REVIEW

Next steps in the identification of gene targets for type 1 diabetes

Struan F. A. Grant [1,2,3] · Andrew D. Wells [1,4] · Stephen S. Rich [5,6]

Abstract
The purpose of this review is to provide a view of the future of genomics and other omics approaches in defining the genetic contribution to all stages of risk of type 1 diabetes and the functional impact and clinical implementations of the associated variants. From the recognition nearly 50 years ago that genetics (in the form of HLA) distinguishes risk of type 1 diabetes from type 2 diabetes, advances in technology and sample acquisition through collaboration have identified over 60 loci harbouring SNPs associated with type 1 diabetes risk. Coupled with HLA region genes, these variants account for the majority of the genetic risk (~50% of the total risk); however, relatively few variants are located in coding regions of genes exerting a predicted protein change. The vast majority of genetic risk in type 1 diabetes appears to be attributed to regions of the genome involved in gene regulation, but the target effectors of those genetic variants are not readily identifiable. Although past genetic studies clearly implicated immune-relevant cell types involved in risk, the target organ (the beta cell) was left untouched. Through emergent technologies, using combinations of genetics, gene expression, epigenetics, chromosome conformation and gene editing, novel landscapes of how SNPs regulate genes have emerged. Furthermore, both the immune system and the beta cell and their biological pathways have been implicated in a context-specific manner. The use of variants from immune and beta cell studies distinguish type 1 diabetes from type 2 diabetes and, when they are combined in a genetic risk score, open new avenues for prediction and treatment.

Abbreviations

eQTL	Expression quantitative trait locus
GRS	Genetic risk score
GWAS	Genome-wide association scan
IRE	Inducible regulatory element
SLE	Systemic lupus erythematosus
T1DGC	Type 1 Diabetes Genetics Consortium
TFH	Follicular helper T cells
WTCCC	Wellcome Trust Case Control Consortium

✉ Stephen S. Rich
ssr4n@virginia.edu

1   Center for Spatial and Functional Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

2   Departments of Pediatrics and Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

3   Divisions of Human Genetics and Endocrinology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

4   Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

5   Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA

6   Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA

Introduction

Diabetes is a clinically heterogeneous, chronic condition characterised by a failure to maintain normal glucose levels through conversion of food into energy via insulin-dependent mechanisms. The most common forms of diabetes have been defined by clinical differences in insulin dependence to maintain glucose homeostasis, the age and abruptness of onset of symptoms, and tendency for ketosis. This review will provide background on the genetic basis of type 1 diabetes, the function of genetic variation, and future work moving to discovery of target genes, pathways and mechanisms, novel interventions and the identification of therapeutic targets.

Definitions and background

- **Single nucleotide polymorphism (SNP):** A variation at a specific location in the DNA sequence involving a change in one of the four nucleotides—adenine (A), guanine (G), cytosine (C) and thymine (T)—to a different nucleotide at that position (e.g. C vs T at a given location).

- SNPs are the most common form of genetic variation in humans, occurring at ~1 out of 1000 of the 3 billion nucleotides of DNA, such that each human has 4–5 million 'common' SNPs.

- An ancestral (reference) genome might have two copies of the same nucleotide at a given site—AA, CC, GG, or TT—called a homozygote. Genetic variation at this site during breeding can create a reference allele and an alternative allele at the site (e.g. from AA in the parent to AC, AG or AT in the offspring).

- The frequencies of specific reference and alternative alleles may vary across populations. When the frequency of the alternative allele (e.g. the C allele at a specific site) is at least 5% in a population, the single nucleotide change is called a 'polymorphism' (SNP).

- Most SNPs are found in the DNA between genes and principally confer small effects on human health or characteristics over the lifespan. Such susceptibility variants may reside in regions that affect a gene's expression, while others alter protein function and can be related to diseases that 'run in families' and responses to drugs and contribute to risk of clinical outcomes.

- **Genetic linkage analysis:** A method that monitors the co-segregation of a trait (e.g. disease risk or measured phenotype) with a measured genotype, and is used to locate the chromosomal region harbouring genetic variation that contributes to the trait.

- Linkage analysis requires a genetic map, based on sites of known genetic markers, that enables 'tracking' of the trait through families with affected and unaffected relatives. This was formally developed in the mid-1950s, when maps were based on protein polymorphism and thus very sparse. However, the advent of microsatellites, then SNPs, high-throughput genotyping technology and understanding of the structure of the human genome, linkage analysis has enabled the discovery of sites contributing to variation in many traits.

- A limitation of linkage analysis is that it depends upon genetic recombination (crossing over), which is relatively rare in one generation. Thus, large numbers of families are required to map traits to relatively narrow regions of a chromosome (i.e. high resolution: distances of tens or hundreds of thousands of nucleotides).

- **Genome-wide association scan (GWAS):** An approach developed in the mid-2000s to map regions of the genome associated with a trait using a large panel of SNPs. The study design is typically focused on unrelated individuals, either comparing cases (with a disease) and controls (absence of disease) or testing for a quantitative phenotype (e.g. BMI).

- GWAS interrogates the frequencies of SNP alleles in the unrelated individuals across the genome to detect significant differences in cases vs controls (or across a phenotypic scale). The ability to collect large numbers of unrelated individuals with extensive traits, coupled with genome-wide coverage with millions of SNPs, has been exceedingly successful in defining regions of the genome robustly associated with thousands of traits.

- With increased genomic coverage in GWAS genotyping panels, statistical imputation (defining the probability of an ungenotyped SNP based upon the local structure of the genome near a genotyped SNP) has further refined the ability to detect sets of likely causal genetic variants associated with the traits investigated.

