Next-Generation Sequencing Reveals Deep Intronic Cryptic *ABCC8* and *HADH* Splicing Founder Mutations Causing Hyperinsulinism by Pseudoexon Activation

Sarah E. Flanagan,[1] Weijia Xie,[1] Richard Caswell,[1] Annet Damhuis,[2] Christine Vianey-Saban,[3] Teoman Akcay,[4] Feyza Darendeliler,[5] Firdevs Bas,[5] Ayla Guven,[6] Zeynep Siklar,[7] Gonul Ocal,[7] Merih Berberoglu,[7] Nuala Murphy,[8] Maureen O'Sullivan,[9,10] Andrew Green,[11,12] Peter E. Clayton,[13] Indraneel Banerjee,[13,14] Peter T. Clayton,[15] Khalid Hussain,[16,17] Michael N. Weedon,[1] and Sian Ellard[1,2,*]

Next-generation sequencing (NGS) enables analysis of the human genome on a scale previously unachievable by Sanger sequencing. Exome sequencing of the coding regions and conserved splice sites has been very successful in the identification of disease-causing mutations, and targeting of these regions has extended clinical diagnostic testing from analysis of fewer than ten genes per phenotype to more than 100. Noncoding mutations have been less extensively studied despite evidence from mRNA analysis for the existence of deep intronic mutations in >20 genes. We investigated individuals with hyperinsulinaemic hypoglycaemia and biochemical or genetic evidence to suggest noncoding mutations by using NGS to analyze the entire genomic regions of *ABCC8* (117 kb) and *HADH* (94 kb) from overlapping ~10 kb PCR amplicons. Two deep intronic mutations, c.1333-1013A>G in *ABCC8* and c.636+471G>T *HADH*, were identified. Both are predicted to create a cryptic splice donor site and an out-of-frame pseudoexon. Sequence analysis of mRNA from affected individuals' fibroblasts or lymphoblastoid cells confirmed mutant transcripts with pseudoexon inclusion and premature termination codons. Testing of additional individuals showed that these are founder mutations in the Irish and Turkish populations, accounting for 14% of focal hyperinsulinism cases and 32% of subjects with *HADH* mutations in our cohort. The identification of deep intronic mutations has previously focused on the detection of aberrant mRNA transcripts in a subset of disorders for which RNA is readily obtained from the target tissue or ectopically expressed at sufficient levels. Our approach of using NGS to analyze the entire genomic DNA sequence is applicable to any disease.

Most intronic mutations occur within or close to conserved splice sites and are readily detected through Sanger sequencing of PCR amplicons targeting exons. Deep intronic mutations located >500 bp from an exon were first reported nearly 30 years ago through studies of cloned human β-globin (*HBB* [MIM 613985]).[1–3] These base substitutions create cryptic splice donor or acceptor sites that result in the inclusion of an out-of-frame pseudoexon, which leads to a frameshift and premature termination codon. Over the past three decades, virtually all deep intronic splicing mutations have been identified through mRNA analysis,[4–6] but this requires access to the target tissue or ectopic expression of aberrant transcripts in leukocytes and is compromised by the degradation of mutant transcripts through nonsense-mediated decay. Next-generation-sequencing (NGS) technology provides the capacity to analyze the entire intronic sequence of

a target gene from a genomic DNA sample rather than just the coding exons and conserved splice sites.

We studied a cohort of hyperinsulinaemic-hypoglycaemia (HH)-affected subjects in whom there was genetic evidence to suggest deep intronic mutations. HH is a genetically heterogeneous disorder that if uncontrolled can cause irreparable brain damage. The hyperinsulinism can affect the entire pancreas or be confined to a focal lesion, but these subtypes are clinically indistinguishable. Focal lesions result from a two-hit mechanism of a paternally inherited *ABCC8* (MIM 256450) or *KCNJ11* (MIM 601820) mutation and paternal uniparental isodisomy of chromosomal region 11p15.[7,8] The majority of these individuals do not respond to medical therapy, but surgical resection of the lesion is curative. In a cohort of 125 subjects with diazoxide-unresponsive HH, we previously used Sanger sequencing and dosage analysis by multiplex

