



**HAL**  
open science

## Pathogenic variants in actionable MODY genes are associated with type 2 diabetes

Amélie Bonnefond, Mathilde Boissel, Alexandre Bolze, Emmanuelle Durand, Bénédicte Toussaint, Emmanuel Vaillant, Stefan Gaget, Franck De Graeve, Aurélie Dechaume, Frédéric Allegaert, et al.

### ► To cite this version:

Amélie Bonnefond, Mathilde Boissel, Alexandre Bolze, Emmanuelle Durand, Bénédicte Toussaint, et al.. Pathogenic variants in actionable MODY genes are associated with type 2 diabetes. *Nature Metabolism*, 2020, 2 (10), pp.1126-1134. 10.1038/s42255-020-00294-3 . hal-03098331

**HAL Id: hal-03098331**

**<https://hal.science/hal-03098331v1>**

Submitted on 5 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Pathogenic variants in actionable MODY genes associate with type 2 diabetes

Amélie Bonnefond<sup>1,2</sup>, Mathilde Boissel<sup>1,16</sup>, Alexandre Bolze<sup>3,16</sup>, Emmanuelle Durand<sup>1,16</sup>,  
Bénédicte Toussaint<sup>1</sup>, Emmanuel Vaillant<sup>1</sup>, Stefan Gaget<sup>1</sup>, Franck De Graeve<sup>1</sup>, Aurélie  
Dechaume<sup>1</sup>, Frédéric Allegaert<sup>1</sup>, David Le Guilcher<sup>1</sup>, Loïc Yengo<sup>1,4</sup>, Véronique Dhennin<sup>1</sup>,  
Jean-Michel Borys<sup>5</sup>, James T. Lu<sup>3</sup>, Elizabeth T. Cirulli<sup>3</sup>, Gai Elhanan<sup>6,7</sup>, Ronan Roussel<sup>8,9,10</sup>,  
Beverley Balkau<sup>11,12</sup>, Michel Marre<sup>9,13</sup>, Sylvia Franc<sup>14,15</sup>, Guillaume Charpentier<sup>14</sup>, Martine  
Vaxillaire<sup>1</sup>, Mickaël Canouil<sup>1</sup>, Nicole L. Washington<sup>3</sup>, Joseph J. Grzymalski<sup>6,7</sup>, Philippe  
Froguel<sup>1,2</sup>

<sup>1</sup>Inserm UMR1283, CNRS UMR8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille University Hospital, Lille, 59000, France; <sup>2</sup>Department of Metabolism, Imperial College London, London, W12 0NN, United Kingdom; <sup>3</sup>Helix, San Mateo, California, 94070, USA; <sup>4</sup>Present address: Institute for Molecular Bioscience, the University of Queensland, 4072, Australia; <sup>5</sup>Fleurbaix Laventie Association, Laventie, 62840, France; <sup>6</sup>Desert Research Institute, Reno, Nevada, 89512, USA; <sup>7</sup>Renown Institute of Health Innovation, Reno, Nevada, 89512, USA; <sup>8</sup>Department of Diabetology Endocrinology Nutrition, Hôpital Bichat, DHU FIRE, Assistance Publique Hôpitaux de Paris, Paris, 75018, France; <sup>9</sup>Inserm U1138, Centre de Recherche des Cordeliers, Paris, 75006, France; <sup>10</sup>UFR de Médecine, University Paris Diderot, Sorbonne Paris Cité, Paris, 75005, France; <sup>11</sup>Inserm U1018, Institut Gustave Roussy, Center for Research in Epidemiology and Population Health, Villejuif, 94800, France; <sup>12</sup>University Paris-Saclay, University Paris-Sud, Villejuif, 94800, France; <sup>13</sup>CMC Ambroise Paré, Neuilly-sur-Seine, 92200, France; <sup>14</sup>CERITD (*Centre d'Étude et de Recherche pour l'Intensification du Traitement du Diabète*), Evry, 91000, France; <sup>15</sup>Department of Diabetes, Sud-Francilien Hospital, University Paris-Sud, Orsay, Corbeil-Essonnes, 91100, France; <sup>16</sup>These authors equally contributed to the manuscript

### **Correspondence to:**

Dr Amélie Bonnefond (amelie.bonnefond@cnr.fr) or Prof Philippe Froguel (p.froguel@imperial.ac.uk)

EGID - UMR 1283/8199

Pôle Recherche - 1er étage Aile Ouest

1 place de Verdun

59045 LILLE CEDEX

France

Genome-wide association studies have identified 240 independent loci associated with type 2 diabetes (T2D) risk, but their knowledge has not advanced precision medicine. In contrast, the genetic diagnosis of monogenic forms of diabetes (including maturity-onset diabetes of the young [MODY]) are textbook cases of genomic medicine. Recent studies trying to bridge the gap between monogenic diabetes and T2D have been inconclusive. Here we show a significant burden of pathogenic variants in genes linked with monogenic diabetes among patients with common T2D, particularly in actionable MODY genes (*i.e.* implying dramatic changes in diabetes care). Among 74,629 individuals, we show that this burden is probably driven by the pathogenic variants found in *GCK*, and to a lesser extent in *HNF4A*, *KCNJ11*, *HNF1B* and *ABCC8*. The carriers with T2D are leaner, which evidences a functional metabolic effect of these mutations. Pathogenic variants in actionable MODY genes are more frequent than previously expected in common T2D. These results open avenues for future interventions assessing the clinical interest of these pathogenic mutations in precision medicine.

## Introduction

Diabetes affects 420 million patients worldwide and is forecasted to increase to 700 million by 2030. Diabetes became in 2016 the sixth leading cause of mortality<sup>1,2</sup>. Among 155 diseases, diabetes care represents the largest and fastest growing portion of health care spending in the USA<sup>3</sup>. Type 2 diabetes (T2D) represents more than 90% of all diabetes presentations. It is a complex polygenic disorder, with an estimated heritability of 72%<sup>4</sup>. Genome-wide association studies (GWAS) have identified 240 frequent loci associated with T2D risk, but they only explain 20% of T2D heritability<sup>5</sup>. Further, the translation of these discoveries into advances in precision medicine has been modest so far<sup>6</sup>. In contrast, the genetic interrogation of monogenic diabetes has yielded insights into key regulators of insulin secretion, leading to actionable examples of genomic medicine<sup>6</sup>.

Monogenic diabetes comprises a broad spectrum of conditions including neonatal diabetes, maturity-onset diabetes of the young (MODY) and diabetes-associated syndromes<sup>7</sup>. Compared to common T2D, monogenic diabetes is much rarer, usually more severe and develops earlier in life<sup>7</sup>. Patients who carry a rare pathogenic activating mutation in *KCNJ11* or *ABCC8*, or who are deficient for *HNF1A* or *HNF4A*, have high sensitivity to oral sulfonylureas, and can efficiently stop insulin injections<sup>8-12</sup>. Furthermore, pharmacological treatments do not improve blood glucose levels of *GCK*-deficient MODY patients<sup>13</sup>, who remarkably do not develop typical microvascular renal, ocular or heart pathologies. Therefore, these patients usually do not require any hypoglycemic treatment. Actionable genes linked with monogenic diabetes also include *HNF1B*, *GATA4* and *GATA6* as they are associated with necessary additional clinical investigations focused on the heart or kidney function<sup>14-16</sup>. In this context, the genetic diagnosis of young patients with a suspicion of monogenic diabetes has proven cost-effective and improved the quality of life of patients<sup>17,18</sup>.

