Article

Common genetic variants modify disease risk and clinical presentation in monogenic diabetes

Check for updates

Jacques Murray Leech [1], Robin N. Beaumont [1], Ankit M. Arni[1],
V. Kartik Chundru [1], Luke N. Sharp [1], Kevin Colclough[2],
Andrew T. Hattersley [1], Michael N. Weedon [1] & Kashyap A. Patel [1]✉

Young-onset monogenic disorders often show variable penetrance, yet the underlying causes remain poorly understood. Uncovering these influences could reveal new biological mechanisms and enhance risk prediction for monogenic diseases. Here we show that polygenic background substantially shapes the clinical presentation of maturity-onset diabetes of the young (MODY), a common monogenic form of diabetes that typically presents in adolescence or early adulthood. We find strong enrichment of type 2 diabetes (T2D) polygenic risk, but not type 1 diabetes risk, in genetically confirmed MODY cases ($n = 1,462$). This T2D polygenic burden, primarily through beta-cell dysfunction pathways, is strongly associated with earlier age of diagnosis and increased diabetes severity. Common genetic variants collectively account for 24% ($P < 0.0001$) of the phenotypic variability. Using a large population cohort ($n = 424,553$), we demonstrate that T2D polygenic burden substantially modifies diabetes onset in individuals with pathogenic variants, with diabetes risk ranging from 11% to 81%. Finally, we show that individuals with MODY-like phenotypes ($n = 300$) without a causal variant have elevated polygenic burden for T2D and related traits, representing potential polygenic phenocopies. These findings reveal substantial influence of common genetic variation in shaping the clinical presentation of early-onset monogenic disorders. Incorporating these may improve risk estimates for individuals carrying pathogenic variants.

Growing evidence suggests that factors beyond the primary mutation play a greater role in rare monogenic conditions than previously recognized[1,2]. Monogenic diseases are classically defined by single, highly penetrant (proportion of carriers who develop the disease) causative mutations. However, many individuals carrying the same pathogenic variant show wide variation in disease expression, suggesting that additional factors influence disease risk[3]. For example, our analysis of pathogenic *HNF1A* mutations, typically associated with young-onset diabetes, revealed striking differences in penetrance: over 90% in

clinically ascertained cohorts versus under 30% in population-based cohorts by the age of 40 years[4]. This pattern of unexpectedly low and variable penetrance has now been documented across multiple monogenic conditions[5,6]. Polygenic background has been proposed as one possible contributor, particularly in age-dependent monogenic disorders[7,8].

Maturity-onset diabetes of the young (MODY) serves as an excellent genetic disease model to investigate how common genetic variants influence young-onset monogenic disease. MODY is the most common

[1]Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK. [2]Exeter Genomics Laboratory, Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK. ✉e-mail: K.A.Patel@exeter.ac.uk

autosomal dominant form of monogenic diabetes contributing up to 3% of all diabetes under the age of 30 years[9]. In this study we focused on the *HNF1A*, *HNF4A* and *HNF1B* genes (collectively referred to as HNF-MODY). The pathogenic variants in these three genes account for >90% of MODY cases[10–12]. These variants cause beta-cell dysfunction leading to age-dependent diabetes typically presenting before age 25 years[11]. The availability of extensive genome-wide association data for both type 1 and type 2 diabetes and related metabolic traits, widely measured diabetes markers such as HbA1c allowing accurate diagnosis, and the availability of large MODY patient cohorts make MODY particularly suitable for studying common and rare disease interplay. Together, these resources provide a robust framework for examining how polygenic factors interact with young-onset monogenic disorders.

Understanding these interactions is crucial both biologically and clinically. It can uncover new biological pathways and enhance disease prediction, knowledge that is essential for family counselling. This becomes increasingly important as genomic screening extends to clinically unselected cases and healthy newborns[13]. Previous studies have demonstrated that polygenic background can modify the penetrance of various monogenic conditions, including familial hypercholesterolaemia, obesity, kidney disease and long QT syndrome[7,14–16]. These studies are important but often lack a defined age of disease onset. Previous studies suggested that polygenic risk for type 2 diabetes (T2D) may influence the age at MODY diagnosis[17,18]. However, those studies used small sample sizes ($n < 410$), focused only on HNF1A-MODY, and did not assess diabetes-related metabolic traits or partition T2D polygenic scores to explore the underlying mechanisms in detail. A more recent study investigated the interaction between T2D polygenic risk and rare intermediate-effect variants in *HNF1A* and *HNF4A* within population cohorts but did not include clinically confirmed MODY cases with pathogenic variants[19]. No previous work has comprehensively analysed how common genetic backgrounds influence diabetes severity in MODY or quantified their overall contribution to the MODY phenotype. Finally, it remains unclear whether the common genetic background also contributes to MODY-like cases without pathogenic variants.

In this study, we investigated the interplay between polygenic background and age-dependent monogenic disorders, using MODY as a model disease. In the largest MODY cohort studied to date, we demonstrate that common genetic variants explain a substantial proportion of phenotypic variation, disease expression and may explain MODY-like phenotypes in individuals without identified monogenic causes.

Results

Polygenic burden of type 2 diabetes is significantly enriched in genetically confirmed MODY

While common variants are known to modify disease expression in other monogenic disorders, their influence on HNF-MODY remains relatively unexplored. We investigated this assumption by analysing polygenic scores (PGSs) for T2D, type 1 diabetes (T1D) and related metabolic traits ($n = 9$) in 1,462 clinically referred patients with HNF-MODY (Supplementary Tables 1–3). We compared these scores with those of 7,645 individuals without diabetes and 4,773 individuals with T2D (Supplementary Table 1). We found significantly higher polygenic scores for T2D, fasting glucose, fasting insulin and waist–hip ratio in HNF-MODY patients compared with non-diabetic controls (0.09–0.42 s.d. increase, all $P < 0.005$) but no enrichment for T1D PGS (Fig. 1a). The T2D PGS remained the strongest contributor (odds ratio (OR) 1.46, 95% confidence interval (CI) 1.36–1.58, $P < 0.0001$) after accounting for other PGSs (Fig. 1b). This enrichment was lower than observed in T2D cases and independent of parental diabetes history (Fig. 1c) and after removing variants in *HNF1A*, *HNF4A* or *HNF1B* genes from the PGS (0.4 s.d. higher than control, 95% CI 0.35–0.46, $P < 0.0001$). To identify which T2D pathways contributed to this enrichment, we analysed eight recently developed T2D

pathway-specific hard cluster PGSs[20] (Fig. 1d). Of these, the metabolic syndrome (0.20 s.d. increase), residual glycaemic (0.17 s.d.) and beta-cell proinsulin-positive (0.16 s.d.) pathway scores showed the strongest enrichments in patients with MODY (all $P < 4 \times 10^{-8}$). Sensitivity analyses by each gene and limited to probands showed consistent findings (Extended Data Figs. 1 and 2). Supporting substantial interplay between rare and common variation, we observed an interaction between T2D PGS and pathogenic variants across the spectrum of predicted deleteriousness. Carriers of less-damaging missense variants showed relatively higher T2D polygenic risk compared with carriers of more deleterious protein-truncating variants (Extended Data Fig. 3). Together, these data suggest that common T2D-associated variants contribute substantially to clinically diagnosed HNF-MODY.

Increased type 2 diabetes polygenic burden was associated with an earlier onset and greater phenotypic severity in patients with genetically confirmed MODY

We next assessed how the polygenic burden of T2D, T1D and related metabolic traits influenced both the age of diagnosis and severity of diabetes in patients with clinically identified HNF-MODY. We defined diabetes severity as either requiring insulin treatment or having HbA1c $\geq 8.5\%$ as proposed previously[21]. Only T2D PGS demonstrated a significant association with age of diagnosis after adjusting for other PGSs ($P < 3.3 \times 10^{-5}$), with 1 s.d. increase linked with a 1.19 years (0.63–1.75) earlier diagnosis (Fig. 2a). In contrast, both the T2D and body mass index (BMI) PGSs were significantly associated with diabetes severity, with ORs of 1.24 (95% CI 1.07–1.44, $P = 0.004$) and 1.32 (95% CI 1.16–1.51, $P < 3.1 \times 10^{-5}$), respectively (Fig. 2b). Our pathway analysis revealed that the beta-cell proinsulin-positive pathway primarily drove the T2D PGS effect on diagnosis age (0.83 years (0.33–1.32) versus 0.67 (0.15–1.18) years for all others combined) (Fig. 2c). Whereas the obesity pathway demonstrated the strongest association with diabetes severity (OR 1.36, 1.19–1.56 versus 1.19, 1.04–1.35 for all other pathways combined) (Fig. 2d). As expected, only the BMI PGS was significantly associated with measured BMI (Supplementary Table 4). Age of diagnosis and severity associations were maintained even after adjusting for clinical features and genetic aetiology. However, we observed strong effects of sex (females diagnosed 2.28 years earlier), maternal diabetes history (diagnosed 3.54 years earlier) and BMI (0.24-year earlier diagnosis) on age of diagnosis (Supplementary Tables 5 and 6). Sensitivity analyses by each gene show directional consistent results (Supplementary Table 7). We also conducted additional sensitivity analyses in 413 HNF-MODY cases with available birthweight data and found that associations between T2D polygenic scores and clinical outcomes remained largely unchanged after adjusting for birthweight (Supplementary Tables 8 and 9). These findings highlight the complex interaction between genetic and clinical factors that shape the clinical presentation of HNF-MODY.

Type 2 diabetes polygenic burden modifies the risk of diabetes in clinically unselected carriers of pathogenic HNF-MODY variants

