Dissecting diabetes/metabolic disease mechanisms using pluripotent stem cells and genome editing tools

Adrian Kee Keong Teo 1,2,4,6,*, Manoj K. Gupta 1, Alessandro Doria 3, Rohit N. Kulkarni 1,*

ABSTRACT

Background: Diabetes and metabolic syndromes are chronic, devastating diseases with increasing prevalence. Human pluripotent stem cells are gaining popularity in their usage for human in vitro disease modeling. With recent rapid advances in genome editing tools, these cells can now be genetically manipulated with relative ease to study how genes and gene variants contribute to diabetes and metabolic syndromes.

Scope of review: We highlight the diabetes and metabolic genes and gene variants, which could potentially be studied, using two powerful technologies — human pluripotent stem cells (hiPSCs) and genome editing tools — to aid the elucidation of yet elusive mechanisms underlying these complex diseases.

Major conclusions: hiPSCs and the advancing genome editing tools appear to be a timely and potent combination for probing molecular mechanism(s) underlying diseases such as diabetes and metabolic syndromes. The knowledge gained from these hiPSC-based disease modeling studies can potentially be translated into the clinics by guiding clinicians on the appropriate type of medication to use for each condition based on the mechanism of action of the disease.

Keywords Diabetes; Metabolic disease; Pluripotent stem cells; Genome editing; CRISPR/Cas; Disease modeling

1. INTRODUCTION

Diabetes and metabolic syndromes are chronic devastating diseases with increasing prevalence. Although a subset of patients develop diabetes and metabolic syndromes due to monogenic mutations, a majority are due to complex gene-environment interactions. Human pluripotent stem cells are gaining popularity in their usage for human in vitro disease modeling. With recent rapid advances in genome editing tools, these cells can now be genetically manipulated with relative ease. In this review, we highlight the diabetes and metabolic genes and gene variants which could potentially be studied using these two powerful technologies (human pluripotent stem cells and genome editing tools) to aid the elucidation of currently elusive mechanisms underlying these complex diseases.

2. INCREASED USAGE OF HUMAN PLURIPOTENT STEM CELLS (hPSCs) FOR HUMAN IN VITRO DISEASE MODELING

Since their discovery in 2007 [1], human induced pluripotent stem cells (hiPSCs) have become a major focus of research [2]. In terms of using hiPSCs for understanding human disease mechanisms, numerous hiPSCs have been generated from patients with a wide variety of diseases [3]. Many of these hiPSCs have been derived from patients with single gene mutations as those monogenic diseases are those that can be modeled with the greatest ease. However, complex diseases such as diabetic cardiomyopathy can also be potentially modeled in vitro [4]. Diabetes, one of the most common metabolic disorders, is a major cause of increased morbidity and mortality for which a cure remains elusive. Classical methodologies for studying diabetes disease mechanisms, such as the use of rodent models, provide significant insights into disease biology. However, it is also known that rodent models do not fully recapitulate the diabetes phenotype in humans, for instance in several types of maturity onset diabetes of the young (MODY) (autosomal dominant conditions) (mice with heterozygous mutations do not develop diabetes) [5] and Wolfram Syndrome [6]. Therefore, hiPSCs could be derived from patients with these forms of diabetes and differentiated into pancreatic beta cells to serve as suitable alternative human models for diabetes research [7,8,105]. To further understand the etiology of...
more common forms of diabetes, hiPSCs have been derived from both type 1 diabetes (T1D) and type 2 diabetes (T2D) patients [9–11]. However, due to its complexity, T1D and T2D have yet to be convincingly modeled in vitro with these hiPSCs [12]. In this review, we propose and discuss the use of genome editing tools to generate hiPSCs carrying specific gene variants associated with diabetes along with isogenic controls. Once these hiPSCs have been derived, they can then be differentiated into various cell types relevant to diabetes, such as pancreatic cells, skeletal myocytes, adipocytes and hepatocytes, in order to study their biology and for drug screening purposes.

3. ADVANCES IN GENOME EDITING TOOLS THAT CAN BE APPLIED TO HUMAN PLURIPOTENT STEM CELL BIOLOGY

Genome editing, the genetic engineering (insertion, deletion or replacement) of DNA, is an approach to understand the function of a gene of interest. Genome editing tools such as the conventional homologous recombination (HR), or the more advanced (and advancing) tools such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) or clustered regularly interspaced short palindromic repeat (CRISPR)/Cas system [13] can be applied to hiPSCs in the presence of a donor template to either 1) create naturally occurring mutations or 2) repair a mutation to generate isogenic controls in hiPSCs. Compared to ZFNs, TALENs are simpler to engineer [14] and may exhibit greater specificity with less cytotoxicity [15]. The CRISPR/Cas9 system has a significant advantage over ZFN and TALEN as it relies on RNA to home onto DNA whereas ZFN and TALEN depend on custom-making proteins for each specific DNA target. Thus, CRISPR/Cas9 is technically easier to use and is more efficient at cutting target DNA [16]. Given these advantages, coupled with the fact that the CRISPR/Cas9 system is cost-effective and easily accessible in repositories such as Addgene (https://www.addgene.org/), this system has led to an exponential increase in genome engineering of cells and model organisms to study disease biology. The subsequent differentiation of these genetically engineered hiPSCs into cell types of interest would thus facilitate the study of disease biology. We will briefly introduce each of these genome editing methodologies.