[1]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter EX2 5DW, UK; [2]Department of Molecular Genetics, Royal Devon and Exeter National Health Service Foundation Trust, Exeter EX2 5AD, UK; [3]Service Maladies Héréditaires du Métabolisme et Dépistage Néonatal, Centre de Biologie et de Pathologie Est, Centre Hospitalier Universitaire, Lyon 69677, France; [4]Department of Pediatric Endocrinology, Sisli Etfal Educational and Research Hospital, Sisli, Istanbul 34377, Turkey; [5]Pediatric Endocrinology Unit, Istanbul Faculty of Medicine, Istanbul University, Istanbul 34394, Turkey; [6]Department of Pediatric Endocrinology, Göztepe Educational and Research Hospital, Kadiköy, Istanbul 34730, Turkey; [7]Department of Pediatric Endocrinology, Ankara University School of Medicine, Ankara 06100, Turkey; [8]Department of Endocrinology, Children's University Hospital, Temple Street, Dublin 12, Ireland; [9]Our Lady's Children's Hospital, Crumlin, Dublin 12, Ireland; [10]University of Dublin, Trinity College, Dublin 12, Ireland; [11]National Centre for Medical Genetics, Our Lady's Children's Hospital, Crumlin, Dublin 12, Ireland; [12]School of Medicine and Medical Science, University College Dublin, Dublin 12, Ireland; [13]Manchester Academic Health Sciences Centre, University of Manchester, Manchester M13 9WL, UK; [14]Department of Paediatric Endocrinology, Royal Manchester Children's Hospital, Manchester M13 9WL, UK; [15]Clinical and Molecular Genetics Unit, University College London Institute of Child Health, in partnership with the Great Ormond Street Hospital for Children National Health Service Trust, London WC1N 1EH, UK; [16]London Centre for Paediatric Endocrinology and Metabolism, in partnership with the Great Ormond Street Hospital for Children National Health Service Trust, London WC1N 1EH, UK; [17]Institute of Child Health, University College London, London WC1N 1EH, UK
*Correspondence: sian.ellard@nhs.net

ligation-dependent probe amplification (MLPA) to detect *ABCC8* or *KCNJ11* mutations in all but two cases with focal hyperinsulinism.[9] The unique mechanism underlying focal hyperinsulinism suggested the possibility of a deep intronic *ABCC8* mutation in these two cases given that *KCNJ11* has only a single exon.

Individuals with recessive *HADH* mutations (MIM 609975) have protein-sensitive HH and are treated with a combination of diazoxide therapy and dietary management.[10,11] Linkage analysis previously led us to identify *HADH* mutations as a common cause of diazoxide-responsive HH in consanguineous pedigrees.[12] The 3.3 Mb region of linkage included *HADH*, but mutations were only found in three of the six homozygous probands despite sequencing of the coding region, conserved splice sites, alternative exons, minimal promoter, and 3′ UTR. In one person, in vitro studies demonstrated a reduction in HADH activity in cultured skin fibroblasts, consistent with an undetected *HADH* mutation (affected individual: 10.9 nmol/min/mg protein; controls [n = 28]: 147.0 ± 29.4 nmol/min/mg protein). Since then, we have identified two further probands with a >3 cM homozygous region spanning *HADH* but no coding mutation. All five probands are of Turkish descent, and genotype analysis with Affymetrix 6.0 SNP chip (Aros Applied Biotechnology, Aarhus, Denmark) showed a shared 1.6 Mb haplotype (chr4: 108,116,751–109,714,172 hg19).

