ARTICLES
https://doi.org/10.1038/s41591-021-01502-7

nature medicine

Check for updates

Mutations and variants of *ONECUT1* in diabetes

Anne Philippi[1,17], Sandra Heller[2,17], Ivan G. Costa[3,17], Valérie Senée[1,17], Markus Breunig[2], Zhijian Li[3], Gino Kwon[2], Ronan Russell[4], Anett Illing[2], Qiong Lin[3], Meike Hohwieler[2], Anne Degavre[1], Pierre Zalloua[5,6], Stefan Liebau[7], Michael Schuster[8], Johannes Krumm[9], Xi Zhang[2], Ryan Geusz[10], Jacqueline R. Benthuysen[10], Allen Wang[10], Joshua Chiou[10], Kyle Gaulton[10], Heike Neubauer[11], Eric Simon[12], Thomas Klein[11], Martin Wagner[2], Gopika Nair[4], Céline Besse[13], Claire Dandine-Roulland[13], Robert Olaso[13], Jean-François Deleuze[13], Bernhard Kuster[9,14], Matthias Hebrok[4], Thomas Seufferlein[2], Maike Sander[10], Bernhard O. Boehm[15], Franz Oswald[2], Marc Nicolino[16,18], Cécile Julier[1,18] ✉ and Alexander Kleger[2,18] ✉

Genes involved in distinct diabetes types suggest shared disease mechanisms. Here we show that One Cut Homeobox 1 (*ONECUT1*) mutations cause monogenic recessive syndromic diabetes in two unrelated patients, characterized by intrauterine growth retardation, pancreas hypoplasia and gallbladder agenesis/hypoplasia, and early-onset diabetes in heterozygous relatives. Heterozygous carriers of rare coding variants of *ONECUT1* define a distinctive subgroup of diabetic patients with early-onset, nonautoimmune diabetes, who respond well to diabetes treatment. In addition, common regulatory *ONECUT1* variants are associated with multifactorial type 2 diabetes. Directed differentiation of human pluripotent stem cells revealed that loss of ONECUT1 impairs pancreatic progenitor formation and a subsequent endocrine program. Loss of ONECUT1 altered transcription factor binding and enhancer activity and NKX2.2/NKX6.1 expression in pancreatic progenitor cells. Collectively, we demonstrate that ONECUT1 controls a transcriptional and epigenetic machinery regulating endocrine development, involved in a spectrum of diabetes, encompassing monogenic (recessive and dominant) as well as multifactorial inheritance. Our findings highlight the broad contribution of ONECUT1 in diabetes pathogenesis, marking an important step toward precision diabetes medicine.

Diabetes affects over 350 million people worldwide[1], and type 2 diabetes (T2D) is the most common form and has a mainly multifactorial etiology. Monogenic diabetes accounts for 1–5% of cases, with a higher prevalence in early-onset patients[2,3]. The discovery and study of genes responsible for monogenic diabetes provide important insights for understanding disease mechanisms, which can enable cost-effective care and improved quality of life[4]. While major progress has been achieved to identify these genes, the rarity and clinical heterogeneity of monogenic cases makes the identification of novel causative genes difficult. This is particularly challenging for cases who are not clinically atypical and are thus generally diagnosed as T2D. In the past decade, large-scale genome-wide association studies (GWASs) have identified many common variants associated with T2D (ref. [5]). The identification of monogenic contribution to T2D using whole exome sequencing (WES) and whole genome sequencing has been limited to date, despite the increasingly large scale of such studies[5,6]. Complementary strategies are therefore needed to increase the power of these studies. Several genes are shared between monogenic diabetes and multifactorial T2D, suggesting that shared disease mechanisms exist. Remarkably, many of these genes encode key proteins for pancreas development (for example, *HNF1A*, *HNF1B*, *HNF4A* and *GLIS3*). Human pluripotent stem cells (PSCs) represent a powerful tool to simulate pancreatic development and facilitate disease modeling[7–9].

Here starting with the study of a consanguineous family presenting cases of neonatal syndromic diabetes and T2D, we used a staged approach combining genetic and in-depth functional studies and identified *ONECUT1*/hepatocyte nuclear factor 6 (*HNF6*) as a gene involved in monogenic recessive and monogenic dominant, as well as multifactorial, diabetes. Using genome-edited human embryonic

[1]Université de Paris, Institut Cochin, INSERM U1016, CNRS UMR-8104, Paris, France. [2]Department of Internal Medicine I, Ulm University, Ulm, Germany. [3]Institute for Computational Genomics, RWTH Aachen University Medical School, Aachen, Germany. [4]Diabetes Center at the University of California, San Francisco, CA, USA. [5]School of Medicine, University of Balamand, Amioun, Lebanon. [6]College of Medicine and Health Sciences, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates. [7]Institute of Neuroanatomy & Developmental Biology, Eberhard Karls University Tuebingen, Tuebingen, Germany. [8]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. [9]Chair of Proteomics and Bioanalytics, Technical University of Munich (TUM), Freising, Germany. [10]Pediatric Diabetes Research Center (PDRC) at the University of California, San Diego, CA, USA. [11]CardioMetabolic Diseases Research, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach, Germany. [12]Global Computational Biology and Digital Sciences, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach, Germany. [13]Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, Commissariat à l'Energie Atomique, Université Paris-Saclay, Evry, France. [14]Bavarian Biomolecular Mass Spectrometry Center (BayBioMS), Technical University of Munich (TUM), Freising, Germany. [15]LKC School of Medicine, Nanyang Technological University, Singapore, Singapore. [16]Service d'Endocrinologie et Diabétologie Pédiatriques et Centre PRISIS, Pathologies Rares de l'Insulino-Sécrétion et de l'Insulino-Sensibilité, Hôpital Femme-Mère-Enfant, Hospices Civils de Lyon, Bron, France. [17]These authors contributed equally: Anne Philippi, Sandra Heller, Ivan G. Costa, Valérie Senée. [18]These authors contributed equally: Marc Nicolino, Cécile Julier, Alexander Kleger. ✉e-mail: cecile.julier@inserm.fr; alexander.kleger@uni-ulm.de

stem cells (ESCs) and patient-specific induced pluripotent stem cells (iPSCs), we dissected the functional consequences of defective ONECUT1 protein in pancreatic development.

Results

A patient with severe neonatal syndromic diabetes. We studied a French boy (Patient 1) born to consanguineous parents, affected by severe neonatal syndromic diabetes following intrauterine growth retardation (IUGR), with pancreatic, hepatic, neurologic and hematologic manifestations (Table 1 and Extended Data Fig. 1a–d (ref. [10])). IUGR was diagnosed at 33 weeks of pregnancy, with hydramnios and fetal abnormalities. Delivery by cesarean section occurred at 37 weeks of gestation. Weight and height at birth were <1st percentile. Diabetes was diagnosed at the first day of life (day 1), with blood glucose at 17 mmol l^{-1}, then above 25 mmol l^{-1} with glycosuria from day 15. Plasma insulin and C-peptide measured at day 18 were undetectable. Insulin treatment was started at day 21, with high doses increased up to 2.3 units per kg per day. Conversely, serum glucagon was elevated (638 ng l^{-1}). Exocrine pancreatic insufficiency was documented by very low fecal chymotrypsin and elastase levels. Imaging showed severe pancreatic hypotrophy and lack of gallbladder. The patient also suffered from poorly regenerative anemia and required blood transfusions from day 1. Signs of cholestasis were also noted, with elevated total plasma bilirubin, and hepatocellular insufficiency with low levels of various plasma components produced by the liver (Table 1). He also presented with facial dysmorphism with microretrognathia and morphological abnormalities of the extremities. Ultrasound imaging of the heart, kidneys and brain showed no anomaly. The clinical course was poor, with no weight gain despite tube feeding. He also showed diffuse hypotonia, limited mobility and reactivity, edema of lower limbs and moderate jaundice with hepatomegaly, as well as neuromuscular respiratory distress (most likely related to central nervous system impairment). Because of the very severe neurological condition, brain magnetic resonance imaging was performed at day 59, but no abnormalities were observed. The patient died at 60 d postpartum.

Patient 1's mother (Individual 2) had a total of seven pregnancies, leading to the birth of three healthy children. Patient 1 was born from the sixth pregnancy. Two pregnancies were terminated by spontaneous miscarriages and one was complicated by hydatidiform mole. During her sixth pregnancy, at age 34 yr, she had gestational diabetes, requiring insulin therapy from the 29th week of gestation. She also had gestational diabetes during her last pregnancy. Additional metabolic explorations, including oral glucose tolerance test (OGTT), were performed in the parents of Patient 1 (Individuals 1 and 2; Supplementary Table 1). Clinical examination of the mother (at age 40 yr) was unremarkable apart from overweight (body mass index (BMI) of 28.1 kg m^{-2}). She did not follow any special treatment. However, metabolic exploration revealed impaired fasting glucose (IFG) and high HbA1c level (6.6%), and OGTT demonstrated abnormal glucose tolerance and overt diabetes, while her plasma insulin level remained low (Supplementary Table 1). Her 30-min incremental insulin to glucose level during OGTT (insulinogenic index) was very low, suggesting impaired β-cell function (Supplementary Table 1). Diet and exercise resulted in considerable weight loss (BMI = 26 at age 41) and almost normal fasting glucose level (5.8 mmol l^{-1}). However, her diabetes persisted (HbA1c = 6.2%) and metformin treatment was initiated. By contrast, the father did not have diabetes at 38 yr, albeit his fasting insulin level was just below normal (34.7 pmol l^{-1}; Supplementary Table 1). In both parents, glutamic acid decarboxylase (GAD), islet antigen 2 (IA2) and human insulin autoantibodies were negative. Based on normal levels for serum lipase; vitamins A, D, E and K; IGF1; bilirubin; lipids; and total proteins, there was no evidence for exocrine pancreas or liver dysfunction. Accordingly, ultrasonography of the pancreas, liver, gallbladder and biliary ducts was normal. Overall, these observations suggested impaired glucose metabolism in Patient 1's parents.

Homozygous *ONECUT1* mutations cause neonatal syndromic diabetes. The familial context suggested a rare autosomal recessive inheritance. To identify the disease-causing gene, we performed linkage analysis combined with candidate genes selection (Fig. 1a). Nine homozygous regions (3.90% autosomes length) were linkage-compatible under a fully penetrant recessive model. Based on the extreme clinical presentation of Patient 1, affecting exocrine and endocrine pancreas as well as gallbladder development, we selected genes specifically involved in early endoderm development as candidates (eight genes: *ONECUT1*, *HNF1B*, *FOXA2*, *HNF4A*, *HHEX*, *GATA4*, *GATA6* and *RFX6*)[11–13] (Fig. 1a). The intersection of linkage regions and these candidates identified a single gene, *ONECUT1*/*HNF6* (Fig. 1a), whose knockout (KO) mice recapitulate the patient's phenotype[14–16]. We sequenced *ONECUT1* exons in all family members and identified a protein-truncating variant (PTV), *ONECUT1*-p.E231X (chr15.hg19:g.53081391C>A), homozygous in the patient, heterozygous in the parents and heterozygous or homozygous wildtype (WT) in healthy siblings (Fig. 1b and Extended Data Fig. 1d). The resulting protein lacks the CUT and HOMEOBOX domains, which are responsible for DNA binding to target genes[17]. Subsequent WES of this patient confirmed *ONECUT1* as the only gene compatible with his rare recessive syndrome.

Independently, in a second patient (Patient 2, Turkish), born to consanguineous parents, diagnosed at 14 months with insulin-requiring diabetes, exocrine pancreatic insufficiency and growth retardation, we identified a homozygous missense variant, *ONECUT1*-p.E231D (chr15.hg19:g.53081389C>G), through targeted candidate genes sequencing (Fig. 1b and Extended Data Fig. 1e). The patient had IUGR associated with neonatal hypotrophy and postnatal failure to thrive, as well as mild anemia. Imaging revealed hypotrophic pancreas and gallbladder, supporting a similar but less severe phenotype than Patient 1 (Table 1). Both *ONECUT1* variants were absent from available public databases (Supplementary Table 4a). Hence, biallelic *ONECUT1* mutations cause a novel syndrome characterized by neonatal and/or very early-onset insulin-requiring diabetes with exocrine pancreas insufficiency and other manifestations. Patient 1's mother and Patient 2's mother, and another *ONECUT1*-p.E231X heterozygous woman with diabetes married to a relative (Family 1, Individual 18), had repeated miscarriages and/or neonatal child mortality, suggesting that homozygous *ONECUT1* mutations are generally lethal or result in early mortality (Fig. 1c,d).

Adult-onset diabetes in *ONECUT1*-p.E231X and p.E231D heterozygotes. Family history of T2D reported in both families suggested that heterozygous carriers for these mutations might be predisposed to adult-onset diabetes. To investigate this hypothesis, we extended the clinical and genetic study of the family members of Patient 1 (Family 1; Fig. 1c). Five of the seven *ONECUT1*-p.E231X carriers aged 30 to 76 had diabetes or IFG. We modeled the transmission of diabetes (neonatal diabetes, and diabetes or IFG (diabetes/IFG hereafter)) with respect to *ONECUT1*-p.E231X in this family, confirming complete penetrance in homozygotes (neonatal diabetes) and incomplete penetrance estimated to 0.63 in heterozygotes (diabetes/IFG) ($P = 0.003$), and showing cosegregation of diabetes with *ONECUT1*-p.E231X under this model (logarithm of odds (LOD) score = 2.35, $P = 0.0005$). There was evidence of increased risk of diabetes/IFG in *ONECUT1*-p.E231X carriers compared with noncarriers when adjusting for the age at examination (logistic regression, one-sided $P = 6 \times 10^{-5}$). Accordingly, the prevalence of diabetes/IFG in *ONECUT1*-p.E231X carriers was increased compared with the French general population[18] ($P = 0.0049$). Similarly, Patient 2's parents, who were heterozygous for *ONECUT1*-p.E231D, had IFG or impaired glucose tolerance (IGT), while the nongenotyped

Table 1 | Clinical, metabolic, biochemical and radiological features of Patient 1 (Family 1) and Patient 2 (Family 2)

Characteristics	Patient 1 (Family 1)	Patient 2 (Family 2)	Reference values
Sex	Male	Male	
Ethnicity	French[a]	Turkish	
Consanguinity	Yes	Yes	
Age at diabetes onset	1 d	14 months	
Anthropometric measurements at birth			
Gestational age (weeks)	37	38	
Weight, g (percentile)	1,540 (<1)	2,660 (10)	
Length, cm (percentile)	42 (<1)	52 (93)	
Head circumference, cm (percentile)	31 (2.6)	32 (5)	
Growth	IUGR during pregnancy, 33rd week, with hydramnios and fetal abnormalities (dilated bowel loops, closed fists); small for gestational age without catch-up growth by 60 d (death)	IUGR during pregnancy, 33rd week; neonatal hypotrophy; postnatal failure to thrive	
Glucose metabolism and diabetes			
Autoantibodies	ND	Negative (ICA, GAD, IA2)	
Fasting glucose, mmol l^{-1}	13.3[b]	ND	<5.6
Fasting insulin, mIU l^{-1}	Undetectable	Undetectable	2–20
Plasma C-peptide, fasting, pmol l^{-1}	Undetectable	Undetectable	0.25–1.70
Plasma glucagon, fasting, ng l^{-1}	638	ND	50–250
Random glucose, mmol l^{-1}	17.5–28[b]	15.9[b]	<11.1
Diabetes treatment	Insulin; very unstable	Insulin; very unstable	
Exocrine pancreas			
Fecal chymotrypsin (stools), U g^{-1}	0.1	ND	>8
Fecal elastase (stools), mg g^{-1}	23	ND	>100
Treatment	Tube feeding	Initially untreated, then oral pancreatic enzymes	
Hematology			
Hemoglobin, g l^{-1}	64[c]	106	120–160
Hepatology			
Plasma bilirubin, µmol l^{-1}	275	ND	<200
HDL cholesterol, mmol l^{-1}	0.52	ND	1.05–1.80
Esterified cholesterol, mmol l^{-1}	0.19	ND	2.8–4.5
Apolipoprotein A1, g l^{-1}	0.22	ND	1.20–1.80
Total proteins, g l^{-1}	35	ND	51–73
Albumin, g l^{-1}	24	ND	27–41
Coagulation factor II, %	20	ND	45–105
Coagulation factor VII, %	18	ND	48–132
Fibrinogen, g l^{-1}	0.5	ND	1.5–3.8
Additional clinical and morphological abnormalities			
Morphological	Microretrognathism; contracture of the fingers and toes, lack of extension of the elbows; prominent heel bone and a convexly rounded sole	NR	
Neurological	Flexion contracture of fingers, toes and elbows; axial hypotonia; limited spontaneous mobility and reactivity; respiratory distress	NR	
Abdominal imaging	Very hypoplastic pancreas (head measuring 2 × 6 mm^2, body and tail not visualized); hepatomegaly; absence of gallbladder	Very hypoplastic pancreas; gallbladder present but small	
Other manifestations	Edema of lower limbs; hydrocoele; moderate jaundice	NR	

Notes: Biochemical explorations were performed between birth and age 45 d (Patient 1) and at diabetes onset (14 months, Patient 2). [a]Family 1 is from a French traveler community, a minority group with a strong tradition of consanguineous marriages. Ancestry analysis confirmed their mainly French European origin (Extended Data Fig. 1a,b and Methods). [b]Diabetes range [c]Hemoglobin value at day 1; Patient 1 was treated by blood transfusions from day 1, then by recombinant human erythropoietin from day 7. Following this treatment, hemoglobin increased to levels between 100 and 110 g l^{-1}. ND, not done; NR, not reported.

grandfather had diabetes (Fig. 1d). Similar to Patient 1's mother and Individual 18 (Family 1), Patient 2's mother had gestational diabetes, with an unusually early start at week 14 of pregnancy, and which required insulin therapy from week 21 of pregnancy. Detailed explorations of six heterozygous *ONECUT1*-p.E231X and *ONECUT1*-pE231D individuals including OGTT in the parents support that they have altered glucose metabolism, resulting in IFG, impaired glucose tolerance or diabetes (Supplementary Table 1).

