Title: Evaluation of evidence for pathogenicity demonstrates that *BLK*, *KLF11* and *PAX4* should not be included in diagnostic testing for MODY

Short title: *BLK*, *KLF11* and *PAX4* do not cause MODY

Thomas W Laver [1], Matthew N Wakeling [1], Olivia Knox [1], Kevin Colclough [2], Caroline F Wright [1], Sian Ellard [1], Andrew T Hattersley [1], Michael N Weedon [1], Kashyap A Patel [1]

1. Institute of Biomedical and Clinical Science, University of Exeter, Exeter, UK
2. Exeter Genomics Laboratory, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK

Corresponding author:
Dr Kashyap A Patel
Institute of Biomedical and Clinical Science, College of Medicine and Health, University of Exeter, RLID building, level 3, Barrack road, Exeter, EX2 5DW,UK
Tel: 01392408237
Email: k.a.patel@exeter.ac.uk

Word count: 4153

Tables: 3
Figures: 1

Abstract

Maturity Onset Diabetes of the Young (MODY) is an autosomal dominant form of monogenic diabetes, reported to be caused by variants in 16 genes. Concern has been raised about whether variants in *BLK* (MODY11), *KLF11* (MODY7) and *PAX4* (MODY9) cause MODY. We examined variant-level genetic evidence (co-segregation with diabetes and frequency in population) for published putative pathogenic variants in these genes and used burden testing to test gene-level evidence in a MODY cohort (n=1227) compared to population control (UK Biobank, n=185,898). For comparison we analysed well-established causes of MODY, *HNF1A* and *HNF4A*. The published variants in *BLK*, *KLF11* and *PAX4* showed poor co-segregation with diabetes (combined LOD scores ≤1.2), compared to *HNF1A* and *HNF4A* (LOD scores >9), and are all too common to cause MODY (minor allele frequency >4.95x10^{-5}). Ultra-rare missense and protein-truncating variants (PTVs) were not enriched in a MODY cohort compared to the UK Biobank (PTVs P>0.05, missense P>0.1 for all three genes) while *HNF1A* and *HNF4A* were enriched (P<10^{-6}). Sensitivity analyses using different population cohorts supported our results. Variant and gene-level genetic evidence does not support *BLK*, *KLF11* or *PAX4* as causes of MODY. They should not be included in MODY diagnostic genetic testing.

Maturity Onset Diabetes of the Young (MODY) is the most common subtype of monogenic diabetes. It is reported to be caused by heterozygous variants in 16 genes [1]. MODY accounts for approximately 3% of all diabetes cases under 30 years of age [2, 3]. The prevalence of MODY is estimated to be 108 cases per million [4]. An accurate genetic diagnosis is important for patients with MODY as it can determine the correct treatment [1, 5] and provides an accurate assessment of the risk of diabetes for future offspring. The advent of next generation sequencing has enabled a paradigm shift in genetic testing from focusing on single gene testing to gene panel tests for diseases [6]. While this can boost diagnostic yield, it also has the potential to increase the risk of reporting variants in genes that are not a cause of MODY if genes are not carefully selected. This is likely to occur with next generation sequencing as it enables large numbers of genes to be tested more easily.. An incorrect genetic diagnosis could result in stopping insulin in a patient with type 1 diabetes. It could lead to inappropriate testing of family members, causing increased anxiety in unaffected relatives and inflicting the psychological burden of having a genetic disease. Therefore, it is crucial that the gene panel only includes genes with robust aetiological evidence to prevent misdiagnosis of MODY.

BLK, *KLF11* and *PAX4* are listed on OMIM as MODY11, MODY7 and MODY9 but there is a need to revaluate whether variants in these genes do cause MODY. Variants in *BLK* and *PAX4* have been reported to cause MODY via haploinsufficiency [7, 8] while variants in *KLF11* were reported to cause the disease, potentially via a gain of function mechanism [9]. These studies were conducted more than 10 years ago, before the availability of variant frequency in large population cohorts [7-9]. *KLF11* and *PAX4* were identified based primarily on biological candidacy rather than the hypothesis-free genetic approach which is now considered to be the most robust method for gene discovery studies. The only *BLK* coding variant (p.A71T) reported to cause MODY was later found to be very common in the population raising doubt over the aetiological role of *BLK* [10]. Rarity of a variant in a large control population as well as enrichment of variants in that gene in a disease cohort compared to a control population have become crucial evidence to support the gene-disease association alongside familial co-segregation [11, 12].

Therefore, the aim of our study was to evaluate genetic evidence for variants in *BLK*, *KLF11* and *PAX4* as a cause of MODY. We evaluated the existing evidence for these genes and assessed the gene-disease association using a large MODY cohort and population cohorts. We demonstrate there is a lack of robust genetic evidence to support the aetiological role of variants in *BLK*, *KLF11* and *PAX4* for MODY.