To investigate the possibility of deep intronic mutations in *HADH* and *ABCC8*, we performed NGS of (1) the 93,929 bp (chr4: 108,874,712–108,968,640 hg19) genomic region of *HADH* in a proband homozygous for the 1.6 Mb shared 4q25 haplotype, as well as his unaffected sibling, and (2) the 116,738 bp (chr11: 17,398,790–17,515,527 hg19) genomic region of *ABCC8* in the two unrelated individuals with a histological diagnosis of focal hyperinsulinism but no identified *ABCC8* or *KCNJ11* mutation. Informed consent was obtained from all participants or their parents, and institutional-review-board approval was received for this study. Peripheral leukocyte DNA was amplified in 10 (*HADH*) or 12 (*ABCC8*) overlapping long-range PCR reactions with the use of the SequalPrep Long PCR kit (Life Technologies). After pooling of amplicons in equimolar quantities, DNA was sheared by sonication (Diagenode Bioruptor), and fragments in the size range of 200–300 bp were isolated for library preparation with custom adapters with a 6 base index sequence tag. Individual libraries were enriched by 12 cycles of PCR amplification and were then pooled in equimolar quantities for 76 bp paired-end sequencing on an Illumina GAII sequencer. After sequencing, reads for each proband were sorted with the unique index sequences, the Burrows-Wheeler Aligner was used for aligning sequence reads to the hg19 reference genome, and duplicate reads were removed.[13] For the *HADH* sequencing, 100% percent of the targeted 93,929 bases were covered with at least 30 reads and had an average coverage of 1,207×. For the *ABCC8* sequencing, 99.6% percent of the targeted 116,738 bases were covered with at least 30 reads and had an average coverage of 969×. GATK was used for calling SNVs and indels.[14]

We identified 123 homozygous variants in the proband with reduced HADH activity and homozygosity across *HADH* (RefSeq accession number NM_005374.4). We then removed SNPs present in dbSNP132 and the 1000 Genomes Project (November 2011 version), indels that occurred in homopolymer tracts, and variants that were homozygous in the unaffected sibling (Table S1, available online). This left two intronic *HADH* variants and an intergenic change within the chromosome 4 region, and none of these occurred at highly conserved bases (the genomic evolutionary rate profiling [GERP] scores were <0). In silico splicing predictions were investigated with Alamut software version 2.1 (Interactive Biosoftware, Rouen, France). No effect on splicing was predicted for the *HADH* c.420-938C>T mutation. However, each of the five splicing programs (NNSpice, SpliceSiteFinder-like, MaxEntScan, Human Splicing Finder, and GeneSplicer) accessed by Alamut predicted the creation of a cryptic splice donor site resulting from the *HADH* c.636+471G>T mutation (Figure 1). The proband was also homozygous for the c.636+385A>G (rs732941) SNP, which creates a predicted cryptic splice acceptor site (Figure 1). In the presence of both variants, a 141 bp pseudoexon was predicted (Alamut software version 2.1, Interactive Biosoftware). In the two cases with focal hyperinsulinism, we found 83 shared heterozygous variants, and additional filtering (as above) yielded a single shared variant, *ABCC8* c.1333-1013A>G (RefSeq NM _000352.3) (Table S2), which was also predicted to create a cryptic splice donor site (Figure 1) and a 76 bp pseudoexon (Alamut software version 2.1n, Interactive Biosoftware). Both mutations were confirmed by Sanger sequencing, and testing of the parents identified heterozygous *HADH* c.636+471G>T mutations in both parents and heterozygous *ABCC8* c.1333-1013A>G mutations in the fathers of both focal-hyperinsulinism cases.

To investigate the effect of these variants on splicing, we derived cell lines from fibroblasts of the person homozygous for the *HADH* c.636+471G>T variant and from Epstein-Barr-virus-transformed peripheral lymphocytes of a parent heterozygous for the *ABCC8* c.1333-1013A>G variant. Total RNA was extracted and cDNA analysis was performed with primers targeted to exons 5 and 6 of *HADH* or exons 8 and 9 of *ABCC8*. Figure 2 shows the sequence electropherograms of the normal and mutant transcripts. Both cryptic splice-donor-site variants result in the inclusion of an out-of-frame pseudoexon (141 bp for *HADH*; 76 bp for *ABCC8*), which leads to a frameshift and premature termination codon. We conclude that both are pathogenic mutations causing HH.

