Diabetologia (2013) 56:1958–1963
DOI 10.1007/s00125-013-2962-5

SHORT COMMUNICATION

Improved genetic testing for monogenic diabetes using targeted next-generation sequencing

S. Ellard · H. Lango Allen · E. De Franco · S. E. Flanagan · G. Hysenaj · K. Colclough · J. A. L. Houghton · M. Shepherd · A. T. Hattersley · M. N. Weedon · R. Caswell

Received: 26 March 2013 / Accepted: 23 May 2013 / Published online: 15 June 2013

Abstract
Aims/hypothesis Current genetic tests for diagnosing monogenic diabetes rely on selection of the appropriate gene for analysis according to the patient's phenotype. Next-generation sequencing enables the simultaneous analysis of multiple genes in a single test. Our aim was to develop a targeted next-generation sequencing assay to detect mutations in all known MODY and neonatal diabetes genes.
Methods We selected 29 genes in which mutations have been reported to cause neonatal diabetes, MODY, maternally inherited diabetes and deafness (MIDD) or familial partial lipodystrophy (FPLD). An exon-capture assay was designed to include coding regions and splice sites. A total of 114 patient samples were tested—32 with known mutations and 82 previously tested for MODY (n=33) or neonatal diabetes (n=49) but in whom a mutation had not been found. Sequence data were analysed for the presence of base substitutions, small insertions or deletions (indels) and exonic deletions or duplications.
Results In the 32 positive controls we detected all previously identified variants (34 mutations and 36 polymorphisms), including 55 base substitutions, ten small insertions or deletions and five partial/whole gene deletions/duplications. Previously unidentified mutations were found in five patients with MODY (15%) and nine with neonatal diabetes (18%). Most of these patients (12/14) had mutations in genes that had not previously been tested.
Conclusions/interpretation Our novel targeted next-generation sequencing assay provides a highly sensitive method for simultaneous analysis of all monogenic diabetes genes. This single test can detect mutations previously identified by Sanger sequencing or multiplex ligation-dependent probe amplification dosage analysis. The increased number of genes tested led to a higher mutation detection rate.

Keywords Genetic diagnosis · MIDD · MODY · Monogenic diabetes · Neonatal diabetes · Next-generation sequencing

Abbreviations

FPLD	Familial partial lipodystrophy
MIDD	Maternally inherited diabetes and deafness
MLPA	Multiplex ligation-dependent probe amplification
TRMA	Thiamine-responsive megaloblastic anaemia

Introduction

A genetic diagnosis of monogenic diabetes enables some patients to stop insulin treatment. Patients with MODY caused by mutations in the transcription factor genes *HNF1A* and *HNF4A* are sensitive to low-dose sulfonylureas [1] and those with *GCK* mutations do not require pharmacological treatment. Most patients with neonatal diabetes caused by mutations in the *KCNJ11* or *ABCC8* genes encoding the K-ATP channel subunits Kir6.2 or SUR1 achieve improved glycaemic control with high-dose sulfonylureas [2].

Current diagnostic testing uses Sanger sequencing as the gold standard to detect base substitutions and small indels (insertions or deletions). An additional assay, often multiplex ligation-dependent probe amplification (MLPA), is required for the identification of partial/whole gene deletions or duplications [3]. Genetic testing is usually

Electronic supplementary material The online version of this article (doi:10.1007/s00125-013-2962-5) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

S. Ellard (✉) · H. Lango Allen · E. De Franco · S. E. Flanagan · G. Hysenaj · K. Colclough · J. A. L. Houghton · M. Shepherd · A. T. Hattersley · M. N. Weedon · R. Caswell
Institute for Biomedical and Clinical Science, University of Exeter Medical School, Barrack Road, Exeter EX2 5DW, UK
e-mail: sian.ellard@nhs.net

restricted to a small subset of genes according to the patient's phenotype [4].

Next-generation sequencing technology provides the potential for simultaneous analysis of all the known disease genes in a single assay at a similar cost to testing a few genes by Sanger sequencing. Targeted assays for gene panels ranging from two to 105 genes have been developed for polycystic kidney disease [5], Bardet–Biedl/Alström syndrome [6] and retinal disease [7]. We developed a targeted next-generation sequencing assay to identify mutations causing monogenic diabetes and tested it in a cohort of patients in whom previous testing for MODY or neonatal diabetes had failed to confirm a genetic diagnosis.