We next investigated how polygenic T2D background influences diabetes risk in carriers of pathogenic HNF-MODY variants. To assess this, we needed to investigate individuals not ascertained clinically to see the clear effect of polygenic background. Therefore, we analysed 424,553 European individuals with whole exome sequencing data from the clinically unselected UK Biobank population cohort. Among these, 100 individuals were identified as carriers of pathogenic variants in *HNF1A* ($n = 34$), *HNF4A* ($n = 51$) or *HNF1B* ($n = 15$) (Supplementary Tables 10 and 11). Using a T2D PGS that did not include UK Biobank in the discovery cohort[19], we found that among mutation carriers, diabetes risk varied substantially by T2D PGS. Compared with non-carriers with intermediate T2D PGS (middle three quintiles), carriers' risk ranged
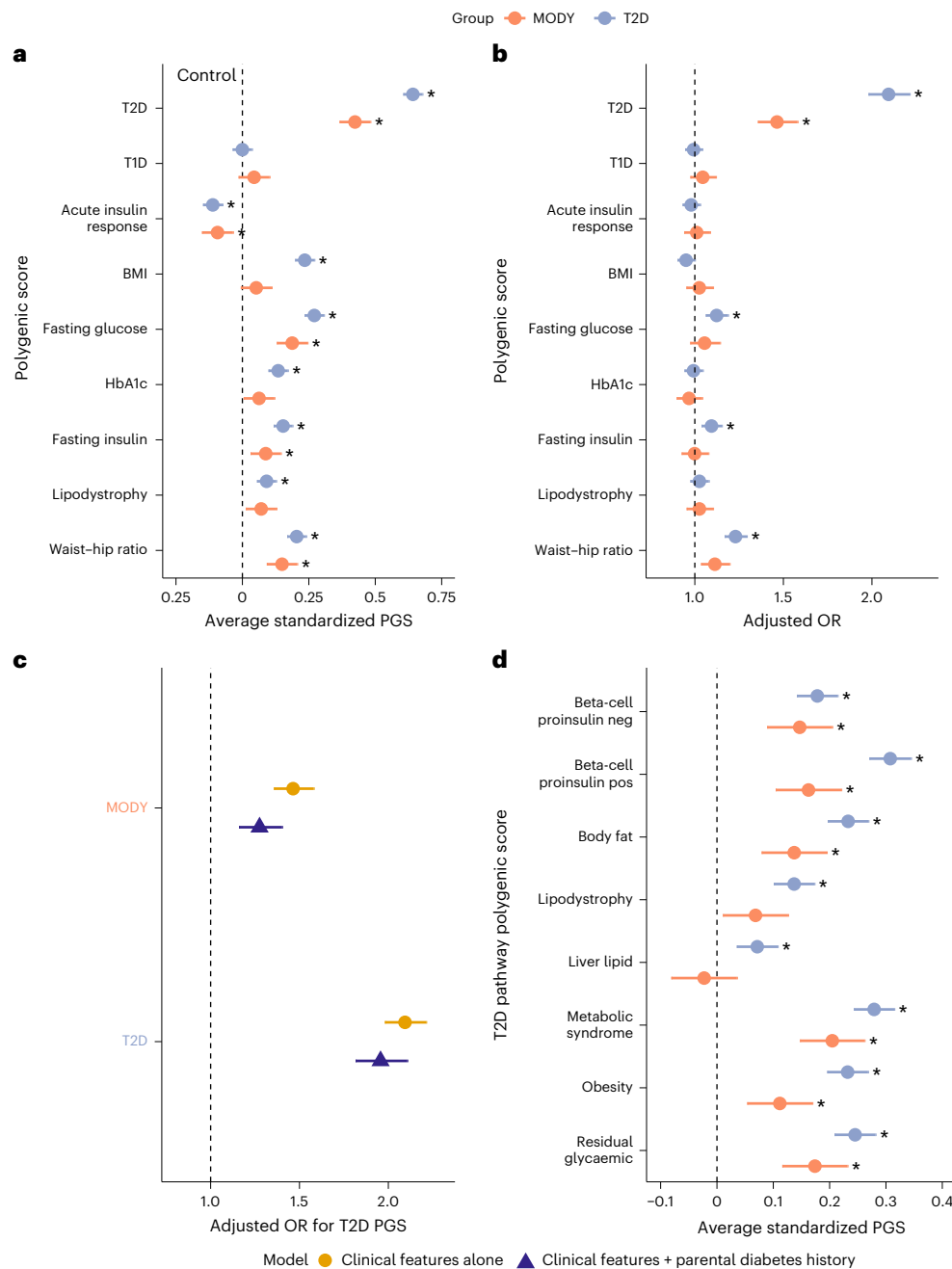
Fig. 1 | Elevated polygenic risk in HNF monogenic diabetes. a, Standardized differences in upper-level diabetes-related polygenic scores, determined by linear regression adjusting for the first ten within-cohort principal components. HNF-MODY carriers (orange, $n = 1,462$) and T2D cases (blue, $n = 4,773$) are compared with controls (dashed black line, $n = 7,645$). **b**, Adjusted ORs for T2D and HNF-MODY cases versus controls, assessed using a logistic regression model including each PGS, sex, age, BMI and the first ten within-cohort principal components as covariates. **c**, Adjusted ORs for T2D PGS enrichment in HNF-MODY and T2D cases under two models: (1) adjusting for covariates as described in **b** (yellow); and (2) adjusting for the same covariates plus family history of diabetes (blue). **d**, Standardized differences in T2D hard cluster partitioned polygenic scores. All scores are standardized to have a mean of 0 and s.d. of 1 in controls. ORs represent the change in risk associated with a 1 s.d. increase in the respective polygenic score. Error bars represent 95% CIs. Asterisks denote Bonferroni-adjusted statistically significant differences from controls. Sample sizes are consistent across **a–d**.

from 8.5-fold (95% CI: 3.65–19.85) in those in the lowest T2D PGS quintile to 40.22-fold (95% CI 14.95–108.24) in those in the highest quintile (Fig. 3a). HNF-MODY carriers had a 6.67-fold higher risk of diabetes (95% CI 4.39–10.12, $P = 4.23 \times 10^{-19}$) compared with non-carriers in the highest quintile, highlighting the strong impact of pathogenic mutations. Despite the limited sample size, diabetes risk seemed to rise consistently across the range of T2D PGS, with diabetes risk ranging from 11.4% (first percentile, 95% CI 6.96–15.88) to 81.7% (99.9th percentile, 95% CI 75.17–88.34) (Fig. 3b). Notably, non-carriers in the 99.9th

T2D PGS percentile showed a 17.7% risk (95% CI 17.3–18.2), which was similar to mutation carriers with lowest T2D PGS. A sensitivity analysis using T2D PGS, which excluded variants within 1 Mb of the three MODY genes, showed similar effect sizes (OR 2.17, 1.2–3.91 per 1 s.d. change whole PGS versus 2.06, 1.16–3.82 without MODY genes). These data together suggest a substantial contribution of T2D polygenic background on diabetes risk in HNF-MODY, while some individuals without MODY mutations but with extreme polygenic risk may reach a similar risk as HNF-MODY.
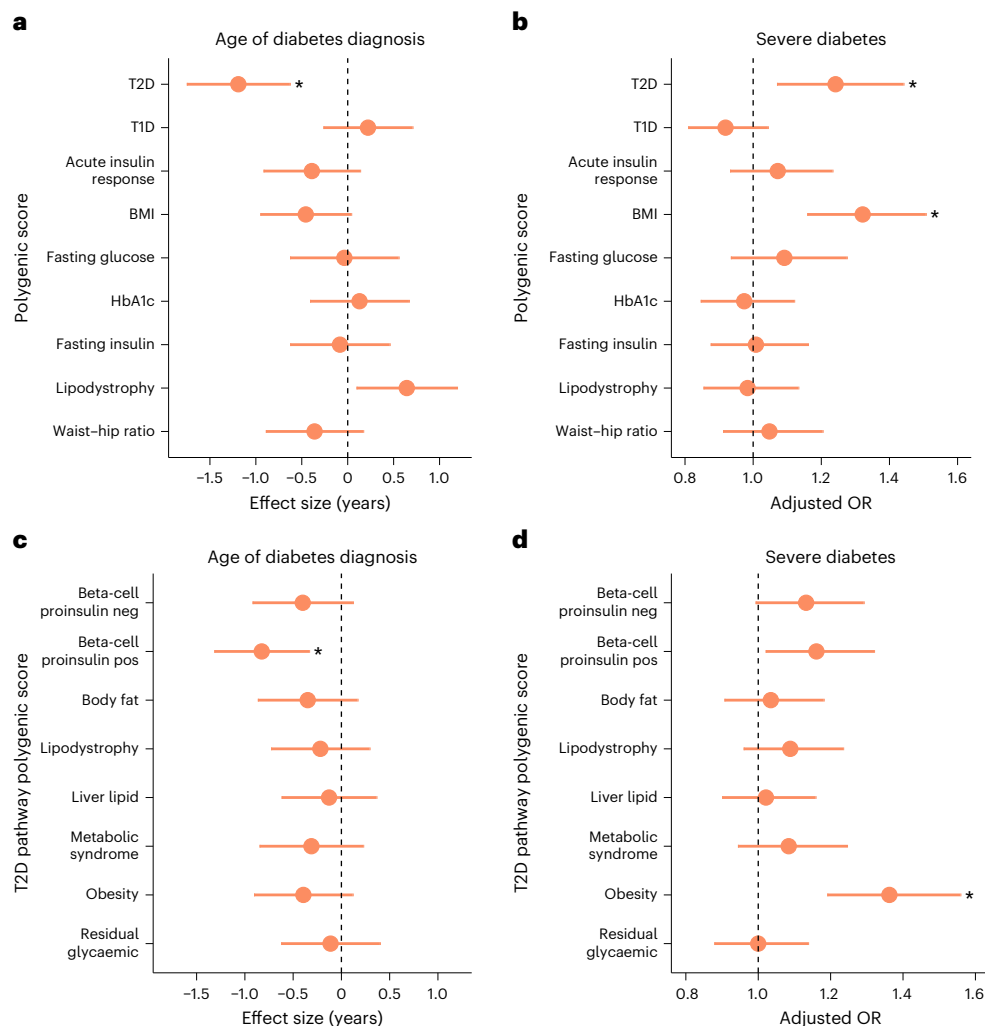
Fig. 2 | Increased polygenic risk associated with earlier and more severe diabetes diagnosis in HNF-MODY. a, Association between polygenic scores for upper-level diabetes-related traits and age of diabetes diagnosis. Estimates were derived using a mixed-effects linear model with family as a random effect and adjusted for other polygenic scores and the first ten within-cohort principal components. **b**, Association between polygenic scores and risk of severe diabetes (defined as HbA1c ≥ 8.5% or insulin treatment at recruitment), using a mixed-effects logistic model with the same covariates as in **a**. In total, 676 out of 1,462 MODY carriers met the criteria for severe diabetes. **c,d**, Association of T2D-partitioned risk scores with age of diabetes diagnosis (**c**) and diabetes severity (**d**), estimated using linear mixed-effects models adjusted for the first ten within-cohort principal components. Dots represent the estimates, with lines indicating 95% CIs. Asterisks highlight significant differences ($P < 0.0056$), after Bonferroni correction. Estimates represent the effect of a 1 s.d. increase in the respective polygenic score. All analyses in **a**–**d** were conducted in 1,462 MODY cases.

Common genetic variants explain 24% of phenotypic variance in MODY

Having observed substantial contribution of polygenic background, we next aimed to quantify the overall contribution of common genetic variants to MODY. Using genome-wide complex trait analysis (GCTA) genome-based restricted maximum likelihood (GREML), we estimated common variant (minor allele frequency > 0.01) single nucleotide polymorphism (SNP) heritability (h^2), on the liability scale. Our analysis revealed a SNP heritability of 23.9% (95% CI 17.2–30.7%, $P < 0.0001$) in individuals with HNF-MODY (Fig. 4). This estimate was only slightly lower than in polygenic T2D cases 30.8% (95% CI 25.08–36.61%, $P < 0.0001$). The heritability estimate remained consistent across multiple approaches, including restricting to HNF1A-related monogenic diabetes, phenotype correlation–genotype correlation regression and applying GREML estimation in linkage-disequilibrium adjusted kinships (LDAK) (Supplementary Table 12). To determine how much of this common variant heritability stems from T2D-associated variants, we calculated SNP-heritability for MODY comparing against 4,461 T2D

cases, both with and without T2D PGS adjustment. The heritability decreased to 20.3% when compared with T2D, and further dropped to 17.2% (95% CI 4.7–29.7%, $P = 0.035$) after T2D PGS adjustment (Fig. 4). These findings reveal that common genetic variants substantially influence MODY's clinical presentation. At least one-third of this influence comes from T2D variants, suggesting the presence of T2D-independent genetic modifiers in HNF-MODY.

