Gene 550 (2014) 136–140

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short Communication

Genotype-based databases for variants causing rare diseases

Barbara Lanthaler [a], Stefanie Wieser [b], Andrea Deutschmann [a], Anna Schossig [a], Christine Fauth [a], Johannes Zschocke [a], Martina Witsch-Baumgartner [a,*]

[a] Division of Human Genetics, Medical University Innsbruck, Innsbruck, Austria
[b] Department of Paediatrics and Adolescent Medicine, Hospital Salzburg, Salzburg, Austria

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 23 April 2014
Accepted 7 August 2014
Available online 8 August 2014

Keywords:
ACAD8
ACADSB
AUH
DHCR7
FKBP14
HMGCS2
HSD17B10
ROGDI
LOVD 3.0

ABSTRACT

Inherited diseases are the result of DNA sequence changes. In recessive diseases, the clinical phenotype results from the combined functional effects of variants in both copies of the gene. In some diseases there is often considerable variability of clinical presentation or disease severity, which may be predicted by the genotype. Additional effects may be triggered by environmental factors, as well as genetic modifiers which could be nucleotide polymorphisms in related genes, e.g. maternal *ApoE* or *ABCA1* genotypes which may have an influence on the phenotype of SLOS individuals. Here we report the establishment of genotype variation databases for various rare diseases which provide individual clinical phenotypes associated with genotypes and include data about possible genetic modifiers. These databases aim to be an easy public access to information on rare and private variants with clinical data, which will facilitate the interpretation of genetic variants.

The created databases include *ACAD8* (isobutyryl-CoA dehydrogenase deficiency (IBD)), *ACADSB* (short-chain acyl-CoA dehydrogenase (SCAD) deficiency), *AUH* (3-methylglutaconic aciduria (3-MGCA)), *DHCR7* (Smith–Lemli–Opitz syndrome), *HMGCS2* (3-hydroxy-3-methylglutaryl-CoA synthase 2 deficiency), *HSD17B10* (17-beta-hydroxysteroid dehydrogenase X deficiency), *FKBP14* (Ehlers–Danlos syndrome with progressive kyphoscoliosis, myopathy, and hearing loss; EDSKMH) and *ROGDI* (Kohlschütter–Tönz syndrome). These genes have been selected because of our specific research interests in these rare and metabolic diseases. The aim of the database was to include all identified individuals with variants in these specific genes. Identical genotypes are listed multiple times if they were found in several patients, phenotypic descriptions and biochemical data are included as detailed as possible in view also of validating the proposed pathogenicity of these genotypes. For *DHCR7* genetic modifier data (maternal *APOE* and *ABCA1* genotypes) is also included.

Databases are available at http://databases.lovd.nl/shared/genes and will be updated based on periodic literature reviews and submitted reports.

1. Introduction

Most mutation databases traditionally focus on individual variants (mutations) and do not take into consideration interactions with additional genetic variants neither in cis nor trans, modifying factors and phenotypic description. While this is generally acceptable for autosomal dominant and X-chromosomal diseases it might be insufficient for autosomal recessive diseases which are caused by the combined effect of variants in both copies of a particular gene. In the case of recessive diseases (6 IEMs (inborn errors of metabolism) in our databases), there is the need to consider genotypes rather than variants for the prediction of clinical consequences, because they often show striking variability of clinical presentation or disease severity which can be at least partly predicted by the genotype. These characteristics should be reflected in specific "genotype databases" that not only cover genetic variants but also report the functional effects of both relevant parental variants as well as possible genetic modifiers. 'Genotype' in this context is referred to as the specific allelic composition of a certain gene.