Methods

We designed a custom Agilent SureSelect exon-capture assay (Agilent Technologies, Santa Clara, CA, USA) with baits for 29 genes (see electronic supplementary material [ESM] Methods). These included 13 known/putative MODY genes (*GCK*, *HNF1A*, *HNF4A*, *HNF1B*, *NEUROD1*, *INS*, *CEL*, *PDX1*, *PAX4*, *BLK*, *KLF11*, *KCNJ11* and *ABCC8*), two genes where mutations cause diabetes through lipodystrophy (*LMNA* and *PPARG*), the m.3243 region of the mitochondrial genome (where the m.3243A>G mutation causes MIDD) and 20 neonatal diabetes genes (*GCK*, *KCNJ11*, *ABCC8*, *INS*, *PDX1*, *PTF1A*, *HNF1B*, *NEUROD1*, *NEUROG3*, *RFX6*, *EIF2AK3*, *FOXP3*, *GLIS3*, *SLC19A2*, *SLC2A2*, *IER3IP1*, *ZFP57*, *WFS1*, *GATA6* and *GATA4*). Bait density and replication were adjusted using coverage data from exome sequencing samples, captured with the Agilent SureSelect Human All Exon v1 (38 Mb) system (see ESM Fig. 1), to achieve greater uniformity of capture across the 66.8 kb target.

DNA samples from a total of 114 patients were tested—32 with known mutations identified by Sanger sequencing, MLPA dosage or array CGH (comparative genomic hybridisation) and 82 previously tested for MODY (n=33, diagnosed at age <35 years, BMI <30 kg/m^2, who had undergone previous testing for *GCK*, *HNF1A* and/or *HNF4A*) or neonatal diabetes diagnosed before 6 months (n=49, all previously tested for mutations in at least *KCNJ11*, *ABCC8* and *INS*) but in whom a mutation had not been found. Study participants gave informed consent and these investigations were carried out in accordance with the Declaration of Helsinki as revised in 2000.

Samples were fragmented using a Bioruptor (Diagenode, Liège, Belgium), indexed for multiplexing and hybridised (in pools of 12 samples) according to the manufacturer's instructions. Sequencing was performed with an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA) (48 samples per lane) and 100 bp paired end reads. Data were processed as described previously [8] to identify potential pathogenic mutations located within 50 bp upstream and 10 bp downstream of each exon. Deletions/duplications >30 bp were identified by relative read depth coverage. All newly identified mutations were confirmed by Sanger sequencing.

Results

The targeted next-generation sequencing assay 'captured' the protein coding regions and conserved splice sites of the 28 monogenic diabetes genes and the m.3243 region of the mitochondrial genome from the patients' DNA samples by hybridisation. These DNA fragments were then amplified and sequenced on an Illumina HiSeq 2000 to generate multiple reads per base. The average read depth across the targeted gene regions in the 114 samples was 257 per base (SD ±85) with ≥30 reads for 97.5% of bases and ≥20 reads for 98.0% of bases. For the 20 genes where testing is currently available in our laboratory by Sanger sequencing, the average depth of coverage was 272 and 99.4% of bases had a minimum read depth of 30 (Fig. 1a). One specific region of low coverage was observed across a ~300 bp GC-rich region of *GATA6* exon 2 (Fig. 1b).