3.1. Homologous recombination (HR)/Gene targeting

Homologous recombination (HR), or gene targeting, is a method for introducing a particular DNA sequence into a host genome (Figure 1) [17]. When DNA double-strand breaks (DSBs) occur in cells (e.g. due to DNA damage), the DNA can be repaired either via HR or via (error-prone) non-homologous end joining (NHEJ) [18]. DNA repair by HR occurs either via the double-strand break repair (DSBR) pathway or the synthesis-dependent strand annealing (SSDSA) pathway. The former commonly results in an exchange of nucleotide sequence(s) between the two strands of DNA (crossover), leading to genetic recombination in contrast to the SDSA pathway, which does not result in a crossover. The natural process of DNA damage and DSBR pathway is the conventional gene targeting method to insert transgenes into the genome without the use of nucleases (ZFNs, TALENs or CRISPR/Cas).

3.2. Zinc finger nuclease (ZFN)

The first advancement in genome editing arose from the discovery of zinc finger nucleases (ZFNs), synthetic restriction enzymes containing DNA-binding domain(s) fused to a DNA-cleavage domain (typically a non-specific cleavage domain from bacterial type IIS FokI restriction enzyme; located at the C-terminus) [19]. These DNA-binding domains, which contain 3–6 zinc finger repeats and are able to recognize 9–18 base pairs, allow homing to target DNA followed by DNA cleavage (Figure 1) [20]. For DNA cleavage to occur, however, the cleavage domain needs to dimerize, necessitating a pair of ZFNs to bind opposite strands of DNA and flank a region of interest. ZFNs induce DNA DSBs and stimulate the endogenous DNA repair machinery [21], which, in the presence of a homologous donor template, can result in a replacement of the target allele. In contrast, NHEJ operates in the absence of a homologous donor template. This strategy may be suitable for disabling autosomal dominant mutations present in hiPSCs derived from diseased patients.

3.3. Transcription activator-like effector nuclease (TALEN)

Following on the heels of the use of ZFNs, transcription activator-like effector nucleases (TALENs) were discovered. TALE is a Xanthomonas bacterial protein containing a domain of tandem repeats (33–34 amino acids), which can target DNA. These tandem repeats contain repeat-variable diresidues (RVDs; 12th and 13th amino acid), which dictate DNA-binding specificity. By combining the appropriate RVDs, TALEs can be custom-made to target DNA. When fused to a DNA-cleavage domain (e.g. FokI endonuclease), TALE nucleases (TALENs) are generated [22]. Like ZFNs, a pair of TALENs induces DSBs and relies upon HR or NHEJ, to either introduce foreign DNA for gene repair or disrupt endogenous gene function, respectively (Figure 1).

3.4. Clustered regularly interspaced short palindromic repeat (CRISPR)/Cas

Very recently, another technological breakthrough led to the discovery of clustered regularly interspaced short palindromic repeats (CRISPRs), a repetitive pattern (between 24 and 48 base pairs) in which DNA sequences would be followed by nearly the same sequence in reverse orientation followed by spacer DNA. This unique pattern, first identified in E. coli bacteria, which integrated the
sequence of phages (spacer DNA) to fend off future attacks, was often found to be associated with CRISPR-associated (Cas) genes. Upon phage attack, the bacterium first transcribes the spacer and palindromic DNA into a long RNA molecule. This RNA molecule is then cleaved by RNase III, involving Cas9 (a nuclease which has two active sites for cutting each strand of the DNA double helix) and trans-activating CRISPR RNA (tracrRNA), into short spacer-derived RNA (CRISPR RNA; crRNA) in the cell [23].

tracrRNA and crRNA can now be intelligently combined into a “guide RNA” molecule, and when combined with Cas9 could target specific DNA and cut accurately (Figure 1) [24]. CRISPR/Cas9 has also been modified into CRISPR interference (CRISPRi), which can reversibly turn off genes by targeting a site without cutting it (Cas9 nuclease rendered inactive), or the inclusion of activators or repressors (synthetic transcription factors) to regulate gene activity at promoter regions [25,26]. Besides the constitutive expression of CRISPR/Cas9, there are alternative systems, which allow for conditional induction via the doxycycline- or Tet-regulated Cas9 system. Current CRISPR/Cas systems are from Streptococcus pyogenes, Streptococcus thermophilus, Neisseria meningitides and Treponema denticola.