Recent large-scale association studies have started to bridge the gap between monogenic diabetes and common T2D, as a way to develop precision medicine in diabetes. Notably, a large meta-analysis based on exome sequencing of 6,504 cases with T2D and 6,436 non-diabetic controls reported a significant signal of association between T2D risk and the aggregation of rare deleterious variants (with minor allele frequency [MAF] below 1%) in 12 MODY genes. However, gene-centric analyses highlighted that this association was not explained by the most frequently mutated genes in monogenic diabetes (*i.e.* *GCK* and *HNF1A*) and age of T2D diagnosis was similar between variant carriers and non-carriers<sup>19</sup>. A last large meta-analysis comprising 20,791 cases with T2D and 24,440 non-diabetic controls confirmed the signal of association between T2D risk and the aggregation of rare deleterious variants in MODY genes, with nominal associations in *PDX1* and *GCK* although the results were not robust according to the models the authors have tested<sup>20</sup>. In both studies, gene-centric analyses were rather disappointing, maybe due to the fact that variants deleteriousness was only assessed through *in silico* software programs and because the authors analyzed unselected (*i.e.* nondiabetic) controls<sup>19,20</sup>.

## Results

### *Burden of rare coding variants in genes linked with monogenic diabetes*

Here, 33 genes linked with monogenic diabetes (**Supplementary Table 1**) were sequenced in 6,348 samples from the French RaDiO study, including 2,178 cases with common T2D and 4,170 controls known to present with normal glucose levels after age 40 (**Supplementary Table 2**). Some genes including *GATA6*, *MNX1* and *PTF1A* were poorly sequenced compared to others despite a high mean depth of coverage of the target (160×). The same sequencing issues of these genes were actually reported in the genome aggregation database (gnomAD) browser that provides 125,748 exome sequences of high quality (**Supplementary Fig. 1**)<sup>21</sup>. Therefore, we applied seven additional quality control (QC) steps. The mean rate of variants and samples filtered out by each QC step across the 33 monogenic diabetes genes and per gene is reported in **Extended Data Fig. 1** and **Supplementary Fig. 2**, respectively. Following QC, 1,408 rare coding variants (including 430 novel variants [30.5%] that were not reported in gnomAD and 983 singletons [69.8%]; **Supplementary Table 3**) were accurately detected across the 33 genes. The burden of rare coding variants of potential interest was significantly higher in patients with T2D (53.8% of carriers among cases *versus* 45.7% of carriers among controls;  $P = 6.38 \times 10^{-3}$  [ $P_{\tau} = 0.0235$ ]; **Supplementary Table 4**), using the mixed effects score test (MiST)<sup>22</sup> adjusted for age, sex, body mass index (BMI) and ancestry. This significant association was partially explained by the variants found in actionable MODY genes listed in **Supplementary Table 1** ( $P = 7.23 \times 10^{-3}$  [ $P_{\tau} = 0.0304$ ]), in particular *GCK* ( $P = 2.62 \times 10^{-3}$  [ $P_{\tau} = 0.0479$ ] with an odds ratio [OR] of 3.21, 95% confidence interval [CI]: 1.36–7.82; **Supplementary Table 4**). In contrast, we found that the burden of rare coding variants of potential interest in *SLC2A2* was higher in normal-glucose controls ( $P = 9.66 \times 10^{-4}$  with an OR of 0.558, 95% CI: 0.387–0.789; **Supplementary Table 4**).

### ***Burden of pathogenic or likely pathogenic (P/LP) variants in actionable MODY genes***

We used the criteria of the American College of Medical Genetics and Genomics (ACMG) to assess the pathogenicity of the 1,408 detected rare coding variants. We identified 167 P/LP variants (**Supplementary Table 3**). The burden of these P/LP variants was markedly higher in patients with T2D (5.62% of carriers among cases *versus* 2.66% of carriers among controls;  $P = 1.25 \times 10^{-4}$ ; **Supplementary Table 5**), using the MiST method adjusted for age, sex, BMI and ancestry. This significant enrichment was mostly explained by the P/LP variants found in actionable MODY genes (3.49% of carriers among cases *versus* 1.39% of carriers among controls;  $P = 3.91 \times 10^{-6}$ ; **Table 1**), particularly *GCK* (0.597% of carriers among cases;  $P = 3.13 \times 10^{-4}$  [ $P_{\tau} = 0.0496$ ] with an OR of 7.08, 95% CI: 2.23–27.3; **Fig. 1a, Supplementary Table 5**), and *HNF1A* (0.872% of carriers among cases;  $P = 2.55 \times 10^{-4}$  with an OR of 7.40, 95% CI: 2.60–26.6; **Fig. 1a, Supplementary Table 5**). In contrast, P/LP variants in *GATA6* were surprisingly associated with decreased T2D risk ( $P = 7.22 \times 10^{-8}$  with an OR of  $8.57 \times 10^{-3}$ , 95% CI:  $7.59 \times 10^{-5}$ –0.372), although the burden effect (based on two variants only) was very heterogeneous ( $P_{\tau} = 1.83 \times 10^{-7}$ ) (**Fig. 1a, Supplementary Table 5**). Importantly, when we analyzed rare variants in actionable MODY genes which were null (e.g. nonsense, frameshift, canonical  $\pm 1$  or 2 splice sites, initiation codon) or deleterious according to *in silico* software programs (*i.e.* PVS1 and PP3 ACMG criteria, respectively), these rare variants were not significantly associated with T2D risk (**Supplementary Table 6**). We only found a significant association between rare deleterious variants in *GCK* and T2D risk which was in fact due to the significantly heterogeneous effect of the cluster ( $P = 0.0214$  [ $P_{\tau} = 0.0221$ ] with an OR of 2.21, 95% CI: 0.748–6.75; **Supplementary Table 6**). This result highlights the potential of ACMG criteria to improve the dissection of rare variants in a common disease like T2D.

In order to confirm these results, we subsequently used whole-exome sequencing data from the UK Biobank (including 2,151 cases with T2D and 33,718 controls known to present with normal glucose levels after age 40; **Supplementary Table 7**), the Healthy Nevada Project (HNP; including 1,518 cases with T2D and 4,171 controls known to present with normal glucose levels after age 40; **Supplementary Table 8**) and the Accelerating Medicines Partnership (AMP) T2D knowledge portal (including 19,852 cases with T2D and 6,871 controls). In these three case-control studies, we interpreted each rare coding variant detected in the actionable MODY genes (**Supplementary Tables 9, 10 and 11**), and we identified 227 additional P/LP variants (**Supplementary Table 12**). The location of P/LP variants that were identified in the four studies (RaDiO, UK Biobank, HNP study and AMP T2D knowledge portal) are shown per gene and coding exon in **Supplementary Fig. 3**. Among all P/LP variants, 43% were novel (*i.e.* not listed in gnomAD v2.1.1) and 22% were null. The burden of the P/LP variants was significantly higher in the patients with T2D from the UK Biobank (2.19% of carriers among cases *versus* 1.24% of carriers among controls;  $P = 1.39 \times 10^{-5}$ ; **Table 1**) and tended to be higher in the HNP study (2.31% of carriers among cases *versus* 1.56% of carriers among controls;  $P = 0.0915$ ; **Table 1**), using the MiST method adjusted for age, sex, BMI and ancestry. In the AMP T2D knowledge portal, the multi-gene analysis was not available, yet the rate of carriers among cases (1.81%) was higher than the rate of carriers among controls (0.815%) (**Table 1**). In the meta-analysis combining data from the RaDiO study, UK Biobank, HNP study and AMP T2D knowledge portal via the generic inverse variance method ( $N = 74,629$  including 25,699 cases and 48,930 controls), we found that the enrichment of P/LP variants in actionable MODY genes among cases was probably driven by the variants found in *GCK* (0.284% of carriers among cases;  $P = 1.41 \times 10^{-9}$  with an OR of 4.52, 95% CI: 2.77–7.37), and to a lesser extent in *HNF4A* (0.148% of carriers among cases;  $P = 8.17 \times 10^{-3}$  with an OR of 2.09, 95% CI: 1.21–3.60), *KCNJ11* (0.152% of carriers



among cases;  $P = 0.0102$  with an OR of 2.38, 95% CI: 1.23–4.62), *HNF1B* (0.0856% of carriers among cases;  $P = 0.0296$  with an OR of 2.08, 95% CI: 1.07–4.02) and *ABCC8* (0.603% of carriers among cases;  $P = 0.0417$  with an OR of 1.35, 95% CI: 1.01–1.81) (**Table 1, Fig. 1e**). In *HNF1A*, we found a significantly heterogeneous effect of the burden on T2D risk ( $P_{\tau^2} = 5.44 \times 10^{-4}$ ; **Table 1**): indeed, we found a strong effect of the burden on increased T2D risk in both RaDiO ( $P = 2.55 \times 10^{-4}$ ; **Table 1, Fig. 1a**) and AMP T2D knowledge portal ( $P = 1.57 \times 10^{-5}$ ; **Table 1, Fig. 1d**), but no significant association was observed in the UK Biobank ( $P = 0.109$ ; **Table 1, Fig. 1b**) and in the HNP study ( $P = 0.628$ ; **Table 1, Fig. 1c**). This might be due to the lower rate of carriers among cases in these two cohorts (**Table 1**) leading to poorer statistical power.