We identified all 34 mutations in the 32 positive control samples (Table 1). These included 19 base substitutions, ten

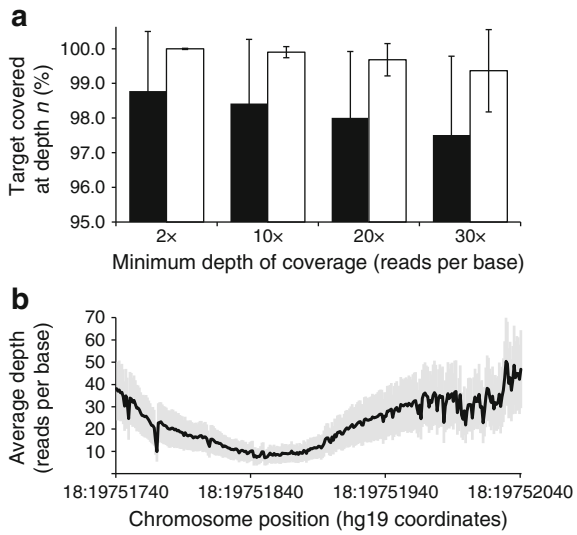

Fig. 1 Depth of coverage for targeted genes. (**a**) The percentage of targeted bases sequenced at a minimum depth of 2, 10, 20 and 30 reads per base for regions of interest for all 28 MODY/neonatal diabetes genes and the m.3243 base (black bars) and for the 20 genes routinely tested by Sanger sequencing in our laboratory (white bars). Data are mean values for the cohort of 114 samples; error bars show 1 SD from the mean. (**b**) Region of low coverage (<30 reads per base) within *GATA6* exon 2. The graph shows the average depth of coverage (black line) in the region chr18:19,751,740–19,752,040 (hg19 coordinates); grey shading indicates 1 SD from the mean. Average coverage depth over the entire *GATA6* exon 2 region of interest (chr18:19,751,056–19,752,250) was 126

Table 1 Mutations identified by targeted next-generation sequencing[a]

Gene	RefSeq	Mutation(s) by nucleotide	Mutation(s) by amino acid	Mutation type	Zygosity	Patient phenotype	Method of detection
ABCC8	U63421 and L78208	c.4261C>T+c.4594A>G	p.R1421C+p.T1532A	Base substitutions	Compound heterozygous	Hyperinsulinism	Sanger sequencing
ABCC8	U63421 and L78208	c.1631T>C+c.3643C>T	p.I544R+p.R1215W	Base substitutions	Compound heterozygous	Neonatal diabetes	Sanger sequencing
ABCC8	U63421 and L78208	c.1333-2374_1671+121dup		Duplication of exons 9–11	Heterozygous	Hyperinsulinism	MLPA
ABCC8	U63421 and L78208	c.580-1G>C+c.2476C>T	p.R826W	Base substitutions	Compound heterozygous	Neonatal diabetes	Next-generation sequencing
CEL	NM_001807.3	c.1742_2170del		Small indel	Heterozygous	MODY	Sanger sequencing
CEL	NM_001807.3	c.1686delT		Small indel	Heterozygous	MODY	Sanger sequencing
CEL	NM_001807.3	c.1785delC		Small indel	Heterozygous	MODY	Sanger sequencing
EIF2AK3	AF110146.1	c.2304_2305delTG		Small indel	Homozygous	Neonatal diabetes	Sanger sequencing
EIF2AK3	AF110146.1	c.818delC		Small indel	Homozygous	Neonatal diabetes	Sanger sequencing
EIF2AK3	AF110146.1	c.1894C>T	p.R632W	Base substitution	Homozygous	Neonatal diabetes	Next-generation sequencing
EIF2AK3	AF110146.1	c.3029G>A	p.G1010D	Base substitution	Heterozygous	Unaffected parent	Sanger sequencing
FOXP3	NM_014009.2	c.1234_1260del27		Small indel	Hemizygous	Neonatal diabetes	Sanger sequencing
FOXP3	NM_014009.2	c.1010G>A	p.R339Q	Base substitution	Hemizygous	Neonatal diabetes	Sanger sequencing
GATA6	NM_005257.3	c.1354A>AG; p.T452A	p.T452A	Base substitution	Heterozygous	Neonatal diabetes	Sanger sequencing
GATA6	NM_005257.3	c.1448_1455delTGAAAAAA		Small indel	Heterozygous	Neonatal diabetes	Sanger sequencing
GATA6	NM_005257.3	c.1296del		Small indel	Heterozygous	Neonatal diabetes	Next-generation sequencing
GATA6	NM_005257.3	c.1397A>G	p.N466S	Base substitution	Heterozygous	Neonatal diabetes	Next-generation sequencing
GATA6	NM_005257.3	c.1303-1G>T		Base substitution	Mosaic	Unaffected parent	Sanger sequencing
GCK	NM_000162.2	IVS4+2_+16del15		Small indel	Heterozygous	MODY	Sanger sequencing
GCK	NM_000162.2	c.544G>A	p.V182M	Base substitution	Heterozygous	MODY	Sanger sequencing
GCK	NM_000162.2	c.1-?_45+?del		Deletion of exon 1	Heterozygous	MODY	MLPA
GCK	NM_000162.2	c.764_767dup		Small indel	Homozygous	Neonatal diabetes	Next-generation sequencing
GCK	NM_000162.2	c.127C>T	p.R43C	Base substitution	Homozygous	Neonatal diabetes	Next-generation sequencing
GCK	NM_000162.2	c.458C>A	p.P153H	Base substitution	Heterozygous	MODY	Next-generation sequencing
GCK	NM_000162.2	c.-557G>C		Base substitution	Heterozygous	MODY	Sanger sequencing
GLIS3	NM_001042413.1	c.1765C>T+c.1-?_2790+?del	p.R589W	Base substitution+ deletion of exons 2–11	Compound heterozygous	Neonatal diabetes	Sanger sequencing and array CGH