Figure 1: Genome editing methodologies which can be applied to human pluripotent stem cells. Homologous recombination (HR), or the more advanced tools such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) or clustered regularly interspaced short palindromic repeat (CRISPR)/Cas system can be applied to human pluripotent stem cells (hPSCs) either to 1) create naturally occurring mutations or 2) repair a mutation to generate isogenic controls in hPSCs, to understand the function of a gene of interest.
Minireview

3.5. Caveats of advanced genome editing tools

**Off-target effects.** The DNA-binding domains of ZFNs and TALENs need to be very specific for the target site to avoid off-target cleavage, which results in unwanted mutations and potentially cytotoxic effects [27]. CRISPR/Cas9 is also known to generate off-target alterations, albeit apparently at low incidence [28,29], since mispairing is allowed between the guide RNA and the genomic DNA. Nonetheless, caution is required in their design and use. Some strategies involving the optimization of the guide RNA/Cas9 include using of software tools to predict potential off-target sites (http://omictools.com/crispr-cas9- c1268-p1.html), truncating the guide RNA (<20 nucleotides) to decrease off-target mutagenesis [30], lowering the dosage of guide RNA and Cas9 plasmids, and decreasing the number of mismatches between the guide RNA and the genomic DNA. A “double nick” system with Cas9 nickase, which contains a single inactive catalytic domain, may also be used [31–33].

**Repeated cleavage of a repaired site.** Due to the high cleavage efficiency of these genome editing tools, there is a possibility of re-cutting a repaired site [34]. One way to mitigate this is to introduce silent mutations in the donor template to minimize subsequent nuclease binding [35].

3.6. Single-stranded oligodeoxyribonucleotide (ssODN)

Since the generation of conventional donor template constructs for conventional HR is time consuming, single-stranded oligodeoxyribonucleotides (ssODNs) are currently being employed in concert with these advanced genome editing tools for high frequency “footprint-free” genome editing [36]. These genome editing tools act by introducing a sequence-specific DSB within the ssODN homology region (thus increasing recombination frequency) and facilitating HR via the DSBR pathway using ssODN as a template [37].

ssODNs are short synthetic nucleotides designed to be either exactly complementary to the target DNA sequence (for gene correction) or contain nucleotide mismatch(es), insertions or deletions. ssODNs are very simple to use, as opposed to conventional donor template constructs (plasmids or viral vectors), predicted to have high safety and fidelity with minimal immunotoxicity, and do not appear to have off-target effects [36]. Studies combining ZFNs with ssODNs were among the first to suggest the possibility of high frequency genome editing [38] and have been successfully applied to Parkinson’s disease-hiPSCs [39].

4. COUPLING HUMAN PLURIPOTENT STEM CELLS AND GENOME EDITING TOOLS TO INVESTIGATE DIABETES AND METABOLIC DISEASE MECHANISMS

4.1. Examples of studying disease mechanisms with hPSCs and genome editing tools

Traditional HR can be used for gene correction to demonstrate genotype-phenotype causality albeit with very low rates of efficiency. ZFNs were first demonstrated in 2007 to be capable of genome editing in human embryonic stem cells (hESCs) [40]. Their ability to enhance HR-mediated gene targeting [41,42] has led to their use in multiple studies including targeted gene disruption at the CCR5 locus (HIV-1 resistance) [43] and correction in the α-synuclein gene (susceptibility variant for Parkinson’s disease) [39], α1-antitrypsin [44], sickle cell anemia mutation [35,45], gp91(phox) (X-linked chronic granulomatous disease) [46], α-thalassemia genes [47] and glucokinase (GCK) [7]. Shortly after, TALENs were employed for genetic engineering in hPSCs [48,49] for correcting the COL7A1 [50] and α1-antitrypsin genes [51]. Given the superior cutting efficiency, CRISPR/Cas9 is increasingly becoming the favored choice for genome editing in hPSCs [16,52].

4.2. Employing hPSCs and genome editing tools to study diabetes and metabolic syndromes

In general, the strategy to carry out in vitro disease modeling of diabetes and related metabolic syndromes with hPSCs and genome editing tools would be to 1) derive hiPSCs from patients with these conditions, 2) generate “repaired/corrected” isogenic controls [53] and then 3) differentiate them into pancreatic cells or target cells of relevance, such as immune cells in the case of T1D or myocytes, adipocytes and hepatocytes in the case of T2D (Figure 2). If patient material is inaccessible, one could introduce (naturally occurring) mutations or gene variants into hESCs and differentiate them accordingly to study disease mechanisms. Since excellent reviews have been published recently, we will provide a brief overview to familiarize the reader with the classification of diabetes and metabolic disorders.