Research Design and Methods
Study populations
MODY cohort

We included 1227 unrelated probands from the UK who were referred for genetic testing for MODY from routine clinical care to the Exeter Genomics Laboratory at the Royal Devon and Exeter Hospital. Cohort characteristics are provided in Supplementary Table 1. None of these individuals were reported to have islet autoantibodies by the referring clinicians. 84% of patients were of self-reported European Ancestry and the overall rate of monogenic diabetes was 22.5%. Informed consent was obtained from the probands or their parents/guardians and the study was approved by the North Wales ethics committee (17/WA/03).

UK Biobank

UK Biobank is a population-based cohort from the UK with deep phenotyping data and genetic data for around 500,000 individuals aged 40-70 years at recruitment [13, 14]. A subset of ~200,000 DNA samples from UK Biobank participants underwent exome sequencing; this dataset was recently made available for research [15]. 94% of individuals were of self-reported white ethnicity. The UK Biobank resource was approved by the UK Biobank Research Ethics Committee and all participants provided written informed consent to participate.

GnomAD

We used GnomAD v.2.1.1 (141,456 individuals) and v3 (76,156 individuals) as alternative population controls in supplementary analyses. The detailed description of the cohort is previously published [16]. GnomAD v2.1.1 contains individuals with exome (n=125,748) and genome (n=15,708) data whereas v3 contains individuals with genome data. 46% of individuals in GnomAD v2.1.1 are of Non-Finnish European ancestry while for v3 this is 45%.

Genetic testing
MODY cohort

We undertook targeted next generation sequencing of *BLK*, *PAX4* and *KLF11* as well as *HNF1A* and *HNF4A* for probands suspected to have MODY, as previously described [6]. Targets were covered at a mean read depth of 460X per base and all bases had a mean coverage depth of at least 30 reads across the cohort. Variants were annotated against Genome Reference Consortium Human Build 37 (GRCh37) using Alamut Batch (Interactive Biosoftware, Rouen, France) using a Refseq transcript: *BLK* NM_001715.3, *KLF11* NM_003597.4, *PAX4* NM_001366110.1, *HNF1A* NM_000545.6, *HNF4A* NM_175914.4.

UK Biobank

We included 185,898 unrelated individuals from all ethnicities with exome sequencing data [17]. Detailed sequencing methodology for UK Biobank samples is provided by Szustakowski *et al.*[15] and is available at https://biobank.ctsu.ox.ac.uk/showcase/label.cgi?id=170. Variants were called against Genome Reference Consortium Human Build 38 (GRCh38). We lifted the variants over to Build 37 (GRCh37) [18] to ensure compatibility with the variants from our MODY cohort then annotated using Alamut Batch (Interactive Biosoftware, Rouen, France)

using a Refseq transcript: *BLK* NM_001715.3, *KLF11* NM_003597.4, *PAX4* NM_001366110.1, *HNF1A* NM_000545.6, *HNF4A* NM_175914.4.

GnomAD

The gnomAD consortium performed joint variant calling of the samples using a standardized BWA-Picard-GATK pipeline [16]. GnomAD was QCed and analysed using the Hail open-source framework for scalable genetic analysis (https://gnomad.broadinstitute.org/about). Variants in v2.1.1 were called against Genome Reference Consortium Human Build 37 (GRCh37), v3 against Build 38 (GRCh38). We lifted over v3 to Build 37 (GRCh37) then annotated all gnomAD variants using Alamut Batch (Interactive Biosoftware, Rouen, France) using a Refseq transcript: *BLK* NM_001715.3, *KLF11* NM_003597.4, *PAX4* NM_001366110.1, *HNF1A* NM_000545.6, *HNF4A* NM_175914.4. As gnomAD is an agglomeration of different sequencing projects, some genomic regions have low coverage in some samples therefore to control for this we removed the variants from both the MODY cohort and gnomAD cohorts if they were in a region of low coverage (≤10x coverage in ≤80% of samples) in either cohort or flagged as low quality in gnomAD.

Co-segregation analysis of putative pathogenic variants

We used author provided LOD (logarithm of the odds) scores where available for the first published variants in *BLK*, *PAX4*, *KLF11*, *HNF1A* and *HNF4A* which suggested the causal role of those variants in MODY. This was only available for *BLK* p.A71T [7]. If the LOD score was not provided, we calculated it based on the Gene Clinical Validity Curation Standard Operating Procedure [19]. We summed the LOD scores for multiple pedigrees where possible based on this guidance to calculate a combined LOD score. Using a binomial test we compared the observed proportion of family members with diabetes and a putative variant to the expected proportion of 0.5 if the variant was not associated with diabetes.

Statistical analysis

For each analysis variant frequency was defined in the MODY cohort plus the control cohort combined. We compared the frequency of ultra-rare (allele count=1) protein truncating variants (PTVs) (essential splice site, stop gain and frameshift variants; excluding those in the last exon) and missense variants in each gene in the MODY cohort to the UK Biobank population cohort. We also provided the evidence of an association in terms of Bayesian false-discovery probabilities (BFDP) as previously described [20]. We replicated our analysis using two alternative population controls: gnomAD v2.1.1 (141,456 individuals) and gnomAD v3 (76,156 individuals) [16].