### ***Phenotypes of the carriers of a P/LP mutation with T2D***

Among the patients with T2D from the RaDiO, UK Biobank and HNP studies, the carriers of a P/LP variant in actionable MODY genes were significantly leaner than non-carriers ( $\beta = -2.04 \pm 0.466$  kg/m<sup>2</sup>;  $P = 1.20 \times 10^{-5}$ ; **Table 2**), using a linear regression adjusted for age, sex and ancestry. This effect was probably driven by the carriers of a P/LP variant in *GCK* ( $\beta = -3.81 \pm 1.06$  kg/m<sup>2</sup>;  $P = 3.08 \times 10^{-4}$ ), *HNF4A* ( $\beta = -4.29 \pm 1.56$  kg/m<sup>2</sup>;  $P = 5.87 \times 10^{-3}$ ) and *HNF1B* ( $\beta = -2.46 \pm 1.25$  kg/m<sup>2</sup>;  $P = 0.0491$ ) (**Table 2**). In the RaDiO study (*i.e.* the only study where age of diagnosis and antidiabetic treatment classes were available for all patients without left-censoring), the association between P/LP variants and decreased BMI had the same magnitude of significance when adjusted for age of T2D diagnosis, fasting glucose levels, or antidiabetic classes (*i.e.* insulin only, insulin associated with other drugs, metformin only, other drugs/combinations or no drug) in addition to age, sex and ancestry (data not shown). In addition, the carriers of a P/LP variant in actionable MODY genes developed T2D earlier when compared to non-carriers (mean age of T2D diagnosis in carriers:  $44.5 \pm 12.1$

years; mean age of T2D diagnosis in non-carriers:  $48.7 \pm 11.0$  years;  $\beta = -2.64 \pm 0.959$  years;  $P = 5.99 \times 10^{-3}$ ; **Extended Data Fig. 2**) using a linear regression adjusted for age, sex, BMI and ancestry, but no carrier developed T2D under the age of 25 (*i.e.* the main criterion for MODY). Furthermore, the carriers of a P/LP variant in actionable MODY genes were respectively three times and twice as likely to be treated by insulin only and insulin plus other antidiabetic drugs than the non-carriers ( $P = 7.67 \times 10^{-5}$  with an OR of 3.09, 95% CI: 1.77–5.42 and  $P = 1.52 \times 10^{-3}$  with an OR of 2.22, 95% CI: 1.36–3.64, respectively; **Extended Data Fig. 3**), using a logistic regression adjusted for age, sex, BMI and ancestry. These results were partly driven by the P/LP variants found in *HNF1A*, *KCNJ11* and *HNF1B* (**Extended Data Fig. 3**). In contrast, the carriers of a P/LP variant were less likely to be treated by metformin ( $P = 0.0316$  with an OR of 0.587, 95% CI: 0.361–0.954; **Extended Data Fig.4**). Sulfonylureas intake between carriers and non-carriers was not statistically different (**Supplementary Table 13**). Importantly, the family history of T2D was not statistically different between carriers and non-carriers (**Extended Data Fig. 5**), using a logistic regression adjusted for age, sex, BMI and ancestry. Therefore, without the next-generation sequencing (NGS) analysis of all the patients, it would have been impossible to distinguish the carriers of a P/LP variant in actionable MODY genes from the non-carriers. The phenotypic characteristics of carriers and non-carriers are recapitulated in **Supplementary Table 14**.

Finally, the prevalence of carriers of a P/LP variant in actionable MODY genes tended to be higher among Europeans compared to non-Europeans in the RaDiO, UK Biobank and HNP studies, but this result was not significant (**Supplementary Table 15**).



## Discussion

This NGS-based investigation combining 74,629 samples demonstrated that the burden of rare P/LP variants in actionable MODY genes was significantly higher in patients presenting with common T2D compared to normal glucose controls, despite a heterogeneity of entry criteria in each population study (**Extended Data Fig. 6**). Our results suggest a direct etiological role of these P/LP variants in the development of common T2D. Among unselected patients with T2D, more than 2% carried a P/LP variant in actionable MODY genes, which was unexpected. Compared to previous NGS-based association studies for T2D risk, we report a much stronger mutational burden in *GCK*, and to a lesser extent in *HNF4A*, *KCNJ11*, *HNF1B* and *ABCC8*<sup>19,20,23</sup>. We believe that these differences were in part due to our highly stringent QC and curation of NGS data in the RaDiO and HNP studies, but also to the time-consuming use of the ACMG criteria to assess the pathogenicity of each detected variant. We have shown that these ACMG criteria can actually be instrumental in the dissection of a common polygenic disease like T2D, beyond rare genetic disorders. Furthermore, sample size and control definition (*i.e.* normal glucose participants after age 40 only in our study *versus* nondiabetic participants including prediabetic individuals in others) could explain the increased significance of results, as shown by our bootstrap stimulation study in RaDiO (**Extended Data Fig. 7**). Indeed, among 273 participants with prediabetes, 4.03% carried a P/LP variant in actionable MODY genes, while the rates of carriers were 3.49% among cases with overt T2D and 1.39% among normal glucose controls (**Table 1**).

In the meta-analysis, the burden of P/LP variants in actionable MODY genes in cases was firstly explained by the variants detected in *GCK* (encoding the glycolysis key enzyme) that was the first identified MODY gene<sup>24</sup>. While MODY due to *GCK* deficiency is thought to be benign, it is remarkable that in RaDiO and HNP studies 83.3% and 85.7% of T2D patients deficient for *GCK* respectively, received oral hypoglycemic agents and/or insulin. Our results

therefore open avenues for future intervention studies assessing the impact of the detection of these P/LP variants on precision medicine of diabetes in the carriers, for example a possible stop of expensive and demanding treatment in T2D patients deficient for *GCK*.

Importantly, the diabetic carriers of a P/LP variant in actionable MODY genes were significantly leaner than the diabetic non-carriers and developed T2D earlier, which evidences a true functional metabolic effect of these variants that we defined P/LP according to the ACMG criteria. This was in line with a previous study that showed an association between the burden of rare deleterious variants in MODY genes and lower BMI<sup>20</sup>. Despite the significant metabolic effects of P/LP variants, no carrier from the RaDiO study has developed T2D under the age of 25, emphasizing that MODY is only the top of the iceberg of monogenic diabetes forms. Importantly, we did not find any difference in family history of T2D between carriers and non-carriers. It may be not surprising knowing that common T2D has more than 70% heritability<sup>4</sup>. Therefore, patients with T2D who carried these P/LP variants could not be clinically distinguished from patients with purely polygenic T2D.

In line with previous studies<sup>19,20,23,25</sup>, we found a substantial number of normal glucose controls carrying P/LP variants in monogenic diabetes genes. We and others previously showed that pathogenic mutations in monogenic diabetes genes were not fully penetrant, even within diabetic families with inherited mutations<sup>26-29</sup>. The impact of coding deleterious variants on glucose values (which define diabetes) can be markedly modified by cis-regulatory variation<sup>30</sup>, suggesting an impact of frequent SNPs in the functional effect of rare variants. Genome-wide polygenic scores have recently showed that part of the general population has a very low risk of T2D development<sup>31</sup>, which might contribute to the lack of full penetrance of pathogenic mutations of actionable genes. Moreover, it is noteworthy that deleterious mutations in several monogenic diabetes genes (e.g. *ABCC8*, *KCNJ11*, *GCK*, *HNF1A*) can actually cause hyperinsulinism and hypoglycemia for a long time (before leading

to the opposite phenotype)<sup>32</sup>, which could have increased the number of normal glucose controls carrying P/LP variants in the present study.