Clinically referred MODY cases without a pathogenic variant have substantially higher polygenic burden of T2D and related traits

Following our observations of substantial common genetic variant contributions in patients with mutation-positive MODY, we investigated whether higher polygenic background could also explain diabetes in individuals with a MODY phenotype but without causative mutations in known monogenic diabetes genes. We studied 300 individuals referred for MODY genetic testing from routine clinical practice with diabetes diagnosis before age 30 and BMI < 30 kg m^{-2}, and without evidence of T1D (positive islet autoantibodies, C-peptide <200 pmol l^{-1},

a

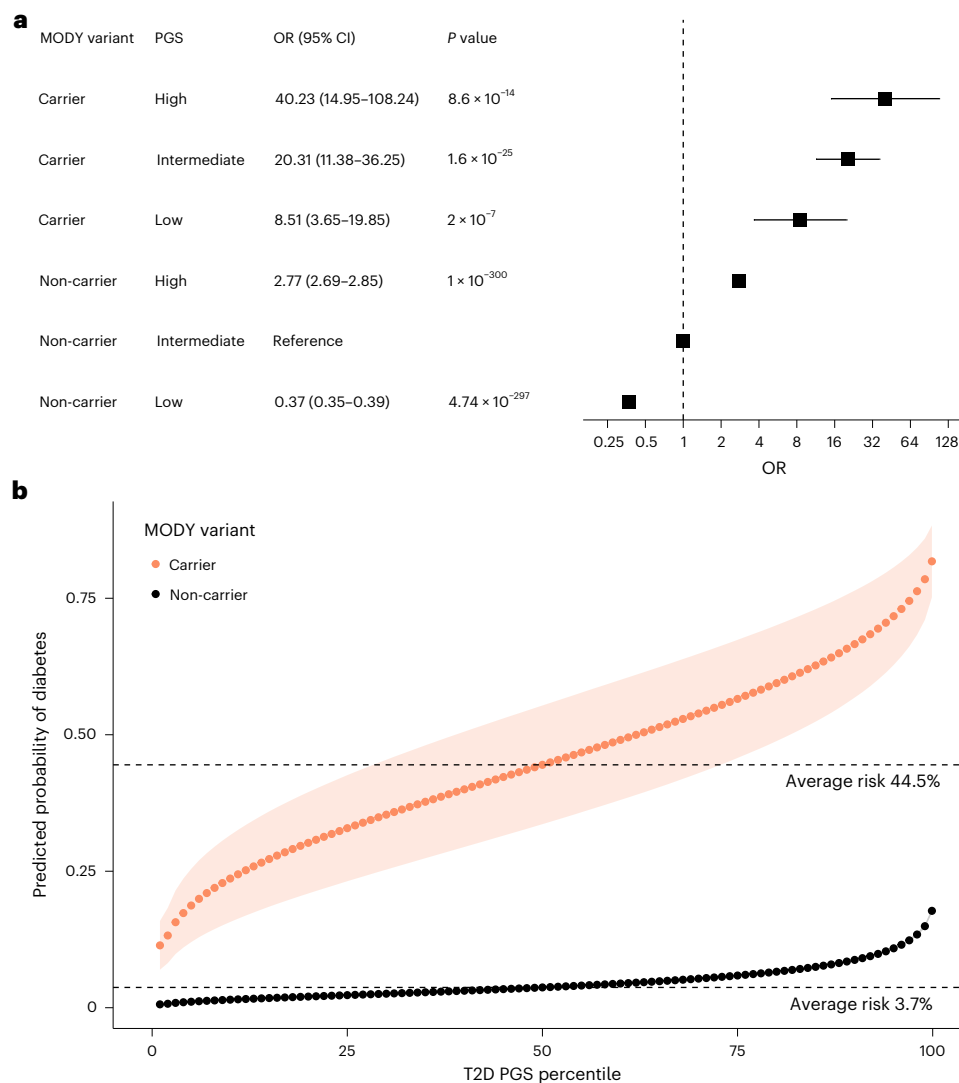| MODY variant	PGS	OR (95% CI)	P value
Carrier	High	40.23 (14.95–108.24)	8.6×10^{-14}
Carrier	Intermediate	20.31 (11.38–36.25)	1.6×10^{-25}
Carrier	Low	8.51 (3.65–19.85)	2×10^{-7}
Non-carrier	High	2.77 (2.69–2.85)	1×10^{-300}
Non-carrier	Intermediate	Reference	
Non-carrier	Low	0.37 (0.35–0.39)	4.74×10^{-297}

Fig. 3 | Polygenic background modifies diabetes risk in clinically unselected HNF-MODY carriers. a, ORs for diabetes risk in HNF-MODY carriers and non-carriers, stratified by T2D polygenic risk levels. Points represent ORs with error bars representing 95% CIs. Low risk is defined as the bottom 20% of T2D polygenic scores, high risk as the top 20% and intermediate risk as the remaining 60%. n = 88,905, 265,079 and 70,469 for non-carriers and n = 28, 55 and 17 for carriers, for low, intermediate and high T2D PGS risk, respectively. Diabetes risk was estimated using a two-sided logistic regression model adjusted for sex, age, family history of diabetes, the first ten ancestry principal components and BMI. **b**, Predicted probability of diabetes at baseline across each percentile of T2D polygenic risk, assessed using a logistic model with T2D polygenic score as a continuous variable and the same covariates as in **a**. Points represent predicted mean probability per percentile, and shaded regions represent 95% CIs. Dashed lines indicate the baseline diabetes risk at the 50th percentile of T2D polygenic scores for HNF-MODY carriers and non-carriers.

or high T1D genetic risk score >50th centile of the T1D population)[22]. These unsolved MODY cases showed similar age of diagnosis and BMI to mutation-positive MODY cases ($P > 0.05$ for both) (Fig. 5a,b and Supplementary Table 13). As expected, these unsolved cases showed no excess T1D PGS but displayed a striking 1.18 s.d. (95% CI 1.07–1.29, $P < 0.0001$) higher T2D PGS than controls (Fig. 5c). This polygenic burden was higher than both mutation-positive MODY cases by 0.73 s.d. and T2D cases by 0.52 s.d. (all $P < 0.0001$) (Extended Data Fig. 4). Compared with controls, we also observed an excess polygenic burden of BMI and waist–hip ratio in these cases (Fig. 5d,e). Unsolved cases demonstrated an enrichment in all T2D partitioned PGSs, with the largest difference from controls in the beta-cell proinsulin-positive cluster (0.62 s.d. increase, 95% CI 0.51–0.74, $P < 0.0001$) (Extended Data Fig. 5). Excess biparental diabetes history further supported the observed excess polygenic enrichment in unsolved cases compared with T2D (53% one parent, 15.7% both parents with diabetes versus 28.9% and 4%, respectively in T2D) (Extended Data Figure 6). These findings

suggest that while some unsolved cases may harbour novel monogenic diabetes mutations, many likely represent polygenic phenocopies driven by an excessive polygenic burden of T2D and related traits.

Discussion

In this study, we demonstrate that HNF-MODY has a significant polygenic component, with common genetic variation substantially influencing disease onset and severity in genetically confirmed MODY cases. The elevated polygenic burden of T2D-related traits may explain MODY phenocopies lacking pathogenic mutations.

MODY's genetic architecture seems more complex than its traditional characterization as a purely monogenic disorder. We found that common genetic variations explain approximately 24% of phenotypic variance in clinically identified cases. This estimate is substantially higher than previously reported in other monogenic disorders (long QT syndrome, 15%[16] and developmental delay, 11% (ref. 23)) and approaches that of T2D. Such high polygenic contribution is unexpected for a
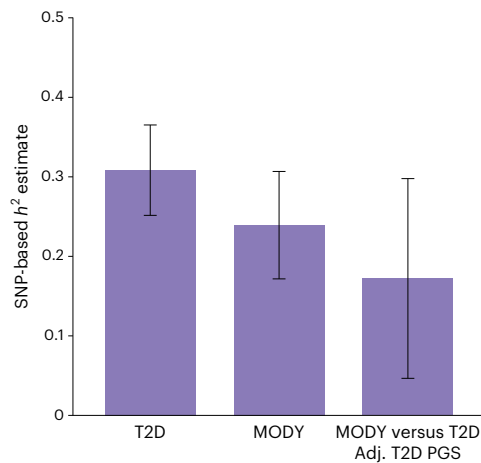
Fig. 4 | SNP-based heritability estimates for MODY. SNP-based h^2 estimates were calculated in unrelated individuals using GCTA GREML-LDMS, stratified into four LD bins of equal size to construct the genetic relationship matrix, with sex, age and the first ten principal components within the cohort as covariates. Bars represent heritability point estimates and error bars represent the 95% confidence intervals. Estimates for HNF-MODY carriers ($n = 864$) and T2D cases ($n = 4,461$) were compared with non-diabetic controls ($n = 6,935$). h^2 is shown on the liability scale for T2D (prevalence, 0.1) and MODY (prevalence, 0.0005). MODY and T2D cases were compared, adjusting for T2D polygenic score.

presumed monogenic disease and may reflect its young-onset nature. T2D-associated variants had the strongest effect among the traits we analysed, likely because of shared pathways in beta-cell dysfunction. We found that T2D polygenic risk influenced age at diagnosis mainly through proinsulin-associated beta-cell pathways, which supports current understanding that HNF-MODY arises primarily from beta-cell dysfunction. Our findings on the relationship between T2D polygenic scores and age at diagnosis are consistent with smaller studies in HNF1A-MODY[17,18]. The absence of interaction with T1D polygenic risk aligns with the current understanding that T1D variants primarily affect autoimmune pathways rather than transcriptional networks disrupted in HNF-MODY[24]. This genetic distinction supports using T1D polygenic risk scores to differentiate MODY from early-onset T1D[22,25].

Our findings reveal distinct genetic pathways modifying different aspects of MODY. Beta-cell proinsulin-related variants predominantly influence age of diagnosis, while obesity-associated variants and beta-cell pathways drive disease progression. This supports a liability threshold model where pathogenic MODY variants drive early-onset disease, with the polygenic background modifying overall disease risk. We observed that the polygenic contribution is not constant but depends on underlying pathogenic variant where less-damaging variants require more contribution for disease expression and clinical diagnosis. Importantly, we show that T2D polygenic risk strongly modifies diabetes risk in individuals carrying pathogenic MODY variants. Previous studies have reported similar effects for intermediate-effect variants in HNF1A and HNF4A in population cohorts, where common T2D risk variants altered the penetrance of those rare alleles[19]. Notably, pathogenic variant carriers with low T2D polygenic risk show substantially lower diabetes risk, with about half remaining disease-free in the population cohort. This explains the disparity between MODY prevalence in clinical referrals (1:10,000) versus genetic screening (1:2,000)[4,26]. Together, these data demonstrate that MODY's pathogenesis involves substantial polygenic interaction rather than following a simple deterministic monogenic model.