We used synonymous variants as a control to assess the difference in sequencing technologies and analysis pipeline. We also compared the frequency of rare variants (MAF<0.0001) and the frequency of all PTVs (no frequency filter) to test if there was an undue influence of ultra-rare variants due to differences in capture platforms.

The most common *HNF1A* pathogenic variant is a frameshift variant (p.G292Rfs*25) in exon 4 due to a duplication of a C nucleotide. This variant is difficult to detect robustly in exome/genome sequencing data due to its location in a repetitive poly-C tract and the presence of a common variant that adds an additional 5' C nucleotide to the tract (rs56348580 G>C, MAF=0.26). Since we were unable to perform confirmatory Sanger sequencing in the UK Biobank or GnomAD cohorts we excluded this variant from our analysis from all study cohorts.

We used Fisher's exact test to assess variant enrichment in our MODY cohort and compute odds ratios with 95% confidence intervals. We used a threshold *P* value of 0.01 (0.05/5) as we tested 5 genes. We used Stata 16 (StataCorp, Texas, USA) for this analysis. BFDP was computed using a 'gap' R package. We used a prior probability of association of 0.99 for *HNF1A* and *HNF4A* to reflect the strong prior evidence for these genes and used 0.2 for *BLK*, *KLF11* and *PAX4* due to their probable disease association. We calculated the variance of the prior log(OR) as described by Wakefield[20], by assuming a 95% probability that the OR was less than 20 for *HNF1A* and *HNF4A,* and 3 for the other genes[20, 21]. We also explored different plausible priors as a sensitivity analysis.

Data and Resource Availability
UK Biobank data is accessible via application: https://www.ukbiobank.ac.uk/enable-your-research. GnomAD data is publically available: https://gnomad.broadinstitute.org/. The MODY cohort data is not publicly available due the limitations of the current ethics and to protect patient confidentiality but is available from the corresponding authors on reasonable request.

No applicable resources were generated or analyzed during the current study.

Results
***BLK, KLF11* and *PAX4* variants had poor co-segregation in the published pedigrees**
Variants that are highly penetrant causes of MODY would be expected to show strong co-segregation with the disease. To evaluate the genetic evidence of co-segregation with disease, we reviewed published pedigrees for putative variants in *BLK*, *KLF11* and *PAX4* causing MODY (Supplementary Table 2). We identified 1 *BLK*, 3 *KLF11* and 1 *PAX4* pedigrees with more than 3 individuals with variants to calculate LOD scores [7-9]. *KLF11* and *PAX4* variants showed poor co-segregation with diabetes in the families, with LOD scores of 1.2 and 0.6 respectively (Table 1). In line with low LOD scores, these variants were not associated with diabetes in family members in these pedigrees (P>0.5, Table 1). The *BLK* variant p.A71T also had a low LOD score of 1.16 and was modestly associated with diabetes in family members (*P*=0.02). In contrast, the variants reported in the first papers for *HNF1A* [22] and *HNF4A* [23], which are well-established causes of MODY, showed strong co-segregation with diabetes with combined LOD scores for the first reported variants of 9.63 and 15.05 respectively (Table 1).

Putative pathogenic variants in *BLK, KLF11* and *PAX4* are common in the population
The frequency of a putative pathogenic variant should not exceed the expected prevalence of the commonest variant in the commonest genetic subtype of the disease. MODY is estimated

to have a population frequency of 1.08 per 10,000 [4]. We used the framework developed by Whiffin *et al* [24, 25] to calculate the maximum tolerated allele count in the population (gnomAD v2.1.1, n=141,456) for a putative pathogenic variant causing MODY. We used *HNF1A,* the most common cause of MODY, as a model to calculate the maximum tolerated allele count in the population. *HNF1A* accounts for 52% of MODY cases [4] and the most common mutation (p.G292Rfs*25) accounts for 19% of *HNF1A* cases [26]. At 50% penetrance the framework suggests that a pathogenic variant causing MODY should be present ≤3 times (frequency <2.1×10^{-5}) in gnomAD v2.1.1 for *HNF1A*. As other genes will account for far fewer MODY cases, the putative pathogenic variants in *BLK*, *KLF11* and *PAX4* should be even rarer. We looked at the frequency of variants in *BLK*, *KLF11* and *PAX4* that were reported to cause MODY before large scale population data was made available publicly in 2016 [27] (Table 2). The variants published since 2016 should have included the frequency of the variant in these databases as part of their screening process and thus would be expected to have only published rare variants (See Supplementary Table 2 for full list of HGMD variants in these genes).

All putative MODY-causing variants in *BLK*, *KLF11* and *PAX4* published prior to 2016 were too common in the population to cause MODY. The allele count in gnomAD v2.1.1 was 4-8608 times higher than the maximum tolerable allele count for the commonest cause of MODY (Table 2). The least common was *PAX4* p.R164W which is seen 14 times in the whole of gnomAD v2.1.1 at a frequency of 4.95×10^{-5} but seen at higher frequency of 1.2×10^{-04} (3/24948) in the African/African American population. In contrast, the first reported variants in *HNF1A* and *HNF4A*, which were reported in the 1990s, are rare in the population with the most common (p.P447L) present 3 times in gnomAD v2.1.1 (1.20×10^{-5}) (Table 2).