In conclusion, by using stringent rules utilized in the molecular diagnosis of rare genetic disorders (including NGS-related QCs and ACMG criteria), we report a significant burden of pathogenic mutations in genes linked with monogenic diabetes among patients with common T2D who could not be identified before without systematic NGS. These data strongly support the hypothesis of a true genetic continuum of monogenic and polygenic forms of non-autoimmune diabetes. The burden was mostly driven by pathogenic variants which were detected in actionable MODY genes, especially in *GCK*, *KCNJ11* and *HNF4A* genes opening avenues for evidence-based precision medicine.

## Methods

**Study participants:** The study design is recapitulated in **Extended Data Fig. 6**. In the RaDiO (*Rare variants involved in Diabetes and Obesity*) study, we investigated 6,348 blood DNA samples (including 2,178 cases with T2D and 4,170 controls) from: 1/ the D.E.S.I.R. 9-year prospective study ( $N = 3,369$ ) including middle-aged men and women from western France<sup>33</sup>; 2/ the Department of Diabetes of the Corbeil-Essonnes Hospital (Corbeil-Essonnes, France;  $N = 1,627$ )<sup>34</sup>; 3/ the CNRS UMR8199 study (Lille, France;  $N = 1,196$ ) including participants recruited and followed-up either by the UMR 8199 unit, by the Department of Nutrition of Hotel-Dieu Hospital (Paris, France), or by the Centre d'Etude du Polymorphisme Humain (CEPH, Saint-Louis hospital, Paris, France)<sup>34,35</sup>; 4/ the French Fleurbaix-Laventie Ville Santé study ( $N = 156$ )<sup>36</sup>. Patients with T2D had fasting glucose  $\geq 7.0$  mmol/L and/or used treatment of hyperglycemia, and were negative for islet autoantibodies and insulin autoantibodies. Control subjects had fasting glucose  $< 6.1$  mmol/L and did not use treatment of hyperglycemia. Importantly, we only included controls known to present with fasting glucose

<6.1 mmol/L after age 40. Family history of T2D was assessed among the parents, siblings and children of each participant with T2D, using the same definition of T2D.

In the UK Biobank (*i.e.* a national and international health resource established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, Scottish Government and the Northwest Regional Development Agency), we analyzed 35,869 individuals including 2,151 cases with T2D and 33,718 controls, with available exome sequencing data<sup>37,38</sup>. This research is part of the research Application Number 40436. The participants were recruited by the UK Biobank from across the United Kingdom between March 13, 2006, and October 1, 2010. To define cases with T2D and controls for this study, we used International Classification of Diseases (ICD)-10 codes, lab results and other fields from the UK Biobank data showcase.

Definition of cases with T2D in the UK Biobank:

- Inclusion criteria:
  - At least two ICD-10 codes starting with E11
  - OR
  - One ICD-10 code starting with E11, AND one of the following codes: diabetes diagnosed by doctor (field #2443), or use of insulin (fields #6153 or #6177 code[3]), or glucose  $\geq 7$  mmol/L (field #30740).
- Exclusion criteria:
  - At least one ICD-10 code starting with E10 (type 1 diabetes mellitus)
  - OR
  - At least one ICD-10 code starting with O24 (diabetes mellitus in pregnancy).

Definition of controls in UK Biobank:

- Inclusion criteria:
  - Blood glucose lower than 6.1 mmol/l (field #30740 first instance) after age 40

- Exclusion criteria:
  - At least one ICD-10 code starting with E10, E11, R73, or O24
  - OR
  - Diabetes diagnosed by doctor (field #2443)
  - OR
  - Diabetes self-reported (field #20002 codes [1220], [1221], [1222], [1223])
  - OR
  - Use of insulin (field #6153 or #6177 code [3])

In the Healthy Nevada Project (HNP), we analyzed 5,689 individuals including 1,518 cases with T2D and 4,171 controls, with available exome sequencing data<sup>39</sup>. HNP is an all-comers human subject research study on health determinants in Northern Nevada with specific recruitment foci in rural and socio-economically depressed areas in Northern Nevada (USA).

Definition of cases with T2D in HNP:

- Inclusion criteria:
  - At least two ICD-10 codes starting with E11
  - OR
  - One ICD-10 code starting with E11 AND one of the following: at least one prescription order for an antidiabetic medication, or glucose test  $\geq 7$  mmol/l
- Exclusion criteria:
  - At least one ICD-10 code starting with E10, or O24

Definition of controls in HNP:

- Inclusion criteria:
  - Median blood glucose results taken after age 40 are lower than 6.1 mmol/l
- Exclusion criteria:
  - At least one ICD-10 code starting with E10, E11, R73, O24



OR

- At least prescription order for an antidiabetic medication

OR

- One glucose test  $\geq 7$  mmol/l at any age

OR

- Less than four records in the electronic medical records available to the study

All blood or saliva DNA samples used were collected with appropriate informed consent consistent with their use in the present study. All cohort studies followed ethical principles defined in the Helsinki declaration (revised in 1996), and they were approved by local ethical committees from Corbeil-Essonnes hospital (France), Comité Consultatif de Protection des Personnes se prêtant à des Recherches Biomédicales (CPPRB) of Lille - Lille Hospital (Lille, France), Hotel-Dieu hospital (France), Bicêtre hospital (France). The HNP study was reviewed and approved by the University of Nevada, Reno Institutional Review Board (IRB, project 956068-12). This research has been also conducted using the UK Biobank Resource under Application Number 40436.

***Target enrichment preparation, NGS and bioinformatics analyses:*** In the RaDiO study, target enrichment was performed according to the manufacturer's protocol (NimbleGen SeqCap EZ Choice XL) for Illumina sequencing on the HiSeq 4000 system. Briefly, 1  $\mu$ g DNA was fragmented through sonication (Covaris E220 Focused-ultrasonicator). The fragmented DNA samples were end-repaired and ligated to adapters using the KAPA HTP Library Preparation Kits, on the Hamilton Microlab STARlet automate. These samples were subsequently amplified by polymerase chain reaction using primers complementary to the adapters. After size selection and sample quantification (Perkin Elmer LabChip GX), 24 samples were combined in a single pool of at least 1  $\mu$ g, and hybridized to the biotin-labeled SeqCap EZ probe pool. After 72 hours at 47 °C, the captures were purified using the SeqCap

Hybridization and Wash Kit on the Agilent Bravo Automated Liquid Handling Platform. Captures were subsequently amplified using the KAPA HiFi HotStart ReadyMix and quantified by both Perkin Elmer LabChip GX and Thermo Fisher Scientific Qubit fluorometric quantitation assays. Then, the samples were sequenced on the Illumina HiSeq 4000 system (with a throughput of one pool including 24 samples per lane), using a paired-end 2×150 bp protocol. The demultiplexing of sequence data (from BCL files generated by Illumina sequencing systems to standard FASTQ file formats) was performed using bcl2fastq Conversion Software (Illumina; version 2.17). Subsequently, sequence reads from FASTQ files were mapped to the human genome (hg19/GRCh37) using Burrows-Wheeler Aligner (version 0.7.13)<sup>40</sup>. The variant calling was performed using Genome Analysis ToolKit (GATK; version 3.3)<sup>41</sup>. Across the samples included in the study, the mean depth of coverage of the target was 160×; 98.6% of bases were covered with at least 8 reads and 96.2% of bases were covered with at least 20 reads. Only variants with a coverage higher than 8 reads were kept for further analyses. The annotation of variants was performed using the Ensembl Perl Application Program Interfaces (version 75) and custom Perl scripts to include data from both dbSNP (version 135) and dbNSFP (version 3.0) databases<sup>42,43</sup>.