4.3. Genetic determinants of diabetes and metabolic syndromes

Diabetes is a metabolic disease characterized by uncontrolled hyperglycemia resulting from the variable combination of dysfunctional insulin secretion by pancreatic beta cells and insulin resistance. It is generally classified into monogenic diabetes (maturity onset diabetes of the young [MODY], neonatal diabetes, mitochondrial diabetes [54,55], syndromes of insulin resistance) [56], type 1 diabetes (T1D) and type 2 diabetes (T2D). The metabolic syndrome is a combination of conditions including central obesity (as measured by waist circumference), dyslipidemia (high triglycerides, reduced high density lipoprotein cholesterol), hypertension and elevated fasting plasma glucose. These conditions increase the risk of diabetes and atherosclerotic vascular disease. The genetic variants responsible for monogenic diabetes are typically rare and have high penetrance whereas the variants involved in the etiology of the complex forms of T1D and T2D are relatively common and have low penetrance. Some of the known genetic determinants of each of these types of diabetes and metabolic syndromes are summarized in Table 1.

4.3.1. MODY, neonatal diabetes and mitochondrial diabetes

Patients with MODY, neonatal diabetes and mitochondrial diabetes (Table 1) commonly exhibit impairments in one or more of the three main beta cell functions: 1) glucose sensing, 2) making and storing insulin, and 3) secreting insulin on demand. For instance, mutations in genes such as GCK [57] and glucose transporter SLC2A2 (also known as GLUT2) [58] result in altered glucose sensing, which, in turn, results in an increased blood glucose threshold for insulin secretion, decreased hepatic glycogen storage, and increased hepatic glucoseogenesis [59–61]. Mutations in several transcription factors result in defects in pancreatic beta cell development, which directly impact insulin synthesis and function while mutations in the insulin gene (INS) obviously affect the key hormone made by pancreatic beta cells [62]. ATP synthesis defect (mitochondrial diabetes) and mutations in ATP-sensitive potassium channel subunits (channel-building Kir6.2 [potassium inwardly-rectifying channel, subfamily J, member 11; KCNJ11] and regulatory SUR1 [ATP-binding cassette transporter subfamily J member 8]; ABC6β) all affect insulin secretion [63]. One strategy to study these monogenic syndromes would be to derive hiPSCs from these patients [105], differentiate them into pancreatic...
progenitors and then transplant these progenitors into immunocompromised (SCID-Beige or NSG) mice for in vivo maturation (Figure 2). This methodology has been recently used to successfully model MODY2, demonstrating that beta cells derived from hiPSCs with GCK mutation are indeed less sensitive to glucose levels [7]. Endoplasmic reticulum (ER) stress-related diabetes in patients with Wolfram syndrome has also been modeled using hiPSC-derived beta cells, demonstrating that WFS1 protein maintains ER function in beta cells by acting upstream of the unfolded protein response (UPR) pathways [8]. These two studies indicate that hiPSC-derived beta cells can indeed recapitulate diabetic phenotypes occurring in humans. Likewise, the stepwise analysis of human pancreatic development with this strategy would likely provide mechanistic insights into the ability of a single gene mutation (PDX1, PTF1A, HNF1B, GATA6 and GATA4) to promote pancreatic agenesis/atrophy. Further, studying mutations in KCNJ11 and ABCC8 using hiPSC-derived beta cells may elucidate the mechanistic differences between permanent and transient neonatal diabetes [64]. Overall, insulin production and secretion could be compared between diseased and gene-corrected pancreatic cells to understand the underlying cause of each type of monogenic diabetes (Figure 2).
Minireview

