www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

ACO2 clinicobiological dataset with extensive phenotype ontology annotation

Khadidja Guehlouz[1], Thomas Foulonneau[2], Patrizia Amati-Bonneau[2,3], Majida Charif[2,4], Estelle Colin[2,3], Céline Bris[2,3], Valérie Desquiret-Dumas[2,3], Dan Milea[5], Philippe Gohier[1], Vincent Procaccio[2,3], Dominique Bonneau[2,3], Johan T. den Dunnen[6], Guy Lenaers[2], Pascal Reynier[2,3] & Marc Ferré[2] ✉

Pathogenic variants of the aconitase 2 gene (*ACO2*) are responsible for a broad clinical spectrum involving optic nerve degeneration, ranging from isolated optic neuropathy with recessive or dominant inheritance, to complex neurodegenerative syndromes with recessive transmission. We created the first public locus-specific database (LSDB) dedicated to *ACO2* within the "Global Variome shared LOVD" using exclusively the Human Phenotype Ontology (HPO), a standard vocabulary for describing phenotypic abnormalities. All the variants and clinical cases listed in the literature were incorporated into the database, from which we produced a dataset. We followed a rational and comprehensive approach based on the HPO thesaurus, demonstrating that *ACO2* patients should not be classified separately between isolated and syndromic cases. Our data highlight that certain syndromic patients do not have optic neuropathy and provide support for the classification of the recurrent pathogenic variants c.220C>G and c.336C>G as likely pathogenic. Overall, our data records demonstrate that the clinical spectrum of *ACO2* should be considered as a continuum of symptoms and refines the classification of some common variants.

Background & Summary

Aconitate hydratase (ACO2; EC# 4.2.1.3) is a ubiquitous human mitochondrial monomeric enzyme composed of 780 amino acids. It catalyses the second reaction of the citric acid cycle by isomerising citrate to isocitrate[1]. It is encoded by the aconitase 2 gene (*ACO2*; MIM# 100850) that extends over 35 kb on chromosome 22q13 and includes 18 translated exons[2]. Biallelic pathogenic variants of this gene have been associated with infantile cerebellar-retinal degeneration (ICRD; MIM# 614599), a severe neurodegenerative disorder with optic neuropathy beginning in childhood and including retinal dystrophy, cerebellar ataxia, seizure, strabismus, axial hypotonia and athetosis[3]. Isolated optic neuropathies have also been associated with *ACO2* pathogenic variants, either with recessive (locus *OPA9*; MIM# 616289)[4] or dominant inheritance, as recently reported[5]. A recent review of the literature has listed a total of 26 individuals reported from 15 families with the syndromic form since 2012, while only seven individuals from four families have been reported with the recessive form of isolated optic neuropathy[6]. More specific clinical presentations, possibly without optic neuropathy, have also extended the clinical spectrum, such as a form of hereditary spastic paraplegia reported in two siblings[7]. The recent description by our team of 116 additional cases belonging to 94 families brings the total to over a hundred cases listed in the literature to date[5].

The high frequency of dominant *ACO2* mutations in our molecular diagnosis experience of optic neuropathies and the diversity of the neurological involvement led us to develop a reliable database dedicated to *ACO2* as part of the Global Variome shared Leiden Open-source Variation Database (LOVD) installation[8,9]. Indeed,

[1]Département d'Ophtalmologie, Centre Hospitalier Universitaire d'Angers, Angers, France. [2]Unité Mixte de Recherche MITOVASC, CNRS 6015/INSERM 1083, Université d'Angers, Angers, France. [3]Département de Biochimie et Génétique, Centre Hospitalier Universitaire d'Angers, Angers, France. [4]Genetics, and immuno-cell therapy Team, Mohammed First University, Oujda, Morocco. [5]Singapore National Eye Centre, Singapore Eye Research Institute, Duke-NUS, Singapore. [6]Human Genetics and Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands. ✉e-mail: marc.ferre@univ-angers.fr

we recently developed such a database to list the genetic variants and clinical presentations of *OPA1*-related disorders, the major cause of dominant optic neuropathies with either isolated (80%; MIM# 165500) or syndromic (20%; MIM# 125250) presentations, and with some rare biallelic cases affected by early severe syndromes (MIM# 605390)[10]. The interoperability and the use of an international clinical thesaurus[11] should, therefore, make it possible to progressively improve understanding of how this increasing number of genes responsible for optic neuropathies contributes to the diversity of the pathophysiological mechanisms responsible for their common clinical phenotype.

In this article, we describe the construction of this *ACO2* dataset, listing all the patients referenced in the literature, and drawing statistical information on gene variants and clinical diversity.

Methods

Nomenclature. All names, symbols, and OMIM numbers were checked for correspondence with current official names indicated by the Human Genome Organization (HUGO) Gene Nomenclature Committee[12] and the Online Mendelian Inheritance in Man database (OMIM)[13]. The phenotype descriptions are based on the standardised Human Phenotype Ontology (HPO)[11], indicating the HPO term name and identifier. *ACO2* variants are described according to both the NCBI genomic reference sequence NG_032143.1 and transcript reference sequence NM_001098.2, including 18 exons encoding a protein of 780 amino acids reference sequence NP_001089.1[14]. The numbering of the nucleotides reflects that of the cDNA, with "+1" corresponding to the "A" of the ATG translation initiation codon in the reference sequence, according to which the initiation codon is codon 1, as recommended by the version 2.0 nomenclature of the Human Genome Variation Society (HGVS; http://varnomen.hgvs.org)[15]. Information concerning changes in RNA levels has been added from the original papers or predicted from DNA mutations if not experimentally studied. Following the HGVS guidelines, deduced changes are indicated between brackets. Protein domains were predicted according to InterPro version 79.0[16] and Pfam version 32.0[17].