- GWAS has focused on the role of common genetic variation (alternative SNP allele frequency >5% in a population) to identify SNPs with ever smaller effect on the trait, and multiple populations have generated summary statistics and results combined using meta-analysis.

- A major limitation of GWAS has been the historical focus on populations of Northern European ancestry. Recent results suggest that findings in these populations are largely not transferable to others and therefore more diverse populations are required in both genetic and risk factor studies.

- **Genetic fine mapping:** A follow-up approach to linkage and GWAS for defining the contribution of SNPs and genes to a trait. The linked or associated SNPs are based upon a small subset of genomic variation and, therefore, are unlikely to represent causal variants, as the regions may harbour hundreds to thousands of SNPs and many genes.

- Fine mapping uses a very dense set of SNPs within a defined region of the genome to refine those SNPs most likely to account for the effect observed. Methods used in fine mapping are similar to those in the original discovery approach; thus, detection of 'true' causal SNPs and genes requires other approaches (e.g. functional genomics).

Definitions and background (cont'd)

- **Genetic risk score (GRS) and the polygenic risk score (PRS):** Approaches to integrate more than a single associated SNP for prediction of risk or phenotype. The GRS is often thought to include only those SNPs that are most associated with the trait in a locus, while the polygenic risk score includes all SNPs interrogated in an analysis.

- The development of the GRS can be calculated as a summation across robustly associated SNPs of genotype score (coding 1 for the risk homozygote, 0 for the heterozygote, and -1 for the non-risk homozygote) or as the summation over SNP genotypes, weighted by the size of the effect on phenotype at each SNP. The PRS extends the concept to the entire set of SNPs in a genome, even those with very small effects.

- Both GRS and PRS capture extensive genomic data in a single value, with subsequent interpretation and utility dependent upon the adjustment of the scores to the appropriate population, selection of SNPs (if a significance threshold is used), thresholds for defining risk and actionable scores, and incorporation of other genetic and non-genetic effects in the score.

- **Quantitative trait locus (QTL):** The result of correlating levels of a phenotype (e.g. height) with genotype across the human genome, mapped using SNPs, often in a GWAS design.

- The phenotypic trait can also be represented by other measurements, including levels of gene expression (eQTL), DNA methylation (methylQTL), chromatin accessibility (caQTL), or amount of specific metabolites (metabQTL) or proteins in plasma (pQTL).

- Genetic variation can be a driver of changes in gene expression and regulation based in regions of open chromatin. In a disease setting, putative causal SNPs would be expected to impact biological pathways that are reflected by QTL effects (e.g. eQTL, caQTL).

- **Chromatin conformation:** A fundamental component of genome regulation based upon the 3D folding of the chromatinised genome in the nucleus. Detection of long-range chromatin interactions using chromosome conformation capture technologies allows examination of physical contacts between distant regulatory elements and gene promoters that are critical for gene expression.

- **Topologically associating domains (TADs):** Regions of DNA (~880 kb in size) in which DNA sequences can physically interact with each other more frequently than with DNA sequences outside the TAD. The function of TADs is not completely known, but it is thought that TADs are largely tissue independent and regulate gene expression by limiting enhancer–promoter interactions to the TAD. Thus, disruption of the TAD boundaries may alter gene expression, phenotype and risk of disease.

- **Genome editing:** A term representing a variety of technologies to add, remove or alter DNA at specific locations in the genome.

- Several approaches of genome editing have been developed, including use of zinc finger nucleases, transcription activator-like effector nucleases (TALENs) and CRISPR/Cas9. Once DNA is cut, the cell's own DNA repair machinery can be used to alter, add or delete pieces.

- Genome editing is an active area of research that can target specific regions of the genome containing putative causal genes or SNPs. Through alteration of these DNA segments, effects on transcription, regulation, and other forms of gene function can be examined to determine the impact of the edited region as it may relate to disease or phenotype.

Genetic basis of type 1 diabetes

Following the discovery of type 1 diabetes associated with HLA [1], the insulin (*INS*) variable number tandem repeat (VNTR) [2], and numerous candidate gene polymorphisms, the development of high-throughput genotyping array technology and analytical methods expanded our knowledge of genetic variation implicated in type 1 diabetes risk (summarised in the Text box 'Genetics of type 1 diabetes: background'). Despite decades of interrogating HLA, much remains to be understood about specific allelic and interaction

effects within that region [3] across populations. The Wellcome Trust Case Control Consortium (WTCCC) established the genome-wide association scan (GWAS) as a primary tool in discovery of genetic variants associated with common disease [4]; however, the WTCCC identified relatively few novel risk loci (*ERBB3*, *SH2B3*), including one simultaneous discovery (*CLEC16A*, formerly known as *KIAA0350*) [5]. Later, the Type 1 Diabetes Genetics Consortium (T1DGC) conducted the largest GWAS meta-analysis of type 1 diabetes in ~7500 cases and ~9000 controls, with replication in ~4000 cases, ~4500 controls and 4300 trio

Genetics of type 1 diabetes: background

- **Defining diabetes subtype by HLA** The initial subtyping of diabetes was defined as 'juvenile-onset' or 'insulin-dependent (at any age)' vs 'adult-onset' or 'non-insulin-dependent (at any age)'.