Rare *ONECUT1* missense variants are associated with diabetes at the population level. To investigate the contribution of *ONECUT1* coding variants to diabetes, we sequenced the coding region of *ONECUT1* in an Ulm (Germany) Diabetes Cohort (UDC; Methods), including 2,165 patients with nonautoimmune diabetes (UDC-T2D thereafter, with age-extended T2D, including a large proportion of patients with early-onset diabetes (age at diagnosis: 10–85 yr, 25% diagnosed before age 35, GAD autoantibodies negative; Supplementary Table 2)), 397 healthy controls and 162 patients with type 1 diabetes (T1D) or latent autoimmune diabetes in adults (LADA). Patients carrying known maturity-onset diabetes of the young (MODY) gene mutations were previously excluded from this cohort (Methods). We identified 13 patients with T2D who were heterozygous for rare *ONECUT1* missense variants (minor allele frequency (MAF) <0.005 in the representative Genome Aggregation Database (gnomAD) North-West Europe (gnomAD-NWE) population) and none in the healthy controls or patients with T1D or LADA (Fig. 1e and Supplementary Tables 3a and 4b; one-sided Fisher exact test $P=0.05$ comparing T2D patients with individuals without T2D). In contrast, rare synonymous variants were equally frequent in patients with and without T2D, as expected (Supplementary Tables 3a and 4c). We also observed one low-frequency missense variant, *ONECUT1*-p.P75A (rs74805019, MAF = 0.03 in gnomAD-NWE), which was not associated with risk of T2D in the UDC (Supplementary Table 3b) or in other cohorts including DIAMANTE (Diabetes Meta-analysis of Trans-ethnic Association Studies)[5], or in previous studies[19,20]. For replication, we performed burden testing for *ONECUT1* coding variants in the AMP-T2D-GENES (Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples) cohort (19,852 T2D cases, 23,273 controls) from the Accelarating Medicines Partnership T2D Knowledge Portal. We found an overall trend for increased risk of T2D (collapsing burden test, odds ratio (OR) = 1.14, $P=0.08$), which reached significance in the European population, consistent with our findings (OR = 1.31, $P=0.002$; Supplementary Table 5). The strongest association with T2D was observed in analyses that allowed variable risks or frequency thresholds (P(sequence kernel association test; SKAT) = 0.00026; P(variable threshold test) = 0.005; Supplementary Table 5), suggesting heterogeneity in risks between variants. Association trends were observed for the less rare variant p.H33Q (MAF(cases) = 0.002, OR = 5.0, $P=0.079$), as well as for several very rare variants: p.G30S (18/5 cases/controls), p.G62C (3/0 cases/controls, absent from gnomAD) and p.V242A, an Asian-specific variant (MAF(gnomAD-East-Asian) = 0.014, OR = 1.4, $P=0.026$). In contrast, p.G96D showed a protective trend or neutral effect (AMP-T2D-GENES: MAF(cases) = 0.0006, OR = 0.34, $P=0.011$; DIAMANTE[5]: OR = 0.81, $P=0.40$; Supplementary Table 6). UDC-T2D Individual 4, a p.G96D carrier, was also heterozygous for the low-frequency p.P75A (compound heterozygous p.G96D rare/p.P75A low-frequency variants), suggesting an additive risk effect of these variants or a modifier effect of p.P75A on diabetes risk. These findings support that rare *ONECUT1* missense variants are overall associated with increased risk of T2D at the population level, although some of these variants might be neutral or even protective. Furthermore, these data suggest that a subset of missense variants, predicted to be the most deleterious, could be associated with higher risk of diabetes, similar to PTVs.

Heterozygotes for rare *ONECUT1* coding variants define a distinctive subgroup of patients with early-onset diabetes. The 13 *ONECUT1* heterozygous UDC-T2D patients had an earlier age at diagnosis than noncarriers (median (IQR) = 29 (25–37) versus 46 (36–55); $P=0.00033$; Fig. 1f), responded well to the initial diabetes treatment and had family histories compatible with dominant transmission (Table 2 and Fig. 1e). The age at diagnosis of relatives with T2D was similar to the probands (median (IQR) = 34 (30–38) versus 46 (36–55); $P=0.37$), and different from noncarriers ($P=0.0015$; Fig. 1f). Kaplan–Meier analysis showed that heterozygous carriers had younger age at diagnosis compared with noncarriers ($P=3.0\times10^{-7}$; Fig. 1g), with a hazard ratio for the median age at diagnosis of 3.75 (95% confidence interval: 2.17–6.48) ($P=2.3\times10^{-6}$). WES performed in these 13 patients confirmed the absence of known MODY gene mutations (11 MODY genes: *HNF1A*, *HNF1B*, *HNF4A*, *GCK*, *ABCC8*, *PDX1*, *INS*, *PAX4*, *KCNJ11*, *NEUROD1* and *RFX6*; all variants predicted to be benign, likely benign or of unknown significance). We also determined HLA-DR risk and the T1D genetic risk score (T1D-GRS) in these patients, confirming that these patients were overall similar to healthy individuals and patients with T2D, but distinct from patients with T1D. We also excluded the presence of the mitochondrial m.3243A>G mutation from these patients (Table 2). Independently, we identified a heterozygous missense variant, *ONECUT1*-p.P215R, by WES in a Lebanese boy with insulin-treated juvenile-onset nonautoimmune diabetes (onset at 12 yr) (Supplementary Table 4d). This variant was absent in the parents and in all public databases, while high-density single-nucleotide polymorphism (SNP) genotyping confirmed family relationships, supporting that it is a diabetes-causing de novo mutation.

Variants identified in our screening of patients with diabetes are shown in Extended Data Fig. 1f. In addition to early age at diabetes diagnosis (median (IQR) =29 (23.5–37), range: 12–47; Supplementary Table 7a), most *ONECUT1* heterozygous patients were normal weight to overweight at diagnosis (median (IQR) = 26.5 (24.6–29.5), range: 20.2–30.1) and normal weight to low-risk (class 1) obesity at recruitment (Supplementary Table 7b)

Fig. 1 | *ONECUT1* mutations cause severe neonatal syndromic diabetes and early-onset diabetes. a, Gene identification in Patient 1 using a combined linkage and selected candidate genes study. **b**, Schematic protein representation of *ONECUT1* mutations: truncated ONECUT1-p.E231X lacking CUT and homeobox (HOX) domains and missense ONECUT1-p.E231D. **c**, Extended pedigree of Patient 1 (arrow, Family 1), showing neonatal diabetes, T2D and IFG status. Individual 21 died at the age of 1 d from unknown causes. His mother (Individual 18) had diabetes and also had gestational diabetes and repeated miscarriages. FPG, fasting plasma glucose. **d**, Pedigree of Patient 2 (arrow, Family 2). **e**, Pedigrees of the 13 patients with diabetes (UDC-T2D cohort) identified with rare *ONECUT1* missense variants, showing *ONECUT1* genotypes and age at diabetes diagnosis. Two of the 13 index cases (*), were also heterozygous for the low-frequency *ONECUT1*-p.P75A variant. Arrows indicate the genotyped index patients. **f**, Age at diagnosis of UCD-T2D patients carrying rare *ONECUT1* missense variants (+/m, $n=13$) and their relatives (T2D relatives of +/m, $n=15$) compared with T2D noncarriers (+/+, $n=2,072$). Boxplots show the median (center), interquartile range (box) and extreme values (whiskers). P values (nonparametric Wilcoxon rank test) are shown. **g**, Kaplan–Meier survival analysis of age at onset of diabetes depending on carrier status for rare *ONECUT1* missense variants (+/m, +/+) in the UDC-T2D patients. NS, not significant.

and had low fasting insulin (6 heterozygous individuals, Families 1 and 2; Supplementary Table 1), further refining the characteristic features of these individuals.

For replication, we performed burden testing for rare *ONECUT1* coding variants in subgroups of patients from the AMP-T2D-GENES cohort selected to best reproduce the distinctive characteristics of our *ONECUT1* heterozygous patients with diabetes based on available phenotypic data: age at recruitment (as a surrogate for age at diagnosis, not available), BMI and fasting insulin. This analysis showed evidence for association with T2D, with increasing ORs

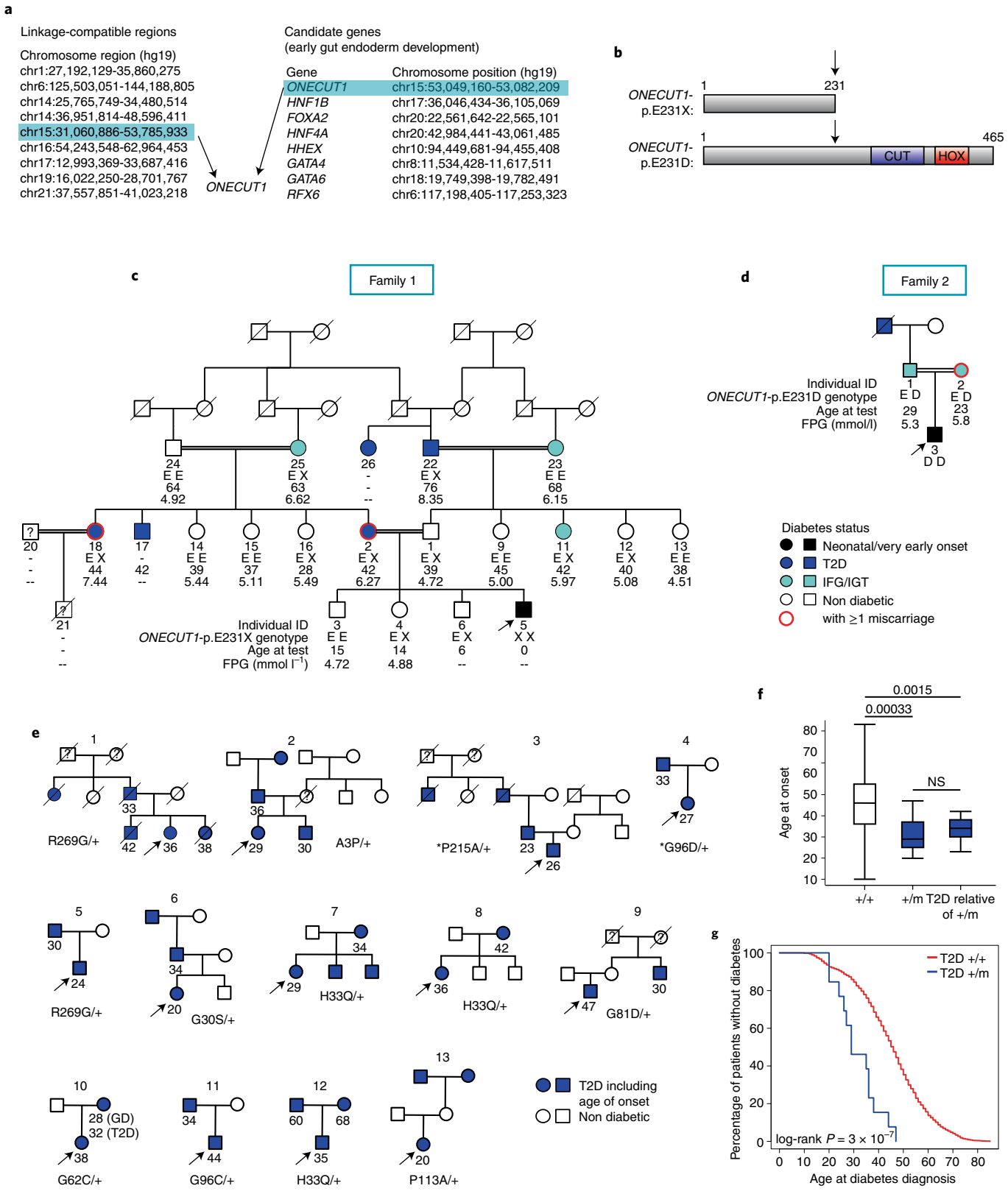


Table 2 | Clinical characteristics of patients with diabetes from the UDC-T2D cohort heterozygous for rare *ONECUT1* missense variants

Patient code	ONECUT1 variant	Sex	Age at onset	Age at recruitment	BMI at onset	BMI at recruitment	GAD antibodies at recruitment	C-peptide at onset (mg ml^{-1})	Treatment at onset	Response to treatment (initial 18 months)	HbA1c at initial follow-up	Age at onset of T2D relatives	PPV for MODY	HLA-DR genotype	10 SNPs T1D-GRS (centile of a reference T1D population)
1	R269G	Female	36	67	30.1	30.1	ND	Positive[a]	Sulfonylurea + metformin	Good; average fasting glucose: 7.8 mmol l^{-1}	NA	NA, 33, 42, 38	NA	DR3/X	0.606 (<25)
2	A3P	Female	29	30	20.5	20.5	0	ND	OAD	Excellent	6.4% after 3 yr	NA, 36, 30	62.4%	DRB1*15/X	0.383 (<25)
3	P215A[b]	Male	26	26	30.1	30.1	0	ND	Diet + exercise	Good; no postprandial glucose levels above 7.8 mmol l^{-1}, fasting 4.4–5.6 mmol l^{-1}	5.6% after 2 yr	NA, NA, 23	62.4%	DR3/ DRB1*15	0.4511 (<25)
4	G96D[c]	Female	27	27	23.3	23.3	0	ND	Diet + exercise	Excellent	6.2% after 4 yr	33	75.5%	DR3/X	0.666 (25–50)
5	R269G	Male	24	34	ND	23.1	0	ND	Diet + exercise, low dose insulin at bedtime	Excellent	6.0% after 5 yr	30	49.4%	DR3/DR3	0.741 (>50)
6	G30S	Female	20	20	29.9	29.9	0	Positive	Sulfonylurea	Excellent	6.1% after 18 months	NA, 34	75.5%	X/X	0.533 (<25)
7	H33Q	Female	29	57	ND	37.3	ND	ND	Metformin	Good; average blood glucose: 6.7 mmol l^{-1}	NA	34, NA, NA	NA	X/X	0.486 (<25)
8	H33Q	Female	36	52	ND	24.1	0	ND	Diet + sulfonylurea	Good; no postprandial glucose levels above 10.0 mmol l^{-1}, fasting 4.4–6.7 mmol l^{-1}	NA	42	NA	X/X	0.532 (<25)
9	G81D	Male	47	47	28	28	0	1.66	Sulfonylurea	Excellent	6.5% after 2 yr	30	NA	DR3/X	0.603 (<25)
10	G62C	Female	38	38	25.6	25.6	0	ND	Sulfonylurea	Excellent	6.2% after 2 yr	28	NA	X/X	0.475 (<25)
11	G96C	Male	44	44	25.6	25.6	0	ND	Metformin + insulin at bedtime	Good; no postprandial glucose levels above 10.0 mmol l^{-1}, fasting 4.4–6.7 mmol l^{-1}	NA	34	NA	X/X	0.486 (<25)
12	H33Q	Male	35	41	ND	ND	0	ND	Metformin (5 yr)	Good; no postprandial glucose levels above 10.0 mmol l^{-1}, fasting 5.0–5.6 mmol l^{-1}	NA	60, 68	NA	DR4-DQ8/X	0.664 (25–50)
13	P113A	Female	20	20	25	25	0	1.44	Insulin	Good; no postprandial glucose levels above 8.9 mmol l^{-1}, fasting 4.4–6.1 mmol l^{-1}	6.3% after 16 yr	NA, NA	4%	X/X	0.513 (<25)