Data collection. Since no locus-specific database (LSDB) dedicated to the *ACO2* gene previously existed (http://www.hgvs.org/locus-specific-mutation-databases, accessed on January 12, 2021), this work was done from scratch. The causative variants were collected from the literature published to date (January 2021)[3–7,18–24] using the NCBI PubMed search tool[25] with the keyword "*ACO2*," and from the classifications of diagnostic laboratories in the Netherlands who recently decided to share them publicly (so-called the VKGL initiative)[26]. The positions of variants in the reference transcript were determined according to the HGVS nomenclature version 2.0[15]. Correct naming at the nucleotide and amino acid levels were verified, and reestablished when necessary, using the Mutalyzer 2.0.32 *Syntax Checker*[27]. All clinical descriptions of the dataset have been strictly and exhaustively translated using exclusively the HPO vocabulary, i.e. each phenotype mentioned having successfully matched an HPO term (or one of the synonyms associated with).

Integrity of the dataset. The work was coordinated by a single ophthalmologist to ensure consistency, with the help of our clinical and research team specializing in hereditary optic neuropathies, which performs monitoring of literature on the subject for several years. No data was rejected, the corresponding molecular biologist or ophthalmologist were contacted when clarification was required. Finally, the consistency and integrity of the entire dataset was validated by the curator of the database, who is a referent specialist, before the technical validation that followed. (Please also refer to the section Author contributions.)

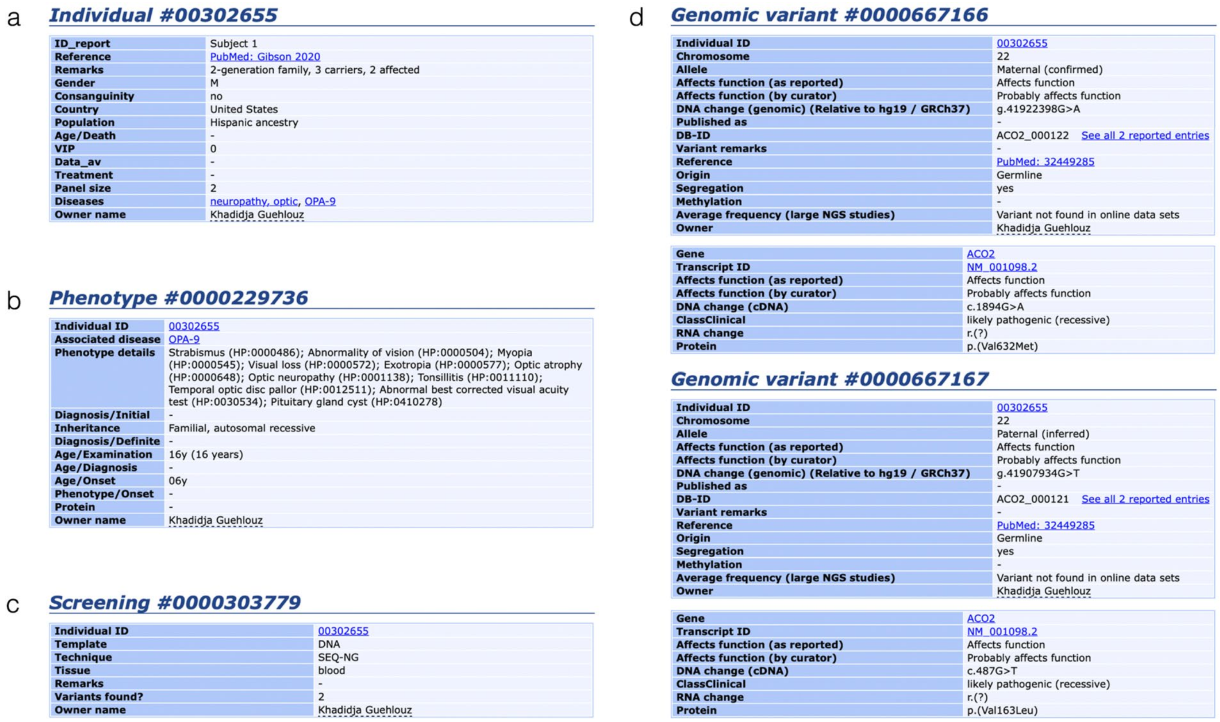
Implementation of the dataset. The *ACO2* dataset belongs to the Global Variome shared Leiden Open-source Variation Database (LOVD) currently running under LOVD v.3.0 Build 23[9], following the guidelines for LSDBs[28] and hosted under the responsibility of the Global Variome/Human Variome Project[8]. The dataset reviews clinical and molecular data from patients carrying *ACO2* variants published in peer-reviewed literature, as well as unpublished contributions that are directly submitted.

Data classification. The criteria of pathogenicity, which depend upon the clinical context and molecular findings, are stated under the headings: *"ClassClinical"* for the classification of the variant based on standardised criteria, and *"Affects function (as reported)"* and *"Affects function (by curator)"* respectively for the pathogenicity reported by the submitter and concluded by the curator (Fig. 1d). As several patients can be registered with different reported pathogenicity or new patients with existing variants added to the database, the status of the variants is reassessed on the basis of the data submitted and stated in the "*SUMMARY record*."