- Role of HLA haplotypes: Significantly increased frequency of HL-Aw15 in individuals with insulin-dependent diabetes vs control individuals or those with non-insulin-dependent diabetes, and no sex difference in the frequency of HL-Aw15 in those with insulin-dependent diabetes. With the discovery of additional HLA region genes, the nomenclature for individual genes was changed to HLA-A (e.g. HLA-A15), with expansion to many class I and class II genes. The advent of DNA sequencing further expanded the number of alleles identified at each gene, with new nomenclature to indicate 'splits' within a gene into multiple alleles, as well as the technology used.

- It was the dependence on insulin, and not the age at onset, that defined a subgroup of individuals with diabetes with a recognised genetic contribution.

- **The insulin (*INS*) gene** The relatively strong association of the 5′ flanking region of the human insulin (*INS*) gene (570 bp class 1, 1320 bp class 2, and 2470 class 3 alleles) permitted 'genotyping' of just 393 unrelated individuals (113 with insulin-dependent diabetes, 76 with non-insulin-dependent diabetes and 83 without diabetes) to identify the association of the *INS* class 1 allele with insulin-dependent diabetes.

- The strength of the association with the *INS* variable number of tandem repeat (VNTR) minisatellite polymorphism is second only to the HLA region.

- **Family studies in type 1 diabetes** Initial studies often were small, focused on affect sibling pairs (ASPs), and now recognised as under-powered.

- An advantage of the family study design was the direct identification of putative causal (co-segregation of genotype with disease) genes, rather than association using cases and controls, using linkage analysis.

- The first type 1 diabetes genome-wide linkage approach utilised a two-stage design of 289 microsatellite markers in 96 UK affected sibling pair families and a second stage of 186 sibling pair families for validation, identifying the HLA and *INS* regions. By 2000, only HLA, *INS* and *CTLA4* had been replicated as bona fide type 1 diabetes susceptibility loci.

- The Type 1 Diabetes Genetics Consortium (T1DGC) was established in 2001 and utilised samples to generate HLA typing, genome-wide linkage scans, with clinical and biochemical data providing samples and data to the research community.

- The T1DGC linkage analyses confirmed the role of HLA in conferring type 1 diabetes risk, but evidence for *INS* and *CTLA4* reached significance only at the largest sample size (~4000 affected sibling pair families), and only the identification of the *UBASH3A* locus was novel.

- **Genome-wide association scans in type 1 diabetes** The Wellcome Trust Case Control Consortium (WTCCC) targeted 2000 cases of seven diseases (including type 1 diabetes) vs 3000 shared controls for gene discovery; at that time, there were six loci associated with type 1 diabetes (MHC/*HLA-DRB1*, *INS*, *CTLA4*, *PTPN22*, *IL2RA* and *IFIH1*).

- The WTCCC identified five loci, two known (MHC and *PTPN22*) and three novel (*ERBB3*, *SH2B3*, *CLEC16A*); neither the *INS* locus (the region was not well-tagged by SNPs) nor the *CTLA4*, *IL2RA*, and *IFIH1* loci (the SNPs failed to meet genome-wide significance) were detected. Follow-up studies reported four additional loci (*BACH2*, *PRKCQ*, *CTSH* and *C1QTNF6*).

- In the same year (2007), a separate study from the Children's Hospital of Philadelphia (CHOP) identified the *CLEC16A* locus (formerly *KIAA0350*). This team was also the first to report the *BACH2* association.

- The T1DGC assembled samples for the largest genome-wide association scan (GWAS) meta-analysis, consisting of ~7500 cases, ~9000 controls, and a replication with ~4000 cases, ~4500 controls and 4300 trio families. Forty-one distinct loci reaching genome-wide significance (26 novel) were identified and replicated.

- **Fine mapping of type 1 diabetes** As GWAS has low resolution, the ImmunoChip was developed for fine mapping, consisting of ~186,000 SNPs covering 186 loci associated with autoimmunity. Analysis of >30,000 T1DGC samples (6808 cases, 12,835 controls, 11,584 family members) of Northern European ancestry yielded 44 regions associated with T1D, including four novel (1q32.1, 2q13, 4q32.3, 5p13.2) and two implicated in type 1 diabetes through strong association with other autoimmune diseases (17q21.31 and 21q22.3).

- Bayesian analysis of T1DGC data established a 99% credible set of SNPs for each of the 44 type 1 diabetes loci, reducing the number of potential candidate genes from about seven to about two per locus.

- T1DGC analysis found enrichment of 99% type 1 diabetes-associated credible SNPs in DNA regulatory regions (enhancers) in immune-relevant tissues (CD4$^+$ and CD8$^+$ T cells, CD19$^+$ B cells and CD34$^+$ stem cells)

families [6]. The T1DGC identified 41 distinct loci, including 26 that were novel.

Fine mapping type 1 diabetes-associated loci identified from GWAS

As with other GWAS, the T1DGC GWAS meta-analysis [6] yielded loci that were large (~250 kb for each locus), with many genes (ranging from 0 to 28) commonly harboured within each corresponding region [7]. To get as close to the underlying causal variants underlying these associations, fine mapping employing genotyping arrays with dense coverage within each locus (ImmunoChip) was performed on >30,000 individuals (cases, controls and families) [8]. Credible sets of SNPs were established for each of the 44 loci, revealing enrichment of SNPs in DNA regulatory regions. These results supported a role for enhancer chromatin states in immune-relevant cell types (CD4$^+$ and CD8$^+$ T cells, CD19$^+$ B cells and CD34$^+$ stem cells) in type 1 diabetes risk. Similar efforts going forward, in larger and more diverse populations, should shed light on additional risk loci and variants contributing to the pathogenesis of type 1 diabetes.