[a]C-peptide for this patient was still positive when tested 20 yr after diabetes onset. [b]This patient was also heterozygous for the low-frequency *ONECUT1*-P75A variant (rare/low-frequency variants in *cis*: P215A(rare)-P75A(low-frequency)) [c]This patient was also heterozygous for the low-frequency *ONECUT1*-P75A variant (rare/low-frequency variants in *trans*: compound heterozygous G96D(rare)/P75A(low-frequency)). HLA-DR genotypes for DR3, DR4 and DR15 were determined using tag SNPs rs2187668 and rs7454108 to tag DR3 (DRB1*0301-DQA1*0501-DQB1*0201) and DR4-DQ8 (DRB1*04-DQA1*0301-DQB1*0302) alleles, respectively, and SNP rs3129889 to tag HLA DRB1*15 (ref. [53]).'X' denotes non-DR3, non-DR4-DQ8, non-DRB1*15. The T1D-GRS was determined using the genotypes of the top ten risk alleles for T1D, according to Oram et al.[54] (Methods). Values shown in parentheses indicate the distribution of these scores compared with a European T1D control population studied by Johnson et al.[55]: <25, below the 25th centile; >50, above the 50th centile; 25–50, between the 25th and 50th centiles. All patients were negative for the mitochondrial m.3243A>G mutation. ND, not done; NA, not available; OAD, oral antidiabetic drug. Positive predictive value (PPV) for MODY was performed using the MODY Probability Calculator (https://www.diabetesgenes.org/mody-probability-calculator/).

under the most selective criteria (Supplementary Table 8), thus replicating our findings. For example, patients with T2D selected for age (12–35 yr) and BMI (20–35) showed increased frequency of rare *ONECUT1* variants compared with unselected controls (3.85% versus 0.81%, OR = 22.3, $P = 0.0015$).

Altogether, these results suggest that heterozygous *ONECUT1* mutations, including loss of function variants and a subset of rare coding variants, could be responsible for a new monogenic diabetes entity characterized by early-onset diabetes (juvenile- to adult-onset). In addition, these heterozygous patients did not have obesity at diabetes onset and had impaired insulin secretion.

T2D and other metabolic traits associated with *ONECUT1* variants. In the recent DIAMANTE GWAS, common variants located upstream of *ONECUT1* showed association with T2D (99% genetic credible set: chr15:53070141-53165681, hg19)[5]. We further explored this region for T2D and other metabolic traits, focusing on the four variants most strongly associated with T2D (credible SNPs; Supplementary Table 9 and Extended Data Fig. 1g (ref. [21])). The strongest T2D association is observed at rs2456530 (OR = 1.09, $P = 4.7 \times 10^{-9}$), which is also associated with BMI and childhood obesity in the same orientation (minor allele associated with increased T2D, increased BMI and childhood obesity). T2D association at this SNP remains highly significant after adjusting for BMI ($P = 1.1 \times 10^{-6}$). Strong association with T2D is also found at rs75332279, which is independent of BMI ($P = 5.8 \times 10^{-9}$ with T2D, $P = 0.0036$ with BMI). In addition, other SNPs in the region are strongly associated with BMI; for example, rs1899730 ($P = 5.6 \times 10^{-10}$) and the low-frequency variant rs16965225 (effect allele frequency = 0.060, $P = 1.7 \times 10^{-10}$). Some of these BMI-associated SNPs are weakly or not associated with T2D; for example, rs10851523 ($P = 1.8 \times 10^{-9}$ with BMI, $P = 0.034$ with T2D adjusted for BMI), which is located distally to *ONECUT1* and is not in linkage disequilibrium with T2D-associated SNPs in this region (Extended Data Fig. 1g). Overall, these observations support the contribution of partly distinct mechanisms involved in T2D and in BMI (Extended Data Fig. 1g), with a T2D-associated region located upstream of *ONECUT1*. SNPs associated with T2D in this region also show suggestive association with decreased fasting insulin adjusted for BMI, increased 2-h insulin and decreased insulin secretion rate, suggesting that insulin secretion is impaired (Supplementary Table 9). These SNPs also show locus-wide significant associations with lipid disorders, and suggestive (nominal significant) association with liver impairment traits (Supplementary Table 9). In addition, SNPs associated with BMI in this region show locus-wide significant association with sleep and circadian traits (Supplementary Table 9), which are known to be altered in several metabolic disorders including T2D and obesity[22]. The distinctive clinical features of *ONECUT1* homozygous and heterozygous patients suggest a pancreatic and/or endocrine developmental defect, and common T2D-associated variants could also affect these traits by modulating the same mechanisms.

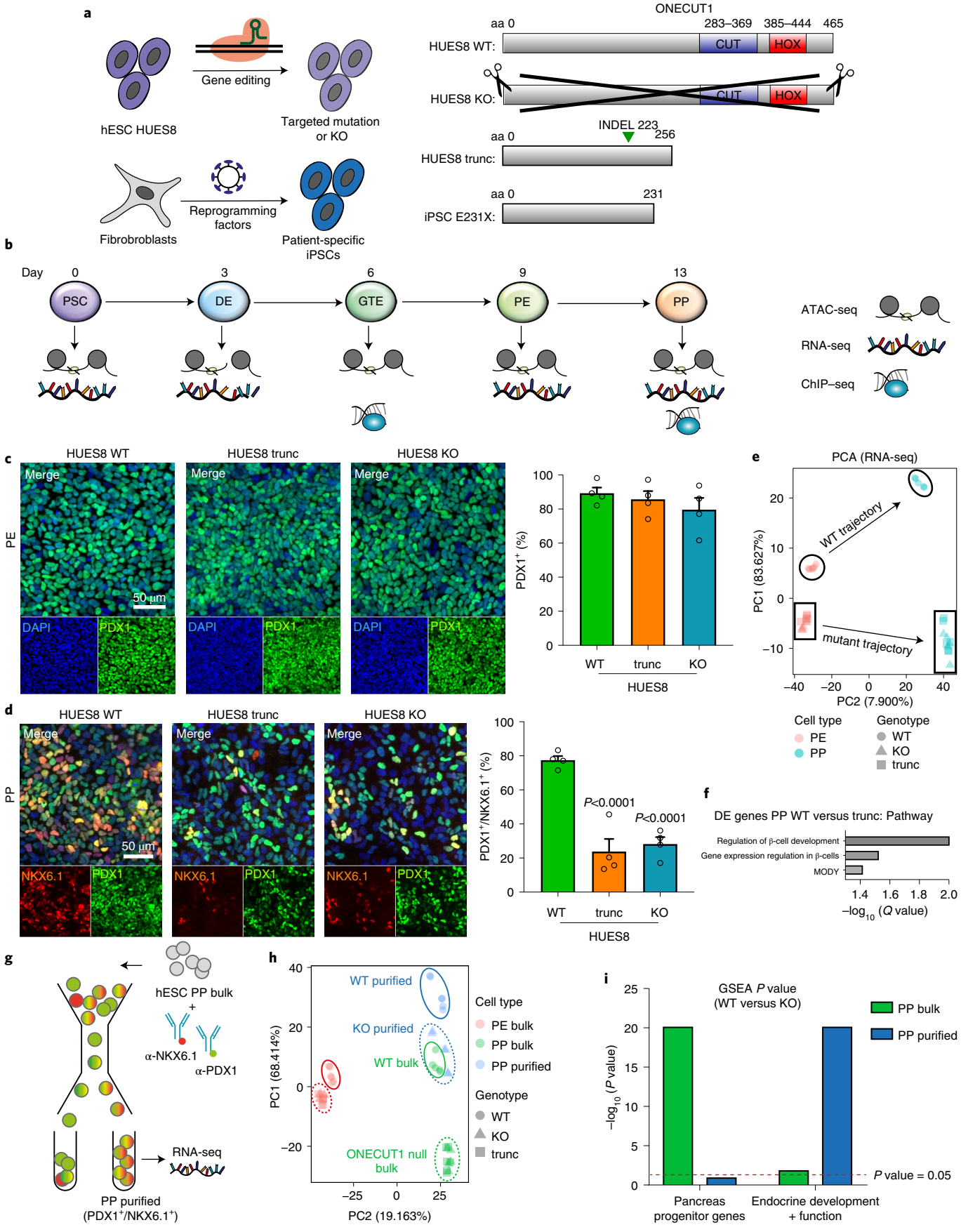
Defective pancreatic progenitor (PP) formation and endocrine priming in *ONECUT1* null PSCs. Next, we used genome engineering to either remove the entire *ONECUT1* gene (KO) or truncate (trunc) the two functional domains in several PSC lines. Moreover, we reprogrammed fibroblasts from *ONECUT1*-p.E231X homozygous Patient 1 (Fig. 2a and Extended Data Fig. 2a–c). Nonsense-mediated messenger RNA decay was excluded. Stage-specific pancreatic differentiation accompanied by stage-specific large-scale sequencing analysis (Fig. 2b) revealed normal definitive endoderm (DE) and pancreatic endoderm (PE) formation (Fig. 2c and Extended Data Fig. 2d–f). In contrast, PP formation was reduced in all *ONECUT1* null genotypes (Fig. 2d and Extended Data Fig. 2g).

Transcriptome-based principal component analysis (PCA) revealed distinct differentiation trajectories in *ONECUT1* null and WT cultures. *ONECUT1* null PP cells showed high similarity to the WT PE cell-stage (Fig. 2e) and gene set enrichment analysis (GSEA[23]) with PP signatures[24] confirmed downregulation of PP programs (Fig. 2i and Extended Data Fig. 3a). Similarly, *ONECUT1* null PP cultures lost programs associated with endocrine cell identity[25] (Fig. 2f,i and Extended Data Fig. 3b,c). Quantitative proteomics confirmed concordance of protein and RNA levels (Extended Data Fig. 3d,e). These data indicate a specific requirement of *ONECUT1* for the transition from PE to PP stage.

Intrinsic defects in *ONECUT1* null PPs to launch the β-cell program. To further characterize the cells successfully activating *NKX6.1* in the absence of *ONECUT1*, we performed fluorescence-activated cell sorting (FACS) purification followed by RNA sequencing (RNA-seq) (Fig. 2g). Purified *ONECUT1* null PPs clustered close to unpurified WT PPs (bulk), while the purified WT PPs clustered further away from all other PPs (Fig. 2h). GSEA indicates that genes downregulated in unsorted *ONECUT1* null PPs are primarily associated with a PP program (Fig. 2i and Extended Data Fig. 3a), while the genes downregulated in purified *ONECUT1* null PPs are more endocrine-specific[25] (Fig. 2i and Extended Data Fig. 3b). This suggests an intrinsically different transcriptional endocrine program of the purified *ONECUT1* KO PPs. Conversely, downregulated genes were enriched for ONECUT1-bound genes as measured by chromatin immunoprecipitation followed by sequencing (ChIP–seq) (Fig. 3a,b). Open chromatin (OC) sequencing (assay for transposase-accessible chromatin using sequencing (ATAC-seq)) also revealed that ONECUT1 ChIP–seq peaks were enriched in OC regions lost upon *ONECUT1* loss at PE and PP stages (Fig. 3c). PCA of OC in WT cells indicated a developmental trajectory culminating and separating genotypes at the PP stage (Extended Data Fig. 4a). Peaks with loss of OC upon *ONECUT1* loss were more abundant at distal regulatory regions (Extended Data Fig. 4b). At PE and PP stages, most regions with loss of OC in *ONECUT1* null cells were bound by ONECUT1 together with other pancreatic transcription factors (TFs) (Fig. 3d and Extended Data Fig. 4c) and were proximal to genes instructive of endocrine specification[26] (Extended Data Fig. 4d,e). A TF-binding activity analysis[27] indicated changes only at PE and PP stages with *ONECUT1* having highest activity loss (Extended Data Fig. 4f,g). Other pancreatic TFs displayed similar OC loss suggesting cooperative binding with ONECUT1 (Extended Data Fig. 4f,g). These data indicate that (1) ONECUT1 shapes chromatin accessibility at the PE to transit to PP stage and (2) regulates transcriptional activity of downstream

Fig. 2 | ONECUT1-depleted PSCs are defective in PP formation. a, ONECUT1 variants in gene-edited HUES8 hESCs and in reprogrammed iPSCs. **b**, Schematic pancreatic PSC differentiation outline and staged sequencing analysis or correspondingly employed dataset[56,57]. **c**,**d**, Differentiation efficiency at the PE (**c**) and PP (**d**) stage. Representative immunofluorescence images at the PE and PP stage, respectively. FACS-based quantification in HUES8 cells showed at PP stage 69% (trunc) and 64% (KO) reduction of efficiency (mean values ± s.e.m.; $n = 4$ independent experiments, one-way ANOVA with Tukey's test). **e**, RNA-seq-based PCA of indicated genotypes and stages (HUES8). Subpopulations and trajectories are indicated as borders and arrows. **f**, Pathway enrichment analysis[43] of differentially expressed genes with decreased expression in trunc versus WT cells (PP stage). **g**, Schematic of the FACS-based PP purification. **h**, RNA-seq-based PCA comprising HUES8 *ONECUT1* null and WT PE and PP cells (bulk) as well as purified PP (PDX1$^+$/NKX6.1$^+$) cells. Dashed circles, *ONECUT1* null; continuous circles, WT cells. **i**, GSEA[23] of contrasting HUES8 WT versus KO of purified PP (PDX1$^+$/NKX6.1$^+$) and bulk PP cells on a specific gene set for PPs[24] as well as for endocrine development and β-cell function[25]. aa, amino acid; hESC, human embryonic stem cell.

factors relevant to employ for β-cell differentiation. Accordingly, *ONECUT1* null PSCs were diminished in forming stage 5 endocrine progenitors and stage 6 immature β-like cells, while many of the C-peptide$^+$ cells lacking ONECUT1 were also negative for the islet-critical TF NKX6.1, indicating an altered transcriptional machinery[28] as confirmed by quantitative PCR (qPCR) (Fig. 3e–h).



Induced insulin secretion was reduced in stage 6 *ONECUT1* null cultures (Fig. 3i). These alterations are consistent with undetectable insulin in *ONECUT1*-mutated homozygous patients and the low fasting insulin in heterozygous patients (Fig. 1c,d, Table 1 and Supplementary Table 1).

An endocrine TF network involving ONECUT1. To further characterize how distinct *ONECUT1* coding variants cause or affect diabetes risk in humans, a set of variants identified in patients with diabetes (diabetes-causing variants hereafter; Extended Data Fig. 5a) were generated. All variants, except *ONECUT1*-p.E231X, showed strong nuclear localization and unchanged DNA binding (Fig. 4a–c and Extended Data Fig. 5b–e). However, transactivation capacities were altered in all *ONECUT1* diabetes-causing variants (Fig. 4d). Except for *ONECUT1*-p.E231X, which lacks the DNA-binding domains, VP16-fusion constructs restored the ability to activate transcription, indicating that these variants do not cause major structural protein impairments (Extended Data Fig. 6a). Previous work defined a network of islet-critical TFs indicative of auto- and cross-regulatory interactions[29]. Similarly, we found that ChIP–seq peaks for some of these TFs overlapped with ONECUT1 binding (Extended Data Fig. 6b,c). Binding is also observed in regions later bound by NKX6.1 and NKX2.2 in human islets (Extended Data Fig. 6c). Physical protein–protein interaction can enhance transcription during isletogenesis[30,31]. ONECUT1 interacted with GATA4, PDX1, GLIS3, NGN3, NKX6.1 and NKX2.2 (Fig. 4e,f and Extended Data Fig. 6d,e). Additionally, ONECUT1 formed homo- and heterodimers involving its C terminus, as ONECUT1-p.E231X shows disrupted interaction with WT ONECUT1 (Extended Data Fig. 7a–c).