Some unsolved MODY cases may represent polygenic phenocopies. Our small cohort of mutation-negative cases shows substantial enrichment of T2D polygenic risk exceeding that seen in typical T2D. This enrichment extends beyond T2D to other related traits,

supporting complex polygenic aetiology. Similar patterns are observed in other monogenic conditions like long QT syndrome[16], developmental delay[23] and familial hypercholesterolaemia[14], where mutation-negative patients show higher polygenic burden than mutation-positive cases. These unsolved cases likely represent a heterogeneous group with multiple underlying causes, including potential overlap with previously defined T2D subtypes[27,28]. Although our sample size limited detailed clustering, it is plausible that some individuals may align with distinct mechanistic pathways, as observed in these subgroups of T2D. While some may resemble the severe insulin-deficient diabetes subgroup, our findings suggest broader enrichment across all T2D risk pathways. This implies that the unsolved MODY group does not map cleanly onto existing subtypes. It likely includes individuals at the extreme tail of the polygenic risk distribution, possibly carrying rare, low-penetrance variants that act additively with high polygenic burden to drive clinical referral. Collectively, these findings suggest the presence of polygenic phenocopies. However, due to the relatively small sample size, these results should be considered as preliminary. Further studies are needed to replicate these observations and elucidate the underlying mechanisms.

Our findings support the hypothesis that monogenic disorders exist on a continuum, where both pathogenic mutations and polygenic background shape disease manifestation[29]. Age-dependent conditions, such as MODY, are likely to have a larger polygenic contribution compared with neonatal-onset disorders. As evidence accumulates, this observation may extend to the majority of monogenic disorders, albeit to varying degrees. However, each condition will require individual evaluation to quantify the relative contributions. With the declining cost of genetic testing and the increasing identification of presymptomatic carriers through incidental findings[30] and newborn screening programmes[31], there is a growing need to refine disease risk prediction. Currently, risk assessment relies solely on the presence of pathogenic mutations. To provide more precise risk stratification, it may be necessary to incorporate non-mutation factors, such as polygenic risk scores or family history, as is already done in conditions such as breast cancer[32]. As whole-genome sequencing moves toward becoming a first-line test, a single assessment could offer comprehensive genetic information, incorporating both monogenic and polygenic risk. However, clinical implementation will require large-scale, multi-ancestry MODY datasets and collaboration across dedicated cohorts to enable robust model development, validation and equitable application. Further research is needed to evaluate the added clinical value of this approach in improving diagnosis and risk prediction.

Although this is the largest MODY study to date, the sample size for individual genes and unsolved cases limited our power to detect subgroup-specific effects. Despite this, the direction and strength of associations were consistent across the HNF-MODY subtypes, supporting the generalizability of our findings. The predominantly UK-based, European ancestry cohorts limit generalizability to other populations. While MODY variants in the UK Biobank were not Sanger-confirmed, we minimized false positives through manual IGV review and strict quality filtering. The UK Biobank's healthy volunteer bias likely underestimates true MODY penetrance in general populations due to underrepresentation of early-onset diabetes. Furthermore, our sample size limited more detailed analysis of non-clinically referred HNF-MODY carriers. Our local MODY cohort is derived from routine clinical referrals across the UK, so case ascertainment may be influenced by environmental factors such as healthcare access, socioeconomic status, clinical practice variability and other unmeasured confounders. We adjusted our analyses of age at diagnosis and disease severity for several known or measurable confounding factors, including variant location, family ID, proband status, BMI, parental diabetes, birthweight and year of diagnosis. However, we could not adjust for lifestyle-related factors such as diet, physical activity, early-life factors, educational attainment, social status, treatment preference or adherence. While these
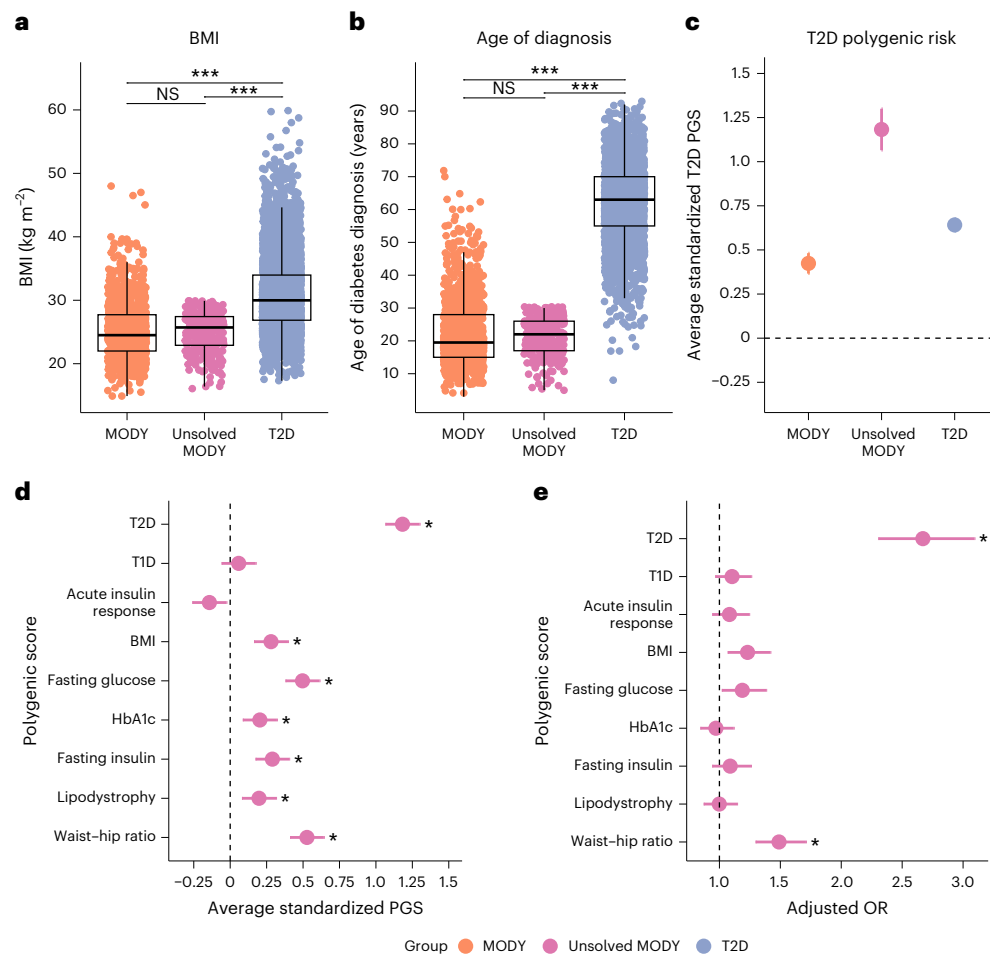
Fig. 5 | Unsolved MODY cases exhibit extreme T2D polygenic risk while phenotypically resembling typical MODY cases. Distribution of key characteristics and polygenic risk across MODY carriers (orange, n = 1,462), T2D cases (blue, n = 4,773) and unsolved MODY cases (pink, n = 300). **a,b**, Distribution of the clinical characteristics BMI (**a**) and age at diabetes diagnosis (**b**), collected at patient referral. Box plots display the median and interquartile range, with individual data points overlaid. Statistical significance between groups is indicated; ***$P < 0.017$; NS, not significant, as determined by two-sided t-tests. **c**, Mean T2D polygenic risk across groups versus controls (n = 7,645), assessed using a two-sided linear regression model adjusting for the first ten within-cohort principal components. **d**, Mean polygenic risk score differences for diabetes-related traits in unsolved MODY cases versus controls, using the same method as in **c. e**, Adjusted ORs for unsolved MODY cases versus controls, assessed using a two-sided logistic regression model including each PGS, sex, age, BMI and the first ten within-cohort principal components as covariates. In **d** and **e**, asterisks denote Bonferroni-corrected significant differences from controls ($P < 0.0056$). In **c**–**e**, dots represent point estimates, with error bars representing 95% CIs.

limitations remain, the large sample size may mitigate some of their effects. The observation of similar associations between T2D risk and age at diagnosis in multigenerational pedigrees from Finland[18] supports the robustness of our findings and suggests that these biases important but does not explain the all the results.

In summary, using MODY as a model disease, we demonstrate substantial interplay between monogenic mutations and polygenic background in young-onset monogenic disorders. Our findings suggest that future approaches to disease prediction will require integration of monogenic, polygenic and environmental factors to improve clinical utility.

Methods
Study populations
This study complies with all relevant ethical regulation and was approved by the appropriate ethics committees. Our study combined three ethically approved cohorts. In our local MODY cohort, all probands or their guardians provided informed consent, and the North Wales Ethics Committee approved the study, with Genetic Beta Cell Research Bank approving sample access. The National Institute for Health Research (NIHR) Exeter Clinical Research Facility management

committee approved access to these samples and genotype data for our T2D and non-diabetic controls. This research also utilized data from the UK Biobank resource carried out under UK Biobank application number 103356. UK Biobank protocols were approved by the National Research Ethics Service Committee.

Exeter MODY cohort
MODY individuals with confirmed pathogenic variants. We analysed individuals referred for monogenic diabetes genetic testing at the Exeter Genomics Laboratory, Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK. These referrals originated from clinical suspicion of MODY during routine clinical care in the UK. These individuals were found to have likely pathogenic or pathogenic variants either by Sanger sequencing or gene panel test performed as part of routine clinical care. Our cohort comprised European individuals with diabetes and carrying pathogenic variants in *HNF1A* (n = 997), *HNF1B* (n = 145) or *HNF4A* (n = 320). We focused on the more commonly diagnosed, age-dependent forms of MODY (*HNF1A*, *HNF4A* and *HNF1B*). We excluded GCK-MODY because it represents a fundamentally different disease: individuals present with lifelong, mild fasting hyperglycaemia that does not progress with age, does not require treatment and is not

associated with excess complications[33]. In this context, age at diagnosis reflects the timing of detection rather than age at disease onset.

Unsolved MODY individuals. We evaluated 300 European individuals referred from routine clinical care in the UK with suspected MODY. All participants received their diabetes diagnosis before age 30 years and lacked clinical features suggestive of T2D (BMI ≥ 30 kg m^{-2}) or T1D (positive islet autoantibodies, C-peptide <200 pmol l^{-1} and a ten-SNP T1D genetic risk score above the 50th centile of the gold-standard T1D population from the WTCCC study)[22]. These individuals underwent comprehensive genetic testing for all known monogenic diabetes genes (*n* = 58) and were not found to have pathogenic variants in these genes. The clinical features of these solved and unsolved MODY cases, at referral for genetic testing, are summarized in Supplementary Table 1.

Type 2 diabetes and non-diabetes control cohort

We analysed participants from two ethically approved population cohorts in Southwest England: the Exeter 10000 study (https://exetercrfnihr.org/about/exeter-10000/)[34] and the Diabetes Alliance for Research in England study (https://www.diabetesgenes.org/current-research/dare/)[35]. These studies recruited unselected participants through primary care practices across the Southwest United Kingdom. At recruitment, participants completed baseline questionnaires and provided fasting blood and urine samples for measurement of diabetes-related markers, including fasting glucose and HbA1c. Our analysis included European individuals who underwent array genotyping as part of these studies. We classified participants as having T2D if they either did not require insulin treatment or initiated insulin treatment after 36 months from diagnosis, thereby excluding potential misclassified T1D cases. We defined controls as individuals without a known diabetes diagnosis and HbA1c ≤ 48 mmol mol^{-1} (6.5%)[36]. The final cohort comprised 7,645 controls and 4,773 individuals with T2D, with their clinical characteristics presented in Supplementary Table 1.