Although GWAS and fine mapping efforts have provided much insight into genetic aetiology, the picture remains incomplete. Those type 1 diabetes risk loci, uncovered by such initial classical approaches, remain dominant factors in the genetic picture of disease; however, they do not explain the entire genetic architecture of type 1 diabetes. In order for the power of genetics to fully contribute to risk prediction and discovery of novel therapeutic avenues, additional approaches need to be employed including expansion to multi-ethnic populations, where novel variants have already begun to emerge [9, 10], and into adults, who account for nearly half of those with type 1 diabetes.

Identifying causal SNPs for type 1 diabetes

A step (of many possible) in determining whether a SNP is a causal variant is to estimate its contribution to gene expression (expression quantitative trait locus, eQTL) [11]. Typical eQTL analysis is equivalent to GWAS but using gene expression as the phenotype. The variant most associated with disease may be near a gene of interest; however, that variant may be regulating the expression of a different, more distal effector gene. This is the situation for a variant in the FTO gene that is most strongly associated with obesity which actually regulates IRX3 [12, 13], and a variant in TCF7L2 most strongly associated with type 2 diabetes that regulates ACSL5 [14].

In type 1 diabetes, the 16p13 locus contains strongly associated SNPs spanning introns 10 and 19 of CLEC16A [4–6]. A single eQTL was identified in the neighbouring DEXI gene, such that the CLEC16A SNPs associated with reduced risk of type 1 diabetes correlated with increased DEXI expression in monocytes [15]. This result was replicated and identified the most strongly associated variant in CLEC16A with expression in B cells, implicating a SNP in CLEC16A alters risk of type 1 diabetes through expression of DEXI [16]. Differential transcriptome analysis of tolerogenic and mature inflammatory dendritic cells, when overlaid with SNPs associated with type 1 diabetes, identified 11 genes with differential expression [17]; three (CCR5, CTSH and RAC2) with higher expression in tolerogenic dendritic cells compared with mature inflammatory dendritic cells, and eight (IKZF4, IKZF1, SH2B3, ORMDL3, TYK2, IL2RA, PTPN2 and ICOSLG) with lower expression. These results implicated a role for these disease-associated variants as activators of the immune response in type 1 diabetes.

Although eQTL analysis from peripheral blood provides some insight into possible causal effects of variants associated with type 1 diabetes, immune cell type-specific evaluation (e.g. T-helper 17 cells [Th17], regulatory T cells [Tregs], monocytes) should enhance our understanding of the impact of these variants on target genes. Microarray data from 92 children (25 seroconverters and 67 nonseroconverters) provided longitudinal change in gene expression profiles with development of islet autoimmunity [18]. Gene expression signatures in the first year of life predicted seroconversion with genes that contribute to T cell, B cell and dendritic cell-related immune responses, primarily through a ubiquitin-proteasome pathway. A protein–protein interaction network was linked to type 1 diabetes-associated genes with differentially expressed seroconversion genes, revealing direct interactions with ERBB3 and GLIS3, two type 1 diabetes susceptibility genes.

Gene regulation from a distance

GWAS have delivered many validated loci associated with novel aetiological pathways. But as mentioned above, these SNP associations do not necessarily implicate the closest gene as causal, even if reasonable hypotheses exist between the SNP location and possible gene function (see the FTO–IRX3 experience, above). Gene expression can be controlled locally or via long-range interactions over large genomic distances. Indeed, many regulatory elements do not control the nearest genes, but, rather, ones residing tens or hundreds of kilobases away. Barriers to detecting the 'true' targets of disease-associated SNPs include the limited, but growing public domain genomic data relevant to individual immune cell types, tissue-specific eQTLs, chromatin conformation capture, and emergent variant-to-gene techniques required to identify causal effector genes. Indeed, the identification of the true gene targets is a crucial precursor to a rational therapeutic and diagnostic development leveraging genetic information.

The majority of type 1 diabetes-associated SNPs map to regions distant from genes [8]; thus, genomic maps are needed that determine how these SNPs might influence chromatin accessibility, transcription factor binding and the physical structure of the genome in order to identify the target genes important in disease. The vast majority (>95%) of the human genome is inaccessible to the machinery that regulates gene expression [19]; thus, essentially all transcription factor and RNA polymerase binding is concentrated at open chromatin regions. Therefore, maps of open, transposase-accessible chromatin (e.g. generated using the assay for transposase-accessible chromatin using sequencing [ATAC-Seq] [20] at the multi- or single-cell level) can identify regions of potential regulatory significance across multiple tissues. One example used the open chromatin landscapes of follicular helper T cells (TFH) from human tonsil to identify functional variants implicated by GWAS of systemic lupus erythematosus (SLE) [21]. The proxies of SLE 'sentinel' SNPs (those SNPs in strong linkage disequilibrium with the most associated SNP from the GWAS) are highly enriched in the open chromatin of TFH cells, a cell type critical for the development of autoantibodies characteristic of SLE, compared with naive CD4+ T cells. These accessible SLE SNPs were more likely to be located in the promoters of genes highly expressed in TFH cells and involved in other systemic autoimmune disorders, including type 1 diabetes. Genetic variation in a promoter can influence expression of its downstream gene, given proximity of the disease-associated SNPs and recognised cis effects.