To seek additional individuals with these cryptic splicing mutations, we undertook testing for the *HADH* variant in a further 56 diazoxide-responsive-HH-affected probands from consanguineous pedigrees (n = 36) and/or of Turkish descent (n = 28), including the four other Turkish cases
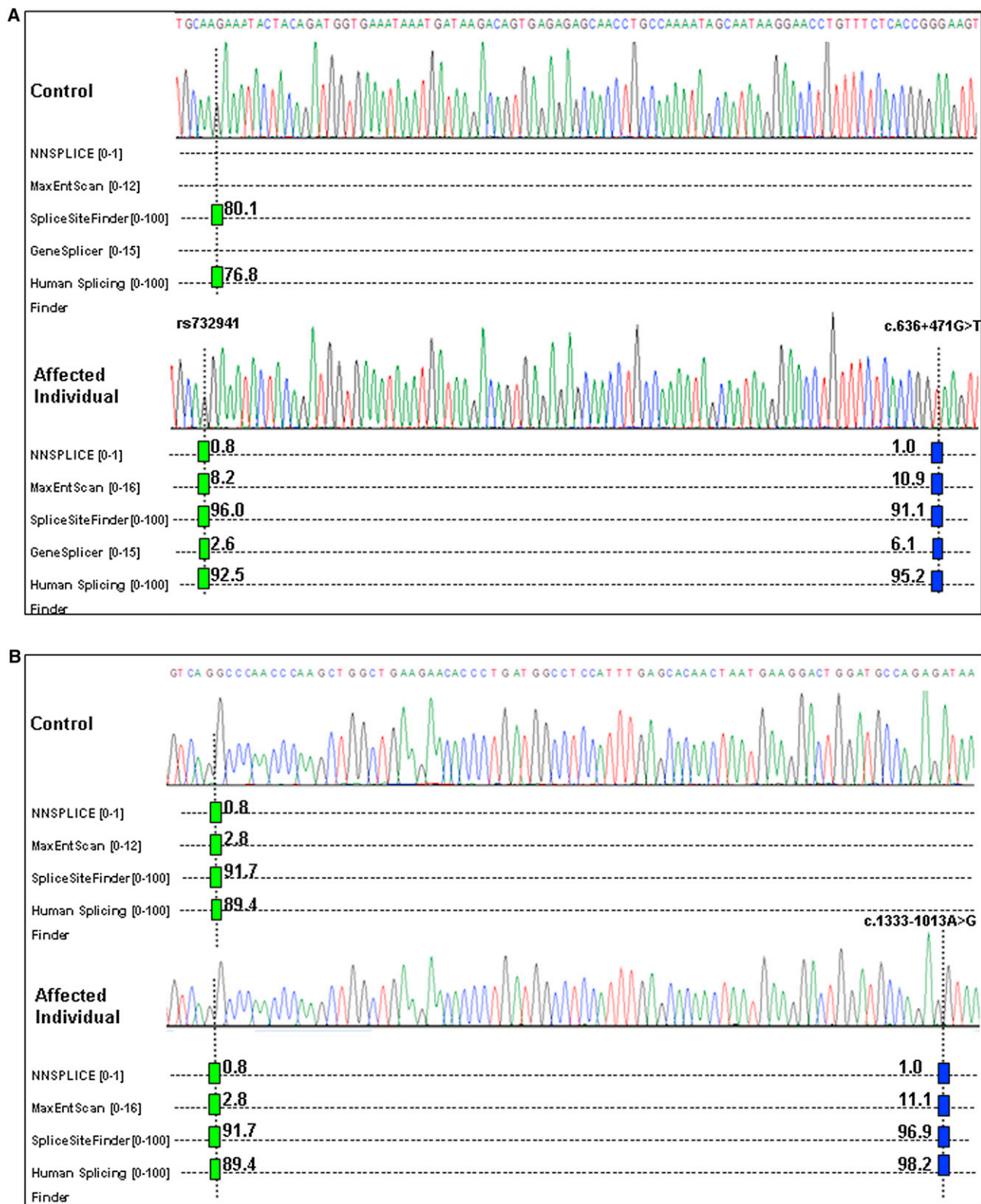
Figure 1. In Silico Splicing Predictions for the Wild-Type and Mutant Genomic Sequences for *HADH* and *ABCC8*
The positions of (A) the c.636+471G>T mutation and c.636+385A>G (rs732941) SNP in *HADH* and (B) the c.1333-1013A>G mutation in *ABCC8* are indicated by dotted lines. Hits from SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, and Human Splicing Finder are displayed as blue vertical bars for 5′ (donor) sites and as green vertical bars for 3′ (acceptor) sites. The prediction score is provided beside each bar, and the range of scores for each program is provided in brackets. A high score predicts a strong splice site.
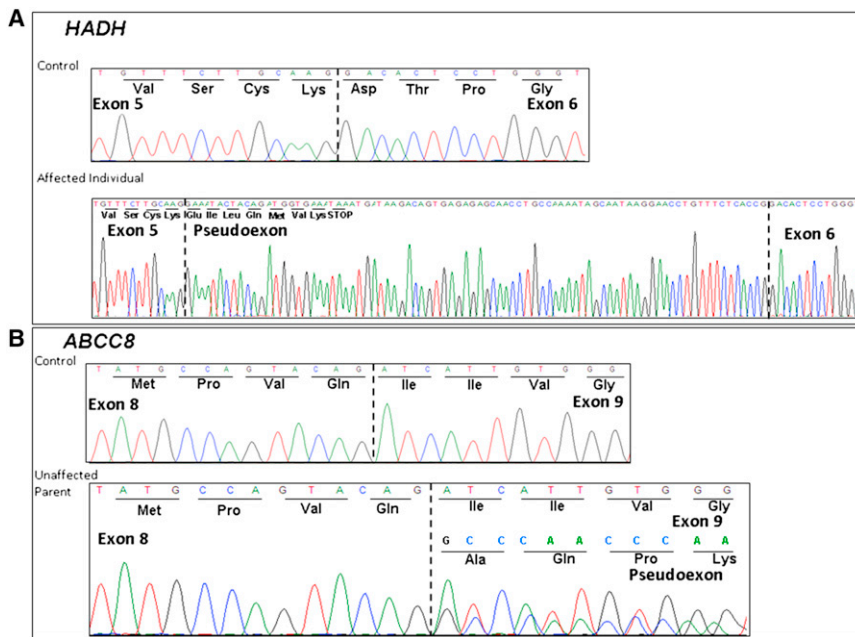
Figure 2. Electropherograms Showing cDNA Sequencing Analysis of PCR Products Amplified with Primers Targeting Exons 5 and 6 of *HADH* and Exons 8 and 9 of *ABCC8*