Abbreviations: ApoE, apolipoprotein E; ABCA1, ATP-binding cassette, sub-family A; ACAD8, acyl-CoA dehydrogenase family, member 8; ACADSB, acyl-CoA dehydrogenase, short/branched chain; AUH, AU RNA binding protein/enoyl-CoA hydratase; DHCR7, 7-dehydrocholesterol reductase; HMGCS2, 3-hydroxy-3-methylglutaryl-CoA synthase 2 (mitochondrial); HSD17B10, hydroxysteroid (17-beta) dehydrogenase 10; FKBP14, FK506 binding protein 14; ROGDI, rogdi homolog (*Drosophila*); IEM, inborn errors of metabolism; IBD, isobutyryl-CoA dehydrogenase deficiency; SCAD, short-chain acyl-CoA dehydrogenase deficiency; 3-MGCA, 3-methylglutaconic aciduria; SLOS, Smith–Lemli–Opitz syndrome; EDSKMH, Ehlers–Danlos syndrome with progressive kyphoscoliosis, myopathy, and hearing loss; MHBD, 2-methyl-3-hydroxybutyryl-CoA dehydrogenation; ERAB, endoplasmic reticulum associated binding protein; ABAD, amyloid beta binding associated alcohol dehydrogenase; PPIase, peptidyl-prolyl cis–trans isomerase; EDS, Ehlers–Danlos syndrome; LOVD, Leiden Open (source) Variation Database; HGVS, Human Genome Variation Society; HGNC, HUGO gene nomenclature committee; LRG, locus reference genomic; NGS, next generation sequencing.

* Corresponding author at: Division of Human Genetics, Innsbruck, Medical University, Schöpfstraße 41, 6020 Innsbruck, Austria.
E-mail address: witsch-baumgartner@i-med.ac.at (M. Witsch-Baumgartner).

http://dx.doi.org/10.1016/j.gene.2014.08.016
0378-1119/

We have aimed to design such databases for six IEMs caused by variants in the genes; *ACAD8*, *ACADSB*, *AUH*, *HMGCS2*, *HSD17B10* (X-chromosomal) and *DHCR7* based on the Leiden Open source Variation Database structure. Additional databases for rare diseases have been established for variants in *FKBP14* and *ROGDI*.

The research focus of our institution is mainly the diagnosis of rare and metabolic diseases. Our special interest is Smith–Lemli–Opitz syndrome (SLOS). We have collected data from many patients with different variants in the *DHCR7* gene, with additional clinical information. We want to share this information not only for SLOS, but also for other genes investigated in our institute.

Submissions of new variants are encouraged, however require registration and are controlled by the curators of the databases who are experts in this field. A detailed concept of how to handle new variants was designed in our laboratory. Curation will be done according to the guidelines offered by the Journal '*Human Mutation*' (Celli et al., 2012).

To provide an overview of all reported variants in the aforementioned genes, we have generated the variation databases. The aim was to create most complete and up-to-date databases, with easy access and possibilities for clinicians and researchers to submit new sequence variants.

This article is intended to provide an overview about the current status of variant data and new possibilities in public resources.

The *ACAD8* gene on chromosome 11 encodes the acyl-CoA dehydrogenase 8 (ACAD8), a mitochondrial enzyme that catalyses the third step in the degradation of the branched chain amino acid 'valine', and has a significant enzyme activity with isobutyryl-CoA. Variants in the *ACAD8* gene lead to defects in valine metabolism named 'isobutyryl-CoA dehydrogenase deficiency' (IBD, OMIM 611283), which is an organic acidemia. Individuals suffering from IBD develop a feeding intolerance, carnitine deficiency and a dilated cardiomyopathy.

The short/branched chain acyl-CoA dehydrogenase (ACADSB) is a mitochondrial enzyme that catalyses the removal of hydrogen atoms from the α- and β-positions of straight and branched chain acyl-CoA esters in the metabolism of amino acids and fatty acids. It shows the greatest activity towards the short branched chain acyl-CoA derivative, (S)-2-methylbutyryl-CoA. '2-methylbutyryl-CoA dehydrogenase deficiency' (SBCAD, OMIM 610006) is another organic acidemia which has a defect in the isoleucine metabolism.

AU-specific RNA-binding protein (AUH) is the key enzyme of the leucine degradation. 3-Methylglutaconyl-CoA hydratase is a mitochondrial enzyme that catalyses the fifth step in leucine catabolism, which is the conversion of 3-methyl-CoA to 3-hydroxy-3-methylglutaryl-CoA. AUH has high 3-methylglutaconyl-CoA hydratase activity. It is likely that AUH regulates the mRNA life-span through binding to the ARE (AU-rich element). '3-Methylglutaconic aciduria type I' (MGA1, OMIM 250950) is an autosomal recessive disorder due to variants in the *AUH* gene which cause an increased urinary excretion of 3-methylglutaconic acid and 3-methylglutaric acid.