ONECUT1 diabetes-causing variants fail to transactivate *NKX6.1* and *NKX2.2* enhancers. *NKX6.1*, *NKX6.2* and *NKX2.2* expression is specifically reduced in *ONECUT1* null PPs (Fig. 4g,h and Extended Data Fig. 8a). As NKX6.1 and NKX2.2 are essential to isletogenesis[28,32], the defective PP and endocrine program in *ONECUT1* null cells could be due to a cooperative interaction. Indeed, physical interaction of the ONECUT1 C terminus with NKX6.1/NKX2.2 was disrupted in the ONECUT1-p.E231X variant (Extended Data Fig. 7b,c). Furthermore, analysis of ATAC-seq and ChIP–seq data revealed putative *NKX6.1*, *NKX6.2* and *NKX2.2* enhancers occupied by ONECUT1 (Extended Data Fig. 8b,c) and GFP-reporter constructs confirmed activation of the ONECUT1-bound regions (Fig. 4i,j and Extended Data Fig. 8b). ONECUT1 diabetes-causing variants showed reduced activation of the *NKX6.1* as well as the *NKX6.2* and *NKX2.2* enhancers (Fig. 4k and Extended Data Fig. 8d). Coexpressing ONECUT1 with NKX2.2 further increased enhancer activity, attenuated by *ONECUT1* diabetes-causing variants (Fig. 4k and Extended Data Fig. 8d).

***ONECUT1*-p.E231D variant in PSCs phenocopies Patient 2.** To further substantiate this, the *ONECUT1*-p.E231D variant was engineered by CRISPR editing in ESCs to specifically resemble Patient 2 (Fig. 1b,d and Extended Data Fig. 9a,b). Pancreatic differentiation of these gene-edited cells confirmed the transition defect from PE to PP stage under modified culture conditions[33] (Extended Data Fig. 9c–e). We finally probed the cooperative action of the *ONECUT1*-p.E231D variant with NKX2.2 and found reduced protein–protein binding (Fig. 4e,f and Extended Data Figs. 6d,e, 7, 9f,g and 10g).

Finally, integrating genetic and functional evidence for rare *ONECUT1* coding variants identified as heterozygous in individuals with diabetes, according to the American College of Medical Genetics and Genomics guidelines, supports that these are likely to be pathogenic whereas the two control variants are not (Supplementary Table 10), further supporting their role in monogenic dominant diabetes.

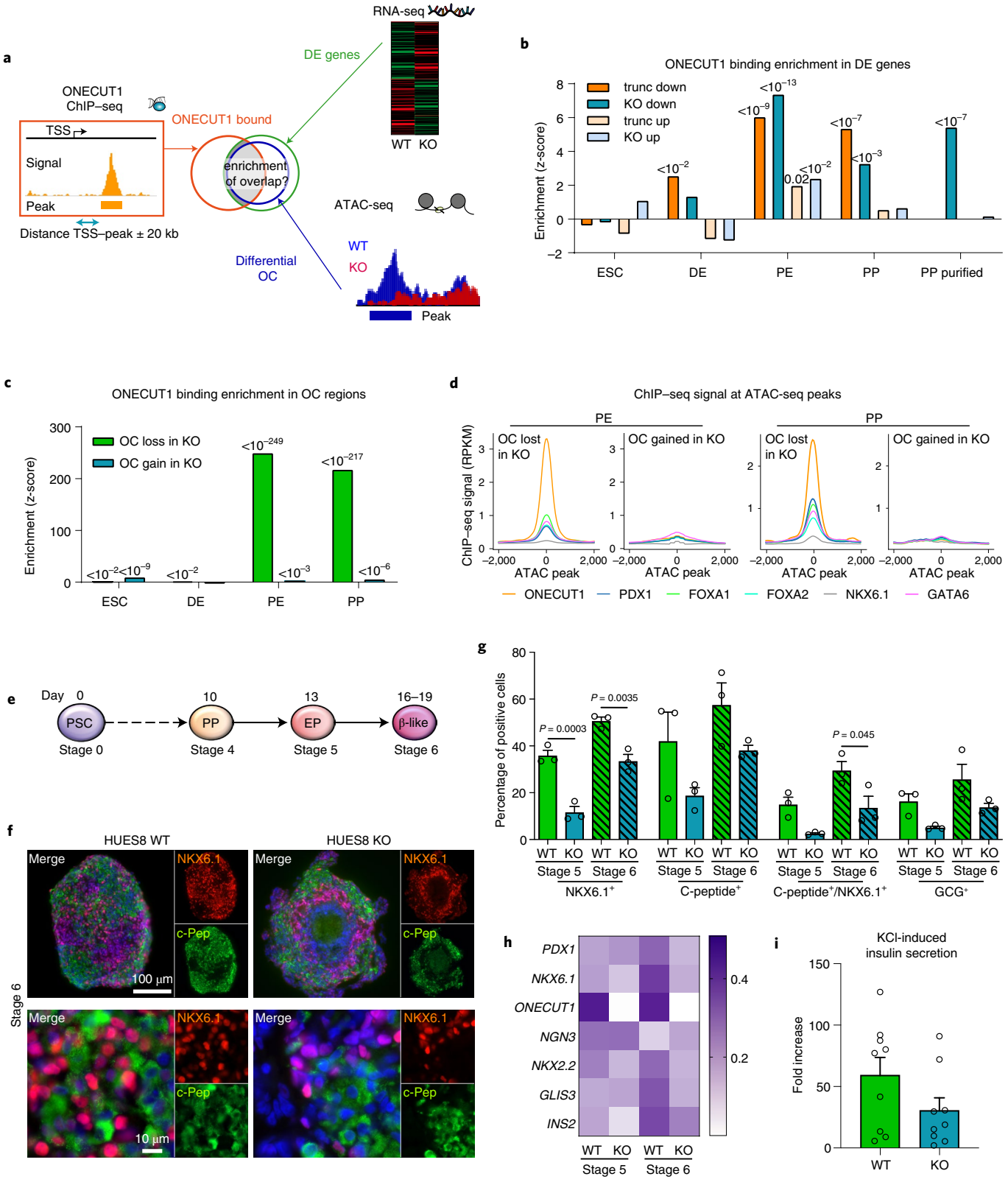
Common T2D-associated variants near *ONECUT1* affect endocrine regulatory elements. We mapped sequencing data onto the *ONECUT1* T2D association region, focusing on the aforementioned four credible SNPs[5] (Extended Data Figs. 1g and 10a). Three of these SNPs (rs2456530, rs2440374 and rs75332279) map to regulatory regions that are exclusively active in pancreatic islets and liver cells (GTEx[34]; Extended Data Fig. 10b), and rs7178476 maps to an enhancer in PP cells. These SNPs are in regions containing relevant functional elements: (1) ONECUT1 binding sites or (2) loss of OC upon *ONECUT1* KO, and (3) enhancer histone marks in PPs (Extended Data Fig. 10a). The T2D association region also includes a large regulatory element containing a long noncoding RNA, *RP11-209K10.2* (*lncRNA-RP11*), with similar tissue-specific expression as *ONECUT1* (Extended Data Fig. 10b,c). rs2440374 overlaps with a ONECUT1 ChIP–seq peak and with an OC region loss upon *ONECUT1* KO (Extended Data Fig. 10a). Moreover, this variant disrupts an NKX2.2 binding motif (Extended Data Fig. 10d) and is an expression quantitative trait locus (eQTL) for *lncRNA-RP11* in pancreas, as are rs2456530 and rs75332279 (Extended Data Fig. 10e). rs2440374 is also an eQTL for *lncRNA-RP11* in the liver ($P=1.2\times10^{-7}$) and for *ONECUT1* in the cerebellum ($P=4.0\times10^{-11}$), suggesting that these variants might also affect additional organs, consistent with the diversity of traits associated with these variants (Supplementary Table 9). *lncRNA-RP11* gets concurrently downregulated in *ONECUT1* KO PPs (Extended Data Fig. 10f). Collectively, these data suggest that SNPs associated with T2D located upstream of *ONECUT1* could affect the regulation of *lncRNA-RP11* expression. Finally, we correlated SNPs associated with T2D at the genome-wide level with TF-binding patterns and dynamically regulated OC upon ONECUT1 loss. This shows that T2D-associated SNPs cluster to ONECUT1-binding peaks in a similar range as for PDX1, but less prominently than the overlap with islet enhancers or NKX2.2-bound regions (Fig. 4l). Accordingly, genome-wide regions with loss of OC upon *ONECUT1* KO were also enriched for SNPs associated with T2D (Fig. 4l).

Discussion

Here we report a role of both rare coding and common regulatory *ONECUT1* variants in human diabetes, similar to several other diabetes genes, including *GCK*, *KCNJ11*, *HNF1A* and *PPARG*[35]. Our findings add *ONECUT1* to a small list of genes (for example, *GCK*)[36–38] involved in monogenic recessive (severe neonatal syndromic)

Fig. 3 | Intrinsic defects in ONECUT1-depleted PP cells disturb the β-cell program. a, Schematic enrichment analysis of ONECUT1-bound genes with either differentially expressed (DE) genes or differential OC peaks (WT versus KO). **b**, Binding enrichment (z-score) test of ONECUT1 (ChIP–seq, PP stage) in up- and downregulated genes of *ONECUT1* null and WT cells. **c**, Binding enrichment (z-score) test of ONECUT1 (ChIP–seq, PP stage) in differential OC regions (WT versus KO, ATAC-seq) of the depicted stages. Bars show enrichment in OC regions lost or gained in ONECUT1-depleted cells. **d**, ChIP–seq signals of key TFs at OC peaks lost or gained in *ONECUT1* KO. **e**, β-like differentiation scheme. **f**, Representative immunofluorescence images (NKX6.1, C-peptide, stage 6). **g**, Marker quantification was performed by FACS at stages 5 and 6 of *ONECUT1* KO and WT cells (mean values ± s.e.m.; $n=3$ independent experiments; one-way ANOVA with Tukey's test). **h**, Heatmap depicting relative marker expression in *ONECUT1* KO cells at stages 5 and 6. Values are normalized to *ONECUT1* WT and scaled by the sum of each row ($n=2$). **i**, Induced insulin secretion of *ONECUT1* KO and WT cells at stage 6 depicted as fold-increase comparing low-glucose-stimulated insulin secretion with subsequent KCl-stimulated insulin secretion (mean values ± s.e.m.; $n=3$ samples examined over 3 independent experiments). TSS, transcription start site.

diabetes and monogenic dominant (adult- and juvenile-onset) diabetes, and with common regulatory variants associated with multifactorial diabetes (T2D). Our human stem cell differentiation assay revealed impaired PP and β-like cell formation, and PP numbers are known to determine pancreas size[39], which is consistent with pancreatic hypoplasia in patients with homozygous mutations. A core action of ONECUT1 is the activation of *NKX6.1*, *NKX6.2* and *NKX2.2* via binding to *cis*-regulatory elements, impaired in diabetes-causing variants. ONECUT1 operates in a feed-forward loop to regulate its own transcription and



chromatin dynamics, consistent with variants in the promoter region being associated with multifactorial T2D. NKX2.2, another critical factor regulated and bound by ONECUT1 at regulatory sites, is also essential for specifying pancreatic islet-cell fates[40]. Diabetes-causing ONECUT1 variants had reduced capacity to activate these three *NKX* genes that compose a previously reported

Fig. 4 | *ONECUT1* mutations disturb the endocrine TF network. a, Subcellular localization of distinct GFP-fused ONECUT1 proteins. **b**, Electromobility shift assay (EMSA) of WT and ONECUT1 variants using a ONECUT1 binding motif (TRANSFAC T03257). **c**, TNT-ONECUT1 proteins as WB-control for **b**. **d**, Luciferase assay with WT and distinct ONECUT1 variants (diabetes-associated: G30S, E231D, E231X, H33Q, G81D, P215R, V242A; control: D26E, K412R) (mean values ± s.e.m.; $n = 6$ for G30S, E231D, E231X, H33Q, G81D; $n = 10$ for D26E, K412R, P215R, V242A; one-way ANOVA with Dunnett's test). **e**,**f**, Co-immunoprecipitation of FLAG-tagged ONECUT1 and interacting factors (NKX6.1; **e**; and NKX2.2; **f**). **g**,**h**, Relative RNA-seq-based expression of *NKX6.1* (**g**) and *NKX2.2* (**h**) at indicated genotypes and states (mean values ± s.e.m.; $n = 2$ samples examined over 3 independent experiments; two-tailed, unpaired *t*-test). **i**, E1-*NKX6.1* enhancer activity (GFP) during pancreatic differentiation in CyT49 cells. **j**, E1-*NKX6.1* enhancer activity in β- (MIN6) and α-cells (αTC). **k**, Luciferase assay with *NKX6.1* and *NKX2.2* enhancer regions overexpressing ONECUT1 variants ± NKX2.2 (mean values ± s.e.m.; $n = 6$ independent experiments; one-way ANOVA with Tukey's test; *compared with WT, #compared with ONECUT1 variant without NKX2.2). **l**, Significance of overlap of variants associated with T2D (DIAMANTE GWAS dataset; $P < 10^{-20}$; one-sided empirical permutation test) and ATAC-seq (loss or gain of OC upon *ONECUT1* KO) as well as ChIP–seq peaks. FPKM, fragments per kilobase of transcript per million mapped reads; IP, immunoprecipitation; WB, western blotting.

islet-specific TF network[29,41]. Physical interactions with other pancreatic TFs occur at the C terminus of ONECUT1, and missense diabetes-causing variants fail to transactivate but allow for DNA binding, which supports the relevance of protein–protein interaction in the endocrine program. We speculate that these variants affect the binding to specific, yet unknown, cofactor interactions involved in ONECUT1 function. Additionally, our chromatin- and *cis*-regulatory binding maps indicate that ONECUT1-induced gene transcription globally resides in OC clusters cobound by physically interacting islet-enriched TFs[28,29,40–42]. This is consistent with results demonstrating that ONECUT1-bound regions are overall associated with T2D in a similar range as NKX-bound regions in human islets[43].

Studies in KO mouse models pioneered by Lemaigre and colleagues revealed a role of Onecut1 for both endocrine specification and proper duct morphology[14,30,31,44,45]. Onecut1 and Pdx1 cooperate to ensure normal endocrine development and functional maturation of β-cells at later stages, in line with our observations[45,46]. Although Ngn3 expression is completely abrogated in *Onecut1* KO mice, this is not the case in our human data. However, the patient with homozygous *ONECUT1*-p.E231X mutations phenocopies the null allele mouse, at least partially[14–16,47]. Our family observations also suggest that homozygous *ONECUT1* loss-of-function mutations lead to early death or miscarriage. In mice the complete loss of *Onecut1* is tolerated in approximately a quarter of the animals[14], underpinning crucial differences between mice and humans. Homozygous *ONECUT1* and *RFX6* mutations in humans have similar clinical phenotypes, including neonatal diabetes with gallbladder agenesis and hypoplastic pancreas, whereas heterozygous *RFX6* mutations are associated with MODY[12,48].

Our findings highlight the power of our staged strategy combining clinical, genetic and in-depth functional approaches, enabling the identification of a monogenic component within a common, mainly multifactorial disease. Here increased power was achieved by different means: (1) reducing genome-wide scale to single-gene level; (2) suggesting the early age at onset of heterozygous patients, and hence screening appropriately designed diabetes cohorts; and (3) performing detailed functional investigations of disease-causing variants. In contrast, genome-wide WES studies in large, unselected T2D cohorts failed to detect a monogenic contribution of *ONECUT1* to T2D (refs. [5,6,49,50]). Hence, our staged strategy provides a powerful complement to these large-scale studies.

Some monogenic forms of diabetes (for example, MODY) appear to respond differently, albeit generally better, to treatment compared with common multifactorial T2D (refs. [4,51,52]). For example, MODY patients with *HNF1A* or *HNF4A* mutations are very sensitive to sulfonylureas, which stimulate insulin secretion[4,51,52]. Most patients heterozygous for rare *ONECUT1* coding variants had HbA1c values close to the normal range upon treatment, suggesting that they also respond generally well to diabetes treatment. Our data suggest that monogenetic profiles may underlie part of the clinical heterogeneity of T2D (ref. [4]). Hence, the identification of *ONECUT1* mutations among patients with T2D or MODY could help to improve treatment responses in these individuals[4,51,52].