**Table 1 — Genetic determinants of diabetes and metabolic syndromes.**

<table>
<thead>
<tr>
<th>Genetic determinants</th>
<th>Gene(s)</th>
<th>Specific associated variant(s)/single nucleotide polymorphisms (SNPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of monogenic diabetes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MODY</td>
<td>MODY1 (HNF4A) (&gt;30 mutations)</td>
<td>rs7901695, rs7901695, rs12255372, rs10885409, rs12573128</td>
</tr>
<tr>
<td></td>
<td>MODY2 (GCK) (&gt;600 mutations)</td>
<td>Pro12Ala (P12A)</td>
</tr>
<tr>
<td></td>
<td>MODY3 (HNF1β) (&gt;200 mutations)</td>
<td>Glu23Lys (E23K)</td>
</tr>
<tr>
<td></td>
<td>MODY4 (FoxR1) (&gt;65 mutations)</td>
<td>Donohue syndrome, Rabson-Mendenhall syndrome, Type A insulin resistance (iNSR)</td>
</tr>
<tr>
<td></td>
<td>MODY5 (MEURID7), MODY7 (KLFL1), MODY8 (CE1), MODY9 (PAX4), MODY10 (INS), MODY11 (IL27), MODY14 (GATA6), MODY19 (ABCB8)</td>
<td>Congenital generalized lipodystrophy (AGPAT2, BSCL2, PTP4A, CAV1)</td>
</tr>
<tr>
<td>Neonatal diabetes</td>
<td>KCNJ1, ABCG8 and many others</td>
<td>Familial partial lipodystrophy (LMNA, PPARG, AKT2, CIDE, PLIN1)</td>
</tr>
<tr>
<td>Mitochondrial diabetes</td>
<td>BWALeu(UU09)(A32430G)</td>
<td></td>
</tr>
<tr>
<td>Syndromes of insulin resistance</td>
<td>Defects in insulin signaling pathway</td>
<td></td>
</tr>
<tr>
<td>Type 1 diabetes (T1D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Candidate gene(s)</td>
<td>HLA DR-DQ, HLA-B, HLA-A, INS</td>
<td>LEP, LEPR, POMC, PCSK1, MC4R, SIM1, BDNF, NTRK2</td>
</tr>
<tr>
<td>IL2A/CD25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTLA4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFIH1/MDA5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLEC16A/MAA0350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERBB3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTPN2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UBASH3A/STS2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL1B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RANTES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CITA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP27B1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGR3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMO7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLK1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC11A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH2B3, IL27, BACH2, PRKCG, IL10, CD69, CYP27B1, CTSH, TNFAIP3, STAT4, IL18RAP, TAGAP, RNLS, IL2, C1QTNF6, CRIB1, CTB2R, ERAP1/ART51, FHOD3, GLS3, Herc2, HTRA1, IL19, IL20, ITPK3, PIP4K2D, RASGRF1, RNF100, RS61, SRPR, SKAP2, STAT3, TNFA, TMR9, Xp22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 2 diabetes (T2D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Candidate gene(s)</td>
<td>TCF7L2 (strongest)</td>
<td>Specific associated variant(s)/single nucleotide polymorphisms (SNPs)</td>
</tr>
<tr>
<td>PPARG</td>
<td></td>
<td>rs7901346, rs7901695, rs12255372, rs10885409, rs12573128</td>
</tr>
<tr>
<td>KCNJ11</td>
<td></td>
<td>Pro12Ala (P12A)</td>
</tr>
<tr>
<td>HHEX/IOE</td>
<td></td>
<td>Glu23Lys (E23K)</td>
</tr>
<tr>
<td>SLC30A8</td>
<td></td>
<td>rs1111875, rs7923837, rs8015480, rs7923837</td>
</tr>
<tr>
<td>CDKAL1</td>
<td></td>
<td>Arg325Trp (rs13266634, rs3802177)</td>
</tr>
<tr>
<td>IGF2BP2</td>
<td></td>
<td>rs756992, rs754840, rs10946398, rs965871, rs7412523, rs7412524, rs6931514</td>
</tr>
<tr>
<td>INS</td>
<td></td>
<td>rs4002960, rs1470579, rs679511</td>
</tr>
<tr>
<td>ABC28</td>
<td></td>
<td>rs10830963, rs1367153, rs2477716, rs1374955, rs2166706, rs10930963</td>
</tr>
<tr>
<td>MTRA1/ADCS5</td>
<td></td>
<td>rs2237892, rs2237895, rs2237897, rs213362, rs2283228, rs163162, rs2299620</td>
</tr>
<tr>
<td>KCNQ1</td>
<td></td>
<td>P446L, rs780094, rs1260326</td>
</tr>
<tr>
<td>GCK</td>
<td></td>
<td>rs1799884, -300/A polymorphism in promoter, rs4607517</td>
</tr>
<tr>
<td>HNF1A</td>
<td></td>
<td>G319S, rs7951797</td>
</tr>
<tr>
<td>ADMT1S9</td>
<td></td>
<td>rs4607103</td>
</tr>
<tr>
<td>GBP2</td>
<td></td>
<td>rs600887, rs552976</td>
</tr>
<tr>
<td>JAZF1</td>
<td></td>
<td>rs684745</td>
</tr>
<tr>
<td>CDC123/CAMK1D</td>
<td></td>
<td>rs12779790, rs10906115</td>
</tr>
<tr>
<td>IRS1</td>
<td></td>
<td>rs2943841</td>
</tr>
<tr>
<td>WFS1</td>
<td></td>
<td>rs4689388, rs1801214</td>
</tr>
<tr>
<td>DUSP9</td>
<td></td>
<td>rs945326</td>
</tr>
<tr>
<td>KLF4</td>
<td></td>
<td>rs172283, rs4731702</td>
</tr>
<tr>
<td>C2CD4A/C2CD4B</td>
<td></td>
<td>rs1436955, rs1370176, rs1436953, rs712432</td>
</tr>
<tr>
<td>PTPRD</td>
<td></td>
<td>rs17584499</td>
</tr>
<tr>
<td>PROX1</td>
<td></td>
<td>rs340874</td>
</tr>
<tr>
<td>BCL11A</td>
<td></td>
<td>rs10490072</td>
</tr>
<tr>
<td>RBMS1</td>
<td></td>
<td>rs7593730</td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Obesity/Metabolic syndrome</th>
<th>Candidate gene(s)</th>
<th>Specific associated variant(s)/single nucleotide polymorphisms (SNPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THADA</td>
<td></td>
<td>Thr1187Ala/rs757897</td>
</tr>
<tr>
<td>ARAP1 (former CENTD2)</td>
<td></td>
<td>rs1552224</td>
</tr>
<tr>
<td>IGFBP2</td>
<td></td>
<td>rs1470579</td>
</tr>
<tr>
<td>PAX4</td>
<td></td>
<td>rs10229583</td>
</tr>
<tr>
<td>PAX6</td>
<td></td>
<td>rs685428</td>
</tr>
<tr>
<td>VEGFA</td>
<td></td>
<td>rs9472138</td>
</tr>
<tr>
<td>VEGF</td>
<td></td>
<td>rs6921438</td>
</tr>
<tr>
<td>HNF1B, HNF4A, DGKB/TMEM195, HMG2A, MADD, AP3S2, CHCHD9, GRB14, HMO20A, LARP8, SSSM2, ST6GAL1, TP53N1P1, VPS26A, ZBED3, UBE2E2, ZFAN6, TSPAN9/SLR5, NOTCH2/ADAMS30, ZFAND3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.3.2. Syndromes of insulin resistance