In the UK Biobank, we utilized the functionally equivalent (FE) version of the UK Biobank PLINK formatted exome files (field #23160)<sup>38,44</sup>. It has been reported that this version is under-calling variants, specifically in regions of the genome with alternate contigs (see note from the UK Biobank: <https://www.ukbiobank.ac.uk/wp-content/uploads/2019/12/Description-of-the-alt-aware-issue-with-UKB-50k-WES-FE-data.pdf>). The analysis of the bed file of the regions impacted (available in the UK Biobank data showcase Resource field #1911) indicated that all coding exons of *HNF1B* were impacted by this issue, resulting in zero variant called mapping to *HNF1B* in the UK Biobank exome dataset. The seven other genes (*i.e.* *GCK*, *HNF1A*, *HNF4A*, *KCNJ11*, *ABCC8*, *GATA4*

and *GATA6*) had no overlap with the problematic regions. Variant annotation was performed with VEP 95.3<sup>45</sup>. Genotype processing and filtering was performed in Hail.

For the HNP, sequencing for all participants was performed at Helix's CLIA (#05D2117342) and CAP (#9382893) accredited facility in San Diego, CA. The Exome+ is a proprietary (Helix) assay that targets ~19,000 genes and ~300,000 non-coding SNPs. The bioinformatics pipeline uses well-established algorithms for alignment and quality control metrics. Helix utilizes a customized version of Sentieon's optimized variant calling software<sup>46</sup>. All returned variants meet Helix's validation criteria for analytical performance required for clinical return. Performance specification can be found in Helix's Exome+ Performance White Paper<sup>47</sup>. Across the samples included in the study, the mean depth of coverage of the target was 101×, and 99.5% of bases were covered with at least 20 reads. Only variants with a coverage  $\geq 20$  reads were kept for further analyses. Variant annotation was performed with VEP 95.3<sup>45</sup>. Genotype processing and filtering was performed in Hail.

***Design of variant analysis:*** We only analyzed rare variants with a MAF below 1% in the present study. We investigated rare coding variants of potential interest (*i.e.* initiator codon variants, stop retained variants, stop lost variants, stop gained variants, missense variants, splice donor variants, splice acceptor variants, frameshift variants and inframe variants) and, among them, P/LP variants according to the standards and guidelines of ACMG for the interpretation of rare variants (**Supplementary Tables 16 & 17**)<sup>48</sup>. For the moderate pathogenic criterion PM2, we used gnomAD browser (v2.1.1)<sup>21</sup>. For the supporting pathogenic criterion PP3, we used PolyPhen-2 (HumDiv), SIFT and Mutation Taster<sup>49–51</sup>. None of the P/LP variants had a MAF higher than 0.1% in gnomAD (v2.1.1), in accordance with the viewpoint of Ellard and colleagues<sup>52</sup>.

***QC of aligned sequence reads:*** Several QCs were performed before association analyses.

In the RaDiO study:

- QC1 based on the analysis subtype: variants were filtered out according to their frequency (variants with a MAF  $\geq$  1% were excluded) and to the analysis subtype.
- QC2 based on QUAL score mean: all variants with a QUAL score mean  $<$ 50 (from vcf files) were excluded.
- QC3 based on homopolymer repeats: all variants in genomic areas with more than five homopolymer repeats were excluded.
- QC4 based on sample call rate: any sample with more than 5% of missing genotypes per cluster was excluded from the analysis.
- QC5 based on variant call rate: any variant with more than 5% of missing genotypes was excluded from the analysis. Missing genotypes in the remaining variants were replaced by 0.
- QC6 based on MAF and variance: any variant with null variance or null MAF (because of previous QC steps) were excluded from the analysis.
- QC7 based on manual curation: among borderline variants with a QUAL score between 50 and 100, some variants were excluded after an inspection of BAM files.

In the HNP study:

- Sample QC. First, the overall quality of the Exome+ sequence is assessed based on metrics measuring contamination (FREEMIX, % bacterial contamination etc.), as well as overall performance (total yield, % of target bases covered by at least 20 reads etc.). Samples from exomes who do not meet the QC criteria were either requeued or re-collected.
- Definition of a reportable range and creation of a blacklist of genome positions causing systematic sequencing error. There was no overlap between the blacklist and the coding exons of the eight actionable MODY genes.

- Variant QC and analytical range. We validated the performance of the Exome+ assay on the following variants: single-nucleotide variants, small insertions and deletions up to 20 bases, and copy number variants (CNVs) of different sizes. Moreover, variants adjacent to homopolymer runs >10bp, dinucleotide repeats of >12bp, or trinucleotide repeats of >21bp were not interpreted.
- Manual review of BAM files was available and occasionally done for complex variants.

**Statistical analyses:** All statistical analyses were performed using the R software (version 3.6.3) or IBM SPSS statistics (version 22.0.0.0).

In the RaDiO study, the UK Biobank and the HNP study, we performed a case-control study for T2D risk adjusted for age, sex, BMI and the first five genetic components (*i.e.* principal components [PCs]). In RaDiO, the genetic components were computed using a principal component analysis (PCA) with the R package `flashpcaR`<sup>53</sup>, on 15,020 single nucleotide polymorphisms (SNPs; with a MAF >5%) available in both RaDiO study and in the 1,000 Genomes project<sup>54</sup>. The first two genetic components were actually sufficient to infer the genetic ancestry stratification (*i.e.* population and super population from 1,000 Genomes project). In the UK Biobank, the first 10 genetic PCs were taken from the field #22009 and ethnicity information was obtained from the field #21000. In HNP, PCs were calculated based on 12,516 SNPs across the genome using Hail `hwe_normalized_pca` function, and ancestry was calculated using ADMIXTURE<sup>55</sup>. The rare variants were analyzed as a single cluster using MiST<sup>22</sup>. MiST provides a score statistic  $S(\pi)$  for the mean effect ( $\pi$ ) of the cluster, and a score statistic  $S(\tau)$  for the heterogeneous effect ( $\tau$ ) of the cluster. The overall  $p$ -value is the combined  $p$ -values  $P_\pi$  and  $P_\tau$ , from the Fisher's procedure. MiST was shown to be the most powerful method across a range of architectures, however it is computationally intensive<sup>56-58</sup>.

Let the equation of the model be:  $Y = \alpha X + \pi GZ$ , where  $Y$ , is the trait of interest and  $X$ , is the matrix of covariates.

For the gene-centric analyses in the RaDiO, UK Biobank and HNP studies,  $Z$  was a vector of ones repeated  $n$  times, with  $n$  the number of rare variants in the gene, leading to:  $\pi GZ = \pi \sum_{i=1}^n G_i$ .

For the multi-gene analyses in the RaDiO, UK Biobank and HNP studies,  $Z$  was a matrix (with variants as rows and genes as columns) giving the variants' characteristics, *i.e.* one when a given variant belonged to gene and zero otherwise, leading to:  $\pi GZ = \pi_1 \sum_{i=1}^{n_1} G_i + \pi_2 \sum_{i=1}^{n_2} G_i + \dots + \pi_k \sum_{i=1}^{n_k} G_i$  (with  $k$  genes in the cluster).

In the AMP T2D knowledge portal, the gene-centric analyses were done through additive burden test adjusted for age, sex, BMI and ancestry (using PCs set by default in the portal). The multi-gene function was not available.

In each study, logistic regressions adjusted for age, sex, BMI and ancestry were performed followed by a meta-analysis of the results from the four studies using the generic inverse variance method from the R package *meta*<sup>59</sup>. The heterogeneous effect ( $\tau^2$ ) of the meta-analysis was reported when relevant.

Among the patients with T2D, additional traits were analyzed in carriers and non-carriers of P/LP variants using either a linear regression for quantitative traits (*i.e.* BMI and age of T2D diagnosis) and a logistic regression for binary traits (*i.e.* use of antidiabetic drugs, family history of T2D and ancestry). Age, BMI, sex, PC1, PC2, PC3, PC4 and PC5 were included as covariates for all traits, when relevant. The meta-analysis of the results was performed using the generic inverse variance method from the R package *meta*<sup>59</sup>, when relevant.