In summary, our study provides a detailed and comprehensive analysis of *ONECUT1* as a diabetes gene. Our findings support the concept that *ONECUT1* variants might contribute to a broad spectrum of diabetes depending on both the risk genotype (homozygous, heterozygous) and the nature of the risk variant (loss of function, missense, regulatory), similar to other pancreatic development genes (for example, *HNF1A*, *PDX1* and *RFX6*). Additional large-scale studies are needed to dissect the precise genotype–phenotype correlation of *ONECUT1* variants in diabetes. Moreover, we studied a selection of these diabetes-causing variants to better understand the underlying mechanisms by which they contribute to diabetes pathogenesis. We also generated a coherent roadmap of *ONECUT1* for human pancreatic and β-cell development. Our study supports an approach that combines clinical studies, human genetics and time-resolved, high-resolution transcriptome and epigenome maps of differentiating human PSCs that might help to develop personalized therapies for diabetes based on molecular knowledge.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-021-01502-7.

Received: 17 July 2020; Accepted: 13 August 2021;
Published online: 18 October 2021

References

1. NCD Risk Factor collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* **387**, 1513–1530 (2016).
2. Bansal, V. et al. Spectrum of mutations in monogenic diabetes genes identified from high-throughput DNA sequencing of 6888 individuals. *BMC Med.* **15**, 213 (2017).
3. Shields, B. M. et al. Population-based assessment of a biomarker-based screening pathway to aid diagnosis of monogenic diabetes in young-onset patients. *Diabetes Care* **40**, 1017–1025 (2017).
4. Hattersley, A. T. & Patel, K. A. Precision diabetes: learning from monogenic diabetes. *Diabetologia* **60**, 769–777 (2017).
5. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505 (2018).
6. Flannick, J. et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71–76 (2019).
7. Heller, S., Melzer, M. K., Azoitei, N., Julier, C. & Kleger, A. Human pluripotent stem cells go diabetic: a glimpse on monogenic variants. *Front. Endocrinol. (Lausanne)* **12**, 648284 (2021).

8. Breunig, M. et al. Modeling plasticity and dysplasia of pancreatic ductal organoids derived from human pluripotent stem cells. *Cell Stem Cell* **28**, 1105–1124 e1119 (2021).
9. Wiedenmann, S. et al. Single-cell-resolved differentiation of human induced pluripotent stem cells into pancreatic duct-like organoids on a microwell chip. *Nat. Biomed. Eng.* **5**, 897–913 (2021).
10. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
11. Oliver-Krasinski, J. M. & Stoffers, D. A. On the origin of the beta cell. *Genes Dev.* **22**, 1998–2021 (2008).
12. Smith, S. B. et al. Rfx6 directs islet formation and insulin production in mice and humans. *Nature* **463**, 775–780 (2010).
13. Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.* **25**, 221–251 (2009).
14. Jacquemin, P. et al. Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene *ngn3*. *Mol. Cell Biol.* **20**, 4445–4454 (2000).
15. Clotman, F. et al. The onecut transcription factor HNF6 is required for normal development of the biliary tract. *Development* **129**, 1819–1828 (2002).
16. Jacquemin, P., Lemaigre, F. P. & Rousseau, G. G. The Onecut transcription factor HNF-6 (OC-1) is required for timely specification of the pancreas and acts upstream of Pdx-1 in the specification cascade. *Dev. Biol.* **258**, 105–116 (2003).
17. Lannoy, V. J., Bürglin, T. R., Rousseau, G. G. & Lemaigre, F. P. Isoforms of hepatocyte nuclear factor-6 differ in DNA-binding properties, contain a bifunctional homeodomain, and define the new ONECUT class of homeodomain proteins. *J. Biol. Chem.* **273**, 13552–13562 (1998).
18. Bonaldi, C. et al. A first national prevalence estimate of diagnosed and undiagnosed diabetes in France in 18- to 74-year-old individuals: the French Nutrition and Health Survey 2006/2007. *Diabet. Med.* **28**, 583–589 (2011).
19. Møller, A. et al. Hepatocyte nuclear factor-6: associations between genetic variability and type II diabetes and between genetic variability and estimates of insulin secretion. *Diabetologia* **42**, 1011–1016 (1999).
20. Zhu, Q. et al. Mutation screening of the hepatocyte nuclear factor (HNF)-6 gene in Japanese subjects with diabetes mellitus. *Diabetes Res. Clin. Pract.* **52**, 171–174 (2001).
21. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).
22. Allada, R. & Bass, J. Circadian mechanisms in medicine. *N. Engl. J. Med.* **384**, 550–561 (2021).
23. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
24. Cebola, I. et al. TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat. Cell Biol.* **17**, 615–626 (2015).
25. Hrvatin, S. et al. Differentiated human stem cells resemble fetal, not adult, β cells. *Proc. Natl Acad. Sci. USA* **111**, 3038–3043 (2014).
26. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495 (2010).
27. Li, Z. et al. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).
28. Schaffer, A. E. et al. Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic beta cell identity. *PLoS Genet.* **9**, e1003274 (2013).
29. Pasquali, L. et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
30. Oliver-Krasinski, J. M. et al. The diabetes gene *Pdx1* regulates the transcriptional network of pancreatic endocrine progenitor cells in mice. *J. Clin. Invest.* **119**, 1888–1898 (2009).
31. Kim, Y. S. et al. Glis3 regulates neurogenin 3 expression in pancreatic beta-cells and interacts with its activator, Hnf6. *Mol. Cells* **34**, 193–200 (2012).
32. Jennings, R. E. et al. Development of the human pancreas from foregut to endocrine commitment. *Diabetes* **62**, 3514–3522 (2013).
33. Thatava, T. et al. Indolactam V/GLP-1-mediated differentiation of human iPS cells into glucose-responsive insulin-secreting progeny. *Gene Ther.* **18**, 283–293 (2011).
34. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
35. Yang, Y. & Chan, L. Monogenic diabetes: what it teaches us on the common forms of type 1 and type 2 diabetes. *Endocr. Rev.* **37**, 190–222 (2016).
36. Fu, D. et al. Genetic polymorphism of glucokinase on the risk of type 2 diabetes and impaired glucose regulation: evidence based on 298,468 subjects. *PloS ONE* **8**, e55727 (2013).
37. Njølstad, P. R. et al. Neonatal diabetes mellitus due to complete glucokinase deficiency. *N. Engl. J. Med.* **344**, 1588–1592 (2001).
38. Vionnet, N. et al. Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 721–722 (1992).
39. Stanger, B. Z., Tanaka, A. J. & Melton, D. A. Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. *Nature* **445**, 886–891 (2007).
40. Churchill, Angela J. et al. Genetic evidence that Nkx2. 2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development. *Elife* **6**, e20010 (2017).
41. Miguel-Escalada, I. et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat. Genet.* **51**, 1137–1148 (2019).
42. Schaffer, A. E., Freude, K. K., Nelson, S. B. & Sander, M. Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. *Dev. Cell* **18**, 1022–1029 (2010).
43. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
44. Tweedie, E. et al. Maintenance of hepatic nuclear factor 6 in postnatal islets impairs terminal differentiation and function of β-cells. *Diabetes* **55**, 3264–3270 (2006).
45. Zhang, H. et al. Multiple, temporal-specific roles for HNF6 in pancreatic endocrine and ductal differentiation. *Mech. Dev.* **126**, 958–973 (2009).
46. Henley, K. D. et al. Threshold-dependent cooperativity of Pdx1 and Oc1 in pancreatic progenitors establishes competency for endocrine differentiation and β-cell function. *Cell Rep.* **15**, 2637–2650 (2016).
47. Zhang, Y. et al. HNF6 and Rev-erbα integrate hepatic lipid metabolism by overlapping and distinct transcriptional mechanisms. *Genes Dev.* **30**, 1636–1644 (2016).
48. Patel, K. A. et al. Heterozygous RFX6 protein truncating variants are associated with MODY with reduced penetrance. *Nat. Commun.* **8**, 888 (2017).
49. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
50. Flannick, J. et al. Sequence data and association statistics from 12,940 type 2 diabetes cases and controls. *Sci. Data* **4**, 170179 (2017).
51. Chung, W. K. et al. Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* **43**, 1617–1635 (2020).
52. Smith, R. J. et al. Individualizing therapies in type 2 diabetes mellitus based on patient characteristics: what we know and what we need to know. *J. Clin. Endocrinol. Metab.* **95**, 1566–1574 (2010).
53. Barker, J. M. et al. Two single nucleotide polymorphisms identify the highest-risk diabetes HLA genotype: potential for rapid screening. *Diabetes* **57**, 3152–3155 (2008).
54. Oram, R. A. et al. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care* **39**, 337–344 (2016).
55. Johnson, M. B. et al. Trisomy 21 is a cause of permanent neonatal diabetes that is autoimmune but not HLA associated. *Diabetes* **68**, 1528–1535 (2019).
56. Geusz, R. J. et al. Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes with roles in development. *eLife* **10**, e59067 (2021).
57. Lee, K. et al. FOXA2 is required for enhancer priming during pancreatic differentiation. *Cell Rep.* **28**, 382–393.e387 (2019).

Methods

Patients and study populations. *Neonatal and very early juvenile-onset diabetes patients and their families (Families 1 and 2).* The index patients, Patients 1 and 2, as well as their families are described in the Results section, in Table 1 (Patients 1 and 2) and in Supplementary Table 1 (heterozygous relatives). The study was explained to the parents of the patients, their other children and relatives (for Family 1). After obtaining written consent, blood samples were collected from family members and DNA extraction was performed using standard procedures. The parents agreed to participate in the subsequent metabolic study. The study was approved by the Hospices Civils de Lyon (Family 1) and by the Institutional Review Board of Ulm University (Family 2). Family 1 belongs to the French traveler community, a minority group with a strong tradition of consanguineous marriages, who in the case of this family was known to be of French descent. PCA confirmed their French ancestry, with possibly a minor level of admixture (Extended Data Fig. 1a,b). Local ancestry analysis at the *ONECUT1* locus shows that the haplotypes from both parents at the *ONECUT1* position are estimated to be of European ancestry (Extended Data Fig. 1c). Family 2 is from Turkish ancestry.

Metabolic exploration of heterozygous parents of Families 1 and 2. OGTT was performed on the parents of both index cases (Individuals 1 and 2 of Families 1 and 2; Supplementary Table 1). Serum insulin concentrations were measured in Patient 1's parents by immunoradiometric assay using a commercial kit (Bi-insulin IRMA, CIS Bio International). OGTT procedure: an unrestricted diet rich in carbohydrates was recommended for 3 d before the test. After an overnight fasting period of 8–14 h, baseline (zero time) blood glucose and insulin were measured. Within 5 min, individuals were required to drink 1.75 g of glucose per kg of body weight as an 18% solution (maximum 75 g). Venous plasma was collected for additional determinations at 30, 60, 90 and 120 min.

Individuals with diabetes mellitus and controls from Ulm, Germany (UDC). All samples were obtained through the Centre of Excellence for Metabolic Disorders, Division of Endocrinology and Diabetes, Ulm University Medical Centre. Diabetes was defined as fasting plasma glucose > 125 mg dl^{-1} or 2-h glucose > 200 mg dl^{-1} after an OGTT. Furthermore, individuals with a history of diabetes or undergoing treatment with oral antidiabetic drugs or insulin were considered as cases. All individuals studied were of Northern European ancestry. In addition, all individuals with diabetes and the controls were tested for the presence of serum autoantibodies, including islet-cell autoantibodies (ICA), GAD antibodies and IA2 antibodies, as previously described[58]. Positivity for ICA, insulin requirement and evidence of ketosis at the time of diagnosis were criteria for exclusion of T2D. Exclusion criteria were also pregnancy and the presence of any other severe disease. All patients with nonautoimmune diabetes (UDC-T2D hereafter) had been previously sequenced for known MODY genes, including *HNF1A*, *HNF1B*, *HNF4A*, *GCK*, *ABCC8*, *PDX1*, *INS*, *PAX4*, *KCNJ11* and *NEUROD1*, and positive cases (carrying pathogenic or likely pathogenic variants) were excluded. Controls had normal fasting glucose (confirmed by HbA1c < 6%) and had no evidence of islet autoimmunity. Overall, 2,165 individuals with T2D were included and 397 controls. A group of 162 patients with adult-onset diabetes positive for ICA, GAD antibodies or IA2 antibodies (T1D/LADA) were used as additional non-T2D controls.

All individuals gave informed consent for use of their DNA samples for genetic studies. The study was approved by the Institutional Review Board of Ulm University, Ulm, Germany (registration numbers 42/2004 and 189/2007) and the Chamber of Physicians, State Baden-Wuerttemberg, Germany (registration number 133-2002), and is in accordance with the ethical principles of the Declaration of Helsinki.

Lebanese patient with nonautoimmune juvenile-onset diabetes. The patient, a boy, was recruited as part of a study of juvenile-onset diabetes in Lebanon[59]. The study was explained to the patient, his parents and his three healthy siblings. After signing written consent, blood samples were collected and DNA extraction was performed using standard procedures. The study was approved by the Research and Ethics Committee of the Chronic Care Center (Lebanon). Diabetes was diagnosed at the age of 12 yr, he was negative for GAD autoantibodies at diabetes onset and he was treated with insulin from diabetes onset. At the time of recruitment in the study (age 18 yr), the patient was obese (BMI = 36, standardized BMI (BMI SDS) = 4.5).

Genotyping and sequencing. We performed a 10,000-SNPs genome scan (Affymetrix) in a nuclear subset of Family 1 composed of the index patient, his two parents and one of the healthy siblings, for subsequent linkage analysis of neonatal diabetes. Linkage regions were intersected with genes involved in early endoderm development (a total of 8 candidate genes). An additional 300,000-SNP genome scan (Affymetrix) was performed in Patient 1's parents for ancestry analysis[10]. We performed mutation screening of *ONECUT1* exons in the index patient of Family 1 (Patient 1), his two unaffected siblings and their parents by sequencing genomic DNA using Sanger sequencing on an ABI-3730 sequencer (Applied Biosystems). We then completed the sequencing of the identified *ONECUT1* PTV variant in the extended Family 1. Sequencing of the German patients with diabetes and controls was also performed by Sanger sequencing of *ONECUT1* exons on genomic DNA. Sequences of primers used for *ONECUT1* sequencing are available in Supplementary Table 11a.

In Family 2, we performed targeted gene sequencing of a panel of 12 candidate genes (*PDX1*, *HNF1A*, *HNF1B*, *HNF4A*, *GCK*, *ABCC8*, *INS*, *GLIS3*, *WFS1*, *PAX4*, *KCNJ11*, *ONECUT1*) in the index case (Patient 2) and his parents. For each gene, exon coordinates were obtained from the RefSeq database to identify the coding and untranslated regions of the target genes. A standard Illumina library preparation process was done using Illumina Library Prep kit (Illumina). Sequencing of the libraries was performed on an Illumina HiSeq System (Illumina). Coverage was greater than 200× at each base.

In Patient 1 as well as in a Lebanese patient diagnosed with juvenile-onset nonautoimmune diabetes, we performed WES on the sequencing platform of the Centre National de Recherche en Génomique Humaine (CNRGH, Evry, France). Exomes were captured using the Agilent Sureselect All Exons Human V5 + UTR (Agilent Technologies). Final libraries were then sequenced on a HiSeq2000 with paired ends, and 100-base pair (bp) reads. Reads were mapped to the reference GRCh37 using BWA-mem 0.7.5a. Variant discovery was done using GATK 3.3 (UnifiedGenotyper) according to GATK Best Practices recommendations. The sequencing coverage over the whole exome was 99.7% and 99.1% for a 10x depth of coverage resulting in a mean sequencing depth of 290x and 121x for Patient 1 and the Lebanese patient, respectively. We also performed WES in the 13 UDC-T2D patients heterozygous for *ONECUT1* missense variants on the same platform. Sequencing coverage was on average 99.2% for a 10x depth, resulting in a mean sequencing depth of 117x on average (ranging from 86x to 139x). The joint variant calling file was annotated with RefGene gene regions using ANNOVAR (16 April 2018). Exome variant analysis was then performed using an in-house python pipeline on genetic variation annotation results. Variants were filtered consecutively based on their quality (variant quality (Phred Q score) > 20, genotype quality > 20 and depth > 5x), the predicted consequence on coding capacity (missense, nonsense, splice-site and coding insertion/deletion-frameshift or in frame) and their rare status based on information available in public databases. All the variants identified by WES were confirmed by Sanger sequencing in the corresponding individuals and genotyped in available family members on PCR-amplified DNA.