A similar prediction, however, is not obvious from maps of open chromatin for more distal SNPs. When disease-associated SNPs are cis eQTLs, they may also physically interact with the promoter (or promoters) that they regulate (for an example, see [22]). These interactions can be detected using chromosome conformation capture, examining not only promoter interactions but also interaction at a distance between promoters, enhancers, silencers and other elements. Chromatin conformation capture-based techniques have the ability to determine whether chromatin 'looping' contributes to human disease at key locations associated with complex traits. In particular, one can now leverage recent findings that have revealed topologically associating domains (TADs) [23],

that are largely tissue-independent chromatin compartments within which most enhancer–promoter contacts occur. Effectively, TADs may establish the boundaries where interactions can occur for a given genomic location, thereby providing a defined shortlist of candidate genes within a locus, among which at least one is highly likely to be a causal effector gene. Whole genome, promoter-focused Capture C, a version of chromatin conformation capture, relates SNPs in the distal regulatory regions to changes in expression of their target genes [24, 25]. High-resolution spatial epigenomic approaches for common complex traits have been able to physically link strongly associated SNPs with their target genes for traits such as SLE [21] and bone mineral density [26, 27], as well as type 2 diabetes and type 1 diabetes (discussed below). These studies demonstrate that 3D regulatory architectures are a consistent feature of highly expressed, lineage-specific genes involved in specialised functions in disease-relevant cell types (Fig. 1).

Type 1 diabetes distant regulators may differ from those in type 2 diabetes

While prior genetic analysis directly implicated the immune system in genetic risk of type 1 diabetes with lack of enrichment in islet regulatory regions [8], other biological pathways are likely to be involved. The impact of type 1 diabetes associated SNPs on islets, through the targeting of the autoimmune attack on beta cells, may occur prior to clinical onset (e.g. at the initiation or progression stage); alternatively, the type 1 diabetes-associated variants may act directly at the beta cell level in response to a perturbation (e.g. inflammation).

In the context of its type 2 diabetes counterpart [28], islet accessible chromatin peaks aided the identification of active enhancers and promoters through the use of islet samples and 3D chromatin maps by identifying chromatin loops enriched at such genomic features. Of the >6000 islet active enhancers that mapped to a chromatin loop anchor, half were in a loop to a gene promoter. Many of these enhancers looped to a promoter over long distances (mean 165 kb, with 14% over 500 kb, and >3% over 1 Mb). These distal islet enhancer chromatin loops were correlated with islet-specific gene expression (as

Fig. 1 Possible promoter interactions in open chromatin regions, suggesting how SNPs may regulate distant genes through physical contact with non-adjacent promoters, defining likely effector genes and targets. This figure is available as a downloadable slide
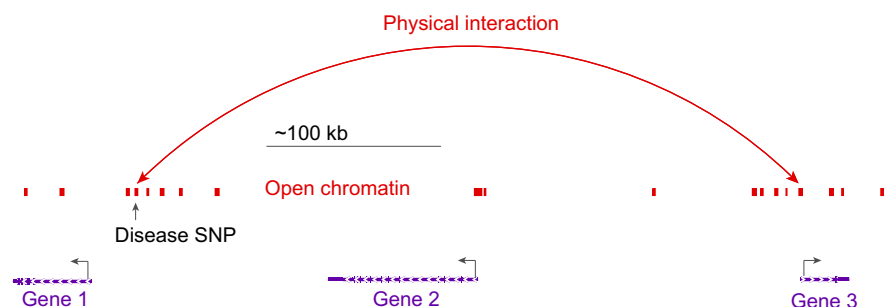
assessed by the presence of eQTLs), with the strongest evidence observed for active promoter and enhancer SNPs proximal to genes. Genome-wide enrichment of SNPs was observed in active islet regulatory elements within chromatin loops. SNPs associated with type 2 diabetes and in active islet enhancers had, on average, two candidate target effector genes, including some that were >500 kb from the SNP. In a different study [29], experimental perturbation (glucose stimulation) in human islets was used to identify over 1300 enhancer hubs that had features of regulatory domains controlling genes involved in islet cell function and differentiation. Factoring in islet hub SNPs in a polygenic risk score improved identification of individuals with risk of type 2 diabetes, possibly acting through islet gene regulation and insulin secretion pathways.

The effect of inflammatory cytokine (IFN-γ and IL-1β) exposure on the beta cell as a model of initiation of type 1 diabetes has been investigated using 3D mapping approaches to detect novel targets [30]. After cytokine exposure, ~12,500 sites were identified that became accessible and correlated with H3K27ac activity (acetylation at the 27th lysine residue of the histone H3 protein, representing evidence of an active enhancer). Inducible regulatory elements (IREs) were identified, with two-thirds becoming both chromatin accessible and showing enhancer activity after cytokine treatment (neo-IREs), and the other third, which were already accessible, gaining only enhancer activity after cytokine treatment. The proinflammatory cytokine exposure was hypothesised to induce a beta cell response by induction of new distal regulatory elements and binding of transcription factors involved in the inflammatory response. In islet 3D chromatin structure studies, the promoters of 13 genes exhibited strong induction of expression by cytokine exposure, with their promoters gaining chromatin interactions. Distal genomic regions formed specific DNA looping events with new human islet cytokine responsive enhancer–promoter interactions. In this system, variants associated with type 2 diabetes (not type 1 diabetes) overlapped human islet responsive regulatory elements that were not cytokine responsive; however, human islet IREs (induced by cytokine exposure) were enriched for SNPs associated with type 1 diabetes (not type 2 diabetes). In two known type 1 diabetes loci, risk SNPs (rs78037977 in 1q24.3 and rs193778 in 16q13.13) directly overlapped IREs in islets. An allele of rs78037977 at 1q24.3 (common in individuals of European ancestry but rare in those of other ancestries) disrupts cytokine exposure-specific enhancer activity and interacts with *TNFSF18*, a gene ~300 kb from this SNP but activated in islets upon cytokine exposure. At 16q13.13, rs193778 is common in most ancestries (yet monomorphic in Asian populations) and increases enhancer activity, having strong chromatin contact with the promoter of *DEXI*, a gene ~300 kb distal to the sentinel SNP and previously implicated in type 1 diabetes [15, 16].