UK Biobank cohort

The UK Biobank represents a large-scale, prospective population-based study comprising approximately 500,000 UK residents aged 40–70 years at enrolment[37]. Recruitment occurred between 2006 and 2010, with comprehensive data collection through multiple channels: participant questionnaires, structured interviews and biomarker measurements[37]. The study supplemented this information with medical history data from Hospital Episode Statistics records coded using ICD-9 and ICD-10 codes. We defined diabetes status using three criteria: self-reported diagnosis, HbA1c levels ≥6.5% at recruitment or active diabetes treatment at recruitment. Our study cohort consisted of 424,553 European individuals who underwent exome sequencing and array genotyping. Clinical characteristics of these individuals can be found in Supplementary Table 10. We analysed the exome sequence data to identify individuals with likely pathogenic and pathogenic variants in *HNF1A*/*HNF4A*/*HNF1B* as described previously[4], with details of variants identified in Supplementary Table 11.

Genetic analysis

MODY pathogenic variants in Exeter MODY cohort and UK Biobank. For the Exeter MODY cohort, all referred patients were screened for potential MODY-associated variants using either Sanger sequencing or gene panel testing, following the methodologies detailed by Ellard et al.[38]. For the UK Biobank participants, we utilized exome sequence data to identify carriers of pathogenic MODY variants. We annotated all variants using clinically validated transcripts: GenBank NM_000545.6 for *HNF1A*, NM_000458.4 for *HNF1B* and NM_175914.4 for *HNF4A*. We classified variants according to the American College of Medical Genetics and Genomics/Association of Molecular Pathology guidelines, designating them as either likely pathogenic (class 4) or pathogenic (class 5)[39]. This classification process followed our established protocols for the

local Exeter cohort and aligned with our recent study's methodology[4]. Supplementary Table 11 presents a comprehensive list of variants identified in the UK Biobank cohort.

Array genotyping. *Exeter MODY, T2D and non-diabetic controls.* We performed array genotyping using the Infinium Global Screening Array platform. Our comprehensive quality control protocol excluded samples with call rates below 98%, sex mismatches, relationship discrepancies or inbreeding coefficients exceeding 0.1. At the variant level, we removed markers with missingness above 2%, minor allele frequency below 5% or deviation from the Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$). We applied these quality control measures both independently for each batch and following batch integration. We then used linkage disequilibrium (LD) pruned markers for genotype imputation through the TOPMed reference panel v.2 (ref. 40) via the Michigan Imputation Server[41]. To determine genetic ancestry, we compared our data with reference populations from the 1000 Genomes Phase 3 and Human Genome Diversity Project, implementing a principal component analysis (PCA) approach within the GenoPred Pipeline (v.2.2.1)[42,43]. For relationship inference, we analysed LD-pruned data using the KING robust algorithm (v.2.2.4) to identify unrelated individuals up to the third degree[44]. To better capture the within-cohort population structure, we conducted PCA using FlashPCA (v.2.0)[45]. Initially, we calculated principal components in unrelated European individuals and then projected these onto related European individuals.

UK Biobank. The UK Biobank individuals were SNP-genotyped using the UK BiLEVE array for the first ~50,000 individuals, with the remaining using the UK Biobank Axiom array. This dataset underwent central quality control by the UK Biobank and was imputed to the TOPMed reference panel[40]. Approximately 450,000 individuals from the UK Biobank Array also underwent exome sequencing using the IDT xGen Exome Research Panel v.1.0. Detailed sequencing methodology for UK Biobank samples has been described previously[46]. In brief, variants were called using GATK v.3.0 filtering variants with an inbreeding coefficient <−0.03 or without at least one variant genotype of DP ≥ 10, GQ ≥ 20 and, if heterozygous, AB ≥ 0.20. For this analysis, we included 424,553 individuals who had both exome and array data and were of European ancestry, inferred from projected PCA using the same approach as for the local cohort.

Polygenic score calculation

We calculated polygenic scores for T2D[20], T1D[47] and seven diabetes-related traits[48–52], alongside eight pathway-specific T2D risk scores[20]. We constructed weighted polygenic scores using genome-wide significant variants. For traits with comprehensive genome-wide association study (GWAS) summary statistics available, we implemented genome-wide calculations to capture additional genetic signal. Our computational pipeline utilized PLINK 1.9's score function for genome-wide significant variant-based scores[53]. For the genome-wide polygenic scores, we implemented the GenoPred v.2.2.1 pipeline with the LDpred2 auto model, enabling comprehensive processing of GWAS summary statistics[43,54]. Further details are available in Supplementary Table 3, including the specific approach used for each trait, including the calculation method, number of variants incorporated and the source GWAS studies.

Heritability estimation

To estimate the common variant contribution to MODY and T2D, SNP-based heritability was estimated in unrelated individuals using GCTA GREML-LDMS, stratifying into four LD bins of equal size to construct the genetic relationship matrix[55]. To test the validity of these estimates we ran phenotype correlation–genotype correlation and restricted maximum likelihood approaches implemented in LDAK, using thinned predictors to construct the kinship matrix[56,57]. We used sex, age

and the first ten within-cohort principal components as covariates for each method. For MODY, disease prevalence was set at 0.0005[4] and 0.00025[26], and for T2D, at 0.1[58] (Table S12). Variants with an imputation quality > 0.9 and minor allele frequency >1% were used to in this analysis.

Statistical analysis

Assessing common variant enrichment in MODY cohort. To assess polygenic risk in MODY carriers and T2D cases, we employed several different approaches. To initially assess whether any common variant pathways contribute to clinically referred HNF-MODY we tested nine PGSs for enrichment. All scores were standardized using the control group as reference (mean = 0, s.d. = 1). To test differences in polygenic scores from controls, we used linear models adjusting within-cohort principal components to control for population structure. We assessed each score individually first, however, due to well-known overlaps of variants across these related metabolic traits, we then performed multivariable logistic regression analysis to identify the key independent common variant pathways contributing to HNF-MODY after adjusting for sex, age, BMI and the first ten within-cohort principal components. We repeated these steps with unsolved MODY cases to examine the hypothesis that they have excess polygenic risk. Owing to the high parental history in MODY that may tag inherited polygenic risk, we then performed further analysis adjusting for parental history of diabetes. We performed sensitivity analysis by limiting to each gene and to probands alone. To investigate whether less-deleterious variants are associated with higher polygenic enrichment, we first grouped variants into missense and protein-truncating variants (PTVs), with PTVs assumed to be the most deleterious due to their likely haploinsufficiency effect. We further stratified missense variants by REVEL[59] (Rare Exome Variant Ensemble Learner) score (<0.9 versus ≥0.9), using it as a proxy for functional severity.

We aimed to include the largest number of MODY cases to maximize the power of the study but were limited by sample and data availability. Based on our final sample size, a post hoc power calculation suggested that we had 80% power to detect minimum differences of 0.08, 0.16 and 0.05 s.d. in polygenic score between controls and genetically confirmed HNF-MODY, unsolved MODY and T2D, respectively. The minimum detectable differences for the clinically referred MODY genetic subgroups were 0.23, 0.16 and 0.094 s.d. for HNF1A, HNF1B and HNF4A, respectively.

Assessing impact of common variants on HNF-MODY phenotype. To investigate how common genetic variants influence the clinical presentation of HNF-MODY, we used mixed-effects models to assess associations between PGSs and key outcomes. Specifically, we applied mixed linear models to evaluate the relationship between PGSs and age at diabetes diagnosis, and mixed logistic models to assess associations with diabetes severity. To account for potential within-family correlations that could bias associations, all models included family ID as a random effect. Initial models included all nine polygenic scores to identify independent genetic pathways contributing to variation in clinical presentation. Further analysis focused on scores that were found to be independently associated with modifying the clinical presentation in HNF-MODY, further adjusting for confounding factors that have been previously reported or suspected to influence clinical outcomes. This included sex, age, BMI, year of diabetes diagnosis, proband or family member, variant location, parental history of diabetes (stratified by mother, father or both to capture potential intrauterine exposure), along with the first ten within-cohort principal components. To account for gene-level differences, we included genetic aetiology (MODY gene) as a covariate and examined outcomes separately by gene.

Assessing impact of common variants on clinically unselected HNF-MODY carriers. HNF-MODY carriers in the UK Biobank allowed us to assess how common variants affect diabetes risk in a clinically unselected setting. We modelled the probability of diabetes using logistic regression, with T2D PGS as a continuous covariate alongside MODY carrier status and relevant clinical characteristics including sex, age, BMI, parental history of diabetes and the first ten ancestry principal components. Among clinically unselected HNF-MODY carriers, we had 80% power to detect an OR greater than 1.58 per s.d. increase in T2D PGS, below the observed effect size of 2.17. To examine how diabetes risk varies across T2D common variant burden, we computed marginal effects per PGS percentile. Additionally, individuals were stratified into low, intermediate or high PGS groups, defined as the bottom quintile, middle three quintiles and top quintile, respectively, using non-MODY carriers with intermediate T2D risk as the reference group. We used logistic regression to assess differences in diabetes risk relative to the reference group, adjusting for the same covariates.

All statistical analyses were performed using R v.4.4.1 and Stata v.18.

Reporting summary

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

Data availability

The data supporting the findings of this study are available within the article, source data and its Supplementary Information. The clinical data, including individual level data, generated and/or analysed as part of this study are not publicly available because of patient confidentiality and ethical approval associated with the data but are available from the corresponding authors upon reasonable request. The UK Biobank dataset is available from https://biobank.ctsu.ox.ac.uk. Source data are provided with this paper.

Code availability

Publicly available software was used for all analysis, as outlined in the Methods. Code supporting this manuscript is available at https://github.com/ExeterGenetics/teamendo_manuscripts/tree/main/Polygenic_MODY_2025.

References

1. Kingdom, R., Beaumont, R. N., Wood, A. R., Weedon, M. N. & Wright, C. F. Genetic modifiers of rare variants in monogenic developmental disorder loci. *Nat. Genet.* **56**, 861–868 (2024).
2. Tukker, A. M., Royal, C. D., Bowman, A. B. & McAllister, K. A. The impact of environmental factors on monogenic Mendelian diseases. *Toxicol. Sci.* https://doi.org/10.1093/toxsci/kfab022 (2021).
3. Kingdom, R. & Wright, C. F. Incomplete penetrance and variable expressivity: from clinical studies to population cohorts. *Front. Genet.* https://doi.org/10.3389/fgene.2022.920390 (2022).
4. Mirshahi, U. L. et al. Reduced penetrance of MODY-associated HNF1A/HNF4A variants but not GCK variants in clinically unselected cohorts. *Am. J. Hum. Genet.* **109**, 2018–2028 (2022).
5. Cho, B. P. H. et al. Association of vascular risk factors and genetic factors with penetrance of variants causing monogenic stroke. *JAMA Neurol.* **79**, 1303–1311 (2022).
6. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* https://doi.org/10.1007/s00439-013-1331-2 (2013).
7. Khan, A. et al. Polygenic risk alters the penetrance of monogenic kidney disease. *Nat. Commun.* **14**, 8318 (2023).
8. Fahed, A. C. et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat. Commun.* **11**, 3635 (2020).