In the two individuals (UDC-T2D Patients 3 and 4) heterozygous for two coding *ONECUT1* variants (one rare coding variant together with the low-frequency P75A variant), we determined the phase of the variants by two distinct methods. For Patient 4, double heterozygous for G96D and P75A, two variants distant by only 64 bp, we determined the phase from the WES data, considering the reads that covered both variants. On a total of 19 reads, the G96D-(D) was associated with the P75A-(P) (11 reads) or the G96D-(G) associated with P75A-(A) (8 reads), establishing that the rare G96D variant is in phase with P75A higher-frequency variant (rare and low-frequency variants in *trans*). For Patient 3, double heterozygous for P215A and P75A, which are distant by 420 bp, we performed allele-specific amplification of each allele followed by sequencing of the amplified fragments using the same primers; primers are shown in Supplementary Table 11b. This showed that the rare variant of P215A-(A) coamplifies with the lower-frequency variant at P75A-(A) (rare and low-frequency variants in *cis*).

Determination of HLA-DR risk alleles and the T1D-GRS. HLA-DR genotypes for DR3, DR4 and DR15 were determined using tag SNPs rs2187668 and rs7454108 to tag DR3 (DRB1*0301-DQA1*0501-DQB1*0201) and DR4-DQ8 (DRB1*04-DQA1*0301-DQB1*0302) alleles, respectively, and SNP rs3129889 to tag HLA DRB1*15 (ref. [53]). The T1D-GRS was determined using the genotypes of the top ten risk alleles for T1D including the two HNS-DR tag SNPs above, according to Oram et al.[54]. The ten-SNPs T1D-GRS obtained was compared with the T1D-GRS distribution obtained with the same SNPs in a European T1D control population studied by Johnson et al.[55]: <25, below the 25th centile; >50, above the 50th centile; 25–50, between the 25th and 50th centiles.

Genotyping of the m.3243A>G mitochondrial mutation. Genotyping of the m.3243A>G mutation was performed by the PCR-restriction fragment length polymorphism (RFLP) method as described by Rong et al.[60].

Characteristics of ONECUT1 variants: allele frequencies and predicted deleterious consequences. Allele and genotype frequencies of the variants were estimated in our study populations (German patients with diabetes and controls) and in publicly available reference populations: gnomAD (https://gnomad.broadinstitute.org), 141,456 individuals; and the Exome Variant Server (NHLBI GO Exome Sequencing Project, Seattle, WA, release ESP6500SI-V2), 6,503 individuals. The damaging consequences of *ONECUT1* missense variants identified by sequencing were predicted using several prediction programs, available through the ANNOVAR web site (http://www.openbioinformatics.org/annovar/; date 15 October 2020).

Statistical analyses. The linkage study was performed by multipoint genetic analysis using Merlin software in a subset of Family 1 (Individuals 1, 2, 3 and 5) under a

rare disease recessive model (allele frequency: 0.000001) with complete penetrance and no phenocopy. The parameters of the genetic model (allele frequencies, phenocopy rate and penetrance) for transmission of diabetes with respect to the *ONECUT1*-p.E231X variant in the extended Family 1 were estimated by likelihood maximization using ILINK software (package fastlink.4.1P). The recombination rate between disease locus and variant was fixed to 0 (causative variant) and allele frequency of the variant was fixed to 0.000001 (absent in public databases). The affection status considered for modeling was diabetes in a broad sense, including neonatal diabetes and T2D/IFG.

Prevalence of T2D or IFG was compared within the extended Family 1 between mutation carriers and normal homozygotes by logistic regression, with IFG defined based on fasting blood glucose $\geq 5.6\,mmol\,l^{-1}$ (American Diabetes Association criteria; https://www.diabetes.org), taking age and sex as covariates. Individuals 3 and 4, aged 14 and 15 yr, were excluded from this analysis.

Prevalence of T2D or IFG was compared between mutation carriers from the extended Family 1 and the general French population, surveyed for fasting blood glucose, using the same criteria[18]. Individuals were stratified in four risk groups based on age (30–54 and 55–74) and sex, and the prevalence in these defined risk groups was obtained from C. Bonaldi (personal communication). The statistical analysis was performed by binomial convolution in these four risk groups. Individuals with ages < 30 (Individuals 3, 4 and 16) were excluded from the analysis. Statistical analyses of the German T2D populations were performed using JMP package (SAS Institute) or R package (library survival). For survival analysis, the log-rank test was used to compare survival functions and the Cox model was used to estimate the effect of rare variant risk factors on age of onset of diabetes.

To determine population ancestry, PCA was performed using EIGENSTRAT and SmartPCA (POPGEN) software from the EIGENSOFT package on the genotype data from the parents (300,000 SNPs, Illumina) and from the child (Patient 1, WES). Genotype data were merged using PLINK v.1.90. Control populations used for the analysis were those from the 1000 Genomes Project[10] (whole genome sequencing) as well as various European groups studied at the CNRGH (a total of more than 3,000 individuals).

Pluripotent stem cell assays and bioinformatics. *Stem cell culture.* Permission to culture and differentiate HUES8 cells into the pancreatic lineage was obtained from the Robert Koch Institute within the '79. Genehmigung nach dem Stammzellgesetz, AZ 3.04.02/0084'. Human ESCs and iPSCs were cultured at 5% CO_2, 5% O_2 and 37 °C on human ESC Matrigel-coated plates in mTESR1 (STEMCELL Technologies) medium with daily change of medium. Cells were split twice a week with TrypLE Express (Invitrogen) to enable feeder-free single-cell cultures. For splitting, cells were washed with PBS and dissociated with TrypLE for 3–5 min at 37 °C. The enzymatic reaction was stopped by diluting with blank medium and the cell suspension was centrifuged at 180*g* for 5 min, supernatant was discarded and cells were carefully resuspended in mTESR1 medium supplemented with 10 µM ROCK inhibitor (Abcam) for improved cell survival.

Reprogramming of fibroblasts. Reprogramming of human fibroblasts was approved by the Ulm University Ethics Committee (no. 232/17) and performed according to ref. [61]. Briefly, fibroblasts were cultured on gelatin-coated cell culture dishes in DMEM (Gibco) supplemented with 2 mM glutamine, 1× nonessential amino acids, 1 mM sodium pyruvate, 10% FCS and penicillin/streptomycin. After reaching 50% confluency, fibroblasts were infected with hOKSM-dTomato lentivirus on 2 consecutive days followed by transfer on irradiated rat embryonic fibroblast feeder cells on day 3. Growing colonies of human iPSCs were mechanically picked, expanded on rat embryonic fibroblast feeder cells and later adapted to feeder-free culture conditions. Reprogramming of patient fibroblasts (Patient 1, *ONECUT1*-p.E231X) yielded one viable iPSC clone. In addition, one clone of iPSCs derived from a healthy proband served as control.

Genome editing in PSCs. For *ONECUT1* KO in HUES8 cells, CRISPR RNAs (crRNAs) were designed using 'http://crispor.tefor.net/' designing tool. Cloning guide oligonucleotides (guide RNA, gRNA) in a gRNA cloning plasmid (a gift from George Church; Addgene plasmid no. 41824) was based on a previously published strategy[62]. PSCs were transfected with XtremeGene 9 DNA transfection reagent (SIGMA) according to the manufacturer's protocol, introducing two gRNA plasmids and pCas9_GFP vector (a gift from Kiran Musunuru; Addgene plasmid no. 44719 (ref. [63])) into the cells. Briefly, 200,000 cells were seeded on a 6-well plate 1 d before transfection and 2 µg of each plasmid was transfected with 18 µl of XtremeGene 9 reagent (3:1 ratio). GFP^+ cells were sorted 24 h after transfection and plated for single-cell seeding in media containing 10 µM ROCK inhibitor and 0.5 µM Thiazovivin (Calbiochem). Single colonies were mechanically isolated, replated and analyzed for *ONECUT1* gene KO. The applied paired guide approach to delete a large DNA fragment by two simultaneous double-strand breaks allowed efficient PCR screening by one external PCR flanking the site of deletion and one internal PCR within the deleted sequence. Homozygous *ONECUT1* KO clones were positive for the external PCR, but negative for the internal PCR. Gene editing yielded one HUES8 clone with homozygous *ONECUT1* KO.

HUES8 cells with truncated *ONECUT1* were generated using a zinc finger nuclease approach. Zinc fingers were designed to target position E231 of *ONECUT1* and respective plasmids were introduced by nucleofection using the Amaxa Nucleofector Kit (Lonza). Following nucleofection, cells were seeded in media containing 10 µM ROCK inhibitor and cultured for 4 d. Finally, cells were seeded in low density to achieve single-cell clones, which were mechanically isolated for genotyping and further expansion. This editing approach led to two homozygous HUES8 clones, where one clone was characterized in more detail.

For editing of HUES8 *ONECUT1* E231D, synthetic single guide RNAs (sgRNAs, Synthego) as well as a single-stranded DNA repair template were introduced with Lipofectamin Stem Reagent according to the manufacturer. Single-cell clones were screened for indels or targeted mutation within *ONECUT1*. Sequences of crRNAs, ZFN arms, sgRNAs and primers used for screening after gene editing are available in Supplementary Tables 12 and 13.

DNA isolation and PCR reaction. After clonal expansion, half of a colony was used for DNA isolation, while the other half of the colony was further cultivated for expansion of successfully edited clones. DNA was isolated using either DNeasy Blood and Tissue Kit (Qiagen) or Tissue Genomic DNA Purification Mini Prep Kit (Genaxxon) according to manufacturer's instructions. For the initial PCR screening, 30–150 ng of DNA was used as template with GoTaq Flexi DNA Polymerase (Promega). For low DNA input (0–30 ng) we used the RedMastermix (2x) Taq PCR Mastermix (Genaxxon). Clonal genotype was validated by Sanger sequencing (Eurofins Genomics). Results were confirmed on DNA isolated after expansion of clones, and KO was validated on transcriptomic and proteomic levels.

Differentiation of PSCs into PP cells. Pancreatic differentiation for HUES8 cells and iPSCs was reported previously[8,9]. Basal media for differentiation culture were: (1) BE1: MCDB131 (Invitrogen) with 0.8 g per liter of cell culture of tested glucose (Sigma), $1.174\,g\,l^{-1}$ sodium bicarbonate (Sigma), 0.5% fatty acid-free BSA (Proliant) and 2 mM L-glutamine. For iPSC differentiation, BSA concentration was reduced to 0.1% in BE1 for the first 3 d. (2) BE3: MCDB131 with $3.32\,g\,l^{-1}$ glucose, $1.754\,g\,l^{-1}$ sodium bicarbonate, 2% fatty acid-free BSA, 2 mM L-glutamine, $44\,mg\,l^{-1}$ L-ascorbic acid and 0.5% ITS-X.

For differentiation, PSCs were seeded on culture plates coated with growth factor-reduced Matrigel (BD, 354230) using 300,000 cells per 24-well plate in mTesR1 containing 10 µM ROCK. The next day, when cells reached 80% confluence, differentiation was initiated after washing with PBS (Sigma) in BE1 medium with 2 µM CHIR99021 (Axon MedChem) and $100\,ng\,ml^{-1}$ Activin A (R&D) (day0-medium). After 24 h, the medium was replaced by BE1 supplemented with $100\,ng\,ml^{-1}$ Activin A and $5\,ng\,ml^{-1}$ bFGF (R&D). Two days later, cells at DE stage were treated with $50\,ng\,ml^{-1}$ FGF10 (R&D), 0.75 µM Dorsomorphin (Sigma) and $3\,ng\,ml^{-1}$ Wnt3a (Peprotech) in BE1 for 3 d. Next, medium was changed to BE3 containing 0.25 µM SANT-1 (Sigma), 200 nM LDN-193189 (Sigma), 2 µM retinoic acid (Sigma) and $50\,ng\,ml^{-1}$ FGF10. After 3 d (at PE stage, from day 9 to 13), the cells received BE3 supplemented with $100\,ng\,ml^{-1}$ EGF (R&D), 200 nM LDN, 330 nM Indolactam V (STEMCELL Technologies) and 10 mM nicotinamide (Sigma) for another 4 d. Of note, differentiation of HUES8 E231D was performed without Indolactam V, a potent activator of protein kinase C (PKC), at PE stage to produce high yields of $PDX1^+/NKX6.1^+$ PPs. During differentiation, cells were cultured at 37 °C in a 5% CO_2 incubator with daily media change.

Differentiation of PSCs into β-like cells. ONECUT1 null and WT PSCs were differentiated across several stages toward β-like cells according to Rezania et al.[64] and Mahaddalkar et al.[65]. Briefly, cells were differentiated to DE and seeded on culture plates coated with growth factor-reduced Matrigel. Subsequently, cells were collected with gentle cell dissociation reagent (STEMCELL Technologies) and 30,000 cells were seeded in round-bottom ultra-low-attachment plates (96-well, Corning) in mTesR containing 10 µM ROCK for aggregation. Cells were cultured at 37 °C in a 5% CO_2 incubator and switched to differentiation media with daily media change the following day. Cells were collected on day 13 (stage 5) and day 16 (stage 6) for marker expression analysis.

Cell culture and preparation of cell extracts. Cell lines HEK293 (ATCC no. CRL 1573) and HeLa (ATCC no. CCL2) were cultivated in DMEM supplemented with 10% FCS and penicillin/streptomycin. For western blotting and immunoprecipitation experiments, whole-cell lysates were prepared essentially as previously described[66]. Protein concentrations were determined using the Bradford assay method (Bio-Rad).

Quantification of insulin secretion. For insulin secretion, from three independent differentiations two to three spheres were collected on day 19 (stage 6) and transferred to one 96-well plate ($n = 3$ biological replicates). Spheres in three 96-well plates ($n = 3$ technical replicates) were washed twice in wash buffer (KRBH buffer containing 0.1% BSA) and incubated for 1 h at 37 °C in 150 µl of KRBH buffer containing 0.1% BSA and 0.1 mM glucose. Spheres were washed, and supernatant was replaced with fresh KRBH buffer containing 0.1% BSA and 0.1 mM glucose and collected after 1 h at 37 °C. Subsequently, spheres were washed and incubated in KRBH buffer containing 0.1% BSA and 30 mM KCl for 15 min at 37 °C. Supernatant was collected and secreted insulin was quantified using an insulin ELISA Kit (ALPCO) normalized to total cell number per well.

NKX6.1 reporter constructs. Coordinates for the candidate NKX6.1 enhancers E1–E3 are: E1: chr4:85354431-85356380; E2: chr4:85356591-85358458; E3: chr4:85358376-85359476, based on hg19. Candidate enhancers were amplified from human genomic DNA by PCR and cloned into the GFP-reporter vector pSinTK (PMID: 21160473) using the polymerase incomplete primer extension cloning method (PMID: 18004753). Lentiviruses were constructed by cotransfecting the pSinTK with pCMV R8.74 and pMD.G helper plasmids into HEK293T cells. Viral supernatant was collected and concentrated by ultracentrifugation for 2 h at 19,400 r.p.m. using an Optima L-80 XP Ultracentrifuge (Beckman Coulter). Undifferentiated CyT49 human ESCs were transduced with the reporter virus on 2 consecutive days and maintained with the addition of 300 μg ml^{-1} Geneticin (G418 antibiotic) for selection of transduced cells. After 1–2 weeks of antibiotic selection, the cells were differentiated as described[67]. At the appropriate stages of differentiation, aggregates were collected for imaging. For cell imaging, 40 μl of cell aggregates were washed in PBS, placed in an optically clear glass-bottom dish (MatTek) and imaged at ×20 magnification. Images from each time point were acquired using the identical exposure time with a Zeiss Axio-Observer-Z1 microscope and a Zeiss AxioCam digital camera. Cell aggregates were also fixed, embedded in Optimal Cutting Temperature Compound (Tissue-Tek), sectioned and analyzed by immunocytochemical analysis as described.