The nucleotide and amino acid sequences are provided. A dashed line denotes an exon-exon boundary.

(A) Sequencing analysis identified a 141 bp pseudoexon between exons 5 and 6 of *HADH* in an affected individual with homozygosity for the c.636+471G>T mutation and the c.636+385A>G (rs732941) SNP. This pseudoexon introduces a premature termination codon after seven amino acids. Sequencing analysis of an unaffected individual homozygous for the wild-type allele demonstrated normal splicing of exons 5 and 6.

(B) Sequencing analysis of cDNA from a parent with the heterozygous c.1333-1013A>G *ABCC8* mutation identified a 76 bp pseudoexon between exons 8 and 9 of *ABCC8*. This pseudoexon results in the addition of 25 amino acids, a shift in the reading frame of exon 9, and a premature termination codon after residue 4 of exon 10.

known to share the 1.6 Mb haplotype. A homozygous c.636+471G>T mutation was found in 8 of the 28 Turkish individuals, all of whom were also homozygous for the c.636+385A>G (rs732941) SNP. Testing for the *ABCC8* c.1333-1013A>G variant was performed in an additional 29 diazoxide-unresponsive-HH-affected individuals who had undergone a pancreatectomy (n = 28) or postmortem (n = 1) and in whom sequencing of *ABCC8* and *KCNJ11* and dosage analysis by MLPA had excluded a mutation. A heterozygous c.1333-1013A>G *ABCC8* cryptic splicing mutation was identified in three further probands with likely focal disease (mutations were paternally inherited in two probands, and the third was not inherited maternally, but no sample was available from that proband's father). A homozygous c.1333-1013A>G *ABCC8* mutation was identified in one person with diffuse disease, confirmed in the postmortem. Both parents are heterozygous. Four of the six probands with the *ABCC8* c.1333-1013A>G mutation were referred from the Republic of Ireland. Analysis of chromosome 11 microsatellite markers demonstrated a minimum shared haplotype of 694 kb (flanking markers D11S921 and D11S1888) in all six individuals, suggesting that this is a founder mutation in the Irish population.