Rare variants in *BLK, KLF11* and *PAX4* are not enriched in a MODY cohort
Having conducted variant level analyses on published variants in these genes we then carried out a gene level analysis to establish if other rare variants in these genes are likely to be pathogenic for MODY. To assess this, we carried out a gene burden test comparing the frequency of ultra-rare coding variants in a cohort of 1227 patients referred for MODY genetic testing with the frequency in the unrelated 185,898 exome-sequenced individuals from the population cohort UK Biobank (Table 3, Figure 1).

Ultra-rare (allele count=1) PTV and missense variants in *BLK*, *KLF11* and *PAX4* are not enriched in our MODY cohort compared to the UK Biobank (all *P* values ≥0.09, Table 3). The bayesian false-discovery probability (BFDP) for ultra-rare PTV and missense variants was ≥0.70 for *BLK*, *KLF11* and *PAX4* (Table 3). The results of BFDP remained ≥0.37 using other plausible priors (Supplementary Table 3). In contrast variants in *HNF1A* and *HNF4A,* which are well established causative genes for MODY, were greatly enriched in our MODY cohort (all *P* values ≤2.79×10^{-6}) with a very low BFDP (all ≤6.74×10^{-5}).

The lack of enrichment of rare variants in *BLK, KLF11* and *PAX4* is not due to technical artefacts
To ensure that our results are not due to differences in sequencing technologies or analysis pipelines between cases and controls, we performed a series of sensitivity analyses. Firstly, we

analysed synonymous variant frequency in our MODY cohort and population control, and showed that the frequency of synonymous variants in all five genes was similar in our MODY cohort and the UK Biobank (all $P>0.05$, Supplementary Table 4).

Secondly, we replicated our gene burden analysis using gnomAD v.2.1.1 and v3 as two alternative population cohorts which were sequenced on different platforms (exome versus genome respectively) and with a different analysis pipeline versus the UK Biobank. Despite these differences, we found similar results with no enrichment in PTV or missense variants in *BLK*, *KLF11* or *PAX4* (Supplementary Tables 5 and 6).

Finally, to remove any undue influence of ultra-rare variants caused by differences in capture platforms, we performed a gene burden analysis for rare PTVs and missense variants (MAF <0.0001). We also compared the frequency of all PTVs in our MODY cohort and population control as all PTVs in these genes are considered to be pathogenic. These analyses showed similar results to our main analysis: rare PTVs and missense variants, and all PTVs, in *BLK*, *KLF11* and *PAX4* were not enriched in our MODY cohort whereas all these variant subsets in *HNF1A* and *HNF4A* showed great enrichment in our MODY cohort. (Supplementary Tables 7 and 8)

Discussion
Variant and gene level genetic evidence presented in this study suggest that variants in *BLK*, *KLF11* and *PAX4* do not cause MODY. The lack of co-segregation of published MODY causing variants, presence in the population at high frequency and lack of enrichment of rare variants in a MODY cohort are consistent with these genes not causing MODY. The robustness of our approach is demonstrated by the results supporting the well-established causality of *HNF1A* and *HNF4A* variants.

Variants in *BLK*, *KLF11* and *PAX4* were reported to cause MODY more than 10 years ago, before large-scale variant population frequency became available [7-9]. Only small numbers of controls were available to rule out variants being present in the population.

BLK was first described in 2009 [7] by following up linkage to the 8p23 region [28] in 6 MODY families and identifying variants in *BLK* in three of the families. The frequency of the *BLK* variants was tested in 336 white control individuals and, for one variant, an additional 577 African American control individuals. *BLK* was identified via a linkage approach – it is possible that another candidate gene within the region of linkage is responsible for the disease in those families. Bonnefond *et al.* [10] found that the only non-synonymous variant in *BLK* reported to cause MODY was common in normoglycaemic individuals. This is the variant (p.A71T) that has a positive LOD score in the published pedigree; however as *BLK* was identified by linkage the LOD score would necessarily be positive regardless of the pathogenicity of the variant and, as also demonstrated by its frequency in gnomAD, the variant is clearly too common to cause MODY. No large MODY pedigrees with co-segregation have been described for *BLK* since the initial report. Non-coding variants in *BLK* were also reported to cause MODY [7], however, as our main cohorts consisted of targeted and exome sequencing

data we were unable to investigate non-coding variants. It is unlikely that non-coding variants would be pathogenic given the lack of evidence for coding variants in *BLK* as a cause of MODY and that both coding and non-coding variants were proposed to cause the disease via loss of function.

KLF11 was proposed as a cause of MODY via a candidate gene approach in 2005 [9]. The frequency of the reported *KLF11* variants was judged in only 313 normoglycemic individuals and 313 type 2 diabetes patients. Functional studies using Gal4 reporter assays suggested a possible mechanism of action for the variants via gain-of-function causing increased *KLF11* repression activity. If pathogenic variants in *KLF11* act via gain-of-function then we would not expect to see enrichment of PTVs in a MODY cohort, however we might expect to see enrichment of missense variants. We did not see enrichment of either type of variant and the previously identified *KLF11* variants are too common in the population to be disease causing.