Dataset and analysis. Starting with version ACO2:200608 of the *ACO2* LSDB (last updated on June 8, 2020), we produced a dataset. To carry out the statistical analysis, the HPO terms have been checked and prepared using the suite of R packages ontologyX[29] within R version 4.0.0[30] to read in the OBO file version hp/releases/2020-06-08[11]. Hierarchical clustering is performed using the *hclust* function from the R-Core package[30].

Data Records

The *ACO2* dataset is available on the Code Ocean cloud-based computational reproducibility platform as a Code Ocean "compute capsule[31]": snapshot of the *ACO2* dataset as of June 8, 2020, */data/LOVD_full_download_ACO2_2020-06-21_18.03.52.txt* (LOVD flat file format); the Human Phenotype Ontology data-version hp/releases/2020-06-08, */data/hp.obo.2020-06-08.txt* (OBO flat file format), which is also available in figshare[32]. The updated dataset is accessible on the Global Variome shared LOVD server (https://www.lovd.nl/ACO2; or through the Mitochondrial Dynamics variation portal: https://mitodyn.org). The data can also be retrieved via an application programming interface (API), i.e. a web service allowing simple queries and retrieval of basic gene and variant information (documentation available on the web page of the database)[9]; as well as serving as a public



a **Individual #00302655**

ID_report	Subject 1
Reference	PubMed: Gibson 2020
Remarks	2-generation family, 3 carriers, 2 affected
Gender	M
Consanguinity	no
Country	United States
Population	Hispanic ancestry
Age/Death	-
VIP	0
Data_av	-
Treatment	-
Panel size	2
Diseases	neuropathy, optic, OPA-9
Owner name	Khadidja Guehlouz

b **Phenotype #0000229736**

Individual ID	00302655
Associated disease	OPA-9
Phenotype details	Strabismus (HP:0000486); Abnormality of vision (HP:0000504); Myopia (HP:0000545); Visual loss (HP:0000572); Exotropia (HP:0000577); Optic atrophy (HP:0000648); Optic neuropathy (HP:0001138); Tonsillitis (HP:0011110); Temporal optic disc pallor (HP:0012511); Abnormal best corrected visual acuity test (HP:0030534); Pituitary gland cyst (HP:0410278)
Diagnosis/Initial	-
Inheritance	Familial, autosomal recessive
Diagnosis/Definite	-
Age/Examination	16y (16 years)
Age/Diagnosis	-
Age/Onset	06y
Phenotype/Onset	-
Protein	-
Owner name	Khadidja Guehlouz

c **Screening #0000303779**

Individual ID	00302655
Template	DNA
Technique	SEQ-NG
Tissue	blood
Remarks	-
Variants found?	2
Owner name	Khadidja Guehlouz

d **Genomic variant #0000667166**

Individual ID	00302655
Chromosome	22
Allele	Maternal (confirmed)
Affects function (as reported)	Affects function
Affects function (by curator)	Probably affects function
DNA change (genomic) (Relative to hg19 / GRCh37)	g.41922398G>A
Published as	-
DB-ID	ACO2_000122 See all 2 reported entries
Variant remarks	-
Reference	PubMed: 32449285
Origin	Germline
Segregation	yes
Methylation	-
Average frequency (large NGS studies)	Variant not found in online data sets
Owner	Khadidja Guehlouz

Gene	ACO2
Transcript ID	NM_001098.2
Affects function (as reported)	Affects function
Affects function (by curator)	Probably affects function
DNA change (cDNA)	c.1894G>A
ClassClinical	likely pathogenic (recessive)
RNA change	r.(?)
Protein	p.(Val632Met)

Genomic variant #0000667167

Individual ID	00302655
Chromosome	22
Allele	Paternal (inferred)
Affects function (as reported)	Affects function
Affects function (by curator)	Probably affects function
DNA change (genomic) (Relative to hg19 / GRCh37)	g.41907934G>T
Published as	-
DB-ID	ACO2_000121 See all 2 reported entries
Variant remarks	-
Reference	PubMed: 32449285
Origin	Germline
Segregation	yes
Methylation	-
Average frequency (large NGS studies)	Variant not found in online data sets
Owner	Khadidja Guehlouz

Gene	ACO2
Transcript ID	NM_001098.2
Affects function (as reported)	Affects function
Affects function (by curator)	Probably affects function
DNA change (cDNA)	c.487G>T
ClassClinical	likely pathogenic (recessive)
RNA change	r.(?)
Protein	p.(Val163Leu)

Fig. 1 Sample recording for a given patient in the *ACO2* dataset. (**a**) individual items; (**b**) phenotype items; (**c**) screening items; and (**d**) molecular items (some uninformative lines were removed to save space). Abbreviations and legends of the fields are given by following the link *"Legend"* on the web page of each table; SEQ-NG: next generation sequencing; M: male. Data as of June 8, 2020.

beacon in The Global Alliance for Genomics and Health Beacon Project[33]. General information is available on the database home page. The process for submitting data begins by clicking the *"Submit"* tab.

The *ACO2*-LOVD dataset contains four main interconnected tables. These tables are visible on a typical web page entry as shown in Fig. 1. The *"Individual"* table contains details of the patient examined, including gender, geographic origin, and patient identification, if applicable (Fig. 1a). The *"Phenotype"* table indicates the clinical phenotypic features, described according to the root of the Phenotypic abnormality subontology (HPO# HP:0000118; Fig. 1b). The *"Screening"* table gives details of the methods and techniques used for investigating the structural variants and the tissue analysed (Fig. 1c). The *"Variants"* table includes information about the sequence variations at the genomic (DNA) and the transcript variant (cDNA) levels, as well as the reported and concluded status for each variant (Fig. 1d). To date, the dataset records 123 patient records and 126 unique variants (Online-only Table 1).