9.   Shepherd, M. H. et al. A UK nationwide prospective study of treatment change in MODY: genetic subtype and clinical characteristics predict optimal glycaemic control after discontinuing insulin and metformin. *Diabetologia* **61**, 2520–2527 (2018).

10. Shields, B. M. et al. Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia* **53**, 2504–2508 (2010).

11. Hattersley, A. T. & Patel, K. A. Precision diabetes: learning from monogenic diabetes. *Diabetologia* https://doi.org/10.1007/s00125-017-4226-2 (2017).

12. Colclough, K., Ellard, S., Hattersley, A. & Patel, K. Syndromic monogenic diabetes genes should be tested in patients with a clinical suspicion of maturity-onset diabetes of the young. *Diabetes* **71**, 530–537 (2022).

13. Downie, L. et al. Gene selection for genomic newborn screening: moving toward consensus? *Genet. Med.* **26**, 101077 (2024).

14. Nomura, A. et al. Polygenic risk scores for low-density lipoprotein cholesterol and familial hypercholesterolemia. *J. Hum. Genet.* **66**, 1079–1087 (2021).

15. Chami, N., Preuss, M., Walker, R. W., Moscati, A. & Loos, R. J. F. The role of polygenic susceptibility to obesity among carriers of pathogenic mutations in MC4R in the UK Biobank population. *PLoS Med.* **17**, e1003196 (2020).

16. Lahrouchi, N. et al. Transethnic genome-wide association study provides insights in the genetic architecture and heritability of long QT syndrome. *Circulation* **142**, 324–338 (2020).

17. Allen, H. L. et al. Polygenic risk variants for type 2 diabetes susceptibility modify age at diagnosis in monogenic HNF1A diabetes. *Diabetes* **59**, 266–271 (2010).

18. Kettunen, J. L. T. et al. A multigenerational study on phenotypic consequences of the most common causal variant of HNF1A-MODY. *Diabetologia* **65**, 632–643 (2022).

19. Huerta-Chagoya, A. et al. Rare variant analyses in 51,256 type 2 diabetes cases and 370,487 controls reveal the pathogenicity spectrum of monogenic diabetes genes. *Nat. Genet.* **56**, 2370–2379 (2024).

20. Suzuki, K. et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347–357 (2024).

21. Zhou, K. et al. Clinical and genetic determinants of progression of type 2 diabetes: a direct study. *Diabetes Care* **37**, 718–724 (2014).

22. Patel, K. A. et al. Type 1 diabetes genetic risk score: a novel tool to discriminate monogenic and type 1 diabetes. *Diabetes* **65**, 2094–2099 (2016).

23. Huang, Q. Q. et al. Examining the role of common variants in rare neurodevelopmental conditions. *Nature* **636**, 404–411 (2024).

24. Onengut-Gumuscu, S. et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).

25. 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).

26. Pang, L. et al. Improvements in awareness and testing have led to a threefold increase over 10 years in the identification of monogenic diabetes in the UK. *Diabetes Care* **45**, 642–649 (2022).

27. Ahlqvist, E. et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369 (2018).

28. Mansour Aly, D. et al. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat. Genet.* **53**, 1534–1542 (2021).

29. Le Collen, L., Froguel, P. & Bonnefond, A. Towards the recognition of oligogenic forms of type 2 diabetes. *Trends Endocrinol. Metab.* **36**, 109–117 (2025).

30. Watson, M. S. ACMG policy statement: Updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genet. Med.* https://doi.org/10.1038/gim.2014.151 (2015).

31. Horton, R. et al. Challenges of using whole genome sequencing in population newborn screening. *Brit. Med. J.* **384**, e077060 (2024).

32. Lee, A. et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet. Med.* **21**, 1708–1718 (2019).

33. Chakera, A. J. et al. Recognition and management of individuals with hyperglycemia because of a heterozygous glucokinase mutation. *Diabetes Care* https://doi.org/10.2337/dc14-2769 (2015).

34. Rodgers, L. R. et al. Choice of HbA1c threshold for identifying individuals at high risk of type 2 diabetes and implications for diabetes prevention programmes: a cohort study. *BMC Med.* **19**, 184 (2021).

35. Thomas, N. J. et al. Type 1 diabetes defined by severe insulin deficiency occurs after 30 years of age and is commonly treated as type 2 diabetes. *Diabetologia* **62**, 1167–1172 (2019).

36. American Diabetes Association Professional Practice Committee. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2022. *Diabetes Care* **45**, S17–S38 (2022).

37. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

38. Ellard, S. et al. Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. *Diabetologia* **56**, 1958–1963 (2013).

39. 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).

40. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

41. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

42. Koenig, Z. et al. A harmonized public resource of deeply sequenced diverse human genomes. *Genome Res.* **34**, 796–809 (2024).

43. Pain, O., Al-Chalabi, A. & Lewis, C. M. The GenoPred pipeline: a comprehensive and scalable pipeline for polygenic scoring. *Bioinformatics* **40**, btae551 (2024).

44. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

45. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776–2778 (2017).

46. Szustakowski, J. D. et al. Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat. Genet.* https://doi.org/10.1038/s41588-021-00885-0 (2021).

47. Sharp, S. A. et al. Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. *Diabetes Care* **42**, 200–207 (2019).

48. Wood, A. R. et al. A genome-wide association study of IVGTT-based measures of first-phase insulin secretion refines the underlying physiology of type 2 diabetes variants. *Diabetes* **66**, 2296–2309 (2017).

49. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

50. Chen, J. et al. The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**, 840–860 (2021).

51. Pulit, S. L. et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).

52. Lotta, L. A. et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17–26 (2017).

53. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

54. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2020).

55. Yang, J. et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).

56. Golan, D., Lander, E. S. & Rosset, S. Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl Acad. Sci. USA* **111**, E5272–E5281 (2014).

57. Weissbrod, O., Flint, J. & Rosset, S. Estimating SNP-based heritability and genetic correlation in case-control studies directly and with summary statistics. *Am. J. Hum. Genet.* **103**, 89–99 (2018).

58. Xue, A. et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* **9**, 2941 (2018).

59. Ioannidis, N. M. et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).

Acknowledgements

Author contributions

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s42255-025-01372-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42255-025-01372-0.

