease models and chromatin maps that are both functionally and developmentally relevant.

The broader applicability of this approach is, in part, determined by the tractability of individual loci. For regions with extensive linkage disequilibrium, the arising complexity can hinder experimental follow-up. Starting from a limited set of credible variants is, thus, essential. More generally, the value in taking an integrative approach is dictated by the extent to which genomic annotations for disease-relevant tissues have been made available (or can be obtained). The construction of a truly integrated framework is an incremental and monumental effort, facilitated by the Encyclopedia of DNA Elements (ENCODE) and the NIH Roadmap Epigenomics project, which together cover hundreds of tissues and epigenetic annotations. Since each dataset is merely a snapshot of a given cell type in a particular metabolic and developmental state, this on-going process will continue to produce an atlas with ever-finer spatiotemporal resolution. For many hard-to-obtain organs, such as pancreatic islets, power to detect relevant genomic features is still limiting.

Even so, chromatin landscapes for tissues relevant to T2D have begun to emerge in recent years, enabling biological inferences to be made. As we have seen, this has successfully uncovered tissue- and species-specific effects of T2D risk variants. The insights have also provided compelling evidence to demonstrate that context-dependent phenotypes are not the exception but, in fact, a fundamental aspect of GWAS biology. In the coming years, we need to build on this paradigm to accelerate the translation of genetic findings into molecular mechanisms.

AUTHOR CONTRIBUTIONS

ST, MM, and AG wrote, edited, and approved the manuscript.

FUNDING

REFERENCES

1. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* (2010) 42:579–89. doi:10.1038/ng.609

2. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè 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

3. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, Mexican American Type 2 Diabetes (MAT2D) Consortium, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, Mahajan A, et al. 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

4. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* (2012) 30:1095–106. doi:10.1038/nbt.2422

5. Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, et al. Extensive variation in chromatin states across humans. *Science* (2013) 342:750–2. doi:10.1126/science.1242510

6. McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, et al. Identification of genetic variants that affect histone modifications in human cells. *Science* (2013) 342:747–9. doi:10.1126/science.1242429

7. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* (2013) 153:307–19. doi:10.1016/j.cell.2013.03.035

8. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets. *Nat Genet* (2010) 42:255–9. doi:10.1038/ng.530

9. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A* (2013) 110:17921–6. doi:10.1073/pnas.1317023110

10. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman I, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* (2014) 46:136–43. doi:10.1038/ng.2870

11. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* (2013) 45:124–30. doi:10.1038/ng.2504

12. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi 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

13. Torres JM, Gamazon ER, Parra EJ, Below JE, Valladares-Salgado A, Wacher N, et al. Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. *Am J Hum Genet* (2014) 95:521–34. doi:10.1016/j.ajhg.2014.10.001

14. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi 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

15. van de Bunt M, Manning Fox JE, Dai X, Barrett A, Grey C, Li L, et al. Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. *PLoS Genet* (2015) 11:e1005694. doi:10.1371/journal.pgen.1005694

16. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proença C, Sparsø 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

17. 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

18. 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

19. Tuomi T, Nagorny CL, Singh P, Bennet H, Yu Q, Alenkvist I, et al. Increased melatonin signaling is a risk factor for type 2 diabetes. *Cell Metab* (2016) 23:1067–77. doi:10.1016/j.cmet.2016.04.009

20. Bonnefond A, Clément 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

21. Lango Allen H, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R, et al. GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet* (2012) 44:20–2. doi:10.1038/ng.1035

22. Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, et al. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* (2014) 46:61–4. doi:10.1038/ng.2826

23. Sellick GS, Barker KT, Stolte-Dijkstra I, Fleischmann C, Coleman RJ, Garrett C, et al. Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat Genet* (2004) 36:1301–5. doi:10.1038/ng1475

24. Grarup N, Rose CS, Andersson EA, Andersen G, Nielsen AL, Albrechtsen A, et al. Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects: validation and extension of genome-wide association studies. *Diabetes* (2007) 56:3105–11. doi:10.2337/db07-0856

25. Krishnamurthy J, Ramsey MR, Ligon KL, Torrice C, Koh A, Bonner-Weir S, et al. p16INK4a induces an age-dependent decline in islet regenerative potential. *Nature* (2006) 443:453–7. doi:10.1038/nature05092

26. 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

27. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab* (2010) 12:443–55. doi:10.1016/j.cmet.2010.09.012

28. Moltke I, Grarup N, Jørgensen 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

29. Baus D, Heermeier K, De Hoop M, Metz-Weidmann C, Gassenhuber J, Dittrich W, et al. Identification of a novel AS160 splice variant that regulates GLUT4 translocation and glucose-uptake in rat muscle cells. *Cell Signal* (2008) 20:2237–46. doi:10.1016/j.cellsig.2008.08.010

30. Hodson DJ, Mitchell RK, Marselli L, Pullen TJ, Gimeno Brias S, Semplici F, et al. ADCY5 couples glucose to insulin secretion in human islets. *Diabetes* (2014) 63:3009–21. doi:10.2337/db13-1607

31. Parton LE, McMillen PJ, Shen Y, Docherty E, Sharpe E, Diraison F, et al. Limited role for SREBP-1c in defective glucose-induced insulin secretion from Zucker diabetic fatty rat islets: a functional and gene profiling analysis. *Am J Physiol Endocrinol Metab* (2006) 291:E982–94. doi:10.1152/ajpendo.00067.2006

32. Eizirik DL, Sammeth M, Bouckenooghe T, Bottu G, Sisino G, Igoillo-Esteve M, et al. The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet* (2012) 8(3):e1002552. doi:10.1371/journal.pgen.1002552

33. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, et al. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* (2010) 464:409–12. doi:10.1038/nature08801

34. Congrains A, Kamide K, Ohishi M, Rakugi H. ANRIL: molecular mechanisms and implications in human health. *Int J Mol Sci* (2013) 14:1278–92. doi:10.3390/ijms14011278

35. Stefflova K, Thybert D, Wilson MD, Streeter I, Aleksic J, Karagianni P, et al. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* (2013) 154:530–40. doi:10.1016/j.cell.2013.07.007

36. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* (2010) 328:1036–40. doi:10.1126/science.1186176

37. Ballester B, Medina-Rivera A, Schmidt D, Gonzàlez-Porta M, Carlucci M, Chen X, et al. Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *Elife* (2014) 3:e02626. doi:10.7554/eLife.02626

38. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, et al. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* (2014) 156:343–58. doi:10.1016/j.cell.2013.10.058

39. Ashcroft FM, Rorsman P. Diabetes mellitus and the beta cell: the last ten years. *Cell* (2012) 148:1160–71. doi:10.1016/j.cell.2012.02.010

40. Ravassard P, Hazhouz Y, Pechberty S, Bricout-Neveu E, Armanet M, Czernichow P, et al. A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. *J Clin Invest* (2011) 121:3589–97. doi:10.1172/JCI58447

41. Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, Caicedo A. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci U S A* (2006) 103:2334–9. doi:10.1073/pnas.0510790103

42. Thomsen SK, Gloyn AL. The pancreatic beta cell: recent insights from human genetics. *Trends Endocrinol Metab* (2014) 25:425–34. doi:10.1016/j.tem.2014.05.001

43. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* (2014) 506:376–81. doi:10.1038/nature12873

44. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* (2015) 518:317–30. doi:10.1038/nature14248

45. Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, et al. 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. doi:10.1038/ng.3245

46. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* (2015) 47:1228–35. doi:10.1038/ng.3404

47. Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, Kong A, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* (2011) 43:561–4. doi:10.1038/ng.833

48. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* (2010) 466:714–9. doi:10.1038/nature09266

49. Adrianto I, Wen F, Templeton A, Wiley G, King JB, Lessard CJ, et al. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat Genet* (2011) 43:253–8. doi:10.1038/ng.766

50. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* (2012) 44:502–10. doi:10.1038/ng.2205

51. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* (2015) 518:337–43. doi:10.1038/nature13835

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer LL declared a past co-authorship with one of the authors MM to the handling Editor, who ensured that the process met the standards of a fair and objective review.