Syndromes due to defects in the insulin signaling pathway as well as lipoatrophic diabetes can be modeled using hiPSCs and genome editing technologies. hiPSCs from patients with defective insulin signaling molecules such as INSR [65], v-akt murine thymoma viral oncogene homolog 2 (AKT2) and TBC1 domain family member 4 (TBC1D4) could be differentiated into skeletal myocytes, adipocytes, hepatic and pancreatic cell types to study abnormalities in development [66] (Figure 2) and to distinguish cell autonomous versus non-cell autonomous effects. Outstanding questions that could be addressed include: 1) studying mechanisms that underlie hyperglycemia and hyperinsulinemia in individuals with INSR mutations, 2) dissecting the mechanistic differences between autosomal dominant and recessive gene mutations that cause insulin resistance of varying degrees, 3) interrogating the role of AKT2 in beta cell function, and 4) exploring why INSR mutations do not result in hepatosteatosis and dyslipidemia in contrast to other forms of severe insulin resistance. hiPSCs from patients with lipoatrophic diabetes could be differentiated into adipocytes to study developmental defects (Figure 2). Monogenic forms of insulin resistance that result into lipoatrophic diabetes can be subclassified into congenital generalized lipodystrophy (CGL) and familial partial lipodystrophy (FPL) [67] (Table 1). In the former, gene mutations typically affect various aspects of adipocyte biology such as the biosynthesis of triglycerides [68] or adipocyte differentiation [69], and hiPSCs from these patients could be a source for investigating the role of the mutated genes in adipocyte biology, lipoatrophic diabetes and insulin resistance. Considering the significant genotypic and phenotypic heterogeneity in FPL, hiPSCs with LMNA and PPARG mutations can be used to study dysfunction in fat metabolism whereas cells with CIDEC and PLIN1 mutations can be used to study triglyceride mobilization and lipid droplet dynamics.

4.3.3. Monogenic forms of obesity

Several gene mutations involved in these syndromes affect satiety and body weight [70], making the brain-fat inter-organ cross-talk challenging to model. However, one could consider examining the effects on adipocyte development using hiPSCs.