Correspondence and requests for materials should be addressed to Kashyap A. Patel.
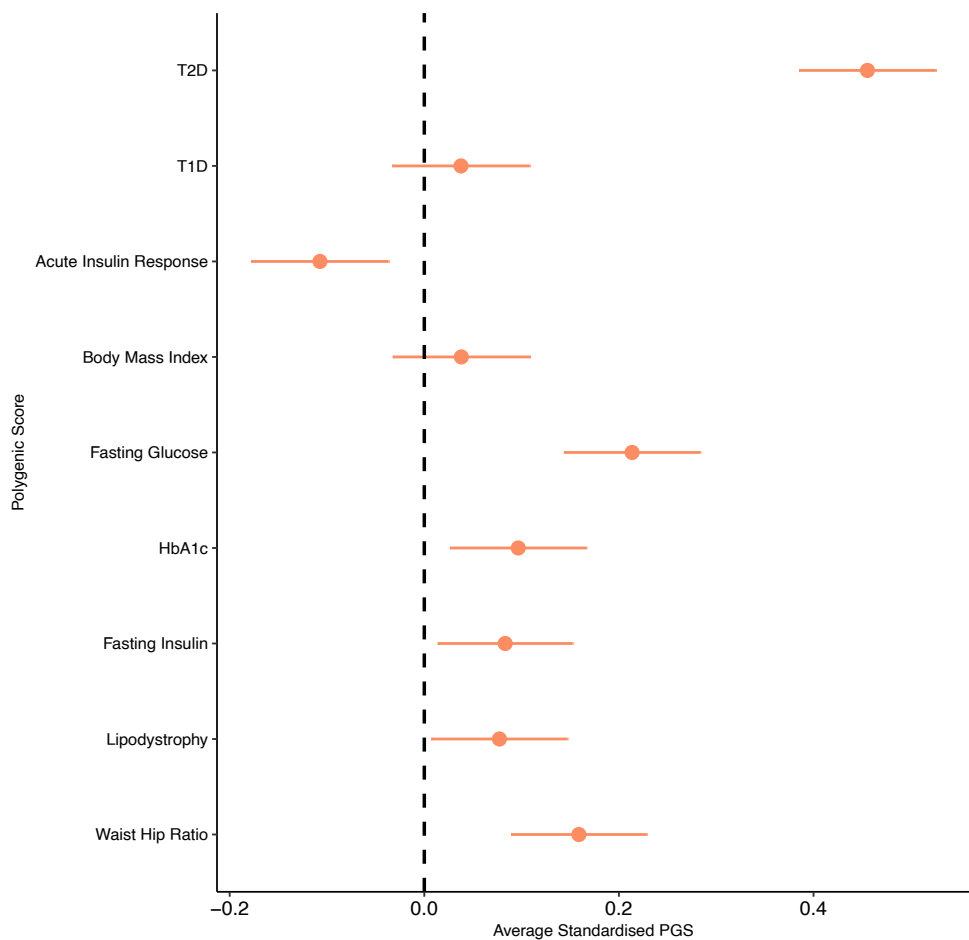
Peer review information *Nature Metabolism* thanks Constantin Polychronakos, Rashmi Prasad and Magdalena Szopa for their contribution to the peer review of this work. Primary Handling Editor: Yanina-Yasmin Pesch, in collaboration with the *Nature Metabolism* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
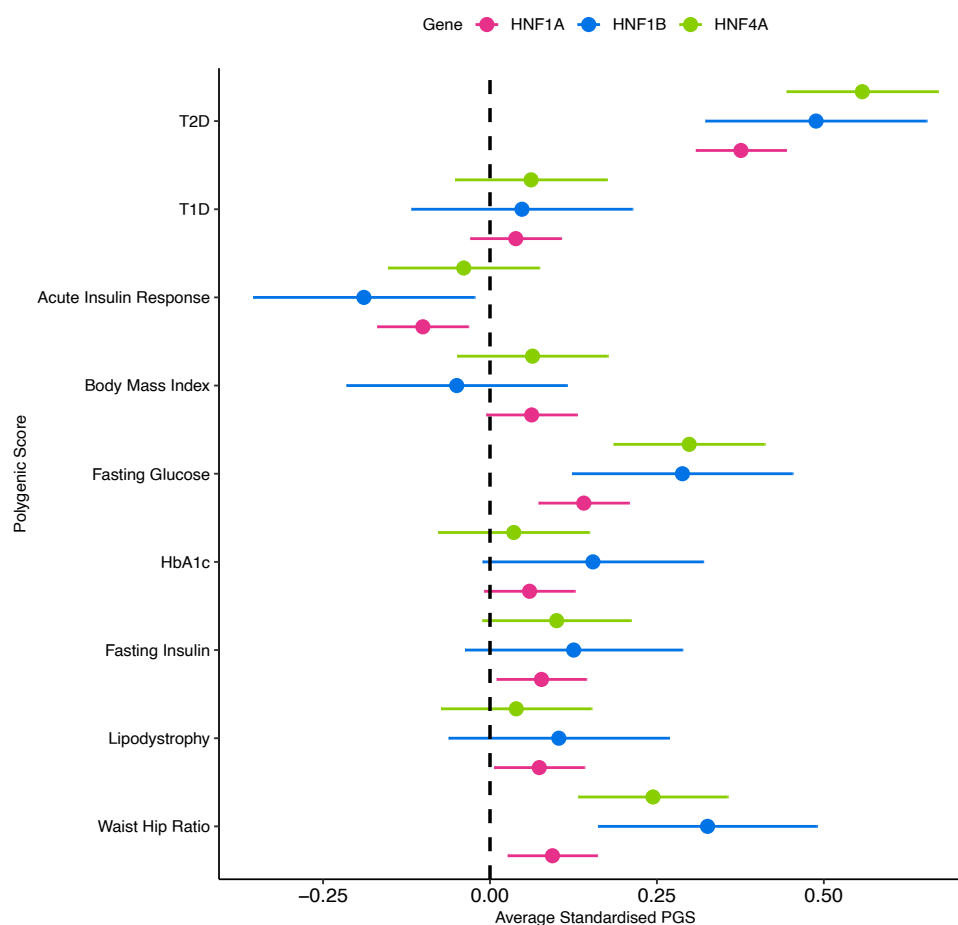
Extended Data Fig. 1 | Polygenic Score difference from controls restricted to probands only. Mean polygenic score difference for MODY probands (orange, N = 924) compared to controls (dashed black line, N = 7645). Scores are standardized, with controls set to a mean of 0 and a standard deviation of 1. Error bars indicate 95% confidence intervals, and dots represent the mean estimates, determined by linear regression models adjusted for the first ten within-cohort principal components.
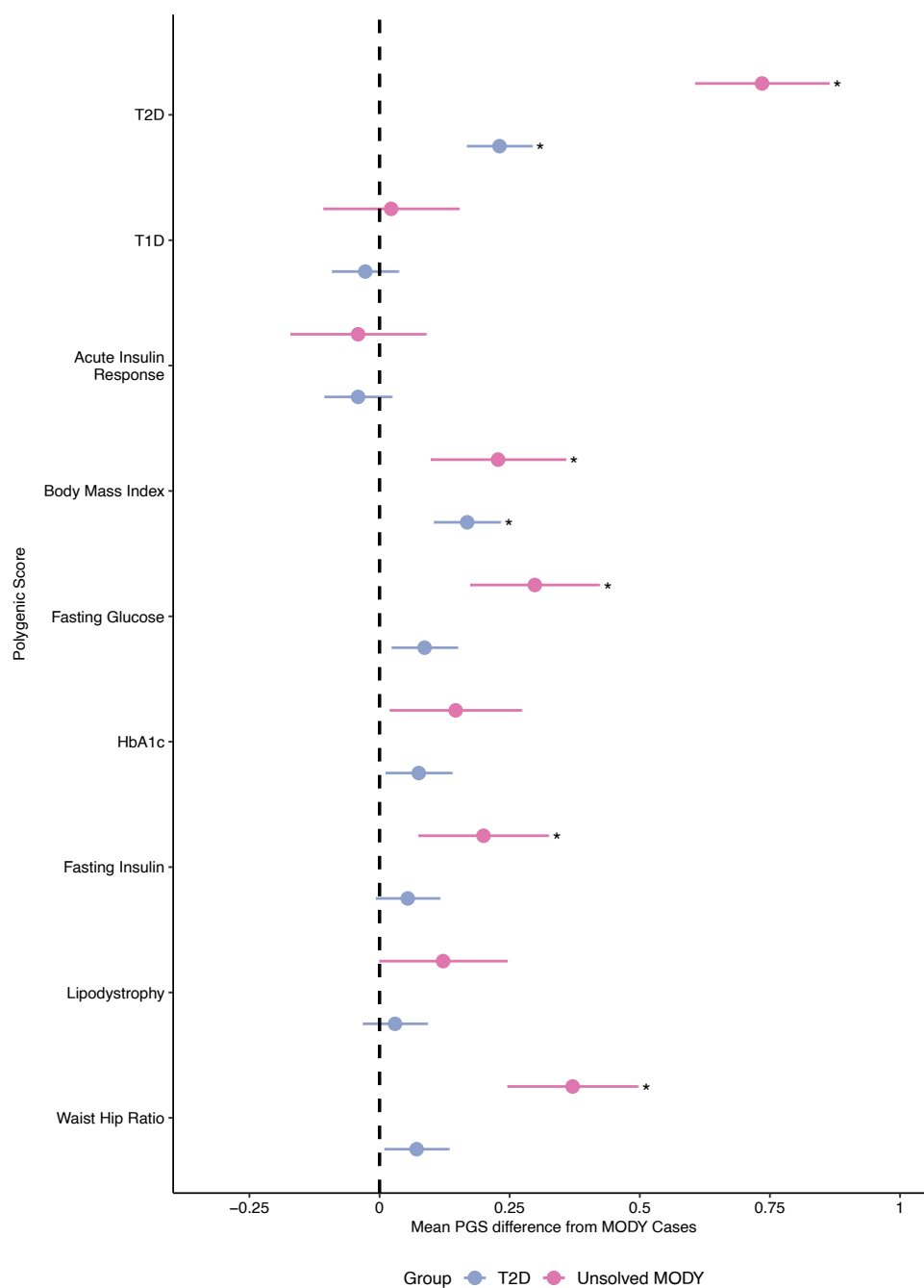
Extended Data Fig. 2 | Polygenic Score difference from controls by HNF MODY subtype. Mean polygenic score differences for HNF1A (pink, N = 997), HNF1B (blue, N = 145), and HNF4A (green, N = 320) carriers compared to controls (dashed black line, N = 7645), as determined by linear regression models adjusted for the first ten within-cohort principal components. Scores are standardized, with controls set to a mean of 0 and a standard deviation of 1. Error bars represent 95% confidence intervals, and dots denote the mean estimates.
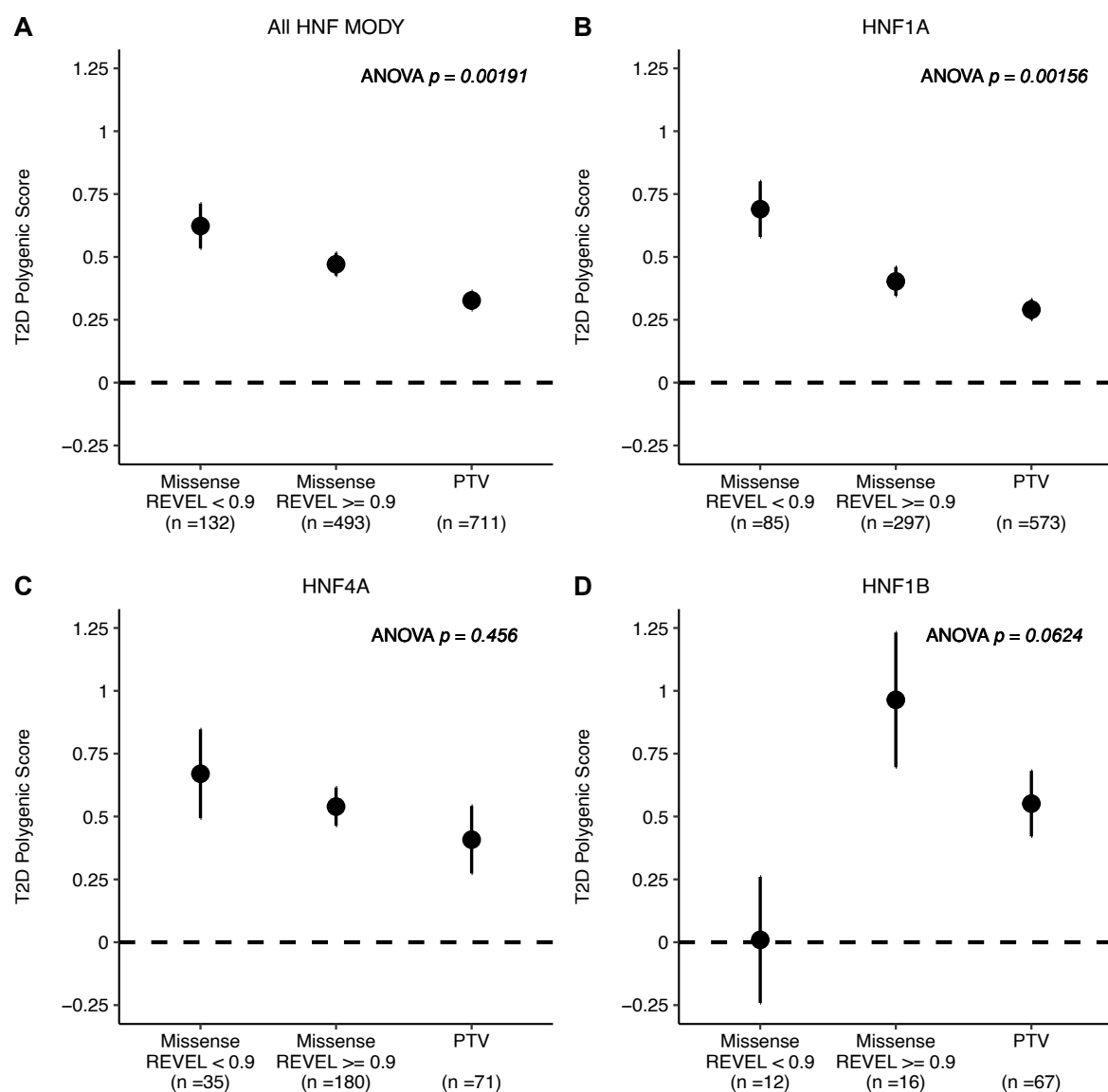
Extended Data Fig. 3 | Type 2 Polygenic Risk across Mutation Severity in HNF-MODY. The graph shows mean type 2 diabetes (T2D) polygenic risk scores with 95% confidence intervals across mutation groups ordered by predicted deleteriousness, from low to high. Panel (**a**) includes all HNF-MODY genes assessed, while panels (**b**), (**c**), and (**d**) focus on *HNF1A*, *HNF4A*, and *HNF1B*, respectively. Scores are standardized, with controls set to a mean of 0 and a standard deviation of 1. We assessed statistical significance using one-way ANOVA.