Table 1 (continued)

Gene	RefSeq	Mutation(s) by nucleotide	Mutation(s) by amino acid	Mutation type	Zygosity	Patient phenotype	Method of detection
HNF1A	NM_000545.3	c.872dup		Small indel	Heterozygous	MODY	Sanger sequencing
HNF1A	NM_000545.3	c.1502-6G>A		Base substitution	Heterozygous	MODY	Sanger sequencing
HNF1B	NM_000458.1	c.1-?_1674+?del		Deletion of exons 1–9	Heterozygous	RCAD	MLPA
HNF1B	NM_000458.1	c.466A>G	p.K156E	Base substitution	Heterozygous	RCAD	Sanger sequencing
HNF1B	NM_000458.1	c.810-344_c.1207-2413dup		Duplication of exons 4–5	Heterozygous	RCAD	MLPA
HNF1B	NM_000458.1	c.475C>G	p.P159A	Base substitution	Heterozygous	MODY	Next-generation sequencing
HNF4A	LRG_483	c.940A>T	p.I314F	Base substitution	Heterozygous	MODY	Sanger sequencing
HNF4A	LRG_483	c.358+5G>A		Base substitution	Heterozygous	MODY	Next-generation sequencing
IER3IP1	NM_010697.3	c.235T>C	c.L78P	Base substitution	Homozygous	Neonatal diabetes	Sanger sequencing
INS	NM_000207.2	c.94G>T	p.32C	Base substitution	Heterozygous	Neonatal diabetes	Sanger sequencing
INS	NM_000207.2	c.265C>T	p.R89C	Base substitution	Heterozygous	Neonatal diabetes	Sanger sequencing
KCNJ11	NM_000525.3	c.602G>A	p.R201H	Base substitution	Heterozygous	Neonatal diabetes	Sanger sequencing
Mitochondrial	NC_012920.1	m.3243A>G		Base substitution	Heteroplasmic	MODY	Next-generation sequencing
Mitochondrial	NC_012920.1	m.3243A>G		Base substitution	Heteroplasmic	MODY	Next-generation sequencing
NKX2-2	NM_002509.3	c.365delC		Small indel	Homozygous	Neonatal diabetes	Sanger sequencing
PDX1	NM_000209.1	c.54C>A	p.C18X	Base substitution	Heterozygous	Neonatal diabetes	Sanger sequencing
PDX1	NM_000209.1	c.508T>C	p.Y170H	Base substitution	Homozygous	Neonatal diabetes	Next-generation sequencing
SLC19A2	NM_006996.2	c.759dup		Base substitution	Homozygous	Neonatal diabetes	Next-generation sequencing
SLC19A2	NM_006996.2	c.958T>G	p.W320G	Base substitution	Homozygous	Neonatal diabetes	Next-generation sequencing
WFS1	NM_006005.3	c.2489A>C	p.E830A	Base substitution	Heterozygous	Neonatal diabetes	Sanger sequencing

[a] These included the 34 mutations in 32 positive control samples identified previously by Sanger sequencing, MLPA or array CGH and the newly discovered mutations in 14 patients in whom previous genetic testing for neonatal diabetes or MODY had not identified a monogenic cause for their diabetes

small insertions or deletions (≤27 bases) and five partial/whole gene deletions/duplications (see ESM Fig. 2). The mosaic *GATA6* c.1303-1G>T mutation was present in 142/598 reads (24%). In these samples a total of 36 different polymorphisms previously identified by Sanger sequencing were confirmed. There were no false-positive variant calls. The sensitivity and specificity for variant identification was 100% (Clopper–Pearson 95% CI 94.9, 100).

Previously unidentified mutations were found in 5/33 patients (15%) referred for MODY testing (Table 1). A mitochondrial m.3243A>G mutation was found in two patients in whom *HNF1A* testing had been requested and confirmed a diagnosis of MIDD. Mutations were identified in the *GCK*, *HNF1B* and *HNF4A* genes in three patients who had previously been tested for *HNF1A* and *HNF4A*. The intronic *HNF4A* mutation, located five bases from the exon (c.358+5G>A), was not detected by genetic testing in 1999 because analysis was restricted to the exons and conserved splice sites (±2 bp). This mutation is predicted to reduce the splicing efficiency of the intron 3 splice donor site and was detected in four additional diabetic relatives.

Mutations were found in 9/49 patients (18%) with neonatal diabetes (Table 1). For eight of these nine patients the mutated gene had not been analysed previously. In three cases the mutations were in genes (*EIF2AK3* or *SLC19A2*) where neonatal diabetes is usually part of a syndrome but the initial testing was performed soon after the diagnosis of diabetes, before other features had developed. Novel mutations were found in *GCK*, *PDX1* and *GATA6* in five patients. The p.R826W *ABCC8* mutation had not been detected previously due to allelic dropout caused by a polymorphism (rs139233603) within a primer binding site that was not listed on variant databases at the time of the Sanger sequencing analysis. This patient and her affected sibling were compound heterozygotes for a previously detected inactivating *ABCC8* mutation (c.580-1G>A). Both have now transferred from insulin to sulfonylurea therapy.

Discussion

This novel targeted next-generation sequencing capture assay provides a sensitive method for simultaneous testing of mutations in 29 known/putative monogenic diabetes genes. The types of mutations identified include base substitutions, small indels and large deletions/duplications. A genetic diagnosis was obtained for 14/82 (17%) patients in whom testing had previously been limited to a subset of these genes.

Six of the 14 newly identified mutations were in genes that had not been tested previously because extrapancreatic features characteristic of the genetic subtype were not present (e.g. no known renal disease in a patient with an *HNF1B* mutation as previously described by Edghill et al [9]), had not yet presented (e.g. skeletal dysplasia in Wolcott Rallison syndrome or deafness and megaloblastic anaemia in TRMA syndrome) or were not noted at referral for genetic testing (e.g. deafness in MIDD). Two mutations were in genes where previous testing had yielded a false-negative result because the mutation was located outside the region of analysis (*HNF4A* c.358+5G>A) or was not detected due to allelic dropout at the PCR stage (*ABCC8* p.R826W). For the two siblings with the *ABCC8* p.R826W mutation the test result allowed them to stop their insulin therapy and replace it with sulfonylurea tablets.

The error rate for next-generation sequencing is estimated to be 1% [10] and therefore multiple reads are required to obtain equivalent sensitivity to Sanger sequencing. We used a bait balancing strategy (increased number of baits within regions of low coverage predicted from exome sequencing using the same SureSelect capture system) to achieve the recommended minimum of 30 reads/base for clinical diagnostic testing for 99.4% for the 20 genes currently tested by Sanger sequencing (97.5% bases with ≥30 reads for all 29 genes). Only one GC-rich region of *GATA6* (<0.3 kb out of the total 66.8 kb sequenced) proved difficult to capture and may require supplementary Sanger sequencing. The high, even coverage enabled us to detect all 70 unique mutations or polymorphisms in the positive controls. Although the assay showed 100% sensitivity and specificity, validation studies of further variants are required to reduce the CIs.