Molecular genetic testing is an essential tool for the clinical management of individuals with HH.[15] Identification of a paternally inherited *KCNJ11* or *ABCC8* mutation suggests the likelihood of focal hyperinsulinism, which can be cured by surgical lesionectomy. In our cohort of 43 individuals with a histological diagnosis of focal hyperinsulinism, a heterozygous *KCNJ11* or *ABCC8* mutation has now been detected in all but one (98%) case, and the deep intronic mutation, c.1333-1013A>G, accounts for 6/43 (14%) cases. We have now included testing for this mutation in our rapid *KCNJ11* and *ABCC8* Sanger sequencing protocol in order to increase the mutation detection rate and distinguish between focal lesions and diffuse disease. The *HADH* c.636+471G>T Turkish founder mutation is the most common *HADH* mutation in our cohort and accounts for 9/28 (32%) subjects with *HADH* mutations. Leucine hypersensitivity is a feature of *HADH* mutations, and a genetic diagnosis highlights the need for dietary management in addition to diazoxide therapy.[10,11,16]

NGS enabled the identification of deep intronic founder mutations in 15 individuals with HH. We achieved this by analyzing the entire genomic region (~100 kb) of each gene in order to find these mutations, which had not previously been detected by Sanger sequencing of the ~5 kb region that includes the exons and conserved splice sites. This direct analysis of genomic DNA is not subject to the limitations of mRNA analysis for finding such mutations. These limitations include access to tissues that express the gene, the presence of multiple mRNA transcripts, and degradation of mRNA transcripts containing premature termination codons. For individuals with HH, any stored pancreatic tissue has usually been fixed in formalin and rendered suboptimal for mRNA analysis. We were able to analyze mRNA from fibroblasts or lymphoblastoid cell lines, but these were created for our research study and are not usually available for clinical diagnostic tests.

To our knowledge, there are no reports of whole-genome sequencing being used for the identification of pathogenic noncoding mutations. Exome sequencing has proved to be a successful strategy for identifying coding mutations in unassociated etiological genes, and our study suggests

that analyzing noncoding regions of known genes by targeted NGS might prove useful before starting to search for mutations in unknown genes. The contribution of deep intronic mutations to the overall mutation burden has yet to be determined given that the current estimate of 0.9% (Human Gene Mutation Database) is based on partial ascertainment. This category of mutations is of therapeutic interest because they might be amenable to correction through antisense therapy.[17,18]

In this study, we demonstrated the utility of NGS for identifying deep intronic splicing mutations causing HH. Sequencing the entire genomic DNA sequence might increase the diagnostic mutation yield for many other genetic diseases and could save futile exome-sequencing studies.

Supplemental Data

Supplemental Data include two tables and can be found with this article online at http://www.cell.com/AJHG.

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://www.1000genomes.org/
Human Gene Mutation Database (HGMD), http://www.hgmd.cf.ac.uk
NCBI dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq

References

1. Dobkin, C., Pergolizzi, R.G., Bahre, P., and Bank, A. (1983). Abnormal splice in a mutant human beta-globin gene not at the site of a mutation. Proc. Natl. Acad. Sci. USA 80, 1184–1188.

2. Treisman, R., Orkin, S.H., and Maniatis, T. (1983). Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature 302, 591–596.

3. Cheng, T.C., Orkin, S.H., Antonarakis, S.E., Potter, M.J., Sexton, J.P., Markham, A.F., Giardina, P.J., Li, A., and Kazazian, H.H., Jr. (1984). beta-Thalassemia in Chinese: Use of in vivo RNA analysis and oligonucleotide hybridization in systematic characterization of molecular defects. Proc. Natl. Acad. Sci. USA 81, 2821–2825.

4. Highsmith, W.E., Burch, L.H., Zhou, Z., Olsen, J.C., Boat, T.E., Spock, A., Gorvoy, J.D., Quittel, L., Friedman, K.J., Silverman, L.M., et al. (1994). A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. N. Engl. J. Med. 331, 974–980.

5. Tuffery-Giraud, S., Saquet, C., Chambert, S., and Claustres, M. (2003). Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy. Hum. Mutat. 21, 608–614.

6. Spier, I., Horpaopan, S., Vogt, S., Uhlhaas, S., Morak, M., Stienen, D., Draaken, M., Ludwig, M., Holinski-Feder, E., Nöthen, M.M., et al. (2012). Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. Hum. Mutat. 33, 1045–1050.