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

## **Data availability**

All relevant data have been included in the manuscript and/or in its supplementary tables and figures. Given the sensitivity and risk of re-identification, all clinical data linked with next-generation sequencing data for this study is only available upon request from the corresponding authors. We used the following web links for publicly available datasets: **1/**

Type 2 Diabetes Knowledge Portal.

<http://www.type2diabetesgenetics.org/gene/geneInfo/XXX>, where XXX is the gene name; **2/**

Genome Aggregation Database (gnomAD). <https://gnomad.broadinstitute.org/>; **3/** dbNSFP.

<https://sites.google.com/site/jpopgen/dbNSFP>; **4/** dbSNP. <https://www.ncbi.nlm.nih.gov/snp/>.

## **Code availability**

Code to perform analyses related to bioinformatics and biostatistics in this manuscript are available following this link: [https://github.com/umr1283/MODY\\_GENES](https://github.com/umr1283/MODY_GENES) (doi: 10.5281/zenodo.4005715).

**To whom correspondence and requests for materials should be addressed:** Amélie Bonnefond ([amelie.bonnefond@cns.fr](mailto:amelie.bonnefond@cns.fr)) and Philippe Froguel ([p.froguel@imperial.ac.uk](mailto:p.froguel@imperial.ac.uk)).

## **Acknowledgements**

We are grateful to all individuals included in the different cohort studies. We thank Olivier Sand and Iandry Rabearivelo for their contribution to the first computer analyses. This research has been conducted using the UK Biobank Application Number 40436. We thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at <https://gnomad.broadinstitute.org/about>. We would like to thank the Type 2 Diabetes

Knowledge Portal and the groups that provided data to this resource. This work was supported by grants from the French National Research Agency (ANR-10-LABX-46 [European Genomics Institute for Diabetes] and ANR-10-EQPX-07-01 [LIGAN-PM]), from the European Research Council (ERC GEPIDIAB – 294785, to P.F.; ERC Reg-Seq – 715575, to A.Bonnefond), and from the National Center for Precision Diabetic Medicine – PreciDIAB, which is jointly supported by the French National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-France Regional Council and by the European Metropolis of Lille (MEL). Funding was also provided to the Renown Institute for Health Innovation by Renown Health and the Renown Health Foundation.

### **Author contributions**

A.Bonnefond and P.F. conceived the idea of the study, supervised the analyses and performed data interpretation; E.D., B.T., E.V. made libraries and next-generation sequencing, with help from A.D. and V.D.; S.G., J.-M.B., J.T.L., E.T.C., G.E., R.R., B.B., M.M., S.F., G.C., M.V., N.L.W., J.J.G and P.F. managed the collection of samples with clinical data; F.A. prepared samples; F.D.G. and D.L.G. performed computer analyses; A.Bonnefond, M.B. and A.Bolze performed data curation and statistical analysis, with help from L.Y. and M.C; A.Bonnefond wrote the first draft of the paper, with help from P.F.; all authors critically reviewed the paper and approved the report for submission.

### **Declaration of interests**

Ronan Roussel is an advisory panel member for AstraZeneca, AbbVie, Sanofi, Merck Sharp & Dohme, Eli Lilly, Janssen, Novo Nordisk, Diabnext, Vaiomer, and Physiogenex; is a speaker for Bayer and Servier; and has received research funding and provided research support to Danone Research, Amgen, Sanofi, and Novo Nordisk. Alexandre Bolze, Elizabeth



T. Cirulli, James Lu and Nicole L. Washington are employees and shareholders of Helix. No other conflicts were reported.

**Table 1. Association between T2D risk and P/LP variants in actionable MODY genes, in the participants from the RaDiO study, the UK biobank, the HNP study and the AMP T2D knowledge portal**

Study	Gene	$n_{\text{variant}}$	Carriers among cases (%)	Carriers among controls (%)	$P_T$	$P_{T^2}$	$P$
RaDiO <i>N</i> = 6,348 (2,178 cases / 4,170 controls)	<i>HNF1A</i>	19	0.872	0.0959	0.220	NA	$2.55 \times 10^{-4}$
	<i>GCK</i>	14	0.597	0.0959	0.0496	NA	$3.13 \times 10^{-4}$
	<i>GATA4</i>	3	0.0918	0.0240	0.0588	NA	0.0389
	<i>KCNJ11</i>	13	0.367	0.168	0.571	NA	0.270
	<i>HNF1B</i>	9	0.689	0.240	0.414	NA	0.239
	<i>ABCC8</i>	32	0.689	0.480	0.196	NA	0.356
	<i>HNF4A</i>	6	0.184	0.144	0.414	NA	0.719
	<i>GATA6</i>	2	0.0459	0.168	$1.83 \times 10^{-7}$	NA	$7.22 \times 10^{-8}$
	<b>Multi-gene</b>	<b>98</b>	<b>3.49</b>	<b>1.39</b>	<b>0.0780</b>	<b>NA</b>	<b><math>3.91 \times 10^{-5}</math></b>
UK Biobank <i>N</i> = 35,869 (2,151 cases / 33,718 controls)	<i>HNF1A</i>	36	0.0930	0.282	0.346	NA	0.109
	<i>GCK</i>	15	0.418	0.0860	0.0269	NA	$3.14 \times 10^{-5}$
	<i>GATA4</i>	7	0.0465	0.0237	0.247	NA	0.532
	<i>KCNJ11</i>	11	0.139	0.0475	0.701	NA	0.168
	<i>HNF1B*</i>	0	NA	NA	NA	NA	NA
	<i>ABCC8</i>	67	0.837	0.433	0.291	NA	0.119
	<i>HNF4A</i>	10	0.418	0.181	0.131	NA	0.0248
	<i>GATA6</i>	5	0.232	0.199	0.671	NA	0.853
	<b>Multi-gene</b>	<b>151</b>	<b>2.19</b>	<b>1.24</b>	<b>0.0767</b>	<b>NA</b>	<b><math>1.39 \times 10^{-5}</math></b>
HNP <i>N</i> = 5,689 (1,518 cases / 4,171 controls)	<i>HNF1A</i>	19	0.395	0.408	0.380	NA	0.628
	<i>GCK</i>	10	0.593	0.120	0.345	NA	$8.72 \times 10^{-3}$
	<i>GATA4</i>	1	0	0.0240	NA	NA	NA
	<i>KCNJ11</i>	5	0.198	0.0959	0.644	NA	0.872
	<i>HNF1B</i>	3	0.461	0.288	0.420	NA	0.155
	<i>ABCC8</i>	14	0.461	0.336	0.481	NA	0.611
	<i>HNF4A</i>	3	0	0.144	1.00	NA	0.320
	<i>GATA6</i>	5	0.264	0.144	0.594	NA	0.397
	<b>Multi-gene</b>	<b>60</b>	<b>2.31</b>	<b>1.56</b>	<b>0.576</b>	<b>NA</b>	<b>0.0915</b>
AMP T2D knowledge portal <i>N</i> = 26,723 (19,852 cases / 6,871 controls)	<i>HNF1A</i>	NA	0.720	0.291	NA	NA	$1.57 \times 10^{-5}$
	<i>GCK</i>	NA	0.212	0.0728	NA	NA	0.0376
	<i>GATA4</i>	NA	0.0302	0	NA	NA	0.897
	<i>KCNJ11</i>	NA	0.126	0.0291	NA	NA	0.145
	<i>HNF1B*</i>	NA	NA	NA	NA	NA	NA
	<i>ABCC8</i>	NA	0.579	0.349	NA	NA	0.323
	<i>HNF4A</i>	NA	0.126	0.0582	NA	NA	0.0569
	<i>GATA6</i>	NA	0.0151	0.0146	NA	NA	0.928
	<b>Multi-gene**</b>	<b>NA</b>	<b>1.81</b>	<b>0.815</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>
Meta-analysis <i>N</i> = 74,629 (25,699 cases / 48,930 controls)	<i>HNF1A</i>	NA	0.662	0.278	NA	$5.44 \times 10^{-4}$	0.399
	<i>GCK</i>	NA	0.284	0.0879	NA	0.661	$1.41 \times 10^{-9}$
	<i>GATA4</i>	NA	0.0350	0.0204	NA	0.805	0.277
	<i>KCNJ11</i>	NA	0.152	0.0593	NA	0.804	0.0102
	<i>HNF1B*</i>	NA	0.0856	0.0450	NA	0.766	0.0296
	<i>ABCC8</i>	NA	0.603	0.417	NA	0.933	0.0417
	<i>HNF4A</i>	NA	0.148	0.157	NA	0.747	$8.17 \times 10^{-3}$
	<i>GATA6</i>	NA	0.0506	0.166	NA	0.273	0.481
	<b>Multi-gene</b>	<b>NA</b>	<b>2.01</b>	<b>1.22</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>

\*Exome data from the UK Biobank and from the AMP T2D knowledge portal were not available for *HNF1B*, therefore the meta-analysis for *HNF1B* only included data from both RaDiO and HNP studies.

\*\*The multi-gene analysis was not available in the AMP T2D knowledge portal.

The number of variants was only available for the whole case-control study from the AMP T2D knowledge portal, while we only analyzed normal glucose participants higher than 40 years ( $N = 6,871$ ) among the non-diabetic controls of the portal ( $N = 23,214$ ). Therefore, the number of variants was not reported in the AMP T2D knowledge portal and in the meta-analysis.

Association analyses were performed using the MiST method adjusted for age, sex, BMI and ancestry. The meta-analyses were performed using the generic inverse variance method.

*NA*, not available;  $\tau$ , heterogeneous effect;  $\tau^2$ , heterogeneous effect of the meta-analyses.

**Table 2. Association between BMI and P/LP variants in actionable MODY genes, in the participants with T2D from the RaDiO study, the UK Biobank and the HNP study**

Study	Gene	$n_{\text{variant}}$	BMI (kg/m <sup>2</sup> ) in carriers (Mean $\pm$ SD)	BMI (kg/m <sup>2</sup> ) in non-carriers (Mean $\pm$ SD)	$\beta \pm SE$	$P$
RaDiO $N = 2,178$ cases	<i>HNF1A</i>	18	25.9 $\pm$ 3.34	28.5 $\pm$ 5.83	-2.24 $\pm$ 1.28	0.0801
	<i>GCK</i>	10	25.6 $\pm$ 3.72	28.5 $\pm$ 5.82	-3.60 $\pm$ 1.55	0.0198
	<i>GATA4</i>	2	25.4 $\pm$ 0.962	28.5 $\pm$ 5.82	-5.36 $\pm$ 3.93	0.173
	<i>KCNJ11</i>	8	27.0 $\pm$ 2.44	28.5 $\pm$ 5.82	-1.93 $\pm$ 1.97	0.327
	<i>HNF1B</i>	6	25.5 $\pm$ 3.25	28.5 $\pm$ 5.82	-2.69 $\pm$ 1.44	0.0616
	<i>ABCC8</i>	15	28.0 $\pm$ 4.70	28.5 $\pm$ 5.82	-0.963 $\pm$ 1.44	0.504
	<i>HNF4A</i>	3	27.4 $\pm$ 5.88	28.5 $\pm$ 5.82	-1.71 $\pm$ 2.78	0.539
	<i>GATA6</i>	1	23.7	28.5 $\pm$ 5.81	-4.40 $\pm$ 5.56	0.429
	<b>Multi-gene</b>	<b>63</b>	<b>26.3 <math>\pm</math> 3.73</b>	<b>28.6 <math>\pm</math> 5.86</b>	<b>-2.47 <math>\pm</math> 0.648</b>	<b>1.45<math>\times</math>10<sup>-4</sup></b>
UK Biobank $N = 2,151$ cases	<i>HNF1A</i>	2	30.4 $\pm$ 1.27	31.9 $\pm$ 5.92	-0.409 $\pm$ 4.02	0.919
	<i>GCK</i>	8	29.4 $\pm$ 4.97	31.9 $\pm$ 5.92	-3.54 $\pm$ 1.89	0.0607
	<i>GATA4</i>	1	38.8	31.9 $\pm$ 5.91	8.04 $\pm$ 5.64	0.154
	<i>KCNJ11</i>	3	36.0 $\pm$ 6.75	31.9 $\pm$ 5.91	3.20 $\pm$ 3.26	0.326
	<i>HNF1B*</i>	0	NA	NA	NA	NA
	<i>ABCC8</i>	15	31.6 $\pm$ 5.80	31.9 $\pm$ 5.91	0.340 $\pm$ 1.34	0.799
	<i>HNF4A</i>	4	26.5 $\pm$ 5.37	32.0 $\pm$ 5.91	-5.47 $\pm$ 1.88	3.69 $\times$ 10 <sup>-3</sup>
	<i>GATA6</i>	3	33.8 $\pm$ 3.43	31.9 $\pm$ 5.92	2.07 $\pm$ 2.53	0.413
	<b>Multi-gene</b>	<b>36</b>	<b>30.8 <math>\pm</math> 5.76</b>	<b>32.0 <math>\pm</math> 5.91</b>	<b>-1.03 <math>\pm</math> 0.832</b>	<b>0.215</b>
HNP $N = 1,518$ cases	<i>HNF1A</i>	6	31.1 $\pm$ 8.46	34.0 $\pm$ 7.27	-2.16 $\pm$ 2.73	0.429
	<i>GCK</i>	8	30.6 $\pm$ 7.76	34.0 $\pm$ 7.27	-4.62 $\pm$ 2.23	0.038
	<i>GATA4</i>	0	NA	NA	NA	NA
	<i>KCNJ11</i>	4	34.1 $\pm$ 1.73	33.9 $\pm$ 7.28	-3.56 $\pm$ 3.88	0.359
	<i>HNF1B</i>	3	32.0 $\pm$ 6.22	34.0 $\pm$ 7.28	-1.76 $\pm$ 2.53	0.486
	<i>ABCC8</i>	7	30.5 $\pm$ 5.94	34.0 $\pm$ 7.28	-1.81 $\pm$ 2.55	0.477
	<i>HNF4A</i>	0	NA	NA	NA	NA
	<i>GATA6</i>	2	33.9 $\pm$ 5.40	33.9 $\pm$ 7.28	-0.967 $\pm$ 3.34	0.772
	<b>Multi-gene</b>	<b>30</b>	<b>31.8 <math>\pm</math> 6.39</b>	<b>34.0 <math>\pm</math> 7.29</b>	<b>-2.62 <math>\pm</math> 1.14</b>	<b>0.0217</b>
Meta-analysis $N = 5,847$ cases	<i>HNF1A</i>	23	NA	NA	-2.09 $\pm$ 1.11	0.0610
	<i>GCK</i>	25	NA	NA	-3.81 $\pm$ 1.06	3.08 $\times$ 10 <sup>-4</sup>
	<i>GATA4</i>	3	NA	NA	-0.980 $\pm$ 3.22	0.761
	<i>KCNJ11</i>	11	NA	NA	-1.03 $\pm$ 1.55	0.503
	<i>HNF1B*</i>	7	NA	NA	-2.46 $\pm$ 1.25	0.0491
	<i>ABCC8</i>	32	NA	NA	-0.464 $\pm$ 0.916	0.612
	<i>HNF4A</i>	5	NA	NA	-4.29 $\pm$ 1.56	5.87 $\times$ 10 <sup>-3</sup>
	<i>GATA6</i>	4	NA	NA	0.339 $\pm$ 1.90	0.858
	<b>Multi-gene</b>	<b>110</b>	NA	NA	<b>-2.04 <math>\pm</math> 0.466</b>	<b>1.20<math>\times</math>10<sup>-5</sup></b>

\*Exome data from the UK Biobank were not available for *HNF1B*, therefore the meta-analysis for *HNF1B* only included data from both RaDiO and HNP studies.