Since the *ACO2* dataset is built on the same central platform (to allow interoperability) and on the same model (for ease of handling) as our previous *OPA1* gene database, the description of these data records extends our related work[10], but the data recorded relates to a new gene and is entirely different.

Technical Validation

Molecular relevance. The dataset contains 126 unique variants, of which 73% (92) are considered pathogenic sequence variants with almost two thirds in a dominant condition and the last third recessive, 7% (9) of unknown significance and 20% (25) benign (Online-only Table 1). The variants considered pathogenic, which affect the coding sequence and exon-intron boundaries of the gene (Fig. 2a), are particularly overrepresented in the aconitase C-terminal domain of the protein (spanning from end of exon 14 to half of exon 17), highlighting the importance of this domain in ACO2 functions (Fig. 2b), as well as the aconitase domain since it spans more than half of the protein sequence (from beginning of exon 5 to beginning of exon 13). Only one mutation considered pathogenic (intron 1 splicing site) and two benign (exon 2) are localized in the N-terminal presequence that is cleaved upon import of ACO2 into the mitochondrial matrix, predicted in the first 28 to 35 amino acids[34]. Among the most frequently observed pathogenic effects on ACO2, 66% are missense variants; 14% are associated with altered splicing, which produces effects that are difficult to assess experimentally; 11% are frameshift variants leading to a premature protein truncation; 8% are nonsense variations; one variant is a deletion of a single amino acid (Fig. 2c). Although only a few mutations are recurrent, two have been significantly more frequently reported, both located in exon 3 (Fig. 2a) with a recessive mode of inheritance: the c.220C>G variant, which induces a missense mutation p.(Leu74Val), has been reported 19 times affecting one allele (of which 14 belongs to compound heterozygous genotypes); the c.336C>G variant, which induces a missense mutation p.(Ser112Arg), has been reported 10 times affecting both alleles (homozygous).

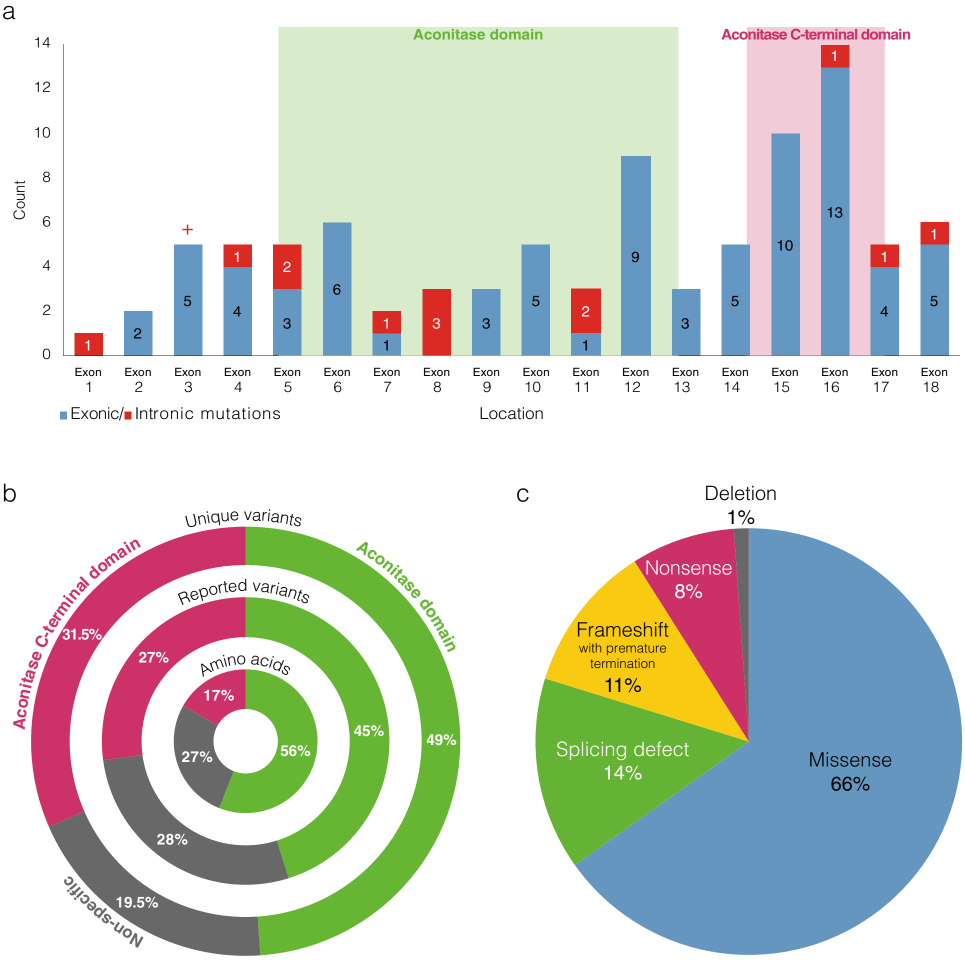


Fig. 2 Distribution of the *ACO2* variants classified as pathogenic or likely pathogenic. (**a**) Distribution of the 89 unique variants. Exons involved in the variants are shown as blue bars; the variants in the intronic neighborhood of the exons are shown as red bars; the location of the two variants significantly more frequently reported being indicated by the "+" symbol. (**b**) Comparison for each region of its size on the sequence (Amino acids), of the distribution of the variants reports in the dataset (i.e. by counting each of the reported case of the same variant; Reported variants), and of the distribution of the unique variants (i.e. by counting only once several reported cases of the same variant). (**c**) Distribution of the different effects on the protein of the *ACO2* variants considered pathogenic. Data as of June 8, 2020.

The Global Variome shared LOVD server has integrated the data from The Genome Aggregation Database (gnomAD), which is the aggregation of the high-quality exome (protein-coding region) DNA sequence data for about 150,000 individuals[35]. It was decided to indicate for each variant its frequency in gnomAD rather than adding it as a new record, in order not to swamp the locus-specific databases (LSDBs) with data not associated to a phenotype. This information is mainly used to assist the curator in estimating the relevance of the classification of variants. In total, only two of the unique variants in our dataset (less than 2%) have assigned a *"likely pathogenic"* status with a frequency slightly below 1% in gnomAD. Interestingly, one of these, the previously mentioned c.220C>G variant, has a frequency from 0.2 to 0.4% (depending on versions of gnomAD) which does not allow its definitive classification as rare or frequent; it is registered in dbSNP (Build 153; dbSNP# rs141772938)[36] referring to ClinVar (record last updated May 9, 2020; ClinVar# VCV000189310.8)[37] for an unclear interpretation (*"Conflicting interpretations of pathogenicity"*). This last-mentioned heterozygous variant has been reported 19 times by at least five independent sources in our dataset, of which 14 are compound heterozygous disease cases. In addition, variant c.220C>G has so far not been reported as a third variant in a case with two other pathogenic variants. These observations provide strong evidence that has allowed us to classify this missense variant as "likely pathogenic (recessive)", strengthening the importance of the LSDB approach and data sharing to support variant classification and DNA diagnostics.

Clinical relevance. The dataset includes 123 patient records (71 males, 49 females, and three records of unspecified gender). Among these, 99 patients had isolated optic neuropathy[4–6,18], 8 had infantile cerebellar-retinal degeneration (ICRD; MIM# 614559)[3], 8 had various forms of encephalopathy (including epileptic encephalopathy

Fig. 3 Visualisation of the Phenotypic abnormality subontology (HPO# HP:0000118) annotation in the *ACO2* dataset. Describing the 113 symptomatic patients' reports with an extended set of full clinical description. Rows are clustered using *hclust* by separating the terms descending from Abnormal eye physiology (HPO# HP:0012373) and Abnormal eye morphology (HPO# HP:0012372); human readable shortened ontological term names were used (where possible). In columns, the identifiers of the patients (8 digits) are prefixed by code of the disease reported (3 letters): ICD: degeneration, cerebellar-retinal, infantile; RDA; dystrophy, retinal, with or without extraocular anomalies; ENM: encephalomyopathy, mitochondrial; ENC: encephalopathy; ENE: encephalopathy, epileptic; ENS: encephalopathy, neonatal, severe; NDG: neurodegeneration; OPN: neuropathy, optic; SPG: paraplegia, spastic; SZR: seizures. A red box indicates the presence of the phenotype. Data as of June 8, 2020.

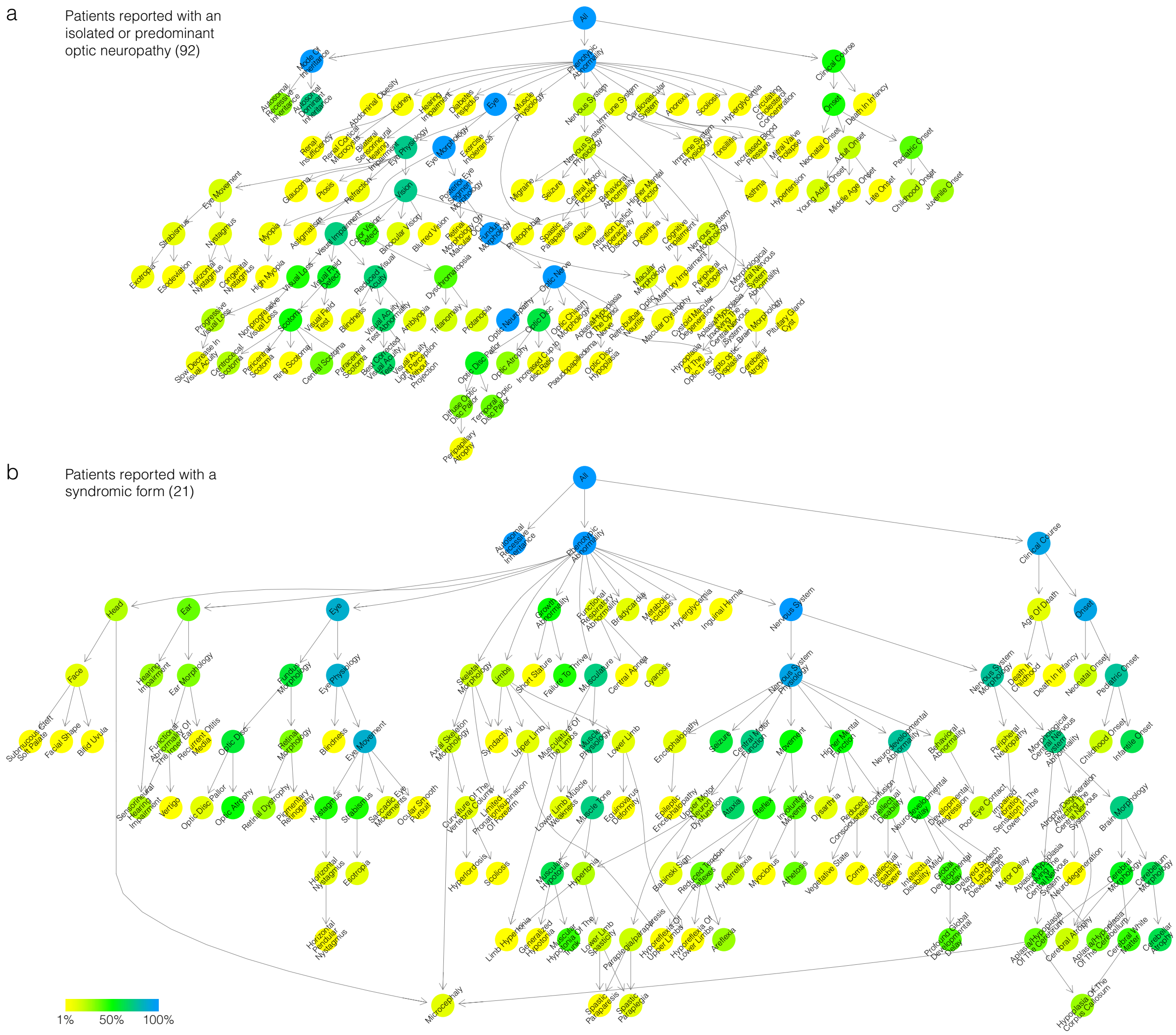


Fig. 4 Visualisation of ontological annotation in the *ACO2* dataset by disease. Subgraphs of the mode of inheritance (HPO# HP:0000005), the phenotypic abnormalities (HPO# HP:0000118) and the natural history of the disease (Clinical Course, HPO# HP:0031797), descending from the root of all terms (All; HPO# HP:0000001) in the Human Phenotype Ontology: (**a**) for the 92 patients reported with an isolated or predominant optic neuropathy; (**b**) for the 21 remaining patients reported with a syndromic form. Terms which are annotated to exactly the same objects as well as all of their children have been removed, showing only informative terms. Arrows indicate relations between terms in the ontology. Colors correspond to the frequency of the phenotypes, from the least frequent in yellow to the most frequent in blue, the green color corresponding to a term present in half of the patients. Human readable shortened ontological term names were used (where possible). Data as of June 8, 2020.

and neonatal severe encephalopathy)[4,19,20], 2 had spastic paraplegia[7], together with single cases of autosomal recessive spinocerebellar ataxia, neurodegeneration[21], retinal dystrophy (RDEOA; MIM# 617175)[22], seizures[23] and unclassified diseases[24]. Of these patients' reports, 113 have an extended set of full clinical description (the remaining ten are either asymptomatic or not described), 90 relating to our Molecular Genetics Laboratory, along with data from 23 retrieved from publications. For the description of all of these phenotypes, use is made exclusively of a standard vocabulary for referencing phenotypic abnormalities, the so-called Human Phenotype Ontology (HPO)[11], confirming the maturity of this ontology to describe eye diseases[38]. Genomic medicine requires the precise and standardized description of phenotypes[39–41]; these HPO annotations are key elements that make possible the development of algorithms for molecular diagnostics and genetic research.

A total of 154 unique HPO terms were used, each assigned from 1 to 92 patients for the most frequent term, optic neuropathy (HPO# HP:0001138); 208 unique HPO terms have been analysed by including the parent terms inferred by the ontological relationships. Figure 3 shows an exhaustive overview of the ontological annotation of the phenotypic abnormalities as a grid (mode of inheritance and natural history of the disease not shown), highlighting that patients reported with an isolated optic neuropathy, especially with a dominant mode of inheritance, have a similar limited phenotypic profile, different from the other patients whose phenotypic variability is particularly wide. We carried out the separate study of these two populations by showing the frequency of phenotypes and removing terms simply linking two terms together to focus the reading on informative phenotypes (Fig. 4): patients reported with an isolated or predominant optic neuropathy almost exclusively have a structural anomaly of the globe of the eye (Abnormal eye morphology; HPO# HP:0012372), in a predominantly dominant but also recessive mode of transmission, with an onset throughout life (Fig. 4a); the other patients reported with a syndromic form have rather a functional anomaly of the eye (Abnormal eye physiology; HPO# HP:0012373), in an exclusively recessive mode of transmission, with a beginning in the first years of life (Fig. 4b).

Overall, the extensive annotation using the phenotype ontology shows that the "isolated" versus "syndromic" separation is actually very relative: it is more likely to be a clinical spectrum with a continuum of symptoms. Figures 3 and 4 show that several cases reported as isolated are, in fact, affected by other symptoms; for the non-recurring symptoms, it is difficult to decide whether they are due to *ACO2* variants or whether they are just associated comorbidities. These figures also reveal that certain syndromic patients do not have optic neuropathy, which is therefore not compulsory in the *ACO2* phenotype as is the case for the *OPA1* gene, with only one exception published[42].

Usage Notes

The databases recording pathogenic variations, i.e. the so-called LSDBs, have proven to be the most complete because they rely on a curator who is a referent specialist for the gene or disease considered[43]. However, they are often based on isolated initiatives which use various interfaces, to the detriment of intuitiveness, and which are hosted on local servers, preventing their interoperability, unlike other types of databases which are central, i.e. encompassing all the genes of an organism, as in sequence databases[44,45] or in databases oriented towards non-pathogenic variations[36,46]. With the aim to overcome this, the Human Variome Project currently favors the centralisation of LSDBs[8,47].

Our current objective with the work reported here is to achieve a cluster of LOVD databases integrating the main genes responsible for optic neuropathies, whether isolated or syndromic, and with recessive or dominant transmission. Interoperability between these databases will be useful for molecular biologists analysing such panels of genes with, in particular, the possibility of detecting digenism. The *ACO2*-dataset, which is, to our knowledge, the only clinicobiological dataset dedicated to *ACO2*, interfaces molecular biology with medicine thanks to a common vocabulary, making it possible to link the phenotypic profiles of *ACO2* patients with those involving mutations in other genes or clinical presentations.

It will also allow a better understanding of the complex and overlapping relationships between isolated optic neuropathies and neurological syndromes involving optic neuropathy. Thus, this open-access dataset should prove useful for molecular biologists, researchers and clinicians.

Code availability

The source codes written in R programming language are available on the Code Ocean cloud-based computational reproducibility platform as a Code Ocean "compute capsule," together with the dataset analysed in this article[31]: snapshot of the *ACO2* dataset as of June 8, 2020, */data/LOVD_full_download_ACO2_2020-06-21_18.03.52.txt* (LOVD flat file format); the Human Phenotype Ontology data-version hp/releases/2020-06-08, */data/hp.obo.2020-06-08.txt* (OBO flat file format). Thus, readers can reproduce and verify the results of this article without having to download or install anything. All the content is available under MIT License.

Received: 21 September 2020; Accepted: 22 June 2021;
Published online: 05 August 2021

References

1. Martins, C. Über den Abbau der Citronensäure. *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **247**, 104–110, https://doi.org/10.1515/bchm2.1937.247.3.104 (1937).
2. Mirel, D. B. *et al.* Characterization of the human mitochondrial aconitase gene (ACO2). *Gene* **213**, 205–218, https://doi.org/10.1016/s0378-1119(98)00188-7 (1998).
3. Spiegel, R. *et al.* Infantile cerebellar-retinal degeneration associated with a mutation in mitochondrial aconitase, ACO2. *Am J Hum Genet* **90**, 518–523, https://doi.org/10.1016/j.ajhg.2012.01.009 (2012).
4. Metodiev, M. D. *et al.* Mutations in the tricarboxylic acid cycle enzyme, aconitase 2, cause either isolated or syndromic optic neuropathy with encephalopathy and cerebellar atrophy. *J Med Genet* **51**, 834–838, https://doi.org/10.1136/jmedgenet-2014-102532 (2014).
5. Charif, M. *et al.* Dominant ACO2 mutations are a frequent cause of isolated optic atrophy. *Brain Communications* **3**, fcab063, https://doi.org/10.1093/braincomms/fcab063 (2021).
6. Gibson, S. *et al.* Recessive ACO2 variants as a cause of isolated ophthalmologic phenotypes. *Am J Med Genet A*, https://doi.org/10.1002/ajmg.a.61634 (2020).
7. Bouwkamp, C. G. *et al.* ACO2 homozygous missense mutation associated with complicated hereditary spastic paraplegia. *Neurol Genet* **4**, e223, https://doi.org/10.1212/NXG.0000000000000223 (2018).
8. Cotton, R. G. *et al.* GENETICS. The Human Variome Project. *Science* **322**, 861–862, https://doi.org/10.1126/science.1167363 (2008).
9. Fokkema, I. F. *et al.* LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* **32**, 557–563, https://doi.org/10.1002/humu.21438 (2011).
10. Le Roux, B. *et al.* OPA1: 516 unique variants and 831 patients registered in an updated centralized Variome database. *Orphanet journal of rare diseases* **14**, 214, https://doi.org/10.1186/s13023-019-1187-1 (2019).
11. Kohler, S. *et al.* Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res* **47**, D1018–D1027, https://doi.org/10.1093/nar/gky1105 (2019).
12. Gray, K. A. *et al.* Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res* **41**, D545–552, https://doi.org/10.1093/nar/gks1066 (2013).
13. Hamosh, A., Scott, A. F., Amberger, J., Valle, D. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat* **15**, 57–61, 10.1002/(SICI)1098-1004(200001)15:1<57::AID-HUMU12>3.0.CO;2-G (2000).