PAX4 was associated with MODY in patients from Thailand [8]. The variants were screened in a maximum of 344 non-diabetic controls. While their controls were from the same population as their cases, using data from gnomAD we now know that p.R192H is common in East Asians and both this variant and p.R164W are too common to cause MODY (p.R192H seen 2214 times in gnomAD v2.1.1 and p.R164W seen 14 times). Plengvidhya *et al.* [8] used luciferase reporter assays to show that p.R164W impairs the repressor activity of PAX4 on the insulin and glucagon promoters. However, they stated that the impairment was relatively small, thus it is possible that the reduction may be insufficient to result in a clinical phenotype. No large MODY pedigrees with co-segregation for a variant in *PAX4* have been described since the initial report.

Our study uses a large cohort of MODY cases and takes advantage of the availability of large population cohorts. The lack of enrichment for *BLK*, *KLF11* and *PAX4* PTV and missense variants in a MODY cohort compared to a population cohort is consistent with these genes not causing MODY. However, alternative explanations may be that the mechanism of action for these genes is not loss of function (as has been suggested for *KLF11* [9]) or they are an extremely rare cause of MODY. However, we did not see enrichment in missense variants (at either AC=1 or MAF<0.0001) suggesting that this is unlikely. In line with our results, gnomAD pLI and missense constraint scores for these genes are low, suggesting these genes are not under strong negative selection, in contrast to *HNF1A* and *HNF4A* which have high constraint scores. This data suggests that variants in these genes do not cause a rare monogenic disorder.

Variants in these genes could still be acting as polygenic risk factors for diabetes. *PAX4* has been reported in the literature as a type 2 diabetes risk factor in East Asian populations [29, 30]. Fuschsberger *et al.* [31], in a study of 6,504 type 2 diabetes cases and 6,436 controls, found a lack of exome-wide enrichment of PTV and deleterious missense variants for *BLK* (*P*~0.001), *KLF11* (*P*>0.05) and *PAX4* (*P*>0.05). However, the *PAX4 p.R192H* variant (a proposed MODY variant) showed association with type 2 diabetes in East Asian cases (OR 1.79) but not with age of diabetes diagnosis (P=0.64), suggesting that this variant influences risk of type 2 diabetes rather than early-onset MODY. Similar to this, a recent large type 2 diabetes case-

control study of 20,791 cases and 24,440 controls did not find an exome wide-significant association in these genes, except for the *PAX4 p.R192H* variant which was associated with type 2 diabetes in East Asians [32]. However, the lack of exome-wide significance may reflect the relatively small size of these studies..

A limitation of our study is that by using publicly available population controls there were cross platform differences between cases and controls. This issue was mitigated by removing genomic positions with low coverage in one cohort from the other and by our sensitivity analyses: using synonymous variants as a negative control and testing alternative population control cohorts. Despite using a large MODY cohort there was still a relatively limited sample size of cases which could tend our gene burden tests of ultra-rare variants towards negative results. To ensure a lack of power was not determining the results we also used sensitivity analyses with MAF <0.001 and these did not suggest there was an association between *BLK*, *KLF11* or *PAX4* and MODY. One other caveat to our burden testing results is the fact that both our MODY cohort and the UK Biobank are predominantly European ancestry. We cannot rule out that an enrichment might be seen in MODY cohorts from other ancestries, particularly for *PAX4* which was originally reported in East Asian ancestry. It must be acknowledged that the power of co-segregation analysis was limited, particularly for *BLK* and *PAX4* as they are only based on one family each. However, detailed review of all the published papers on putative pathogenic variants did not identify additional large published pedigrees for co-segregation analysis.

Our study results have important implications for genetic diagnostic laboratories worldwide who offer testing for MODY. Based on our results, we recommend that *BLK*, *KLF11* and *PAX4* should not be included in the gene panels for genetic testing for MODY and should not be reported as a cause of MODY. Studies are still reporting variants in these genes as a cause of MODY and they are routinely tested in clinical practice [33-37]. Our systematic review of the NCBI gene testing registry showed that 19 of 25 panels offered by diagnostic genetic laboratories still have at least one of these genes on their panel. Our study removes the ambiguity of the etiological role of these genes for MODY and provides the clearest results to date that refute their role as causative genes for MODY. Excluding these genes from diagnostic panels will prevent misdiagnosis of MODY and reduce workload for laboratories. The results from our study provide much needed evidence to gene curation efforts such as ClinGen and the Gene Curation Coalition to support the removal of these three genes from MODY genetic panels [38, 39]. The ClinGen curation panel also came to a similar conclusion using their own scoring system [12] independently to our study. They classified *BLK* and *PAX4* as 'Refuted' genes (https://search.clinicalgenome.org/kb/genes/HGNC:1057; https://search.clinicalgenome.org/kb/genes/HGNC:8618) and *KLF11* as a 'Disputed' gene (https://search.clinicalgenome.org/kb/genes/HGNC:11811). However, it is important to note that in addition to their own approach, they used our current work as previously published as a conference abstract to reach their conclusion. We also strongly recommend that variants in *BLK*, *KLF11* and *PAX4* should be removed as a cause of MODY on databases such as HGMD [40], OMIM [41], ClinVar and panelapp [42] that are widely used by diagnostic laboratories and geneticists worldwide.

In conclusion, we present evidence from re-analysis of published variants in *BLK*, *KLF11* and *PAX4* that they are too common to cause MODY, have poor co-segregation with diabetes in those families and since the initial description no large MODY families with co-segregation of a variant have been published. We have then shown a lack of enrichment of rare variants in these genes in a MODY cohort compared to a population cohort providing evidence that rare variants in these genes do not cause MODY. Overall, the evidence does not support *BLK*, *KLF11* or *PAX4* as causes of MODY and they should not be included in diagnostic genetic testing.

References
1.   Owen, K.R., *Monogenic diabetes in adults: what are the new developments?* Current Opinion in Genetics & Development, 2018. **50**: p. 103-110.
2.   Pihoker, C., et al., *Prevalence, Characteristics and Clinical Diagnosis of Maturity Onset Diabetes of the Young Due to Mutations in HNF1A, HNF4A, and Glucokinase: Results From the SEARCH for Diabetes in Youth.* The Journal of Clinical Endocrinology & Metabolism, 2013. **98**(10): p. 4055-4062.

3.   Shepherd, M., et al., *Systematic Population Screening, Using Biomarkers and Genetic Testing, Identifies 2.5% of the U.K. Pediatric Diabetes Population With Monogenic Diabetes.* Diabetes Care, 2016. **39**(11): p. 1879-1888.

4.   Shields, B.M., et al., *Maturity-onset diabetes of the young (MODY): how many cases are we missing?* Diabetologia, 2010. **53**(12): p. 2504-2508.

5.   Hattersley, A.T. and K.A. Patel, *Precision diabetes: learning from monogenic diabetes.* Diabetologia, 2017. **60**(5): p. 769-777.

6.   Ellard, S., et al., *Improved genetic testing for monogenic diabetes using targeted next-generation sequencing.* Diabetologia, 2013. **56**(9): p. 1958-1963.

7.   Borowiec, M., et al., *Mutations at the BLK locus linked to maturity onset diabetes of the young and β-cell dysfunction.* Proceedings of the National Academy of Sciences, 2009. **106**(34): p. 14460-14465.

8.   Plengvidhya, N., et al., *PAX4 Mutations in Thais with Maturity Onset Diabetes of the Young.* The Journal of Clinical Endocrinology & Metabolism, 2007. **92**(7): p. 2821-2826.

9.   Neve, B., et al., *Role of transcription factor KLF11 and its diabetes-associated gene variants in pancreatic beta cell function.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(13): p. 4807-4812.

10.   Bonnefond, A., et al., *Reassessment of the putative role of BLK-p.A71T loss-of-function mutation in MODY and type 2 diabetes.* Diabetologia, 2013. **56**(3): p. 492-496.

11.   MacArthur, D.G., et al., *Guidelines for investigating causality of sequence variants in human disease.* Nature, 2014. **508**(7497): p. 469-476.

12.   Strande, N.T., et al., *Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource.* Am J Hum Genet, 2017. **100**(6): p. 895-906.

13.   Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.* PLoS Med, 2015. **12**(3): p. e1001779.

14.   Bycroft, C., et al., *The UK Biobank resource with deep phenotyping and genomic data.* Nature, 2018. **562**(7726): p. 203-209.

15.   Szustakowski, J.D., et al., *Advancing Human Genetics Research and Drug Discovery through Exome Sequencing of the UK Biobank.* medRxiv, 2020: p. 2020.11.02.20222232.

16.   Karczewski, K.J., et al., *The mutational constraint spectrum quantified from variation in 141,456 humans.* Nature, 2020. **581**(7809): p. 434-443.

17.   *Exome sequences - Genomics.* [cited 2021 28.07.2021]; Available from: https://biobank.ctsu.ox.ac.uk/showcase/label.cgi?id=170.

18.   Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

19.   ClinGen. *Gene Clinical Validity Curation Process.* 2019; Available from: https://www.clinicalgenome.org/site/assets/files/3975/gene-disease_validity_standard_operating_procedures_version_7-1.pdf.

20.   Wakefield, J., *A Bayesian measure of the probability of false discovery in genetic epidemiology studies.* Am J Hum Genet, 2007. **81**(2): p. 208-27.

21.   *Breast Cancer Risk Genes — Association Analysis in More than 113,000 Women.* New England Journal of Medicine, 2021. **384**(5): p. 428-439.

22.   Yamagata, K., et al., *Mutations in the hepatocyte nuclear factor-1α gene in maturity-onset diabetes of the young (MODY3).* Nature, 1996. **384**(6608): p. 455-458.

23.   Yamagata, K., et al., *Mutations in the hepatocyte nuclear factor-4[alpha] gene in maturity-onset diabetes of the young (MODY1).* Nature, 1996. **384**(6608): p. 458-460.

24.   Whiffin, N., et al., *Using high-resolution variant frequencies to empower clinical genome interpretation.* Genetics in Medicine, 2017. **19**(10): p. 1151-1158.

25.   Ware, J. *alleleFrequencyApp.* 2016; Available from: https://www.cardiodb.org/allelefrequencyapp/.

26. Colclough, K., et al., *Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha and 4 alpha in maturity-onset diabetes of the young and hyperinsulinemic hypoglycemia.* Human mutation, 2013. **34**(5): p. 669-685.

27. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-291.

28. Kim, S.-H., et al., *Identification of a Locus for Maturity-Onset Diabetes of the Young on Chromosome 8p23.* Diabetes, 2004. **53**(5): p. 1375-1384.

29. Shimajiri, Y., et al., *A Missense Mutation of Pax4 Gene (R121W) Is Associated With Type 2 Diabetes in Japanese.* Diabetes, 2001. **50**(12): p. 2864-2869.

30. Kwak, S.H., et al., *Nonsynonymous Variants in PAX4 and GLP1R Are Associated With Type 2 Diabetes in an East Asian Population.* Diabetes, 2018. **67**(9): p. 1892-1902.

31. Fuchsberger, C., et al., *The genetic architecture of type 2 diabetes.* Nature, 2016. **536**(7614): p. 41-47.

32. Flannick, J., et al., *Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls.* Nature, 2019. **570**(7759): p. 71-76.

33. Kleinberger, J.W., et al., *Monogenic diabetes in overweight and obese youth diagnosed with type 2 diabetes: the TODAY clinical trial.* Genetics In Medicine, 2017. **20**: p. 583.

34. Chapla, A., et al., *Maturity onset diabetes of the young in India – a distinctive mutation pattern identified through targeted next-generation sequencing.* Clinical Endocrinology, 2014. **82**(4): p. 533-542.

35. Doddabelavangala Mruthyunjaya, M., et al., *Comprehensive Maturity Onset Diabetes of the Young (MODY) Gene Screening in Pregnant Women with Diabetes in India.* PLOS ONE, 2017. **12**(1): p. e0168656.

36. Anık, A., et al., *Molecular diagnosis of maturity-onset diabetes of the young (MODY) in Turkish children by using targeted next-generation sequencing*, in *Journal of Pediatric Endocrinology and Metabolism*. 2015. p. 1265.

37. Ming-Qiang, Z., et al., *Maturity onset diabetes of the young (MODY) in Chinese children: genes and clinical phenotypes.* J Pediatr Endocrinol Metab, 2019. **32**(7): p. 759-765.

38. Rehm, H.L., et al., *ClinGen — The Clinical Genome Resource.* New England Journal of Medicine, 2015. **372**(23): p. 2235-2242.

39. *The Gene Curation Coalition*. Available from: https://thegencc.org/.

40. Stenson, P.D., et al., *Human Gene Mutation Database (HGMD): 2003 update.* Hum Mutat, 2003. **21**(6): p. 577-81.

41. Amberger, J.S., et al., *OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders.* Nucleic Acids Res, 2015. **43**(Database issue): p. D789-98.

42. Martin, A.R., et al., *PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels.* Nat Genet, 2019. **51**(11): p. 1560-1565.

43. Furuta, H., et al., *Organization and partial sequence of the hepatocyte nuclear factor-4 alpha/MODY1 gene and identification of a missense mutation, R127W, in a Japanese family with MODY.* Diabetes, 1997. **46**(10): p. 1652-1657.

44. Lindner, T., et al., *Hepatic function in a family with a nonsense mutation (R154X) in the hepatocyte nuclear factor-4alpha/MODY1 gene.* J Clin Invest, 1997. **100**(6): p. 1400-5.

45. Chapla, A., et al., *Maturity onset diabetes of the young in India - a distinctive mutation pattern identified through targeted next-generation sequencing.* Clin Endocrinol (Oxf), 2015. **82**(4): p. 533-42.

Table 1: Co-segregation of *BLK*, *KLF11*, *PAX4* published variants with diabetes

Table shows the LOD scores and association of variants with diabetes in family members for variants where there were families with 3 or more people with the variant. We used author provided LOD scores where available for the first published variants which suggested the causal role of those variants in MODY. If the LOD score was not provided, we calculated it based on the Gene Clinical Validity Curation Standard Operating Procedure [19]. We summed the LOD score for each pedigree to calculate the combined LOD score. *two pedigrees with p.T220M were included in the combined LOD score calculation.

Gene	No. Pedigree used in analysis	Variants	Combined LOD score	Number of family members with diabetes and variant / total number of family members with diabetes and genotype information	Proportion of family members with diabetes and variant (95%CI)	*Binomial test P value (against expected proportion of 0.5)*
BLK	1	p.A71T[7]	1.16	9/10	0.9 (0.55-1)	0.02
KLF11	3	p.A347S[9], p.T220M[9]*	1.2	2/4	0.5 (0.068-0.93)	1
PAX4	1	p.R164W[8]	-	2/2	1 (0.16-1)	0.5
HNF1A	7	p.G292Rfs*25[22], p.P447L[22], p.V380Sfs*4[22], p.E548Rfs*112[22], p.R131Q[22], c.1768+1G>A[22], c.1108-2A>G[22]	9.63	38/39	0.97 (0.87-1)	1.4×10^{-10}
HNF4A	2	p.Q255*[43],p.R141*[44]	15.05	49/50	0.98 (0.89-1)	9×10^{-14}

Table 2: Population frequency of variants in *BLK*, *KLF11* and *PAX4* published as MODY causing.