Flow cytometry. Differentiation efficiency was determined by flow cytometry. For DE stage, surface markers c-Kit (CD117) and CXCR4 (CD184) were quantified; for PE and PP stage, TFs PDX1 and NKX6.1; and, for endocrine cells, C-peptide, glucagon and NKX6.1 were analyzed using FACSDiva software v.8.0.1 (BD Biosciences) and FlowJo 10.5.0.

Surface marker staining. After collecting cells from a 24-well plate with TrypLE Express (Invitrogen), cells were washed with FACS buffer (2% FCS in PBS) followed by blocking (10% FCS in FACS buffer) for 20 min on ice. After washing in FACS buffer, cell pellets were resuspended in FACS buffer containing PE-conjugated CXCR4 antibody (Life Technologies). First, cells were incubated on ice for 30 min. Second, APC-conjugated c-Kit antibody (Life Technologies) was added for an additional 15 min. Cells were washed and resuspended in FACS buffer supplemented with 100 ng ml^{-1} DAPI to assess cell viability. Before analysis on a BD LSC II flow cytometer, samples were filtered through a 50-μm mesh.

Intracellular marker staining. For intracellular marker staining, HUES8 cells from a 24-well plate were collected as described above, washed with PBS and fixed in PFA solution (PBS with 4% PFA and 10% sucrose) for 25 min on ice. After fixation, cells were washed twice with PBS and blocked (5% donkey serum and 0.1% Triton X-100 in PBS) for 30 min on ice. Following blocking, cells were washed twice with Wash Solution (2% donkey serum and 0.1% Triton X-100 in PBS) and incubated overnight at 4 °C with primary antibodies PDX1 (R&D) and NKX6.1 (DSHB) or C-peptide (Cell Signaling) and glucagon (Sigma) in Blocking Solution. The next day, cells were washed three times and incubated with donkey Alexa Fluor secondary antibodies (Invitrogen) in Blocking Solution for 90 min on ice. For endocrine cells, an additional staining step using NKX6.1-APC (BD Biosciences) was included. Finally, cells were washed three times with Wash Solution, filtered through a 50-μm mesh into FACS tubes and analyzed on a BD LSR II flow cytometer.

Immunofluorescence staining. For In-Well immunofluorescence staining, cells were seeded and differentiated on μ-Plate 24-well plates (Ibidi). After washing with PBS, cells were fixed in 4% PFA solution for 20 min at room temperature and washed three times with PBS. Quenching with 50 mM NH_4Cl for 10 min was followed by washing with PBS. Subsequently, cells were incubated in Blocking Solution (5% donkey serum and 0.1% Triton X-100 in PBS) for 45 min at room temperature followed by staining with primary antibodies overnight at 4 °C. The next day, cells were washed twice with Wash Solution (2% donkey serum and 0.1% Triton X-100 in PBS) and incubated with respective secondary antibodies for 1.5 h at room temperature. Finally, after PBS washing nuclei were counterstained with 500 ng ml^{-1} DAPI and images were acquired on a Keyence Biozero BZ-9000 microscope. The following antibodies were used: OCT3/4 (Santa Cruz), NANOG (Cell Signaling), SOX17 (R&D), PDX1 (R&D) and NKX6.1 (DSHB), together with Alexa-conjugated secondary antibodies from Invitrogen.

Pancreatic spheres were fixed in PFA solution (PBS with 4% PFA and 10% sucrose) overnight and subsequently kept in 1 M sucrose in PBS overnight on a rotating platform. Spheres were embedded in O.C.T. freezing compound (Tissue-Tek) and cryoblocks were sectioned at 7 μm. Immunofluorescent stainings were performed similarly to In-Well staining with NKX6.1 (DSHB) and C-peptide (Cell Signaling) primary antibody. Slides were mounted with Fluoromount-G (Southern Biotech). Images were acquired using a Zeiss ApoTome.

Western blot for ONECUT1 and Actin. Proteins of pancreatic cells generated from ESCs were extracted with radioimmunoprecipitation assay buffer supplemented with protease inhibitor cocktail. Proteins (30–40 μg of protein lysate) were separated with 10% SDS–PAGE and transferred onto an Immobilon-P PVDF membrane (Millipore) using the Transblot semidry transfer system (Bio-Rad). Membranes blocked in TBS + 0.1% Triton-X (TBS-T) with 3% BSA for 1 h at room temperature were incubated overnight at 4 °C with primary antibody in blocking solution, washed in TBS-T and subsequently incubated with secondary antibody for 1 h (Supplementary Tables 14 and 15). Detection of horseradish peroxidase (HRP) was performed with SuperSignal West Dura Kit (Thermo Scientific) and Chemiluminescence Imaging–Fusion SL system (VILBER).

The following primary and secondary antibodies were used: anti-ONECUT1 (Santa Cruz), anti-β-Actin (Sigma), anti-mouse-HRP and anti-rabbit-HRP (ECL anti-rabbit or mouse IgG, GE Healthcare).

RNA isolation of fixed and permeabilized cells after cell sorting. Differentiated cells were collected, washed in PBS and filtered through a 40-μm mesh followed by fixation for 30 min on ice in 4% PFA containing 0.1% saponin and RNasin Plus Ribonuclease Inhibitors (Promega). Cells washed in PBS supplemented with 0.2% BSA, 0.1% saponin and RNasin were stained with anti-PDX1-PE and anti-NKX6.1-Alexa647 (BD Biosciences) in PBS with 1% BSA, 0.1% saponin and RNasin for 60–90 min at 4 °C on a rotating shaker, and after two washing steps were sorted for PDX1$^+$/NKX6.1$^+$ cells. Subsequently, sorted cells were lysed for 3 h at 50 °C in digestion buffer containing protease, then incubated for 15 min at 80 °C. Further purification steps were performed according to the RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies).

RNA isolation, reverse transcription and qPCR. Total RNA from ESC-derived pancreatic cells was isolated using GeneJET RNA Purification Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. For qPCR, 500 ng to 1 μg of total RNA was reverse-transcribed with the iScript cDNA Synthesis Kit (Bio-Rad) and 40 ng of complementary DNA was utilized for PCR reactions with the SensiMix SYBR Kit (Bioline) and QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific). QuantiTect primers (Qiagen) were used with hydroxymethylbilane synthase as endogenous control gene.

RNA-seq. After quality check with samples reaching RNA integrity number (RIN) values > 8, up to 1 μg of total RNA was used for poly-A enrichment and subsequent library preparation with the TrueSeq Stranded mRNA Kit from Illumina (HUES8 $n = 6$). Subsequent RNA-seq was performed on a HiSeq 3000 system (Illumina, single read, 1 × 50 bp) at the Biomedical Sequencing Facility (BSF) of the CeMM in Vienna, Austria.

Reads were aligned with STAR aligner (v.2.5.2b) on human genome hg38 and using Ensembl Transcriptome Annotation e87 as transcriptome reference. STAR parameters followed options suggested by the ENCODE project. Next, DESeq2 (v.1.22.1) was used for normalization and differential expression analysis. Only genes with at least 10 reads, an adjusted P value < 0.05 and absolute fold-change (FC) > 1.5 were considered. Clustering of gene expression is based on a fuzzy c-means algorithm (R package e1071). This pipeline was performed to cluster data from HUES8 (ESC, DE, PE and PP stages on WT, *ONECUT1* KO and truncated *ONECUT1* cells).

GSEA (Broad Institute). Lists of differentially regulated genes and corresponding $\log_2$-FC from RNA-seq were analyzed with GSEA 3.0 using GSEAPreranked with 1,000 permutations in default settings. For analysis, we used the following gene sets: PP genes were retrieved from Cebola et al.[24] with 500 pancreatic multipotent progenitor cell-specific genes. For endocrine development, we used 152 endocrine lineage genes from Hrvatin et al.[25] for GSEA analysis. We also used gene signatures of pancreas cells from the single-cell RNA-seq study for deriving signatures of alpha, acinar, beta, delta, ductal and mesenchymal cells obtained from GEO GSE81547.

ATAC-seq. For ATAC-seq, 50,000 ESC-derived pancreatic cells were collected, washed with PBS and directly lysed in tagmentation buffer containing TDE1 Tagment DNA enzyme, Digitonin and protease inhibitor cocktail (Nextera DNA Library Preparation Kit, Illumina). After 30 min of incubation at 37 °C, samples were purified with MicroElute Kit (Qiagen) and subsequently processed and analyzed at the BSF in Vienna ($n = 3$).

ChIP–seq. For ChIP–seq, the ChIP-IT High-Sensitivity Kit (Active Motif) was used according to the manufacturer's instructions. Briefly, aggregates containing approximately 10^7 cells were fixed for 15 min in an 11.1% formaldehyde solution, and chromatin was extracted by lysing cells in a Dounce homogenizer followed by shearing via sonication in a Bioruptor Plus (Diagenode), on high for 3 × 5 min (30 s on, 30 s off). For immunoprecipitation, 10–30 μg of the sheared chromatin was incubated with 4 μg of primary antibody overnight at 4 °C on an end-to-end rotator, followed by incubation with Protein G agarose beads for 3 h at 4 °C on the rotator. Reversal of crosslinks and DNA purification were performed according to the ChIP-IT High-Sensitivity instructions with an incubation at 65 °C for 2 h. DNA libraries were constructed using KAPA DNA Library Preparation Kits for Illumina (Kapa Biosystems) and library sequencing was performed using a HiSeq 4000 System (Illumina) with single-end reads of 50 bp in the Institute for Genomic Medicine core research facility at the University of California at San Diego. Additional details on the employed datasets are reported also in refs. [56,57]

Analysis of ATAC-seq and ChIP–seq. Reads were trimmed with skewer (v.0.2.1) and aligned with Bowtie2 to the human hg19 genome (v.2.3.4.2). For ATAC-seq, only properly mapped pairs were further considered, and duplicates were removed. We used MACS2 (2.1.2) with false discovery rate (FDR) of 5% to find condition-specific peaks using corresponding input DNA and replicates. For ONECUT1 ChIP–seq, we performed de novo analysis with MEME-ChIP (v.4.12). The same pipeline was also used to re-analyze public TF ChIP–seq data (FOXA1, FOXA2, PDX1, NKX6.1) and activating histone ChIP–seq (H3K4me1, H3K27ac) deposited at GEO GSE54471 and promoter associated ChIP–seq histone marks (H3K4me3) deposited at ArrayExpress E-MTAB-1086 for PPs and E-MTAB-1919 for islets.

Concerning ATAC-seq, we used THOR (v.0.11.6 (ref. [68])) to detect regions with differential peaks between consecutive differentiation steps in WT cells (ESC versus DE, DE versus gut tube endoderm (GTE), GTE versus PE and PE versus PP) and to compare WT and *ONECUT1* KO condition for each day ($P < 10^{-5}$). We combined all differential peaks in WT condition and quantified the number of reads per peak, which were first normalized by library size and then by removing the mean and scaling to unit variance, and performed clustering with fuzzy c-medoids using Pearson correlation as similarity metric. Next, we used HINT-ATAC (v.0.11.8) to find footprints at each condition analyzed after combining replicate libraries (ESC WT, ESC KO, DE WT and so on). We then performed motif matching of footprint sequences using JASPAR 2018 and MotifMatching from the Regulatory Genomics Toolbox (RGT; www.regulatory-genomics.org/rgt). Finally, differential footprint analysis was performed by comparing WT and *ONECUT1* KO conditions using HINT-ATAC. We only considered TFs with a high change in TF activity ($P < 0.05$).

For both ATAC-seq and ChIP–seq, genomic algebra operations were performed with bedtools and peak enrichment analysis was performed with GREAT. Deeptools was used to create histone and ATAC-seq signal heatmaps. Finally, T2D and BMI variants ($P < 10^{-8}$) were obtained from the DIAMANTE meta-analysis study. We also obtained OC regions from distinct pancreas cells (alpha, beta, acinar, ductal) and liver from the Diabetes Epigenome Atlas (https://www.diabetesepigenome.org/). Binding enrichment test (Fig. 3b) comparing gene sets with peaks was based on a *z*-score test, which estimates if the number of genes close to a peak (transcription start site ± 20 kb) is higher than in random generated peaks (1,000 randomizations). Enrichment tests between two peak sets (Fig. 3c) are based on an intersection test. First, we perform random permutations of two sets of peaks (5,000 randomizations) to obtain a null distribution of peak intersection. The *P* value is calculated as the fraction of intersection values that are at least as extreme as the intersection observed in the two peak sets. Both tests are implemented at RGT. *P* values are adjusted for multiple test correction using the Bonferroni method.

Functional enrichment analysis. Gene Ontology enrichment analysis of differentially downregulated genes (RNA-seq) or differentially closed chromatin (ATAC-seq) in ONECUT1-depleted cells was performed with ToppFun (FDR correction, *P* value cut off at 0.5) of ToppGene Suite. Adjusted *P* values for terms related to GO: biological function, mouse phenotype or pathway are shown.

Mass spectrometry. *In-solution digest.* Triplicates from PP stage HUES8 WT, KO and trunc were lysed in urea lysis buffer (8 M urea, 40 mM Tris (pH 7.6), 1× EDTA-free protease inhibitor (cOmplete, Sigma Aldrich) and 1× phosphatase inhibitor mixture). After lysing, cells were centrifuged for 20 min at 20,000*g* and 4 °C. For each sample, 50 µg of protein was further processed for digestion. Therefore, samples were reduced with 10 mM dithiothreitol for 45 min at 37 °C on a Thermoshaker (700 r.p.m.) and alkylated with 55 mM chloroacetamide (CAA) for 30 min at room temperature in the dark. After diluting the samples with 40 mM Tris (pH 7.6) to a urea concentration <1.6 M, trypsin (Trypsin Sequencing Grade, Roche) was added at 1:100 (enzyme/protein) and incubated for 3 h at 37 °C and 700 r.p.m. After preincubation, trypsin was added again at 1:100 (enzyme/protein) and incubated overnight at 37 °C and 700 r.p.m.

StageTip desalting. After digestion, samples were acidified to ~1% formic acid (FA) and desalted via StageTips as described earlier[69]. Briefly, five C18 discs (3 M Empore) were packed into a 200-µl pipette tip and put through the lid of a 1.5-ml Eppendorf tube. After wetting the column with 0.1% FA in 60% acetonitrile (ACN) and equilibration with 0.1% FA, acidified samples (pH 2) were loaded onto the StageTip. After washing off unspecific binders by two rounds of washing with 0.1% FA, peptides were eluted using two times 40 µl of 0.1% FA in 60% ACN. Samples were further frozen at −80 °C and dried using a Speed-Vac.

TMT labeling. TMT labeling was performed as previously described[70]. Briefly, digests were reconstituted in 20 µl of 50 mM HEPES (pH 8.5) and 5 µl of 11.6 mM TMT 10-plex (Thermo Fisher) in 100% anhydrous ACN was added to each sample. After incubating the samples for 1 h at 25 °C and 400 r.p.m., the reaction was stopped using 2 µl of 5% hydroxylamine. All nine TMT-labeled samples were pooled and the reaction vessels were rinsed with 20 µl of 10% FA in 10% ACN for 5 min at 400 r.p.m., and also added to the pooled samples. The samples were frozen at −80 °C and dried.

Sep-Pak desalting. Pooled samples were further desalted using 50-mg Sep-Pak columns (Waters). The columns were first wetted with 100% ACN, followed by 0.1% FA in 50% ACN, and further equilibrated with three washes of 0.1% FA. The dried pooled samples were reconstituted in 1 ml of 0.1% FA and loaded twice onto the column. After washing off unspecific binders with three times 0.1% FA, peptides were eluted using 200 µl of 0.1% FA in 50% ACN. Samples were frozen at −80 °C and dried.

High-pH reversed-phase fractionation (48 fractions). For high-pH reversed-phase fractionation, dried samples were reconstituted in mass spectrometry-grade water with 10% fractionation buffer A (25 mM ammonium bicarbonate, pH 8) and centrifuged for 5 min at 20,000*g* and 4 °C. The supernatant was then loaded on a C18 column (XBridge BEH130, 3.5 µm, 2.1 × 150 mm^2, Waters), which was connected to a Dionex Ultimate 3000 HPLC system (Thermo Fisher). After injecting 200 µg of protein digest of the sample at a flow rate of 200 µg min^{-1}, the system was equilibrated for 5 min with 85% fractionation buffer B (mass spectrometry-grade water), 10% fractionation buffer A and 5% fractionation buffer C (ACN). Peptides were eluted in a three-step linear gradient from 5% to 9% buffer C in 2 min with a constant amount of 10% buffer A. Then, a linear gradient from 9% to 47% buffer C in 91 min and from 47% to 55% buffer C in 3 min (with buffer A being constant at 10%) was used.