The *DHCR7* gene encodes the enzyme 3ß-hydroxysteroid-Δ7-reductase, which catalyses the last step of cholesterol biosynthesis. The Δ7 sterol reductase is responsible for reduction of the Δ7 double bond of 7-DHC and the formation of cholesterol. Variants in the *DHCR7* gene cause the 'Smith–Lemli–Opitz syndrome' (SLOS, OMIM 270400), which is an autosomal recessive metabolic malformation and mental retardation disorder.

3-Hydroxy-3-methylglutaryl coenzyme A (CoA) synthase 2 (HMGCS2) catalyses the condensation of acetyl-CoA and acetoacetyl CoA into 3-hydroxy-3-methylglutaryl CoA. HMGCS2 is involved in hepatic ketogenesis, a metabolic pathway that provides lipid-derived energy for various organs during times of carbohydrate deprivation. 'HMGCS2 deficiency' (OMIM 605911) is an inherited autosomal recessive disorder which originates from a defect in the enzyme that regulates the formation of ketone bodies. It causes hypoketotic hypolgycemia after prolonged fasting periods.

HSD10 (also known as ERAB or ABAD) is the gene product of the hydroxysteroid (17β) dehydrogenase type 10 gene. It is a mitochondrial multifunctional enzyme which shows enzymatic activity against a broad range of substrates including fatty acids and steroids. The primary function of the protein is the catalysis of the 2-methyl-3-hydroxybutyryl-CoA dehydrogenation (MHBD) reaction in isoleucine metabolism. Variants in this gene lead to '17-beta-hydroxysteroid dehydrogenase X deficiency' (OMIM 300438).

FKBP14 belongs to the FK506-binding protein family, known as immunophilins. The gene product of FKBP14 exhibits peptidyl-prolyl cis–trans isomerase (PPIase) activity, and is thought to accelerate protein folding. Defects in this gene are a cause of a type of 'Ehlers–Danlos syndrome' (EDSKMH, OMIM 614557).

ROGDI (rogdi homolog (*Drosophila*)) is a gene on chromosome 16 which encodes a protein of uncertain function. Loss-of-function variants in this gene cause 'Kohlschütter–Tönz syndrome' (KTZS OMIM 226750), a rare autosomal recessive disease characterized by amelogenesis imperfecta.

2. Material and methods

We used LOVD version 2.0 for the establishment of the databases (Fokkema et al., 2011). Recently the databases have been transferred to LOVD 3.0, the newest software version. All of the mentioned databases are deposited on the HGVS homepage (http://www.hgvs.org/dblist/glsdb.html) and on the LOVD homepage (http://grenada.lumc.nl/LSDB_list/lsdbs).

Gene names were verified using the HGNC (HUGO gene nomenclature committee) official gene symbol. Variants were defined according to HGVS (http://www.hgvs.org/mutnomen/recs.html). NCBI transcript reference sequences (NM-number for each gene, see Table 2) were used for variant numbering; and variants taken from the literature were corrected according to current nomenclature.

Submissions were made for the generation of an LRG (locus region genomic http://www.lrg-sequence.org/) number for the gene transcript used in the corresponding database, in order to provide a stable reference DNA sequence along with all relevant transcript and protein sequences essential for the description of gene variants (Dalgleish et al., 2010).

Each variant listed in the database was reviewed with the Mutalyzer 2.0 software (freely available https://mutalyzer.nl/).

The likely functional effects of variants were assessed with Alamut® v.2.1 software (http://www.interactive-biosoftware.com/alamut-visual), which includes the prediction software SIFT, PolyPhen-2 and Align GVGD as well as splice site prediction software SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer and Human Splice Finder (HSF). This software allows estimating the influence of variants on splicing as well as the functional relevance of missense variants, thereby enabling a prediction of likely pathogenicity.

3. Results

3.1. DHCR7 (http://databases.lovd.nl/shared/genes/DHCR7)

The individuals' phenotype is best described in the *DHCR7* database which includes descriptions of behavioural peculiarities, dysmorphisms, organ malformations, biochemical features, patient's origin (geographic and ethnic origin) and parental inheritance. To date, 453 patients have been submitted to the SLOS database. In addition to the patients analysed in our institute, all patients that have been included in the literature 1999 to 2012 (publications in English) were included in the database. Plain listings of variants without clinical information were excluded from the databases. There are 133 unique variants listed and a total of 889 variants are stated in the database. Out of the 133 listed unique variants, 92% are amino acid substitutions and 8% account for deletions. Database analysis shows the frequencies of the most common *DHCR7* variants,