Extended Data Fig. 4 | Mean polygenic score difference from solved MODY cases. Mean polygenic score difference between unsolved MODY cases (N = 300, pink) and T2D cases (N = 4773, blue) versus MODY cases (N = 1462, assessed using a logistic regression model including the first ten within-cohort principal components as covariates. Asterisks denote significant differences (P < 0.0028). Polygenic Scores were standardized so that the control population has a mean of 0 and standard deviation of 1. Error bars represent 95% confidence intervals.
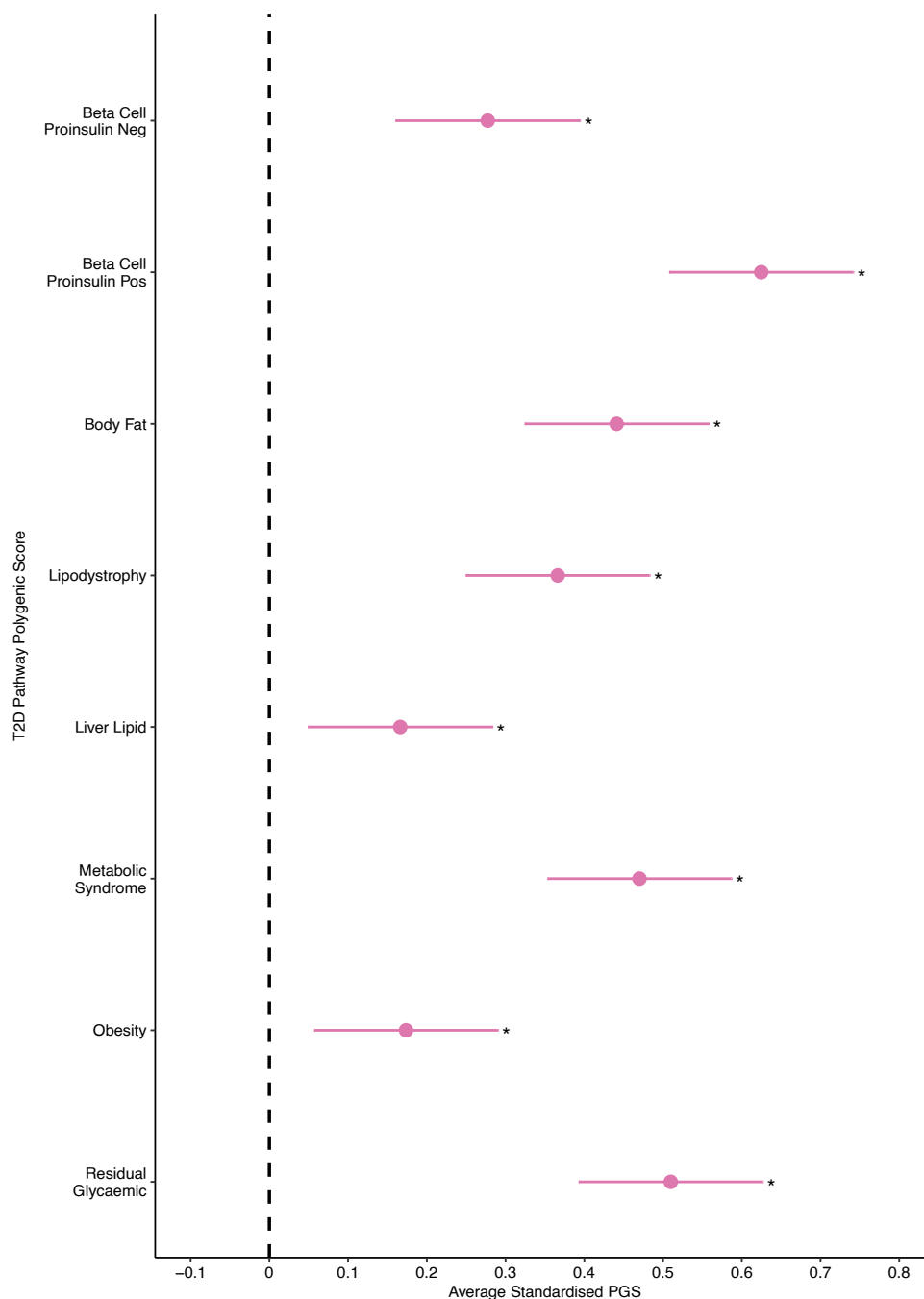
Extended Data Fig. 5 | T2D Hard Cluster Polygenic Score Differences in Unsolved MODY Cases Compared to Controls. Mean polygenic score differences for T2D hard cluster partitioned scores in Unsolved MODY cases (pink, N = 300) compared to non-diabetic controls (dashed black line, N = 7,645). Scores are standardized with controls set to a mean of 0 and a standard deviation of 1. Error bars represent 95% confidence intervals, and dots indicate the mean estimates. Asterisks denote statistically significant differences from controls (P < 0.006), based on the Bonferroni significance threshold, as determined by linear regression models adjusted for the first ten within-cohort principal components.
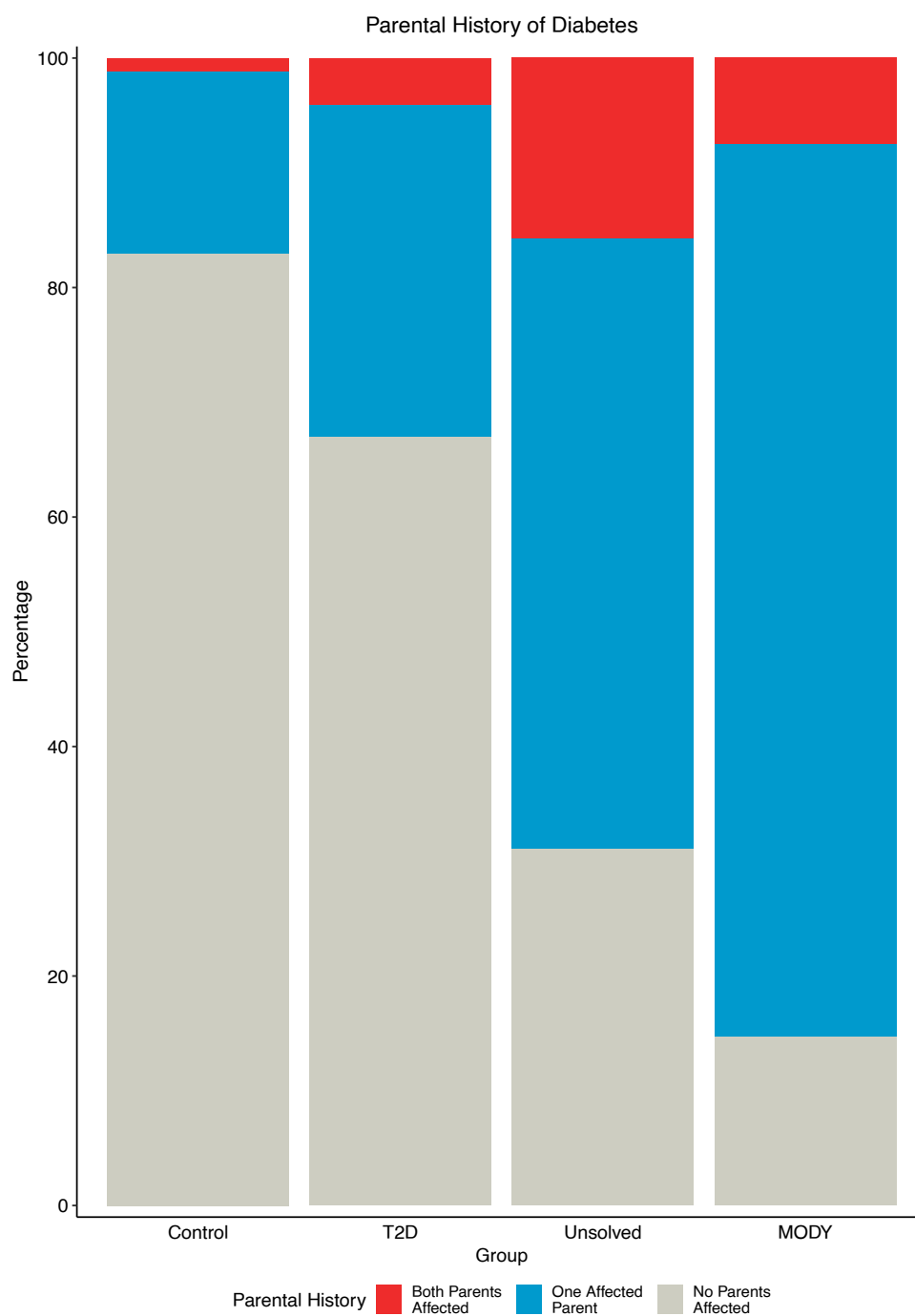
Extended Data Fig. 6 | Unsolved MODY cases have a high percentage of familial diabetes history. The plot shows the distribution of parental family history of diabetes for four groups: non-diabetic controls (N = 7,645), genetically confirmed MODY cases (N = 1,462), Unsolved MODY cases (N = 300), and type 2 diabetes cases (N = 4,773). Family history is categorized into three groups: no parent with diabetes (grey), one parent with diabetes (blue), and both parents with diabetes (red). Percentages are calculated within each group.

nature portfolio

Corresponding author(s): Kashyap Patel

Last updated by author(s): Aug 4, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No Software was used for data collection
Data analysis	In this analysis we used publicly available software including: Genopred (v2.2.1), KING robust algorithim (v2.2.4), FlashPCA(v2.0), PLINK (v1.9), GCTA (v1.94.1) LDAK (v6.0), R (v4.4.1) and Stata (v18.0). Further detail is provided in the Methods Section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data supporting the findings of this study are available within the article, source data, and its supplemental information. The clinical data, including individual level data, generated and/or analysed as part of this study are not publicly available because of patient confidentiality and ethical approval associated with the data but are available from the corresponding authors upon reasonable request. The UK Biobank dataset is available from https://biobank.ctsu.ox.ac.uk

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender (identity/presentation), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Information on sex was collected at patient recruitment to the UK Biobank or for genetic testing for our local cohorts. We removed participants when their sex inferred by genetic data did not match reported data. We performed analysis with both sex combined, correcting for sex as a covariate. Sample sizes by sex can be found in supplementary tables.
Reporting on race, ethnicity, or other socially relevant groupings	We only used individuals who had an inferred European ancestry, defined in the methods section. We also used genetic and/or within cohort principal components to further adjust for population structure
Population characteristics	We have outlined the population characteristics for our cohorts in detail in the methods section, with more detail contained in the supplementary tables. Population characteristics e.g.. sex, age, parental history of diabetes. were included as covariates when appropriate.
Recruitment	In our local MODY cohort we analysed individuals referred for monogenic diabetes genetic testing at the Exeter Genomics Laboratory, Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK. These referrals originated from clinical suspicion of MODY during routine clinical care in the UK. We also analysed unselected T2D and non diabetic participants from two ethically approved population cohorts in Southwest England: The Exeter 10000 study and The Diabetes Alliance for Research in England (DARE) Study. For non-clinically referred individuals, recruitment was performed by UK Biobank, between 2006 and 2010.
Ethics oversight	This study complies with all relevant ethical regulation and was approved by the appropriate ethics committees. Our study combined three ethically approved cohorts. In our local MODY cohort, all probands or their guardians provided informed consent, and the North Wales Ethics Committee approved the study, with Genetic Beta Cell Research Bank approving sample access. The NIHR Exeter Clinical Research Facility management committee approved access to these samples and genotype data for our T2D and non-diabetic controls. This research also utilised data from the UK Biobank resource carried out under UK Biobank application number 103356. UK Biobank protocols were approved by the National Research Ethics Service Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](#)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample Size is reported in the manuscript and with further detail in supplementary tables. In the UK Biobank we included 424,553 individuals who had both exome and array data and were of genetically inferred European ancestry.
Data exclusions	In this study we excluded individuals who were not of genetically inferred European ancestry. We also applied detailed sample quality quality control, such as removing sex mismatches, from our local array data. Quality control on UK Biobank data is provided in text and in refferenced publications.
Replication	We demonstrate the effect of polygenic effects in MODY in two cohorts. Clinically referred MODY patients and clinically unselected MODY carriers in the UK Biobank.
Randomization	Not relevant in this study design
Blinding	Not relevant in this study design

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern
☒ ☐	Plants

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA