BMC Medical Genetics

RESEARCH ARTICLE   **Open Access**

Comprehensive genomic analysis identifies pathogenic variants in maturity-onset diabetes of the young (MODY) patients in South India

Viswanathan Mohan[1*†], Venkatesan Radha[1†], Thong T. Nguyen[2†], Eric W. Stawiski[2,3†], Kanika Bajaj Pahuja[2], Leonard D. Goldstein[2,3], Jennifer Tom[3], Ranjit Mohan Anjana[1], Monica Kong-Beltran[2], Tushar Bhangale[3,4], Suresh Jahnavi[1], Radhakrishnan Chandni[5], Vijay Gayathri[1], Paul George[6], Na Zhang[2], Sakthivel Murugan[6], Sameer Phalke[6], Subhra Chaudhuri[2], Ravi Gupta[6], Jingli Zhang[2], Sam Santhosh[6], Jeremy Stinson[2], Zora Modrusan[2], V. L. Ramprasad[6], Somasekar Seshagiri[2*] and Andrew S. Peterson[2*]

Abstract

Background: Maturity-onset diabetes of the young (MODY) is an early-onset, autosomal dominant form of non-insulin dependent diabetes. Genetic diagnosis of MODY can transform patient management. Earlier data on the genetic predisposition to MODY have come primarily from familial studies in populations of European origin.

Methods: In this study, we carried out a comprehensive genomic analysis of 289 individuals from India that included 152 clinically diagnosed MODY cases to identify variants in known MODY genes. Further, we have analyzed exome data to identify putative MODY relevant variants in genes previously not implicated in MODY. Functional validation of MODY relevant variants was also performed.

Results: We found MODY 3 (*HNF1A*; 7.2%) to be most frequently mutated followed by MODY 12 (*ABCC8*; 3.3%). They together account for ~ 11% of the cases. In addition to known MODY genes, we report the identification of variants in *RFX6*, *WFS1*, *AKT2*, *NKX6–1* that may contribute to development of MODY. Functional assessment of the *NKX6–1* variants showed that they are functionally impaired.

Conclusions: Our findings showed *HNF1A* and *ABCC8* to be the most frequently mutated MODY genes in south India. Further we provide evidence for additional MODY relevant genes, such as *NKX6–1*, and these require further validation.

Keywords: MODY, Diabetes, Exome, Genomics analysis, NKX6–1

---

* Correspondence: drmohans@diabetes.ind.in; sekar@gene.com; andrewp@gene.com
†Equal contributors
[1]Madras Diabetes Research Foundation & Dr. Mohan's Diabetes Specialities Centre, No. 4, Conran Smith Road, Gopalapuram, Chennai, Tamil Nadu 600 086, India
[2]Department of Molecular Biology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA
Full list of author information is available at the end of the article

Mohan et al. BMC Medical Genetics (2018) 19:22

Page 2 of 10

Background

Maturity-onset diabetes of the young (MODY) refers to a group of monogenic, non-insulin dependent forms of diabetes that affect non-obese young adults with a prevalence of ~ 100 per million in European populations [1, 2]. MODY has a high inheritance rate and represents one end of a continuum of monogenic forms of diabetes that includes neonatal diabetes [1–3]. Seven of the known MODY genes, the largest functional class, encode transcription factors which are involved in development, maturation and maintenance of β-cells while the remaining are primarily involved in the process of insulin release in response to glucose [1–3]. Although molecular genetic studies have identified heterozygous causal variants in 14 MODY genes, it is believed that additional MODY genes remain to be identified [1, 4].

MODY is often misdiagnosed as either type 1 or type 2 diabetes [5]. Thus, an accurate molecular genetic diagnosis of monogenic diabetes forms like MODY can have a significant impact on patient care. For example, individuals with mutations in GCK do not ordinarily need pharmacological treatment whereas those with mutations in KCNJ11 or ABCC8 are often best managed with high dose sulfonylureas rather than with insulin, and those with mutations in HNF1A or HNF4A benefit from treatment with low dose sulfonylureas [6, 7]. A recent study on diabetes in youth reported that MODY was misdiagnosed in 36% of the cases as type 1 (T1DM) and in 51% of the cases as type 2 (T2DM), highlighting the need for the inclusion of molecular genetic diagnosis in the management of young diabetic patients [8].

Given the opportunity for accurate diagnosis and informed treatment, genetic testing for MODY is an ideal test case for effective implementation of precision health care [1, 7]. For those individuals whose age of onset is less than 30 years, molecular genetic testing for variants in known MODY genes has the potential to provide significant patient benefit [6]. For those with later onset or for those who are young and asymptomatic, it is not clear whether genotyping would be beneficial since most studies of MODY have focused on affected individuals and closely related family members, leading to inflated estimates of effect sizes. A recent study of 7 MODY genes in ~ 4000 individuals of European and African-American ancestry identified variants that are predicted to be damaging at a general population frequency of 0.5–1.5% [9]. Damaging variants were however substantially enriched in cohorts of MODY patients or in young, lean subjects with diabetes thus reiterating the potential value of genetic diagnosis in guiding treatment for those diagnosed with diabetes by conventional criteria. A significant caveat to the interpretations outlined above is that the overwhelming majority of sequencing studies that seek to understand disease-causing variants and estimate disease burden have focused on individuals of Northern European descent [10]. These studies have found that ~ 70–80% of causal variants are in either GCK or HNF1A. Small-scale studies of South Asian MODY patients have indicated that substantially fewer MODY cases can be explained by variants in HNF1A [11, 12], HNF4A [13], GCK [14] and HNF1B [15]. Given the distinct clinical profile of diabetes in the Indian subcontinent [16, 17] it is not surprising that a different pattern of MODY inheritance may exist and indeed South Asians may harbor variants in yet to be identified MODY loci. The large population, the significant burden of Mendelian disorders, the very high incidence of diabetes and the earlier age of onset in the Indian subcontinent [18] make it particularly important to understand the pattern of MODY gene variation in India.

In this study, we have performed a comprehensive characterization of the spectrum of MODY gene variation in a cohort of patients from South India. Our genomic analysis and functional testing indicate NKX6–1 to be a potential MODY gene.

Methods

Samples, DNA and RNA preps

Diabetic subjects were recruited from a large diabetes center in Chennai (formerly Madras) city in southern India. All patients underwent a structured assessment including detailed family history and met MODY clinical criteria of Fajans and Tattersal [19]: age at diagnosis of diabetes 30 years or less, control of hyperglycemia for a minimum period of 2 years without insulin, negative for auto antibodies, absence of ketonuria at any time and evidence of autosomal dominant inheritance including a three-generation family history of diabetes.

Clinical characteristics of the MODY patients and those of the normal glucose tolerant controls are given in Additional file 1: Table S1a. Written informed consent was obtained from all the study participants. The study was approved by the Institutional Ethics Committee of the Madras Diabetes Research Foundation. The reported investigations have been carried out in accordance with the principles of the Declaration of Helsinki.

Consented and de-identified patient blood samples were used for extraction of DNA. EDTA anti-coagulated venous blood samples were collected from all study subjects, and the genomic DNA was isolated from whole blood by proteinase K digestion followed by phenol-chloroform extraction. Subsequently genomic DNA was precipitated in ethanol. The quality and quantity were assessed spectrophotometrically.

A total of 289 samples were used in this study. This consisted of 152 clinically diagnosed MODY cases and 137 normal glucose tolerant (NGT) subjects used for variant filtering.

Mohan *et al. BMC Medical Genetics* (2018) 19:22

Page 3 of 10

Whole genome, exome and targeted sequencing

Whole genome libraries were constructed using TruSeq-Nano DNA Library Preparation Kit (Illumina, CA) and sequenced on Illumina HiSeq2500 (Illumina, CA) to generate 2×75 bp read.

The whole exome analysis was performed using Agilent SureSelect (Santa Clara, CA) Human All Exome kit v5 (50 Mb). Exome capture libraries were sequenced on HiSeq 2500 (Illumina, CA). Targeted exome sequencing of 258 samples (includes 121 MODY samples and 137 normal glucose tolerant subjects) was performed using custom probes corresponding to 1965 genes that included established MODY genes and those that were implicated in pancreatic cell biology and/or diabetes (Additional file 1: Table S3).

Sequence data processing

All sequencing reads were evaluated for quality using the Bioconductor ShortRead package [20]. An all-against-all sample comparison was done on germline variants to confirm the patient exome-WGS data pairing and establish sample relatedness prior to additional data analysis.

Variant calling

Whole genome and whole exome sequencing data were combined and then processed using the Genome Analysis Toolkit (GATK) (version v3.5–0-g36282e4) best practices recommendations [21, 22]. Briefly, reads were mapped to the human reference genome GRCh37 using BWA-MEM (version 0.7.10; http://arxiv.org/abs/1303.3997). Duplicate alignments were marked and removed using Picard tool (version 1.126) (http://broadinstitute.github.io/picard/) followed by indel realignment and base quality score recalibration. The GATK Haplotype Caller algorithm was used to generate gVCFs for all samples. Joint variant calling was performed separately for discovery cohort samples and validation cohort samples using GATK Genotype GVCFs. Variant quality score recalibration was carried out to estimate the confidence of variants called in the discovery cohort. For the validation cohort, standard hard filters were applied using a set of criteria and parameters as recommended by GATK [22]. Variant annotation was carried out using SnpEff (version 4.1) [23].

Variant filtering

Our variant filtering strategy is outlined in Fig. 2a. It involved sequential filtering steps followed by manual review. We filtered for variants that meet the following three criteria: variants that (1) were rare (MAF < = 1% in 1000 genomes [24] and the NHLBI Exome Sequencing Project (ESP) [25]; (2) were protein-altering or potentially protein-altering; (3) and were in the curated list of 35 genes that included known MODY genes and others

implicated in early onset diabetes including neonatal diabetes (Additional file 1: Table S3). In the manual review process, the set of variants obtained following filtering were manually checked for read evidence using Integrative Genomics Viewer (IGV) [26]. We retained ultra-rare variants, which have an allele-frequency < 0.01% in ExAC and were present in at the most only one sample in the ExAC South Asian cohort. Finally, we checked if any of our filtered variants were reported in previous MODY studies as disease relevant and annotated them.

Plasmids, stable cell lines and expression studies

Mutant *NKX6-1* constructs were generated using a site-directed mutagenesis kit (Stratagene, USA) in TOPO-TA vector backbone. The mutant and wildtype sequences were sub-cloned from TOPO-TA vector into pLVX-TetOne-Puro vector (Clonetech, USA; Cat No. 631847) or pRK5 vector (Genentech, CA). Lenti-X 293T packaging cells (Clontech, USA; Cat No. 632180) were transfected with clones carrying variants in the pLVX-TetOne-Puro vector using Lenti-X Packaging Single Shots (VSV-G) (Clontech, USA; Cat No. 631275). At 48h post transfection, lentivirus containing supernatants were used to infect β-TC-6 cells according to the standard protocol in the Lenti-X Tet-One Inducible Expression Systems User Manual (Clontech, USA) for 24 h. 2 μg/mL puromycin (Gibco, USA; Cat No. A11138–03) containing β-TC-6 complete growth media with Tet-free FBS (Clontech, USA; Cat No. 631101) was used to select infected cells for 7 days. Post selection, β-TC-6 stable cell lines were maintained in complete growth media supplemented with 0.5 μg/mL puromycin. All the stable cell lines used in the subsequent studies were passaged for at least 3 times and tested with Lenti-X p24 Rapid Titer Kit (Clontech, Cat No. 632200) for the clearance of lentivirus. Stable cell lines were expanded and cultured in complete growth media with 15% Tet-free FBS for 24 h. We seeded 50,000 cells in a 12 well plate and induced gene expression using 100 ng/ml doxycycline (dox). Cells were harvested immediately following induction (0 h) and at 48 h post induction. Cells were lysed and total RNA from cell lines was extracted using RNeasy Plus Mini kit (Qiagen, CA). For each stable cell line expressing a particular gene we prepared three biological replicates for each condition (wt and variants) at 0 h (no Dox) and 48 h (with Dox) time point.

RNA-Seq and data analysis

We used 0.5μg of total RNA to generate RNA-seq libraries using TruSeq RNA Sample Preparation kit (Illumina, CA). The libraries were multiplexed and sequenced on HiSeq 2500 to obtain on average ~ 30 million single end 50 bp reads per sample.

RNA-seq data from mouse cell lines with human expression constructs were mapped to the mouse genome

(GRCm38/mm10) and respective wildtype human construct. Mapping was performed with GSNAP (2013–10-10) [27]. Only uniquely mapped reads were considered for downstream analysis. Gene models were based on RefSeq transcripts and NCBI gene annotation. Differential expression was performed with the R/Bioconductor package DESeq2 [28]. For clustering, count data were transformed by variance stabilization and genes were centered to have mean zero. Clustering was performed using 1-Pearson correlation as distance metric and average linkage. Heat maps were generated using the R/Bioconductor package NMF.
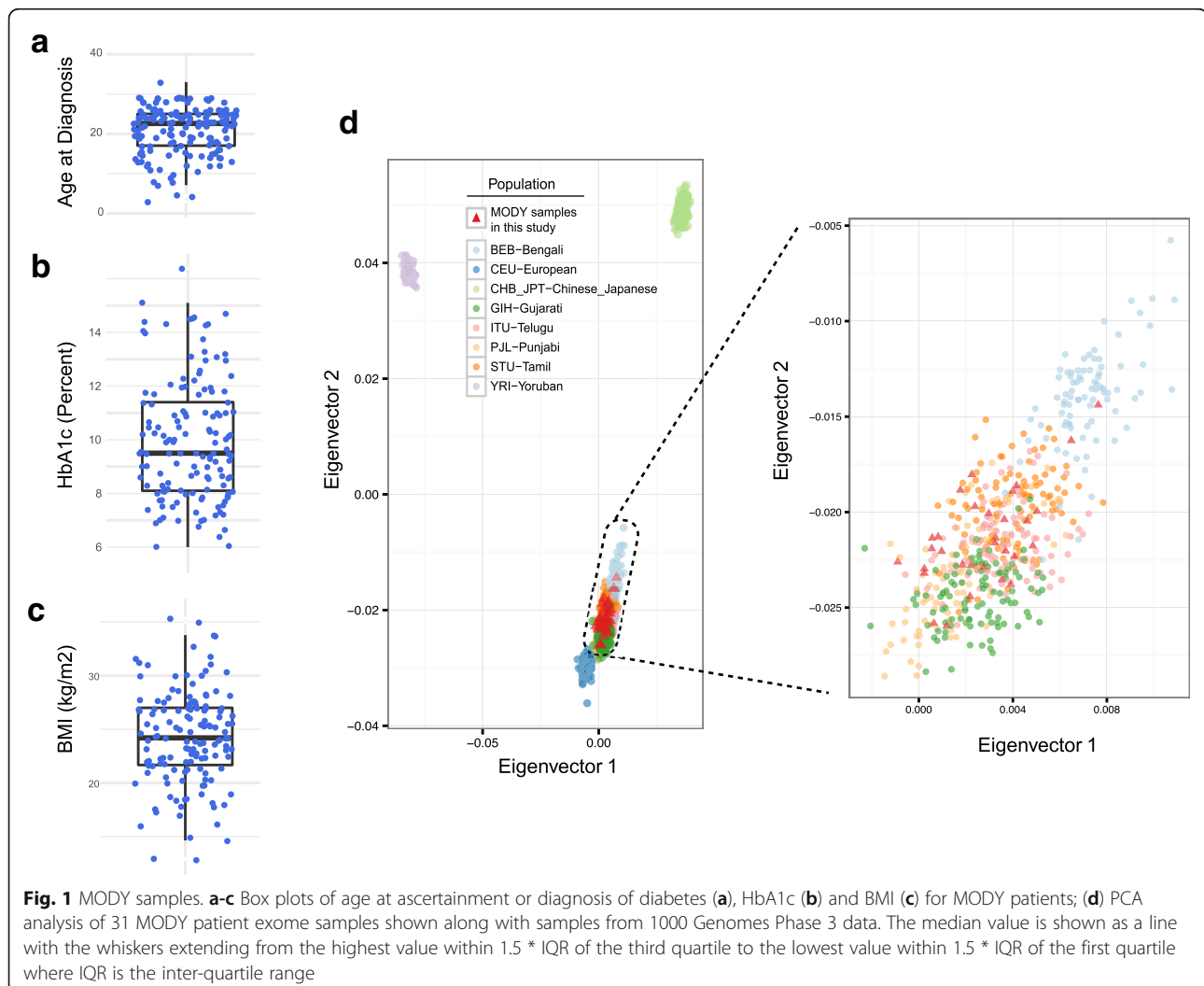
Western blot analysis

Expression of NKX6–1 was tested using western blot as previously described [29]. Briefly, β-TC-6 stable cells (50,000) expressing NKX6–1 were harvested at 48 h post induction with 100 ng/ml dox. Lysates were prepared for western blotting as previously described [29]. We used anti-NKX6–1 (Novus Biologicals; NBP-149672) and anti-

Hsp90 (Santa Cruz Biotechnology; sc-7947) antibodies for protein detection. Appropriate HRP-conjugated secondary antibodies (Pierce Biotechnology, IL) were used along with the primary antibodies in the western blot assay.

Results

Clinical characteristics

Our genomic analysis included 152 clinically diagnosed MODY patients (Additional file 1: Table S1a). The mean age of diabetes ascertainment for patients in our cohort was 20.85 ± 5.9 (±SD) years (Fig. 1a). This included 56 female and 96 male patients. Mean HbA1c glycated haemoglobin (HbA1c) values were 8.95 ± 2.26% (Fig. 1b) and mean BMI of these patients was 24.29 ± 4.37 kg/m^2 (Fig. 1c). Additional clinical characteristics of the discovery cohort (Additional file 1: Table S1b and Additional file 2: Figure S1) include fasting plasma glucose (mean: 166.2 ± 66.85 mg/dL) and fasting C-peptide levels (mean: 0.668 ± 0.40 ng/mL) and post breakfast stimulated C-peptide level (1.60 ± 0.73 ng/mL).

Fig. 1 MODY samples. **a-c** Box plots of age at ascertainment or diagnosis of diabetes (**a**), HbA1c (**b**) and BMI (**c**) for MODY patients; (**d**) PCA analysis of 31 MODY patient exome samples shown along with samples from 1000 Genomes Phase 3 data. The median value is shown as a line with the whiskers extending from the highest value within 1.5 * IQR of the third quartile to the lowest value within 1.5 * IQR of the first quartile where IQR is the inter-quartile range

Identification of MODY relevant variants

We performed whole or targeted exome sequencing of our 152 MODY cases to assess the presence of MODY relevant variation. All samples were confirmed to be unrelated (Additional file 2: Figure S2). We initially performed whole exome sequencing on 31 MODY probands (Additional file 1: Table S1a). We also performed low-pass whole genome sequencing (mean coverage ~ 6×) of the 31 probands. We combined reads from exome and low-pass WGS and achieved > 46× coverage for the 31 samples (Additional file 1: Table S2a). Principal component analysis [30] using the genotypes called from low-pass WGS, merged with those from 1000 Genomes [24] project phase 3 samples, showed that the 31 samples overlap with Indian reference samples, as expected (Fig. 1d) confirming the ethnicity of the samples used in the study. We assessed the remaining 121 MODY patient samples for presence of pathogenic variants using targeted exome sequencing (see

Methods). We obtained an average coverage of 82× for the 121 samples (Additional file 1: Table S2b).

To identify MODY relevant variants we started with 12,635,966 variants that resulted from the joint variant calling across 31 probands (see Methods; Fig. 2a). We filtered out common variants present at MAF >= 1% frequency in 1000 genomes [24] or the NHLBI Exome Sequencing Project (ESP) [25] to obtain 2,589,012 variants. We next focused on 24,117 variants predicted to have an impact on protein function and/or disease relevance. We discarded variants present in the 137 NGT samples. To facilitate variant interpretation, we focused on analysis of variants in a curated list of 35 genes that included known MODY genes and others implicated in early onset diabetes including neonatal diabetes. This resulted in 18 candidate MODY variants in 12 genes in 14 of the 31 probands (Fig. 2a). Finally, we performed manual review (see Method) on this variant list and
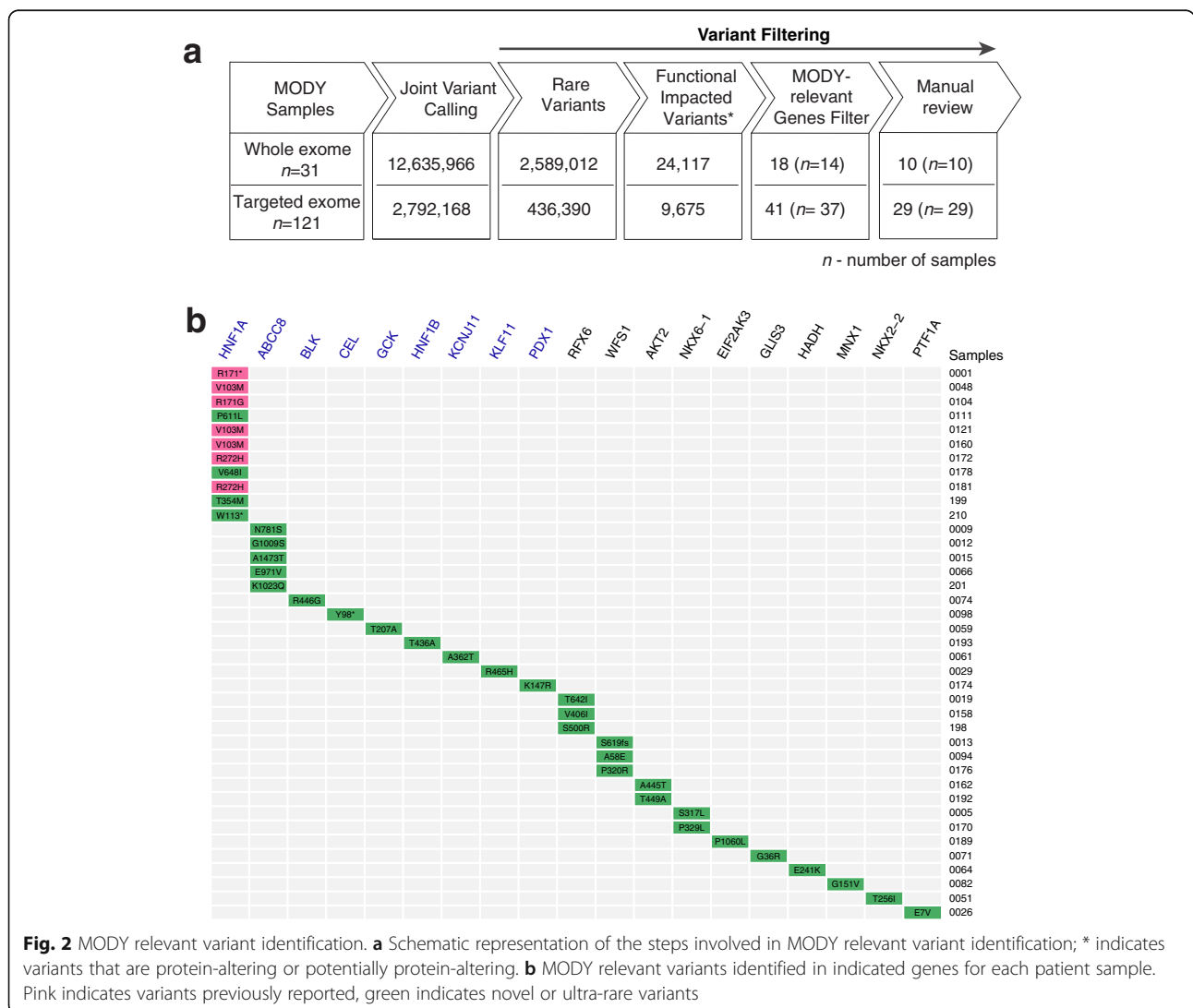

Fig. 2 MODY relevant variant identification. **a** Schematic representation of the steps involved in MODY relevant variant identification; * indicates variants that are protein-altering or potentially protein-altering. **b** MODY relevant variants identified in indicated genes for each patient sample. Pink indicates variants previously reported, green indicates novel or ultra-rare variants

arrived at 10 candidate variants (Fig. 2a; Additional file 1: Table S4).

Targeted exome sequencing and analysis of the 121 MODY samples resulted 2,792,168 variants. We applied the same filtering strategy as done for the whole exome samples above and finally arrived at 29 variants (Fig. 2a; Additional file 1: Table S4).

Combined together, we identified 39 candidate variants in 39 of the 152 MODY samples (Fig. 2b; Additional file 1: Table S4). Of these, variants R272H and V103M found in HNF1A were observed in two and three patient samples, respectively (Fig. 2b). Of the 14 known MODY genes [1], 9 were found mutated in our samples (Fig. 2b). *HNF1A* (MODY 3) was most frequently mutated at 7.2% (11/152), followed by *ABCC8* at 3.3% (MODY 12; 5/152). Other MODY genes were found to be mutated in one sample and this includes *BLK* (MODY 11), *CEL* (MODY 8), *GCK* (MODY 2), *HNF1B* (MODY 5), *KCNJ11* (MODY 13), *KLF11* (MODY 7), and *PDX1* (MODY 4). In addition to the known MODY genes, we found candidate variants in 10 MODY-relevant genes. This includes *RFX6* (*n* = 3), *WFS1* (*n* = 3), *NKX6–1* (*n* = 2), *AKT2* (n = 2) (Fig. 2b; Additional file 1: Table S4). All variants identified, except for *WFS1* P320R, were heterozygous, consistent with the dominant inheritance pattern of MODY.

Besides single-nucleotide variants, we assessed the WGS data for copy loss in known MODY genes in the 32 patients with WGS data. We detected copy loss in proband MDX 0042 that spans a 1.4 Mb region of 17q12 that contains *HNF1B* and is flanked by *TBC1D3G* at the 5′ end *TBC1D3F* at the 3′ end (Fig. 3). This patient had no other MODY candidate variants, suggesting that the loss of one copy of HNF1B is responsible for the phenotype.

NKX6–1 variants found in MODY patients are functionally impaired

Although the variants in *NKX2–2* [31], *RFX6* [32] and *MNX1* [31] have been found in recessively inherited neonatal diabetes, variants in *NKX6–1* have not thus far been found in neonatal diabetes or MODY. To understand these *NKX6–1* variants, we functionally assessed the transcriptional activity of the *NKX6–1* variants by stably expressing them or the wild-type gene in a mouse insulinoma cell line (β-TC-6), using a tetracycline inducible expression system (Additional file 2: Figure S3). We used RNA-seq to assess transcriptional activity of the two NKX6–1 variants, S317L and P329L (Fig. 4c). As a control, we included a previously described and functionally impaired control variant (EEDD321RPPR) [33] (Fig. 4). We confirmed the expression of the NKX6–1 variants (Additional file 2: Figure S3 and Additional file 2: Figure S4).

For each of the two NKX6–1 variants, and the functionally impaired control mutant, we found significant differences in gene expression between cells expressing the variant as compared to cells expressing wild type NKX6–1 (adjusted $p < 0.05$, fold-change > 1.5; Additional file 1: Table S5). Notably, clustering of samples based on genes identified as differentially expressed in at least one of the three comparisons showed that the transcriptional activity of the two MODY relevant variants was similar to the functionally impaired control mutant (Fig. 4c). This suggests that S317L and P329L NKX6–1 variants were functionally impaired and that disruption of the carboxy-terminal domain of NKX6–1 is likely responsible for MODY in these individuals (Fig. 4c). While the function of the carboxy terminus of NKX6–1 is poorly understood, deletions and substitutions have been shown to disrupt
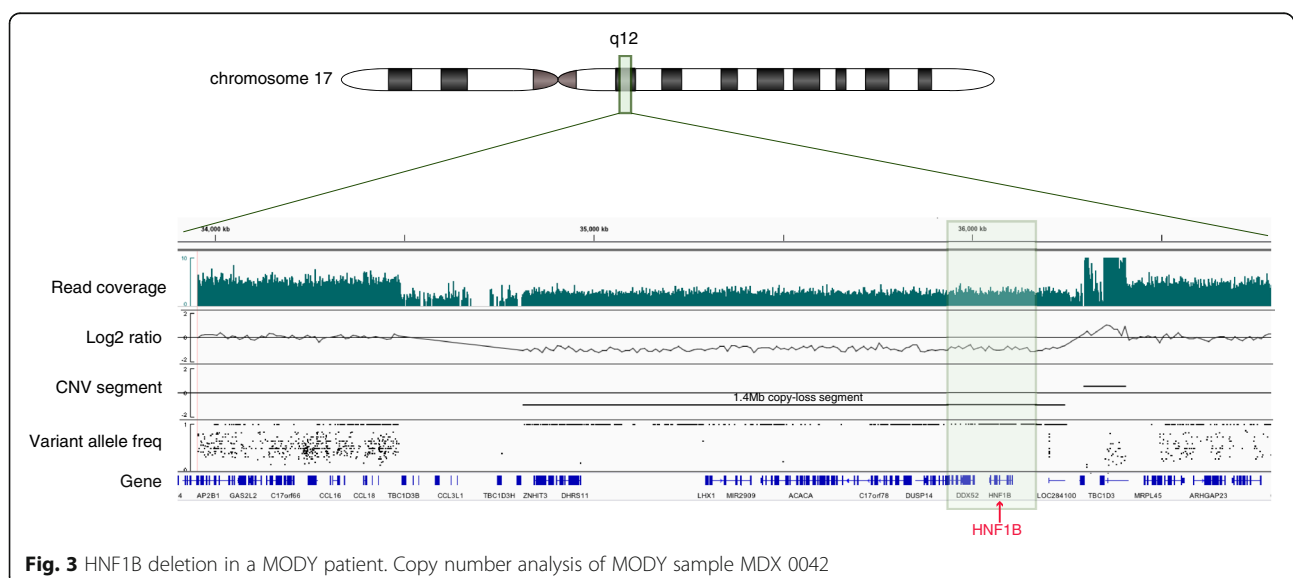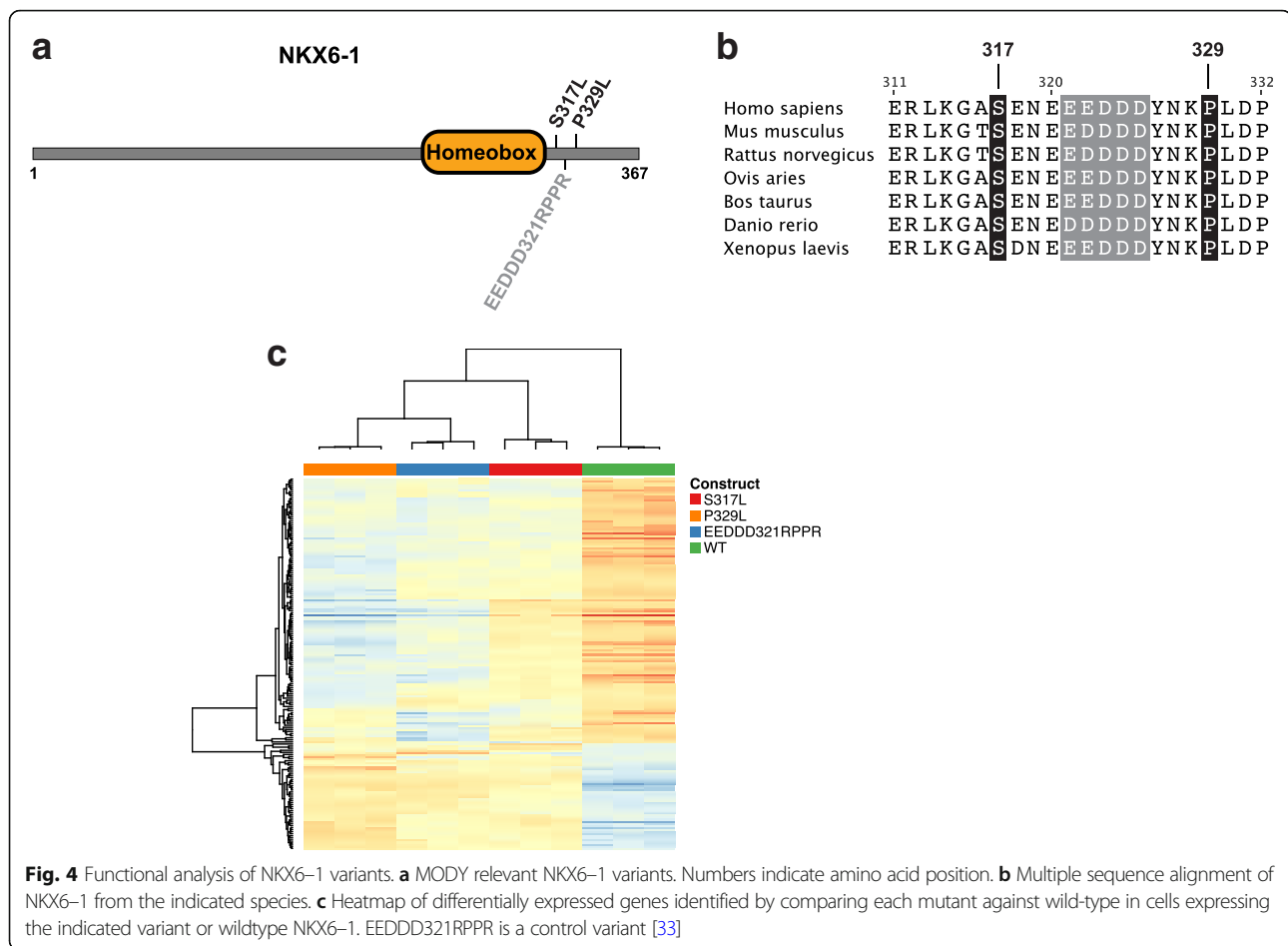

Fig. 3 HNF1B deletion in a MODY patient. Copy number analysis of MODY sample MDX 0042

Fig. 4 Functional analysis of NKX6–1 variants. **a** MODY relevant NKX6–1 variants. Numbers indicate amino acid position. **b** Multiple sequence alignment of NKX6–1 from the indicated species. **c** Heatmap of differentially expressed genes identified by comparing each mutant against wild-type in cells expressing the indicated variant or wildtype NKX6–1. EEDDD321RPPR is a control variant [33]

the ability of the nearby homeobox domain to bind DNA with appropriate specificity [34].

We found 39 genes that were differentially regulated in common between the two NKX6-1 MODY mutants (Additional file 1: Table S5), suggesting that some of these may be related to the development of diabetes in these patients. Further, inspection of the genes that are dysregulated by the two variants and comparison to NKX6–1 binding sites in the β-TC-6 cell line (MKB, JZ and ASP, unpublished) revealed a number of genes which may be related to the development of diabetes in these patients (Additional file 1: Table S5). For example, *BHLAHA15* (MIST1), with a NKX6–1 binding site upstream of the transcription start site, is significantly down-regulated by both variants. *BHLHA15* knockout mice have decreased expression of Glut2 in pancreatic β-cells and age-related impairment in glucose clearance [35], suggesting that its down-regulation could contribute to the development of diabetes in these patients. Interestingly, the most significantly up-regulated gene for all three variants was *MAP3K15* (ASK3). The function of *MAP3K15* has not been studied in any detail but it is highly homologous to Apoptosis-Signal-regulated Kinase 1 (*ASK1*), a gene that

stimulates apoptosis in response to ER stress and other noxious stimuli [36], suggesting that an increased rate of apoptotic β cell loss may also be involved in the development of diabetes in patients carrying the *NKX6–1* variants.

Discussion

Overall, we analyzed 152 clinically diagnosed unrelated MODY patients from southern India for the presence of variants in known MODY genes and additional loci of potential MODY-relevance. We found relevant variants in previously described MODY genes in ~ 15% (23/152) of the cases analyzed (Fig. 2b). As has been previously reported, the locus diversity of MODY in India is unlike that in European populations where up to 80% of patients carry variants in *HNF1A* and *GCK*. We found *HNF1A* variants in ~ 7% of the patients and *GCK* alterations in < 1% of the patient samples, a substantially lower rate than in Europeans and consistent with previous reports of Indian MODY patients [11, 12, 14, 37–39]. We found ~ 3% of the samples carried an *ABCC8* variant (MODY 12). The *ABCC8* variants have important treatment implications, as it has been suggested that these patients may be best managed with the use of sulfonylureas [40, 41], similar to

Mohan *et al. BMC Medical Genetics* (2018) 19:22

Page 8 of 10

the situation with cases of neonatal diabetes caused by mutations in *KCNJ11* or *ABCC8*.

Importantly, we found 6 variants, including three PDX1 variants P33T [42], E224K [12, 43], P242L [44], HNF4A V169I [12, 45], BLK A71T [12, 46] and NEUROD1 H241Q [12, 47], previously reported to be linked to MODY, to occur at a similar frequencies in the MODY cases and the general population. This suggests that these variants most likely have little to no effect on risk for developing MODY. The findings underscore the difficulty in interpreting relevance of variants where population specific common variant data are not available and highlights the need for such data to enable the application of genomics in the clinic.

In addition to known MODY genes we found rare variants of functional consequence in genes implicated in other forms of diabetes or beta-cell biology. This includes variants in NKX6-1 (S317L, P329L). A role for NKX6–1 in MODY is consistent with its role in β-cell function, maintenance and/or development and the phenotypes described in knockout mice studies [48, 49]. The NKX6-1 S317 and P329 are highly conserved residues across orthologs from multiple organisms and substitution at these positions is likely to impair its normal function. Consistent with this, both S317L and P329 were found to be functionally impaired in their transcriptional activity (Fig. 4). These findings while suggest a plausible role NKX6–1 in MODY further studies would be needed to confirm its role as a MODY gene.

In addition to *NKX6–1* we found variants in *WFS1*, *RFX6* and 7 other genes associated with monogenic forms of diabetes (*AKT2, EIF2AK3, GLIS3, HADH, MNX1, NKX2–2,* and *PTF1A*) and they may be relevant to development of MODY. Additional genetic or functional studies will be required to evaluate and establish the relevance of these genes in MODY.

Conclusion

In conclusion, our study reports that *HNF1A* and *ABCC8* are among the most frequently mutated MODY genes in south India. Further we provide evidence for additional MODY relevant genes, such as *NKX6–1*, and these require further validation. Our findings have important implications for MODY gene testing.

Additional files

Additional file 1: Table S1a. Sample information and clinical characteristics. **Table S1b.** Summary of clinical characteristics of MODY patient samples. **Table S2a.** Sequencing statistics of combined exome and WGS. **Table S2b.** Sequencing statistics of targeted exome samples. **Table S3.** Targeted gene panel. **Table S4.** Candidate variants identified in MODY samples. **Table S5.** Differentially expressed genes - NKX6–1 (XLS 292 kb)

Additional file 2: Figure S1. Box plot showing (a) fasting plasma glucose, (b) fasting insulin, (c) C-peptide fasting, (d) C-peptide stimulated and (e) creatinine in MODY and control samples. The median value is shown as a line with the whiskers extending from the highest value within 1.5 * IQR of the third quartile to the lowest value within 1.5 * IQR of the first quartile where IQR is the inter-quartile range. **Figure S2.** Heatmap depicting the genotype based identity of the discovery and validation MODY cohort and control samples. Genomic regions for which we obtained data for the validation cohort samples and corresponding regions from the discovery set samples using GATK joint-variant caller. The sample identity was computed based on the high-confidence set of single nucleotide variants (SNVs) that passed GATK Hard-Filtering criteria. **Figure S3.** Expression level of mouse *Nkx6–1* (top) or human *NKX6–1* (bottom) following induction in cells stably expressing the indicated variant or wildtype. **Figure S4.** Western blot showing the expression of NKX6–1 48 h post dox induction. Hsp90 was used as a loading control. (ZIP 5136 kb)

Authors' contributions
ASP, SS, TTN, VR and VM designed the study. VR, VG and CR helped in collection of samples.VM and RMA carried out clinical evaluations of MODY patients and controls. VR and SJ performed DNA extraction of study samples. PG and JT managed and analysed clinical data. TB carried out genetic analyses of data. SM, SP, SSanthosh, SJ, ZM, JAS, and VLR designed the plan for generation of sequence data and participated in its analysis. TTN, ES, RG,SJ, VR, ZM and ASP managed and analysed sequence data. KBP, MKB, NZ, SC and JZ carried out functional studies. LG analysed RNA-seq data. SS and ASP co-wrote the manuscript with the assistance and contributions of all of the other authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Written informed consent was obtained from all the study participants. Consent from the Parents or legal guardian was obtained wherever samples from minors were included in the study. The study was approved by the Institutional Ethics Committee of the Madras Diabetes Research Foundation. The reported investigations have been carried out in accordance with the principles of the Declaration of Helsinki.

Consent for publication
Not applicable

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Madras Diabetes Research Foundation & Dr. Mohan's Diabetes Specialities Centre, No. 4, Conran Smith Road, Gopalapuram, Chennai, Tamil Nadu 600

Mohan *et al. BMC Medical Genetics* (2018) 19:22

Page 9 of 10

086, India. ²Department of Molecular Biology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA. ³Department of Bioinformatics and Computational Biology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA. ⁴Department of Human Genetics, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA. ⁵Department of General Medicine, Govt. Medical College, Kozhikode 673008, India. ⁶MedGenome, Bangalore, Karnataka 560 099, India.

References

1. Anik A, Catli G, Abaci A, Bober E. Maturity-onset diabetes of the young (MODY): an update. J Pediatr Endocrinol Metab. 2015;28(3–4):251–63.
2. Schwitzgebel VM. Many faces of monogenic diabetes. J Diabetes Investig. 2014;5(2):121–33.
3. Ashcroft FM, Rorsman P. Diabetes mellitus and the beta cell: the last ten years. Cell. 2012;148(6):1160–71.
4. Prudente S, Jungtrakoon P, Marucci A, Ludovico O, Buranasupkajorn P, Mazza T, Hastings T, Milano T, Morini E, Mercuri L, et al. Loss-of-function mutations in APPL1 in familial diabetes mellitus. Am J Hum Genet. 2015;97(1):177–85.
5. Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. Nat Clin Pract Endocrinol Metab. 2008;4(4):200–13.
6. Naylor R, Philipson LH. Who should have genetic testing for maturity-onset diabetes of the young? Clin Endocrinol. 2011;75(4):422–6.
7. Thanabalasingham G, Pal A, Selwood MP, Dudley C, Fisher K, Bingley PJ, Ellard S, Farmer AJ, McCarthy MI, Owen KR. Systematic assessment of etiology in adults with a clinical diagnosis of young-onset type 2 diabetes is a successful strategy for identifying maturity-onset diabetes of the young. Diabetes Care. 2012;35(6):1206–12.
8. Pihoker C, Gilliam LK, Ellard S, Dabelea D, Davis C, Dolan LM, Greenbaum CJ, Imperatore G, Lawrence JM, Marcovina SM, 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. J Clin Endocrinol Metab. 2013;98(10):4055–62.
9. Flannick J, Beer NL, Bick AG, Agarwala V, Molnes J, Gupta N, Burtt NP, Florez JC, Meigs JB, Taylor H, et al. Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. Nat Genet. 2013;45(11):1380–5.
10. Johansson S, Irgens H, Chudasama KK, Molnes J, Aerts J, Roque FS, Jonassen I, Levy S, Lima K, Knappskog PM, et al. Exome sequencing and genetic testing for MODY. PLoS One. 2012;7(5):e38050.
11. Radha V, Ek J, Anuradha S, Hansen T, Pedersen O, Mohan V. Identification of novel variants in the hepatocyte nuclear factor-1alpha gene in south Indian patients with maturity onset diabetes of young. J Clin Endocrinol Metab. 2009;94(6):1959–65.
12. Chapla A, Mruthyunjaya MD, Asha HS, Varghese D, Varshney M, Vasan SK, Venkatesan P, Nair V, Mathai S, Paul TV, et al. Maturity onset diabetes of the young in India - a distinctive mutation pattern identified through targeted next-generation sequencing. Clin Endocrinol. 2015;82(4):533–42.
13. Anuradha S, Radha V, Mohan V. Association of novel variants in the hepatocyte nuclear factor 4A gene with maturity onset diabetes of the young and early onset type 2 diabetes. Clin Genet. 2011;80(6):541–9.
14. Kanthimathi S, Jahnavi S, Balamurugan K, Ranjani H, Sonya J, Goswami S, Chowdhury S, Mohan V, Radha V. Glucokinase gene mutations (MODY 2) in Asian Indians. Diabetes Technol Ther. 2014;16(3):180–5.
15. Kanthimathi S, Balamurugan K, Mohan V, Shanthirani CS, Gayathri V, Radha V. Identification and molecular characterization of HNF1B gene mutations in Indian diabetic patients with renal abnormalities. Ann Hum Genet. 2015;79(1):10–9.
16. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009;461(7263):489–94.
17. Majumder PP, Basu A. A genomic view of the peopling and population structure of India. Cold Spring Harb Perspect Biol. 2015;7(4):a008540.
18. Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, Bhansali A, Joshi SR, Joshi PP, Yajnik CS, et al. Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: phase I results of the Indian Council of Medical Research-INdia DIABetes (ICMR-INDIAB) study. Diabetologia. 2011;54(12):3022–7.
19. Tattersall RB, Fajans SS. A difference between the inheritance of classical juvenile-onset and maturity-onset type diabetes of young people. Diabetes. 1975;24(1):44–53.
20. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics. 2009; 25(19):2607–8.
21. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8.
22. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11 10 11–33.
23. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila Melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.
24. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. A global reference for human genetic variation. Nature. 2015; 526(7571):68–74.
25. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013; 493(7431):216–20.
26. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6.
27. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010;26(7):873–81.
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
29. Jaiswal BS, Kljavin NM, Stawiski EW, Chan E, Parikh C, Durinck S, Chaudhuri S, Pujara K, Guillory J, Edgar KA, et al. Oncogenic ERBB3 mutations in human cancers. Cancer Cell. 2013;23(5):603–17.
30. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–9.
31. Flanagan SE, De Franco E, Lango Allen H, Zerah M, Abdul-Rasoul MM, Edge JA, Stewart H, Alamiri E, Hussain K, Wallis S, et al. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. Cell Metab. 2014; 19(1):146–54.
32. Sansbury FH, Kirel B, Caswell R, Lango Allen H, Flanagan SE, Hattersley AT, Ellard S, Shaw-Smith CJ. Biallelic RFX6 mutations can cause childhood as well as neonatal onset diabetes mellitus. Eur J Hum Genet. 2015;23(12):1744–8.
33. Taylor DG, Babu D, Mirmira RG. The C-terminal domain of the beta cell homeodomain factor Nkx6.1 enhances sequence-selective DNA binding at the insulin promoter. Biochemistry. 2005;44(33):11269–78.
34. Mirmira RG, Watada H, German MS. Beta-cell differentiation factor Nkx6.1 contains distinct DNA binding interference and transcriptional repression domains. J Biol Chem. 2000;275(19):14743–51.
35. Fazio EN, Everest M, Colman R, Wang R, Pin CL. Altered Glut-2 accumulation and beta-cell function in mice lacking the exocrine-specific transcription factor, Mist1. J Endocrinol. 2005;187(3):407–18.
36. Shiizaki S, Naguro I, Ichijo H. Activation mechanisms of ASK1 in response to various stresses and its significance in intracellular signaling. Advances in biological regulation. 2013;53(1):135–44.
37. Anuradha S, Radha V, Deepa R, Hansen T, Carstensen B, Pedersen O, Mohan V. A prevalent amino acid polymorphism at codon 98 (Ala98Val) of the hepatocyte nuclear factor-1alpha is associated with maturity-onset diabetes of the young and younger age at onset of type 2 diabetes in Asian Indians. Diabetes Care. 2005;28(10):2430–5.
38. Chavali S, Mahajan A, Tabassum R, Dwivedi OP, Chauhan G, Ghosh S, Tandon N, Bharadwaj D. Association of variants in genes involved in pancreatic beta-cell development and function with type 2 diabetes in north Indians. J Hum Genet. 2011;56(10):695–700.

Mohan *et al. BMC Medical Genetics* (2018) 19:22

Page 10 of 10

39. Sahu RP, Aggarwal A, Zaidi G, Shah A, Modi K, Kongara S, Aggarwal S, Talwar S, Chu S, Bhatia V, et al. Etiology of early-onset type 2 diabetes in Indians: islet autoimmunity and mutations in hepatocyte nuclear factor 1alpha and mitochondrial gene. J Clin Endocrinol Metab. 2007;92(7):2462–7.

40. Artuso R, Provenzano A, Mazzinghi B, Giunti L, Palazzo V, Andreucci E, Blasetti A, Chiuri RM, Gianiorio FE, Mandich P, et al. Therapeutic implications of novel mutations of the RFX6 gene associated with early-onset diabetes. Pharmacogenomics J. 2015;15(1):49–54.

41. Bowman P, Flanagan SE, Edghill EL, Damhuis A, Shepherd MH, Paisey R, Hattersley AT, Ellard S. Heterozygous ABCC8 mutations are a cause of MODY. Diabetologia. 2012;55(1):123–7.

42. Gragnoli C, Stanojevic V, Gorini A, Von Preussenthal GM, Thomas MK, Habener JF. IPF-1/MODY4 gene missense mutation in an Italian family with type 2 and gestational diabetes. Metabolism. 2005;54(8):983–8.

43. Cockburn BN, Bermano G, Boodram LL, Teelucksingh S, Tsuchiya T, Mahabir D, Allan AB, Stein R, Docherty K, Bell GI. Insulin promoter factor-1 mutations and diabetes in Trinidad: identification of a novel diabetes-associated mutation (E224K) in an indo-Trinidadian family. J Clin Endocrinol Metab. 2004;89(2):971–8.

44. Ang SF, Lim SC, Tan C, Fong JC, Kon WY, Lian JX, Subramanium T, Sum CF. A preliminary study to evaluate the strategy of combining clinical criteria and next generation sequencing (NGS) for the identification of monogenic diabetes among multi-ethnic Asians. Diabetes Res Clin Pract. 2016;119:13–22.

45. Gragnoli C, Menzinger Von Preussenthal G, Habener JF. Triple genetic variation in the HNF-4alpha gene is associated with early-onset type 2 diabetes mellitus in a philippino family. Metabolism. 2004;53(8):959–63.

46. Borowiec M, Liew CW, Thompson R, Boonyasrisawat W, Hu J, Mlynarski WM, El Khattabi I, Kim SH, Marselli L, Rich SS, et al. Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. Proc Natl Acad Sci U S A. 2009;106(34):14460–5.

47. Gonsorcikova L, Pruhova S, Cinek O, Ek J, Pelikanova T, Jorgensen T, Eiberg H, Pedersen O, Hansen T, Lebl J. Autosomal inheritance of diabetes in two families characterized by obesity and a novel H241Q mutation in NEUROD1. Pediatr Diabetes. 2008;9(4 Pt 2):367–72.

48. Schaffer AE, Taylor BL, Benthuysen JR, Liu J, Thorel F, Yuan W, Jiao Y, Kaestner KH, Herrera PL, Magnuson MA, et al. Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. PLoS Genet. 2013;9(1):e1003274.

49. Sander M, Sussel L, Conners J, Scheel D, Kalamaras J, Dela Cruz F, Schwitzgebel V, Hayes-Jordan A, German M: Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. Development 2000, 127(24):5533-5540.