Detection and validation of targets of SNPs

Multiple levels of evidence are necessary to determine which SNPs in a locus are likely to be causal and how these variants regulate target effector genes and their products. Candidate SNPs may influence gene expression in appropriate cell types (e.g. detected by applying RNA-seq in immune cells and beta cells) and on transcription (e.g. detected using ATAC-seq for evidence of transposase-accessible chromatin). These and other types of evidence provide a prioritisation for mapping interactions between promoters and distal regulatory elements, with increasing resolution [31]. As discussed, the target gene may not be the nearest neighbour to the causal SNP; furthermore, SNP-connected putative effector genes may have been implicated in other diseases (for example, see [32]), not only providing additional evidence for causality but also providing new therapeutic options.

CRISPR/Cas9 genome editing can be used to confirm that accessible SNPs in one gene reside in novel, *cis*-regulatory elements for other genes with known roles in function disease risk [21]. In type 2 diabetes, the most strongly associated SNP lies within the *TCF7L2* gene [33], with the rs7903146 T allele in intron 3 widely implicated as the causal variant [34]. Informed by observation of chromatin conformation, in addition to influencing *TCF7L2* expression itself, CRISPR/Cas9-mediated editing of rs7903146 dramatically reduced *ACSL5* gene expression and protein levels [14], thus implicating a putative additional effector gene at this locus. *ACSL5* is three genes away from *TCF7L2* and encodes an enzyme (acyl-CoA synthetase long chain family, member 5) with known roles in mammalian fatty acid metabolism. In addition, the knockout mouse for *ACSL5* has increased insulin sensitivity [35]. A similar approach has been employed for epigenome editing of enhancer–promoter assignments in a cell model for type 2 diabetes [29]. With the increasing number of targets being generated, there is a need to validate such variant-to-gene connections at scale. Emerging techniques, such as massively parallel reporter assays [36] and wholesale CRISPR-based perturbation of implicated enhancers, are growing areas that will meet this need.

From omics to therapeutic targets

Using multiple lines of evidence (genomics, transcriptomics, DNA methylation, perturbation, gene editing), selection and prioritisation of potential therapeutic targets from validated effector gene lists can proceed using a translational 'bench-to-bedside' rationale. Gene products not previously implicated in type 1 diabetes, but currently targeted with therapeutics approved by the US Food and drug administration in autoimmune disease settings, could make excellent drug repurposing candidates. Future drug repurposing candidates would include

targets with modalities in overlapping biological pathways. Gene products in need of more potent and/or selective agonists or antagonists could be the targets of future drug development efforts.

Genetics as predictors of stages of type 1 diabetes

Variants associated with type 1 diabetes in prevalent case–control or affected family studies (primarily of young-onset, Northern European ancestry) may not translate to other ancestries, adults or to the initiation and progress of islet autoimmunity. The T1DGC characterised affected family members for genetic contributions to the presence of islet and other organ-specific autoantibodies [37]. HLA alleles (*DRB1*0101* and *DRB1*0404*) and the *PTPN22* rs2476601 (R620W) locus were associated with autoimmunity, while variants in *IFIH1*, *PTPN22*, *SH2B3*, *BACH2* and *CTLA4* were associated with occurrence of multiple autoantibodies [38]. However, this study was conducted in those with existing disease.

Rather than consider risk in terms of single SNPs, genetic risk scores (GRS) sum the risk alleles for each associated SNP (0, 1 or 2), weighted by the effect of the SNP on the phenotype. The use of the GRS in type 1 diabetes permits an assessment of 'global' impact of SNPs as a single value, although the composition of the GRS can vary by the number of SNPs included, the population tested and phenotypic definition. In The Environmental Determinants of Diabetes in the Young (TEDDY) study, a type 1 diabetes GRS (T1D-GRS) in the upper quartile increased the risk of developing multiple auto-antibodies by the age of 6 years from 5.8% to 11.0% (compared with 4.1% in the lower T1D-GRS quartile) [39]. The risk of developing type 1 diabetes by age 10 years increased from 3.7% to 7.6% in those with a high T1D-GRS (compared with 2.7% in those without). Children in the highest T1D-GRS quartile had an earlier age of onset of islet autoimmunity, a faster progression from single to multiple autoantibodies, and were more likely to develop type 1 diabetes [40]. A high T1D-GRS also predicts proliferation responses to one or more islet antigens [41]. A T1D-GRS has the potential for use in newborn screening for genetic risk of type 1 diabetes [42], classification of adult-onset disease and progression of islet autoimmunity [43]. Furthermore, a T1D-GRS has the potential power to predict when those with type 2 diabetes may require insulin administration [44].

From bench to bedside to community

Given the low prevalence of type 1 diabetes in the general population (~4/1000), even a highly sensitive and specific test will likely yield low predictive values. While

knowledge of associated risk variants and their function and target effector genes offers the opportunity to identify novel therapeutic pathways [45], there is uncertainty as to how genetics can drive risk prediction. The overall risk of type 1 diabetes is, in part, due to genetic factors, so a high T1D-GRS does not mean one is destined to develop type 1 diabetes per se, just as a low T1D-GRS is not necessarily protective from the disease. Nonetheless, a relatively simple T1D-GRS can identify >10% risk of developing autoimmunity before the age of 6 years [42], making genetic screening a real possibility [46]. Currently, genetics is the only tool detecting those at risk prior to development of islet autoimmunity, until environmental factors (or other novel biomarkers) that trigger the autoimmunity are identified. It is likely that population screening will use a combination of genetics with emergent risk factor testing to determine those eligible for intervention (e.g. intervention trials prior to disease onset, such as with oral insulin therapy) and is being tested now [47]. With the ever-expanding reliance on the merger of electronic health records with biobanks, these research directions could be directly applied to prediction, intervention and treatment in diverse and previously under-served populations.

Conclusions

The genetic basis of type 1 diabetes is becoming increasingly clear, particularly in Northern European paediatric populations. These gains have yet to impact prediction, prevention and treatment strategies. The vast majority of genetic variants associated with type 1 diabetes reside in regulatory regions of the genome (not in coding regions of genes). Thus, integration of genomics with gene expression, epigenetics and 3D mapping of interactions within the genome are needed to determine the likely target effector genes involved in type 1 diabetes pathogenesis. Identification of new classes of genetic variants associated with type 1 diabetes may enhance the application of genetic risk scores in many ways, from prediction of risk to the need for insulin treatment in type 2 diabetes. Many needs remain, including studies in ethnically diverse populations and in adults, all of which may provide the biological insights needed to translate genomic findings into precision diabetes medicine.

References

1. Singal DP, Blajchman MA (1973) Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue autoantibodies in patients with diabetes mellitus. Diabetes 22(6):429–432. https://doi.org/10.2337/diab.22.6.429

2. Bell GI, Horita S, Karam JH (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33(2):176–183. https://doi.org/10.2337/diab.33.2.176

3. Lenz TL, Deutsch AJ, Han B et al (2015) Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. Nat Genet 47(9):1085–1090. https://doi.org/10.1038/ng.3379

4. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447(7145):661–678. https://doi.org/10.1038/nature05911

5. Hakonarson H, Grant SF, Bradfield JP et al (2007) A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 448(7153):591–594. https://doi.org/10.1038/nature06010

6. Barrett JC, Clayton DG, Concannon P et al (2009) Genome-wide association study and meta-analysis find that over 40 loci affect type 1 diabetes. Nat Genet 41(6):703–707. https://doi.org/10.1038/ng.381

7. Concannon P, Rich SS, Nepom GT (2009) Genetics of type 1A diabetes. N Engl J Med 360(16):1646–1654. https://doi.org/10.1056/NEJMra0808284

8. Onengut-Gumuscu S, Chen WM, Burren O et al (2015) Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet 47(4):381–386. https://doi.org/10.1038/ng.3245

9. Noble JA, Johnson J, Lane JA, Valdes AM (2013) HLA class II genotyping of African American type 1 diabetic patients reveals associations unique to African haplotypes. Diabetes 62(9):3292–3299. https://doi.org/10.2337/db13-0094

10. Onengut-Gumuscu S, Chen WM, Robertson CC et al (2019) Type 1 diabetes risk in African-ancestry participants and utility of an ancestry-specific genetic risk score. Diabetes Care 42(3):406–415. https://doi.org/10.2337/dc18-1727

11. Nica AC, Dermitzakis ET (2013) Expression quantitative trait loci: present and future. Philos Trans R Soc Lond Ser B Biol Sci 368(1620):20120362. https://doi.org/10.1098/rstb.2012.0362

12. Smemo S, Tena JJ, Kim KH et al (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507(7492):371–375. https://doi.org/10.1038/nature13138

13. Claussnitzer M, Dankel SN, Kim KH et al (2015) FTO obesity variant circuitry and adipocyte browning in humans. N Engl J Med 373(10):895–907. https://doi.org/10.1056/NEJMoa1502214

14. Xia Q, Chesi A, Manduchi E et al (2016) The type 2 diabetes presumed causal variant within TCF7L2 resides in an element that controls the expression of ACSL5. Diabetologia 59(11):2360–2368. https://doi.org/10.1007/s00125-016-4077-2

15. Davison LJ, Wallace C, Cooper JD et al (2012) Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. Hum Mol Genet 21(2):322–333. https://doi.org/10.1093/hmg/ddr468

16. Tomlinson MJ 4th, Pitsillides A, Pickin R et al (2014) Fine mapping and functional studies of risk variants for type 1 diabetes at chromosome 16p13.13. Diabetes 63(12):4360–4368. https://doi.org/10.2337/db13-1785

17. Nikolic T, Woittiez NJC, van der Slik A et al (2017) Differential transcriptome of tolerogenic versus inflammatory dendritic cells points to modulated T1D genetic risk and enriched immune regulation. Genes Immun 18(3):176–183. https://doi.org/10.1038/gene.2017.18

18. Mehdi AM, Hamilton-Williams EE, Cristino A et al (2018) A peripheral blood transcriptomic signature predicts autoantibody development in infants at risk of type 1 diabetes. JCI Insight 3(5): e98212. https://doi.org/10.1172/jci.insight.98212

19. Thurman RE, Rynes E, Humbert R et al (2012) The accessible chromatin landscape of the human genome. Nature 489(7414): 75–82. https://doi.org/10.1038/nature11232

20. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10(12):1213–1218. https://doi.org/10.1038/nmeth.2688

21. Su C, Johnson ME, Torres A et al (2020) Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. Nat Commun 11(1):3294. https://doi.org/10.1038/s41467-020-17089-5

22. De Laat W, Dekker J (2012) 3C-based technologies to study the shape of the genome. Methods 58(3):189–191. https://doi.org/10.1016/j.ymeth.2012.11.005