7. de Lonlay, P., Fournet, J.C., Rahier, J., Gross-Morand, M.S., Poggi-Travert, F., Foussier, V., Bonnefont, J.P., Brusset, M.C., Brunelle, F., Robert, J.J., et al. (1997). Somatic deletion of the imprinted 11p15 region in sporadic persistent hyperinsulinemic hypoglycemia of infancy is specific of focal adenomatous hyperplasia and endorses partial pancreatectomy. J. Clin. Invest. 100, 802–807.

8. Damaj, L., le Lorch, M., Verkarre, V., Werl, C., Hubert, L., Nihoul-Fékété, C., Aigrain, Y., de Keyzer, Y., Romana, S.P., Bellanne-Chantelot, C., et al. (2008). Chromosome 11p15 paternal isodisomy in focal forms of neonatal hyperinsulinism. J. Clin. Endocrinol. Metab. 93, 4941–4947.

9. Flanagan, S., Damhuis, A., Banerjee, I., Rokicki, D., Jefferies, C., Kapoor, R., Hussain, K., and Ellard, S. (2012). Partial ABCC8 gene deletion mutations causing diazoxide-unresponsive hyperinsulinaemic hypoglycaemia. Pediatr. Diabetes 13, 285–289.

10. Heslegrave, A.J., Kapoor, R.R., Eaton, S., Chadefaux, B., Akcay, T., Simsek, E., Flanagan, S.E., Ellard, S., and Hussain, K. (2012). Leucine-sensitive hyperinsulinaemic hypoglycaemia in patients with loss of function mutations in 3-Hydroxyacyl-CoA Dehydrogenase. Orphanet J. Rare Dis. 7, 25.

11. Kapoor, R.R., James, C., Flanagan, S.E., Ellard, S., Eaton, S., and Hussain, K. (2009). 3-Hydroxyacyl-coenzyme A dehydrogenase deficiency and hyperinsulinemic hypoglycemia: Characterization of a novel mutation and severe dietary protein sensitivity. J. Clin. Endocrinol. Metab. 94, 2221–2225.

12. Flanagan, S.E., Patch, A.M., Locke, J.M., Akcay, T., Simsek, E., Alaei, M., Yekta, Z., Desai, M., Kapoor, R.R., Hussain, K., and Ellard, S. (2011). Genome-wide homozygosity analysis reveals HADH mutations as a common cause of diazoxide-responsive hyperinsulinemic-hypoglycemia in consanguineous pedigrees. J. Clin. Endocrinol. Metab. 96, E498–E502.

13. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

14. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

15. Banerjee, I., Skae, M., Flanagan, S.E., Rigby, L., Patel, L., Didi, M., Blair, J., Ehtisham, S., Ellard, S., Cosgrove, K.E., et al. (2011). The contribution of rapid KATP channel gene

mutation analysis to the clinical management of children with congenital hyperinsulinism. Eur. J. Endocrinol. *164*, 733–740.

16. Li, C., Chen, P., Palladino, A., Narayan, S., Russell, L.K., Sayed, S., Xiong, G., Chen, J., Stokes, D., Butt, Y.M., et al. (2010). Mechanism of hyperinsulinism in short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency involves activation of glutamate dehydrogenase. J. Biol. Chem. *285*, 31806–31818.

17. Mancini, C., Vaula, G., Scalzitti, L., Cavalieri, S., Bertini, E., Aiello, C., Lucchini, C., Gatti, R.A., Brussino, A., and Brusco, A.

(2012). Megalencephalic leukoencephalopathy with subcortical cysts type 1 (MLC1) due to a homozygous deep intronic splicing mutation (c.895-226T>G) abrogated in vitro using an antisense morpholino oligonucleotide. Neurogenetics *13*, 205–214.

18. Nakamura, K., Du, L., Tunuguntla, R., Fike, F., Cavalieri, S., Morio, T., Mizutani, S., Brusco, A., and Gatti, R.A. (2012). Functional characterization and targeted correction of ATM mutations identified in Japanese patients with ataxia-telangiectasia. Hum. Mutat. *33*, 198–208.