4.4. Employing hPSCs and genome editing tools to study type 1 diabetes (T1D)

Patients with T1D are unable to secrete insulin due to near complete destruction of their pancreatic beta cells. More than 50 risk variants/susceptibility alleles have been found to be associated with susceptibility to this disease [71] (https://www.niddkrepository.org/studies/t1dgc) (Table 1). The strongest association is with the human leukocyte antigens (HLAs), which accounts for a large proportion of the genetic risk for T1D [71]. Most of the T1D genes affect adaptive and innate immunity leading to incomplete self-tolerance to beta cell antigens and immune-mediated destruction of beta cells [71]. T1D-hiPSCs can be differentiated into T lymphocytes [72–74] and pancreatic beta cells [75–77] to allow co-culture experiments aimed at progressively evaluating their interactions in vitro (Figure 2) [78]. A similar strategy can be applied to hiPSCs derived from T1D-susceptible patients to examine the impact of susceptible gene variants (Table 1) on the vulnerability of pancreatic beta cells to immune attack. For instance, hiPSCs derived from patients with a gene variant in PTPN22 can be differentiated into lymphocytes to study lymphocyte function [79–81]. hiPSCs from subjects with gene variants in ERBB3, which is expressed in monocytes and dendritic cells, and may affect antigen presenting cell (APC) function [82], can be differentiated into selective immune cells to study how they affect APC function. hiPSCs from patients with gene variants in UBAH3A (also known as STS2), which is specifically expressed in lymphocytes [83], are well suited for differentiation into lymphocytes to study the function of this gene.

4.5. Employing hPSCs and genome editing tools to study type 2 diabetes (T2D) and metabolic syndromes

T2D results from a combination of genetic and environmental effects. T2D patients, characterized by early insulin resistance followed by a failure of pancreatic beta cells to compensate, develop uncontrolled hyperglycemia. The association of obesity with insulin resistance suggests that excess adipose tissue could contribute to the insulin resistance [84]. The poor ability of skeletal myocytes and adipocytes to utilize glucose and fatty acids, and to store them as glycogen and triglycerides (respectively) in states of insulin resistance lead to hyperglycemia and dyslipidemia. Insulin resistant hepatocytes respond poorly to insulin, fail to synthesize and store glycogen, and, with rising glucagon levels, continue to produce glucose leading to hyperglycemia. Patients with T2D exhibit defects in glucose-stimulated insulin secretion and a gradual decrease in pancreatic beta cell mass that, combined, fail to sustain the increased demands of insulin resistance. Thus, T2D-hiPSCs can be differentiated into skeletal myocytes, adipocytes and hepatocytes to study insulin resistance (independent of organ cross-talk) in each of these tissues, and into pancreatic beta cells to study mechanisms underlying cell death, de-differentiation and/or altered secretory function in T2D [85,86] (Figure 2). To date, at least 100 risk variants/susceptibility alleles have been associated with T2D and/or quantitative glycemic traits (fasting glucose, fasting insulin, HOMA-B and HOMA-IR) [87] (Table 1). Although the cumulative effects of these variants appear to explain only 5–10 % of the familial aggregation of T2D [88], understanding the mechanisms by which they increase T2D risk has important implications for the development of new therapies. The use of hPSCs and advanced genome editing tools can potentially facilitate this goal. For example, the transcription factor-7 like 2 (TCF7L2) gene, a transcription factor reported to be important for beta cell survival and function [89], contains variants that exhibit the strongest association with T2D [90]. Gene variants in TCF7L2 influence the therapeutic response to sulfonylureas [91] and have been demonstrated to modulate other aspects of pancreatic islet function [92]. The poorly understood mechanisms underlying these effects potentially could be investigated by using hPSCs and genome editing. For instance, one can leverage upon hiPSCs derived from T2D-susceptible patients based on their TCF7L2 genotype, and differentiate them into pancreatic beta cells to answer whether TCF7L2 gene variants affect development and provide clues to the initial beta cell mass at adulthood. Related to the two cases of hiPSC-based modeling of MODY2 [7] and Wolfram Syndrome [9], gene variants found in GCK, GCKR and WFS1 have been linked to T2D risk. Thus, it may be possible to extrapolate these findings and begin investigating the functional effects of associated gene variants. A similar approach could be adapted to investigate the role of other variants (Table 1) with appropriate readouts. For instance, hiPSCs with a defined gene variant and their gene-corrected isogenic control hiPSCs could be differentiated simultaneously and then subjected to high throughput analyses to determine developmental or expression profile differences. This is expected to reveal gene targets affected by the presence of the specific gene variant and could thereby shed light on mechanistic insights into the predisposition towards T2D. One caveat, though, would be the specific cell type of action of the identified genome-wide association study (GWAS) gene. For example, gene variants in the HHEX gene could possibly affect pancreatic beta
cell, delta cell or even hepatocyte biology, thereby predisposing individuals toward developing T2D.