Methods for detecting large deletions or duplications (also known as copy number variants) in targeted next-generation sequencing data have lagged behind calling of base substitutions and small indels. Difficulties arise from the low and variable coverage in many other targeted assays. Our high, even coverage enabled the detection of two multi-exonic duplications and three single or multi-exon deletions by cross-sample normalisation and comparison. If further studies confirm the sensitivity of this approach it might be possible to replace both Sanger sequencing and dosage analysis by MLPA with a single targeted next-generation sequencing assay for monogenic diabetes.

Our study suggests that the implementation of targeted next-generation sequencing for clinical diagnostic testing will increase the number of patients with confirmed monogenic diabetes. A genetic diagnosis is important since it defines the diagnostic subtype, determines the most appropriate treatment and informs the sibling recurrence risk or risk of diabetes in offspring. This targeted next-generation sequencing assay may also prove to be a useful pre-screen before exome or genome sequencing for disease gene discovery.

Acknowledgements We thank H. John (University Hospitals Bristol, Bristol, UK), J. Houghton (Central Manchester Hospitals, Manchester, UK), J. Beermann (Sachsska Barn- & ungdomssjukhuset, Stockholm, Sweden), R. Clissold (University of Exeter, Exeter, UK) and K. Owen (University of Oxford, Oxford, UK) for providing clinical data. The technical assistance of A. Damhuis, A. Parrish, K. Stals and M. Day (Department of Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK) is much appreciated. S. Johansson (University of Bergen, Bergen, Norway) provided DNA samples with *CEL* variants/mutations.

Funding S. Ellard and A. T. Hattersley are employed as core members of staff within the National Institute for Health Research funded Exeter Clinical Research Facility. M. N. Weedon is supported by the Wellcome Trust as part of the WT Biomedical Informatics Hub funding. This study received funding from the European Community's Seventh Framework Programme (FP7/2008-2012) under grant agreement No. 223211 (Collaborative European Effort to Develop Diabetes Diagnostics, CEED3), BOLD (EU FP7-PEOPLE-ITN-2008 Biology of Liver and Pancreatic Development and Disease), Diabetes UK and a Wellcome Trust Senior Investigator Award to S. Ellard and A. T. Hattersley.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement SE designed the study. RC designed the targeted capture assay. RC, EDF, SEF, JALH, KC, MS, ATH, HLA, MNW, GH and SE acquired and/or analysed the data. SE wrote the manuscript. All authors reviewed and revised the manuscript critically. All authors approved the final version of the manuscript. SE is the guarantor of this work and, as such, had full access to all of the study data and takes full responsibility for the integrity of the data.

References

1. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT (2003) Genetic cause of hyperglycaemia and response to treatment in diabetes. Lancet 362:1275–1281
2. Pearson ER, Flechtner I, Njolstad PR et al (2006) Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med 355:467–477
3. Ellard S, Thomas K, Edghill EL et al (2007) Partial and whole gene deletion mutations of the GCK and HNF1A genes in maturity-onset diabetes of the young. Diabetologia 50:2313–2317
4. Ellard S, Bellanne-Chantelot C, Hattersley AT (2008) Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. Diabetologia 51:546–553
5. Rossetti S, Hopp K, Sikkink RA et al (2012) Identification of gene mutations in autosomal dominant polycystic kidney disease through targeted resequencing. J Am Soc Nephrol 23:915–933
6. Redin C, Le Gras S, Mhamdi O et al (2012) Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet–Biedl and Alström syndromes. J Med Genet 49:502–512
7. O'Sullivan J, Mullaney BG, Bhaskar SS et al (2012) A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. J Med Genet 49:322–326
8. Lango Allen H, Flanagan SE, Shaw-Smith C et al (2011) GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nature Genetics 1:20–22
9. Edghill EL, Stals K, Oram RA, Shepherd MH, Hattersley AT, Ellard S (2013) HNF1B deletions in patients with young-onset diabetes but no known renal disease. Diabet Med 30:114–117
10. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145