Allele frequency taken from gnomAD v2.1.1. The table provides coding variants reported before 2016 which are reported to cause MODY as the release of ExAC [27] that year meant variants published since then have had access to a large population control as part of their screening process. The *HNF1A* and *HNF4A* variants included here for comparison are those from the original papers used in the LOD score calculations in Table 1.

Gene	Variant	Allele count /total alleles in gnomAD v2.1.1	Allele frequency in gnomAD v2.1.1	Allele count in ancestry with maximum frequency/total alleles in the ancestry	Maximum Allele frequency in a single ancestry in gnomadv2.1.1 (ancestry)	Reference for variants
BLK	p.A71T	3281/282812	0.012	420/10368	0.041 (Ashkenazi Jewish)	[7]
KLF11	p.Q62R	25823/282778	0.091	1497/10370	0.144 (Ashkenazi Jewish)	[9]
	p.T220M	1207/282762	4.27×10^{-3}	1098/24958	0.044 (African/African American)	[9]
	p.A347S	36/282304	1.28×10^{-4}	17/35410	4.80×10^{-4} (Latino/Admixed American)	[9]
PAX4	p.R31L	105/250972	4.18×10^{-4}	102/30616	0.003 (South Asian)	[45]
	p.R164W	14/282800	4.95×10^{-5}	3/24948	1.2×10^{-4} (African/African American)	[8]
	p.R192H	2214/282856	7.83×10^{-3}	2182/19946	0.109 (East Asian)	[8]
HNF1A	p.P447L	3/249186	1.20×10^{-5}	1/20812	4.81×10^{-5} (European Finnish)	[22]
	p.V380Sfs*4	0	0	0	0	[22]
	p.E548Rfs*112	0	0	0	0	[22]
	p.R131Q	1/251390	3.98×10^{-6}	1/113698	8.80×10^{-6} (European non-Finnish)	[22]
	c.1768+1G>A	0	0	0	0	[22]
	c.1108-2A>G	0	0	0	0	[22]
HNF4A	p.Q255*	0	0	0	0	[43]
	p.R141*	0	0	0	0	[44]

Table 3: Results of gene burden tests comparing the frequency of variants in *BLK*, *KLF11* and *PAX4* in a disease cohort to a population cohort

The frequency of ultra-rare (allele count=1) PTV and missense variants in a MODY cohort (n=1227) were compared to the frequency in the population cohort UK Biobank (n=185,898).

Variant type	Gene	Allele count in MODY cohort	Allele frequency in MODY cohort	Allele count in Population cohort (UK biobank)	Allele frequency in Population cohort (UK Biobank)	Odds ratio (95%CI)	*P*	Prior Probability	Bayesian false-discovery probability (BFDP)
Ultra rare PTVs	*BLK*	1	4.10×10^{-4}	14	3.77×10^{-5}	10 (0.26-71)	0.09	0.2	0.71
	KLF11	0	0	15	4.03×10^{-5}	0 (0-39)	1	0.2	0.8
	PAX4	0	0	6	1.61×10^{-5}	0 (0-97)	1	0.2	0.8
	HNF1A	15	6.11×10^{-3}	3	8.07×10^{-6}	762 (215-4108)	1.37×10^{-30}	0.2	1.20×10^{-09}
	HNF4A	3	1.22×10^{-3}	2	5.38×10^{-6}	228 (26-2724)	2.79×10^{-6}	0.2	0.03
Ultra rare Missense	*BLK*	1	4.10×10^{-4}	141	3.79×10^{-4}	1 (0.03-6)	0.6	0.2	0.82
	KLF11	2	8.10×10^{-4}	122	3.28×10^{-4}	2.5 (0.3-9)	0.2	0.2	0.78
	PAX4	2	8.10×10^{-4}	84	2.26×10^{-4}	3.6 (0.43-13)	0.1	0.2	0.70
	HNF1A	21	8.56×10^{-3}	102	2.74×10^{-4}	31 (19-51)	1.70×10^{-23}	0.2	5.80×10^{-38}
	HNF4A	14	5.70×10^{-3}	83	2.23×10^{-4}	26 (13-46)	4.41×10^{-15}	0.2	1.20×10^{-25}

Figure 1: Graph showing the odds ratios for ultra-rare (allele count=1) protein truncating variants (PTVs) (black lines) and missense variants (grey lines) in a MODY cohort (n=1227) compared to the population cohort UK Biobank (n=185,898). Dotted line shows an odds ratio of 1. Bar not shown for *KLF11* and *PAX4* PTVs as odds ratio is 0 which cannot be plotted on a log axis.