More than 30 risk variants/susceptibility alleles have been associated with waist circumference, waist-hip ratio, BMI and obesity (Table 1), and are estimated to account for 2% of the occurrence of obesity [87]. The fat mass and obesity-associated (FTO) locus exhibits a strong association with BMI and obesity [89]. Other variants including those found near melanocortin 4 receptor (MC4R) which can explain 2–3% of cases of severe obesity [94] and proopiomelanocortin (POMC) [95], are strong biological candidates for BMI and obesity. Compared to gene variants that affect saliety and body weight [96], those affecting adipocyte development and function are relatively easier to model because of a definitive readout in the latter. For instance, FTO is a potential candidate for studies in hiPSC-derived adipocytes since its expression has been reported to decrease during adipocyte differentiation [97]. In common obesity, circulating leptin (LEP) levels are high, indicating leptin resistance [98,99]. Differentiating hiPSCs derived from those with risk variants in the LEP gene into adipocytes can aid in increased understanding of leptin regulation of adipocyte function. Genes such as TMEM18, KCCTD15, MTC12, BCDIN3, SULT1A1, PTER, TUFM, FAIM2, BDNF, NEGR1 and NPC1 with potential regulatory roles in the adipose tissue [100,101] could be studied by independently differentiating hiPSCs into adipocytes to study the specific gene function(s). In addition, TMEM18, PTER, SULT1A1 and TUFM may be studied in hepatocytes whereas SH2B1 and TUFM could be studied in skeletal myocytes [102]. However, it should be cautioned that genes currently associated with the risk-conferring polymorphisms may not actually be the only genes involved in the polymorphism, as in the case of FTO and I RX3 [103].

5. CONCLUSIONS

hiPSCs and the advancing genome editing tools appear to be a timely and potent combination for probing molecular mechanism(s) underlying diseases such as diabetes and metabolic syndromes. Studying monogenic forms of diabetes and syndromes of insulin resistance using these tools would be extremely useful given the lack of an autoimmune attack and confounding effects of insulin resistance and obesity. One caveat of this methodology at the moment is the “low” efficiency of deriving human beta cells in vitro [75,76], possibly due to our incomplete knowledge on human pancreatic development. Another explanation would be the lack of in vivo environmental cues emanating from proximal tissues such as the vasculature. Nonetheless, successful disease modeling of MODY2 [7] and Wolfram Syndrome [8] already suggests a high possibility of success. These technologies have the potential to elucidate the underlying pathophysiology that stem from defects in 1) beta cell development, metabolism or survival or 2) development of adipocyte. For instance in the case of MODY2, it is now clear that GCK mutation affects glucose-stimulated insulin secretion but not insulin synthesis or beta cell proliferation [7]. With the latest advances in the derivation of mature and functional human pancreatic beta-like cells from hiPSCs in vitro [75–77], eventually circumventing the requirement for in vivo maturation, disease modeling of diabetes is expected to progress exponentially. The knowledge gained from these hiPSC-based disease modeling studies can potentially be translated into the clinics by guiding clinicians on the appropriate type of medication to use for each condition based on the mechanism of action of the disease.

Findings from these proposed studies could also offer clues to the pathophysiology of the “garden variety” of type 2 diabetes which is known to manifest defects in each of these tissues. hiPSCs and genome editing tools may also provide an opportunity to better understand the relevance of gene variants identified from GWAS studies, in causing T1D, T2D, obesity and metabolic syndromes, given that they exhibit only modest effects and ~85% of the variants map onto non-coding regions such as enhancers or regulatory elements [104]. Investment into hiPSCs and genome editing would allow a better mechanistic understanding of the pathophysiology of monogenic and complex diseases relevant for organismal homeostasis and therefore an improved approach to stratified personalized medicine. By identifying the impact of gene variants on disease predisposition, prophylactic measures in the form of lifestyle alterations or medication could be adopted early on in life to delay or even prevent the onset of diabetes and/or metabolic diseases. It is also likely that these hiPSC-based disease modeling studies would provide insights into approaches to predict the susceptibility of disease. Henceforth, the translational potential of studying human diabetes and metabolic syndrome disease mechanisms is huge, with opportunities for early prophylactic intervention that could have long-term implications for global health care and reduction of economic burden. While the derivation of hiPSCs from human tissues is relatively easier and gaining popularity compared to just a few years ago [2], it is likely that the modern technology of generating site-specific nucleases will also rapidly mature to make in vitro disease modeling a routine approach.

CONTRIBUTION STATEMENTS

A.K.K.T. conceived the idea and wrote the manuscript. M.K.G. contributed to the manuscript. A.D. edited and commented on the manuscript. R.N.K. edited and approved the manuscript.

ACKNOWLEDGMENTS

A.K.K.T. was supported by a Juvenile Diabetes Research Foundation Postdoctoral Fellowship (3-2013-236) and is currently supported by the Institute of Molecular and Cell Biology (IMCB), A*STAR, NHG-KTPH SIG/14033 and NHUSR2014/033. A.D. is supported by the National Institutes of Health grant R01 DK 55523. R.N.K. is supported by the Harvard Stem Cell Institute, the National Institutes of Health grants R01 DK 67536, DK 55523 and DK103215, and a grant from AstraZeneca. The funding sources had no involvement in the study design, writing of report or submission of article for publication.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Minireview


Minireview