23. Dixon JR, Siddarth S, Yue F et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398):376–380. https://doi.org/10.1038/nature11082

24. Hughes JR, Roberts N, McGowan S et al (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet 46(2):205–212. https://doi.org/10.1038/ng.2871

25. Mifsud B, Tavares-Cadete F, Young AN et al (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 47(6):598–606. https://doi.org/10.1038/ng.3286

26. Chesi A, Wagley Y, Johnson ME et al (2019) Genome-scale Capture C promoter interactions implicate effort genes at GWAS loci for bone mineral density. Nat Commun 10(1):1260. https://doi.org/10.1038/s41467-019-09302-x

27. Cousminer DL, Wagley Y, Pippin J et al (2020) Genome-wide association study implicates novel loci and reveals candidate effector genes for longitudinal pediatric bone accrual through variant-to-gene mapping. medRxiv https://www.medrxiv.org/content/10.1101/2020.02.17.20024133v1

28. Greenwald WW, Chiou J, Yan J et al (2019) Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. Nat Commun 10(1):2078. https://doi.org/10.1038/s41467-019-099975-4

29. Miguel-Escalada I, Bonas-Guarch S, Cebola I et al (2019) Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. Nat Genet 51(7):1137–1148. https://doi.org/10.1038/s41588-019-0457-0

30. Ramos-Rodriguez M, Raurell-Vila H, Colli ML et al (2019) The impact of proinflammatory cytokines on the β-cell regulatory landscape provides insights into the genetics of type 1 diabetes. Nat Genet 51(11):1588–1595. https://doi.org/10.1038/s41588-019-0524-6

31. Kumasaka N, Knights A, Gaffney DJ (2019) High-resolution genetic mapping of putative causal interactions between regions of open chromatin. Nat Genet 51(1):128–137. https://doi.org/10.1038/s41588-018-0278-6

32. Javierre BM, Burren OS, Wilder SP et al (2016) Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 167(5):1369–1384. https://doi.org/10.1016/j.cell.2016.09.037

33. Grant SF, Thorleifsson G, Reynisdottir I et al (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38(3):320–323. https://doi.org/10.1038/ng1732

34. Helgason A, Palsson S, Thorleifsson G et al (2007) Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. Nat Genet 39(2):218–225. https://doi.org/10.1038/ng1960

35. Bowman TA, O'Keeffe KR, D'Aquila T et al (2016) Acyl CoA synthetase 5 (ACSL5) ablation in mice increases energy expenditure and insulin sensitivity and delays fat absorption. Mol Metab 5(3):210–220. https://doi.org/10.1016/j.molmet.2016.01.001

36. Ernst J, Melnikov A, Zhang X et al (2016) Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. Nat Biotechnol 34(11):1180–1190. https://doi.org/10.1038/nbt.3678

37. Rich SS, Concannon P (2015) Role of type 1 diabetes-associated SNPs on autoantibody positivity in the Type 1 Diabetes Genetics Consortium: overview. Diabetes Care 38(Suppl 2):S1–S3. https://doi.org/10.2337/dcs15-2001

38. Rich SS, Concannon P (2015) Summary of the Type 1 Diabetes Genetics Consortium Autoantibody Workshop. Diabetes Care 38(Suppl 2):S45–S48. https://doi.org/10.2337/dcs15-2008

39. Bonifacio E, Beyerlein A, Hippich M et al (2018) Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: a prospective study in children. PLoS Med 15(4):e1002548. https://doi.org/10.1371/journal.pmed.1002548

40. Beyerlein A, Bonifacio E, Vehik K et al (2019) Progression from islet autoimmunity to clinical type 1 diabetes is influenced by genetic factors: results from the prospective TEDDY study. J Med Genet 56(9):602–605. https://doi.org/10.1136/jmedgenet-2018-105532

41. Claessens LA, Wesselius J, van Lummel M et al (2020) Clinical and genetic correlates of islet autoimmune signatures in juvenile-onset type 1 diabetes. Diabetologia 63(2):351–361. https://doi.org/10.1007/s00125-019-05032-3

42. Sharp SA, Rich SS, Wood AR et al (2019) Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. Diabetes Care 42(2):200–207. https://doi.org/10.2337/dc18-1785

43. Redondo MJ, Geyer S, Steck AK et al (2018) A type 1 diabetes genetic risk score predicts progression of islet autoimmunity and development of type 1 diabetes in individuals at risk. Diabetes Care 41(9):1887–1894. https://doi.org/10.2337/dc18-0087

44. Thomas NJ, Jones SE, Weedon MN et al (2018) Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank. Lancet Diabetes Endocrinol 6(2):122–129. https://doi.org/10.1016/S2213-8587(17)30362-5

45. Fang H, ULTRA-DD Consortium, De Wolf H et al (2019) A genetics-led approach defines the drug target landscape of 30 immune-related traits. Nat Genet 51(7):1082–1091. https://doi.org/10.1038/s41588-019-0456-1

46. Ziegler AG, Achenbach P, Berner R et al (2019) Oral insulin therapy for primary prevention of type 1 diabetes in infants with high genetic risk: the GPPAD-POInT (global platform for the prevention of autoimmune diabetes primary oral insulin trial) study protocol. BMJ Open 9(6):e028578. https://doi.org/10.1136/bmjopen-2018-028578

47. Winkler C, Haupt F, Heigermoser M et al (2019) Identification of infants with increased type 1 diabetes genetic risk for enrollment into primary prevention trials-GPPAD-02 study design and firsts results. Pediatr Diabetes 20(6):720–727. https://doi.org/10.1111/pedi.12870