14. O'Leary, N. A. *et al*. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–745, https://doi.org/10.1093/nar/gkv1189 (2016).
15. den Dunnen, J. T. *et al*. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat* **37**, 564–569, https://doi.org/10.1002/humu.22981 (2016).
16. Mitchell, A. L. *et al*. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* **47**, D351–D360, https://doi.org/10.1093/nar/gky1100 (2019).
17. El-Gebali, S. *et al*. The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427–D432, https://doi.org/10.1093/nar/gky995 (2019).
18. Kelman, J. C. *et al*. A sibling study of isolated optic neuropathy associated with novel variants in the ACO2 gene. *Ophthalmic Genet* **39**, 648–651, https://doi.org/10.1080/13816810.2018.1509353 (2018).
19. Sadat, R. *et al*. Functional cellular analyses reveal energy metabolism defect and mitochondrial DNA depletion in a case of mitochondrial aconitase deficiency. *Mol Genet Metab* **118**, 28–34, https://doi.org/10.1016/j.ymgme.2016.03.004 (2016).
20. Abela, L. *et al*. Plasma metabolomics reveals a diagnostic metabolic fingerprint for mitochondrial aconitase (ACO2) deficiency. *PLoS One* **12**, e0176363, https://doi.org/10.1371/journal.pone.0176363 (2017).
21. Fukada, M. *et al*. Identification of novel compound heterozygous mutations in ACO2 in a patient with progressive cerebral and cerebellar atrophy. *Mol Genet Genomic Med* **7**, e00698, https://doi.org/10.1002/mgg3.698 (2019).
22. Srivastava, S. *et al*. Increased Survival and Partly Preserved Cognition in a Patient With ACO2-Related Disease Secondary to a Novel Variant. *J Child Neurol* **32**, 840–845, https://doi.org/10.1177/0883073817711527 (2017).
23. Helbig, K. L. *et al*. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. *Genet Med* **18**, 898–905, https://doi.org/10.1038/gim.2015.186 (2016).
24. Narang, A. *et al*. Frequency spectrum of rare and clinically relevant markers in multiethnic Indian populations (ClinIndb): A resource for genomic medicine in India. *Hum Mutat* **41**, 1833–1847, https://doi.org/10.1002/humu.24102 (2020).
25. Sayers, E. W. *et al*. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **38**, D5–16, https://doi.org/10.1093/nar/gkp967 (2010).
26. Fokkema, I. *et al*. Dutch genome diagnostic laboratories accelerated and improved variant interpretation and increased accuracy by sharing data. *Hum Mutat*, https://doi.org/10.1002/humu.23896 (2019).
27. Wildeman, M., van Ophuizen, E., den Dunnen, J. T. & Taschner, P. E. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat* **29**, 6–13, https://doi.org/10.1002/humu.20654 (2008).
28. Vihinen, M., den Dunnen, J. T., Dalgleish, R. & Cotton, R. G. Guidelines for establishing locus specific databases. *Hum Mutat* **33**, 298–305, https://doi.org/10.1002/humu.21646 (2012).
29. Greene, D., Richardson, S. & Turro, E. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics* **33**, 1104–1106, https://doi.org/10.1093/bioinformatics/btw763 (2017).
30. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2020).
31. ACO2 locus-specific database with extensive phenotype ontology annotation to unravel clinical spectrum. *Code Ocean* https://doi.org/10.24433/CO.5810290.V1 (2020).
32. Ferré, M. ACO2 clinicobiological dataset with extensive phenotype ontology annotation. *figshare* https://doi.org/10.6084/m9.figshare.14915682.v1 (2021).
33. Global Alliance for Genomics and Health. GENOMICS. A federated ecosystem for sharing genomic, clinical data. *Science* **352**, 1278–1280, https://doi.org/10.1126/science.aaf6162 (2016).
34. Calvo, S. E. *et al*. Comparative Analysis of Mitochondrial N-Termini from Mouse, Human, and Yeast. *Molecular & cellular proteomics: MCP* **16**, 512–523, https://doi.org/10.1074/mcp.M116.063818 (2017).
35. Karczewski, K.J., Francioli, L.C., Tiao, G. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 https://doi.org/10.1038/s41586-020-2308-7 (2020).
36. Sherry, S. T. *et al*. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308–311 (2001).
37. Landrum, M. J. *et al*. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062–D1067, https://doi.org/10.1093/nar/gkx1153 (2018).
38. Kohler, S. *et al*. The Human Phenotype Ontology in 2017. *Nucleic Acids Res* **45**, D865–D876, https://doi.org/10.1093/nar/gkw1039 (2017).
39. Deans, A. R. *et al*. Finding our way through phenotypes. *PLoS Biol* **13**, e1002033, https://doi.org/10.1371/journal.pbio.1002033 (2015).
40. Robinson, P. N. Deep phenotyping for precision medicine. *Hum Mutat* **33**, 777–780, https://doi.org/10.1002/humu.22080 (2012).
41. Biesecker, L. G. Phenotype matters. *Nat Genet* **36**, 323–324, https://doi.org/10.1038/ng0404-323 (2004).
42. Nasca, A. *et al*. Not only dominant, not only optic atrophy: expanding the clinical spectrum associated with OPA1 mutations. *Orphanet journal of rare diseases* **12**, 89, https://doi.org/10.1186/s13023-017-0641-1 (2017).
43. Brookes, A. J. & Robinson, P. N. Human genotype-phenotype databases: aims, challenges and opportunities. *Nat Rev Genet* **16**, 702–715, https://doi.org/10.1038/nrg3932 (2015).
44. The UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158–D169, https://doi.org/10.1093/nar/gkw1099 (2017).
45. Benson, D. A. *et al*. GenBank. *Nucleic Acids Res* **41**, D36–42, https://doi.org/10.1093/nar/gks1195 (2013).
46. Lek, M. *et al*. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291, https://doi.org/10.1038/nature19057 (2016).
47. Smith, T. D., Vihinen, M. & Human Variome, P. Standard development at the Human Variome Project. *Database (Oxford)* **2015**, https://doi.org/10.1093/database/bav024 (2015).

Acknowledgements
We thankfully acknowledge grants from the following foundations and patients' associations: ACO2 GENE, Association contre les Maladies mitochondriales, Fondation Visio, Kjer France, Ouvrir les Yeux, Retina France, and Union nationale des Aveugles et Déficients visuels.

Author contributions
M.F. designed supervised the project. J.T.D. and P.R. participated in the design and supervision. K.G. collected the data with the help from P.A.B., M.C., E.C., C.B., V.D., V.P., D.B., D.M., P.G. and G.L. T.F. participated in the data acquisition; M.F. performed the statistical analysis. K.G., P.R. and M.F. wrote the manuscript with inputs from all authors. All authors have read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.F.

Reprints and permissions information is available at www.nature.com/reprints.

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