Association analyses were performed using a linear regression adjusted for age, sex, and genetic ancestry. The meta-analyses were performed using the generic inverse variance method. None significant heterogeneous effect was observed in the meta-analyses.

$\beta$ , mean effect; *NA*, not available; *SD*, standard deviation; *SE*, standard error.

## Figure Legends

**Figure 1. Association between T2D risk and P/LP variants in each actionable MODY genes, detected in the participants from (a) the French RaDiO study including 2,178 cases and 4,170 controls, (b) the UK Biobank including 2,151 cases and 33,718 controls, (c) the HNP study including 1,518 cases and 4,171 controls, (d) the AMP T2D knowledge portal including 19,852 cases and 6,871 controls, and (e) the meta-analysis including 25,699 cases and 48,930 controls**

Association analyses were performed using the MiST method adjusted for age, sex, BMI and ancestry. The meta-analyses were performed using the generic inverse variance method.

Estimated ORs (bars, left axis) and *P*-values (right axis) for carriers of variants in each actionable MODY gene. Orange bars, OR >1; Purple bars, OR <1; Red dotted line, *P* = 0.05;

Green diamonds represent the overall *P*-values.

\*Exome data from the UK Biobank and from the AMP T2D knowledge portal were not available for *HNF1B*, therefore the meta-analysis for *HNF1B* only included data from both RaDiO and HNP studies.

\*\*Only found in one control.

†Results with OR >10<sup>4</sup>.

## References

1. Mathers, C. D. & Loncar, D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine* **3**, e442 (2006).
2. Abajobir, A. A. *et al.* Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **390**, 1084–1150 (2017).
3. Dieleman, J. L. *et al.* US Spending on Personal Health Care and Public Health, 1996–2013. *JAMA* **316**, 2627–2646 (2016).
4. Willemsen, G. *et al.* The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. *Twin Res Hum Genet* **18**, 762–771 (2015).
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**, 1505–1513 (2018).
6. Bonnefond, A. & Froguel, P. Rare and common genetic events in type 2 diabetes: what should biologists know? *Cell Metab.* **21**, 357–368 (2015).
7. Vaxillaire, M. & Froguel, P. Monogenic diabetes: Implementation of translational genomic research towards precision medicine. *J Diabetes* **8**, 782–795 (2016).
8. Babenko, A. P. *et al.* Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. *N. Engl. J. Med.* **355**, 456–466 (2006).
9. Pearson, E. R. *et al.* Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N. Engl. J. Med.* **355**, 467–477 (2006).
10. Shepherd, M., Shields, B., Ellard, S., Rubio-Cabezas, O. & Hattersley, A. T. A genetic diagnosis of HNF1A diabetes alters treatment and improves glycaemic control in the majority of insulin-treated patients. *Diabet. Med.* **26**, 437–441 (2009).
11. Pearson, E. R. *et al.* Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* **362**, 1275–1281 (2003).
12. Pearson, E. R. *et al.* Molecular genetics and phenotypic characteristics of MODY caused by hepatocyte nuclear factor 4alpha mutations in a large European collection. *Diabetologia* **48**, 878–885 (2005).
13. Stride, A. *et al.* Cross-sectional and longitudinal studies suggest pharmacological treatment used in patients with glucokinase mutations does not alter glycaemia. *Diabetologia* **57**, 54–56 (2014).
14. Garg, V. *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**, 443–447 (2003).
15. Bockenhauer, D. & Jaureguiberry, G. HNF1B-associated clinical phenotypes: the kidney and beyond. *Pediatr. Nephrol.* **31**, 707–714 (2016).
16. Kodo, K. *et al.* GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13933–13938 (2009).
17. Naylor, R. N. *et al.* Cost-effectiveness of MODY genetic testing: translating genomic advances into practical health applications. *Diabetes Care* **37**, 202–209 (2014).
18. GoodSmith, M. S., Skandari, M. R., Huang, E. S. & Naylor, R. N. The Impact of Biomarker Screening and Cascade Genetic Testing on the Cost-Effectiveness of MODY Genetic Testing. *Diabetes Care* **42**, 2247–2255 (2019).
19. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
20. Flannick, J. *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71–76 (2019).

21. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
22. Sun, J., Zheng, Y. & Hsu, L. A unified mixed-effects model for rare-variant association in sequencing studies. *Genet. Epidemiol.* **37**, 334–344 (2013).
23. 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).
24. Froguel, P. *et al.* Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 162–164 (1992).
25. Flannick, J. *et al.* Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat. Genet.* **45**, 1380–1385 (2013).
26. Bonnefond, A. *et al.* Whole-Exome Sequencing and High Throughput Genotyping Identified KCNJ11 as the Thirteenth MODY Gene. *PLoS ONE* **7**, e37423 (2012).
27. Bonnefond, A. *et al.* GATA6 inactivating mutations are associated with heart defects and, inconsistently, with pancreatic agenesis and diabetes. *Diabetologia* **55**, 2845–2847 (2012).
28. De Franco, E. *et al.* GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency. *Diabetes* **62**, 993–997 (2013).
29. Meur, G. *et al.* Insulin gene mutations resulting in early-onset diabetes: marked differences in clinical presentation, metabolic status, and pathogenic effect through endoplasmic reticulum retention. *Diabetes* **59**, 653–661 (2010).
30. Castel, S. E. *et al.* Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat. Genet.* **50**, 1327–1334 (2018).
31. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219–1224 (2018).
32. Baier, L. J. *et al.* ABCC8 R1420H Loss-of-Function Variant in a Southwest American Indian Community: Association With Increased Birth Weight and Doubled Risk of Type 2 Diabetes. *Diabetes* **64**, 4322–4332 (2015).
33. Balkau, B. [An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome]. *Rev Epidemiol Sante Publique* **44**, 373–375 (1996).
34. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
35. Meyre, D. *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.* **41**, 157–159 (2009).
36. Romon, M. *et al.* Relationships between physical activity and plasma leptin levels in healthy children: the Fleurbaix-Laventie Ville Santé II Study. *Int. J. Obes. Relat. Metab. Disord.* **28**, 1227–1232 (2004).
37. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
38. Van Hout, C. V. *et al.* Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. <http://biorxiv.org/lookup/doi/10.1101/572347> (2019) doi:10.1101/572347.
39. Cirulli, E. T. *et al.* Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat Commun* **11**, 542 (2020).
40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
41. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

42. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
43. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* **37**, 235–241 (2016).
44. Regier, A. A. *et al.* Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun* **9**, 4038 (2018).
45. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
46. Kendig, K. I. *et al.* Sentieon DNaseq Variant Calling Workflow Demonstrates Strong Computational Performance and Accuracy. *Front Genet* **10**, 736 (2019).
47. Helix [https://cdn.shopify.com/s/files/1/2718/3202/files/Helix-WP\\_\\_Performance\\_of\\_the\\_Helix\\_Exome\\_TM\\_Assay.pdf?13203](https://cdn.shopify.com/s/files/1/2718/3202/files/Helix-WP__Performance_of_the_Helix_Exome_TM_Assay.pdf?13203). Performance of the Helix Exome+TM Assay.
48. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
49. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
50. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat Protoc* **11**, 1–9 (2016).
51. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
52. Ellard, S., Colclough, K., Patel, K. A. & Hattersley, A. T. Prediction algorithms: pitfalls in interpreting genetic variants of autosomal dominant monogenic diabetes. *J. Clin. Invest.* **130**, 14–16 (2020).
53. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS ONE* **9**, e93766 (2014).
54. Siva, N. 1000 Genomes project. *Nat. Biotechnol.* **26**, 256 (2008).
55. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
56. Chen, H. *et al.* Efficient Variant Set Mixed Model Association Tests for Continuous and Binary Traits in Large-Scale Whole-Genome Sequencing Studies. *Am. J. Hum. Genet.* **104**, 260–274 (2019).
57. Moutsianas, L. *et al.* The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS Genet.* **11**, e1005165 (2015).
58. Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23 (2014).
59. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health* **22**, 153–160 (2019).