Starting from minute 3, 96 fractions (1 fraction per minute) were collected in a 96-well plate and pooled to 48 fractions. For that, column 7 was pooled to column 1, column 8 was pooled to column 2 and so forth. All fractions were frozen at −80 °C and dried using a Speed-Vac.

Liquid chromatography tandem mass spectrometry (LC-MS/MS) data acquisition. Fractionated samples were measured in data-dependent acquisition mode using a nanoflow LC-MS/MS by coupling a Dionex Ultimate 3000 UHPLC+ system to a Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). Dried samples were reconstituted in 0.1% FA and approximately 1 µg of peptides were injected. The sample was loaded to a trap column (75 µm × 2 cm, packed in-house with 5-µm C18 resin; Reprosil PUR AQ, Dr. Maisch) with a flow rate of 5 µl min^{-1} and washed for 10 min with 0.1% FA. Subsequently, peptides were separated on an analytical column (75 µm × 40 cm, packed in-house with 3-µm C18 resin; Reprosil PUR AQ) with a flow rate of 300 nl min^{-1} and a linear 50-min gradient from 8% to 34% liquid chromatography buffer B (0.1% FA, 5% dimethylsulfoxide in ACN) in liquid chromatography buffer A (0.1% FA, 5% dimethylsulfoxide in mass spectrometry-grade water). The eluate was sprayed via a stainless-steel emitter into the mass spectrometer, which was run in positive-ion mode. Fullscan MS1 spectra were acquired in the Orbitrap with 60,000 resolution and a scan range from 360 to 1,300 *m*/*z* (automatic gain control target of 4×10^5 charges, maximum injection time of 50 ms). A cycle time of 2 s and a dynamic exclusion of 90 s were used. MS2 spectra were recorded in the Ion Trap in rapid mode via sequential isolation of up to 10 precursors and the following settings: an automatic gain control target of 2×10^4, maximum injection time of 60 ms, isolation window of 0.7 *m*/*z* and fragmentation via collision-induced dissociation (CID) (normalized collision energy (NCE) of 35%). For the following MS3 scan, the 10 most intense precursors were further fragmented via higher-energy C-trap dissociation (HCD) (normalized collision energy (NCE) of 55%) and acquired in the Orbitrap with 50,000 resolution, scan range of 100–1,000 *m*/*z*, automatic gain control target of 1.2×10^5 charges, maximum injection time of 120 s and a charge-dependent isolation window from 1.3 (2+) to 0.7 (5–6+).

Data analysis. Acquired raw files were searched with Maxquant v.1.5.7.4 against the UniProtKB human reference list (downloaded 22 July 2013). For the search settings, up to two missed cleavages were allowed; carbamidomethylation was defined as a fixed modification; and oxidation of methionine, phosphorylation (STY) as well as N-terminal protein acetylation were set as variable modifications. Reporter ion MS3 was set as quantification type and TMT 10-plex as isobaric labels. The first search peptide tolerance was set to 20 ppm and the main search peptide tolerance was set to 4.5 ppm. Results were filtered by setting the protein and peptide FDR to 1% using a classical target-decoy approach. Data analysis was mostly done with Perseus (v.1.6.1.1 (ref. [71])). First, all reverse database hits, nonhuman contaminants and peptides that were only identified by site were excluded from the results. Then, peptides with less than two valid values in each triplicate were removed and significant changes between WT, KO and trunc were determined. For that, the S0 value was calculated as previously described[72] and used to perform a two-sample Student's *t*-test with an FDR of 0.05 in Perseus.

Plasmids and cell transfection. Generated expression constructs are listed in Supplementary Table 16.

HEK293 and HeLa cells were transfected using Lipofectamine 2000 transfection reagent (Invitrogen) according to the manufacturer's instructions.

Fluorescence microscopy. HeLa cells were plated (1×10^5 cells per cm^2) on chamber coverslips (Nunc). After 18 h, cells were transfected with 150 ng of expression plasmids for the specific GFP-fusion proteins. After 24 h, cells were rinsed with PBS, fixed with 4% paraformaldehyde and permeabilized with 0.1% Triton X-100. Specimens were embedded in ProLong Gold antifade reagent

(Invitrogen) supplemented with DAPI and stored at 4 °C overnight. Cells were imaged using a fluorescence microscope (IX71, Olympus).

In vitro protein translation. The in vitro protein translation was performed as previously described[66]. Quality of in vitro translations was analyzed by western blotting.

Electromobility shift assay (EMSA). In vitro-translated Flag-tagged ONECUT1 proteins (TnT Quick Coupled Transcription/Translation System, Promega) were used for electromobility gel shift assays in a binding buffer consisting of 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.1 mM EDTA, 0.5 mM dithiothreitol and 4% glycerol. For binding reaction, 2 µg of poly(dI-dC) (GE healthcare) and approximately 0.5 ng of ^{32}P-labeled oligonucleotides were added. The sequence of the double-stranded oligonucleotide (ONECUT1_E_UP: 5´-CCT GGT TTT GAA ATC AAT ATG GAA TCG-3´; ONECUT1_E_DO: 5´-CTC GCG ATT CCA TAT TGA TTT CAA AAC-3´) corresponds to a high-affinity ONECUT1 binding site[73]. Super shifting of complexes was achieved by adding 1 µg of anti-Flag (M5, Sigma) antibody. The reaction products were separated using 5% polyacrylamide gels with 1× Tris-glycine-EDTA at room temperature. Gels were dried and exposed to X-ray films (Kodak).

Coimmunoprecipitation experiments. Coimmunoprecipitation experiments were carried out essentially as described previously[74]. Briefly, 24 h after transfection, HEK293 (ATCC no. CRL 1573) cells were lysed with 600 µl of CHAPS lysis buffer (10 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate (CHAPS, Merck), 50 mM Tris-HCl (pH 7.8), 150 mM NaCl, 5 mM NaF, 1 mM dithiothreitol (Merck), 0.5 mM PMSF (Merck) and 40 µl ml^{-1} 'Complete Mix' protease inhibitor cocktail (Roche)). The extracts were incubated with 40 µl of agarose-conjugated anti-Flag antibody (M2, Sigma) at 4 °C overnight, and precipitates were washed 6–8 times with CHAPS lysis buffer and finally resuspended in SDS–polyacrylamide gel loading buffer. For western blotting the proteins were separated in SDS–polyacrylamide gels and transferred electrophoretically to PVDF membranes (Millipore) for 1 h at room temperature and 50 mA using a Tris-glycine buffer system. After blotting, the membranes were preblocked for 1 h in a solution of 3% milk powder in PBS-T (0.1% Tween20 in PBS) before adding antibodies. The following antibodies were used: anti-GFP (Sigma), anti-Flag (M5, Sigma) and HRP-linked secondary antibody (GE Healthcare).

Luciferase assay. HeLa (ATCC no. CCL2) cells were seeded in 48-well plates at a density of 2×10^5 cells. After 16 h, transfection was performed with Lipofectamine 2000 (above) using 250 ng of reporter plasmid alone or together with various amounts of expression plasmid (given in the corresponding figure). After 24 h cells were lysed, and luciferase activity was determined from at least six independent experiments with 10 µl of cleared lysate in a Centro LB960 luminometer (Berthold Technology) by using the luciferase assay system from Promega.

Cell line summary. Details on the employed cell lines are shown in Supplementary Table 17.

Statistics and reproducibility. Bar graphs are depicted as mean ± s.e.m. If not stated otherwise, statistical analysis for comparison of two groups was performed by unpaired *t*-test (two-tailed) and for more than two groups by ordinary one-way analysis of variance (ANOVA) (Tukey's post test) using GraphPad Prism software v.8.4.1. For immunofluorescence staining, coimmunoprecipitation, EMSA and western blot, representative images of at least two independent experiments with similar results have been provided.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequencing data (RNA-seq, ATAC-seq and ChIP–seq) generated for this work have been deposited at the Gene Expression Omnibus (GEO) (GSE131817). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[75] with the dataset identifier PXD018887. Moreover, gene signatures obtained from publicly available data deposited at GEO GSE81547 were used. We re-analyzed public TF ChIP–seq data and activating histone ChIP–seq data deposited at GEO GSE54471 and promoter associated ChIP–seq histone marks deposited at ArrayExpress E-MTAB-1086 for pancreatic progenitors and E-MTAB-1919 for islets. Source data are provided with this paper.

Code availability

For raw data processing off the instruments, code for two custom programs based on Picard tools (2.19.2) is available at https://github.com/DanieleBarreca/picard/ and https://broadinstitute.github.io/picard/. Further programs used for transcriptome analysis are described in Methods section.

References

58. Howson, J. M. et al. Genetic analysis of adult-onset autoimmune diabetes. *Diabetes* **60**, 2645–2653 (2011).
59. Zalloua, P. A. et al. WFS1 mutations are frequent monogenic causes of juvenile-onset diabetes mellitus in Lebanon. *Hum. Mol. Genet.* **17**, 4012–4021 (2008).
60. Rong, E. et al. Heteroplasmy detection of mitochondrial DNA A3243G mutation using quantitative real-time PCR assay based on TaqMan-MGB probes. *BioMed Res. Int.* **2018**, 1286480 (2018).
61. Hohwieler, M. et al. Human pluripotent stem cell-derived acinar/ductal organoids generate human pancreas upon orthotopic transplantation and allow disease modelling. *Gut* **66**, 473–486 (2017).
62. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
63. Ding, Q. et al. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**, 393–394 (2013).
64. Rezania, A. et al. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–1133 (2014).
65. Mahaddalkar, P. U. et al. Generation of pancreatic β cells from CD177$^+$ anterior definitive endoderm. *Nat. Biotechnol.* **38**, 1061–1072 (2020).
66. Salat, D., Liefke, R., Wiedenmann, J., Borggrefe, T. & Oswald, F. ETO, but not leukemogenic fusion protein AML1/ETO, augments RBP-Jκ/SHARP-mediated repression of Notch target genes. *Mol. Cell. Biol.* **28**, 3502–3512 (2008).
67. Wang, A. et al. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386–399 (2015).
68. Allhoff, M., Seré, K., F. Pires, J., Zenke, M. & G. Costa, I. Differential peak calling of ChIP-seq signals with replicates with THOR. *Nucleic Acids Res.* **44**, e153 (2016).
69. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896 (2007).
70. Zecha, J. et al. TMT labeling for the masses: a robust and cost-efficient, in-solution labeling approach. *Mol. Cell. Proteomics* **18**, 1468–1478 (2019).
71. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731 (2016).
72. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
73. Conforto, T. L., Steinhardt, G. F. IV & Waxman, D. J. Cross talk between GH-regulated transcription factors HNF6 and CUX2 in adult mouse liver. *Mol. Endocrinol.* **29**, 1286–1302 (2015).
74. Wacker, S. A. et al. RITA, a novel modulator of Notch signalling, acts via nuclear export of RBP-J. *EMBO J.* **30**, 43–56 (2011).
75. Vizcaíno, J. A. et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2012).

Acknowledgements

We thank R. Gowdru Bijegatte, K. Köhn, R. Köhntop, S. Schirmer, R. Rittelmann and J. Krüger from the Department of Internal Medicine I, Ulm University, Germany for their technical support. We thank S. Warth at the Core Facility Cytometry, Ulm University Medical Faculty, Germany for FACS-mediated cell sorting and M. Groth at Leibnitz Institute of Aging in Jena, Germany for performing RNA sequencing of our samples. We are also grateful to Kuhn Elektro-Technik GmbH for supporting our work. We thank S. Hays from the Neonatal Unit, Lyon, France for clinical management of the newborn patient and I. Plotton from the Biochemistry Laboratory, Lyon, France for her contribution to the hormonal evaluations. We thank C. Bonaldi and the Institut de Veille Sanitaire for providing data on IFG and T2D prevalence in the French population and the Centre National de Recherche en Génomique Humaine for providing access to their genomic platform. We thank the patients and their families for participating in this study. This project was funded by the Boehringer Ingelheim Ulm University BioCenter (BIU) as well as by the ANR-DFG collaborative research project (grant no. ANR-18-CE92-0031, DFG KL 2544/5-1) to C.J. and A.K. and via additional funding by the Deutsche Forschungsgemeinschaft (DFG) grant nos. KL 2544/6-1, KL 2544/7-1, 'Fokus-Förderung COVID-19' KL 2544/8-1 and KL 2544/1-2 to A.K.; by the Agence Nationale pour la Recherche (grant no. ANR-09-GENO-021), the European Foundation for the Study of Diabetes/JDRF/Novo Nordisk, the Assistance Publique-Hôpitaux de Paris Programme Hospitalier de Recherche Clinique (project DIAGENE) and France Génomique (project DIAPED) to C.J.; by a grant from the E: MED Consortia Fibromap funded by the German Ministry of Education and Science (BMBF) and by the DFG grant no. GE 2811/3-1 to I.G.C.; as well as grant no. SFB1074/A3, OS287/4-1 to F.O., NIH grants no. DK068471 and no. DK105541 as well as NIH T32 grant no. GM008666 to M. Sander and grant no. DFG-GrK1041, Centre of Excellence Metabolic Disorders Baden-Wuerttemberg, Germany as well as Ministry of Education, Singapore, MOE2018-T2-1-085 to B.O.B. Work in M. Hebrok's laboratory was supported by a grant from the NIH (grant no. DK105831). This work was also supported by the France Génomique National infrastructure, funded as part of the Investissements

d'Avenir program managed by the Agence Nationale pour la Recherche (grant no. ANR-10-INBS-09).

Author contributions

A.P., S.H., I.G.C. and V.S. contributed equally to this work. A.P., S.H., I.G.C. and V.S. acquired, analyzed and interpreted data, and drafted and revised the work. A.P. performed statistical, genetic and bioinformatics analysis of the human genetic part of the project. S.H. performed functional studies of PSCs and prepared samples for RNA-seq, ATAC-seq and mass spectrometry. I.G.C. performed and directed bioinformatics analysis of the project. V.S. performed sequencing and genotyping of diabetic patients and their families as well as of the German cohort. M.B., Z.L. and G.K. acquired data and performed data analysis. A.D., P.Z., H.N., E.S., T.K., M.W., C.B., R.O., J.-F.D., B.K. and C.D.-R. acquired data for the project. M.B., R.G., M. Hebrok and T.S. revised the manuscript. Specifically, M.B. and A.I. performed gene editing of hESCs and initial functional analysis. Z.L., G.K. and X.Z. did ChIP–seq; Z.L., G.K. and M. Schuster did ATAC-seq; and Z.L., G.K., M. Schuster and Q.L. did RNA-seq bioinformatics analysis. R.R. acquired data and performed substantial revision of the work. M. Hohwieler performed reprogramming of patient fibroblasts and iPSC analysis. A.D. performed bioinformatics analysis of WES data and P.Z. identified and clinically characterized the Lebanese patient and his family. S.L. interpreted data and revised the work. M. Sander interpreted data, provided materials and revised the work. J.K. acquired and analyzed mass spectrometry data. J.R.B. generated reporter ESC lines. A.W. acquired and R.G. provided ChIP–seq data. K.G. and J.C. interpreted genetics data and G.N. provided RNA from differentiated Mel1 hESCs. B.O.B., F.O., M.N., C.J. and A.K. were responsible for acquisition and analysis of data, and drafting and revision of the work. Also, B.O.B. directed studies regarding German patient cohorts and interpreted genetics data, F.O. expressed and analyzed TFs and ONECUT1 variants and M.N. identified and clinically characterized Patient 1 and his extended family and interpreted the human genetic and clinical data. In addition, C.J. and A.K. designed the work, interpreted data and drafted the manuscript with input from all authors. C.J. directed the genetic part of the project and performed human genetic analyses. A.K. directed the functional studies of the project.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-021-01502-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-021-01502-7.

Correspondence and requests for materials should be addressed to Cécile Julier or Alexander Kleger.

Peer review information *Nature Medicine* thanks Timo Otonkoski, Anna Gloyn, Katherine Owen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.