Genetics in Medicine | ARTICLE

Biallelic variants in *HPDL*, encoding 4-hydroxyphenylpyruvate dioxygenase-like protein, lead to an infantile neurodegenerative condition

Shereen G. Ghosh, MSc[1,2], Sangmoon Lee, MD, PhD[1,2], Rudy Fabunan, MS[3], Guoliang Chai, PhD[1,2], Maha S. Zaki, MD, PhD[4], Ghada Abdel-Salam, PhD[4], Tipu Sultan, MSc, FRCPCH[5], Tawfeg Ben-Omran, MD, FRCPS[6], Javeria Raza Alvi, MBBS, FCPS[5], Jennifer McEvoy-Venneri, BS[1,2], Valentina Stanley, BS[1,2], Aakash Patel, BS[1,2], Danica Ross, BS[1,2], Jeffrey Ding, BS[7], Mohit Jain, MD, PhD[7], Daqiang Pan, PhD[8], Philipp Lübbert, PhD[9], Bernd Kammerer, PhD[8], Nils Wiedemann, PhD[9,10], Nanda M. Verhoeven-Duif, PhD[11], Judith J. Jans, PhD[11], David Murphy, MSc[12], Mehran Beiraghi Toosi, MD[13], Farah Ashrafzadeh, MD[14], Shima Imannezhad, MD[15], Ehsan Ghayoor Karimiani, MD, PhD[16,17], Khalid Ibrahim, MD, FRCPCH[18], Elizabeth R. Waters, PhD[3], Reza Maroofian, PhD[19,20] and Joseph G. Gleeson, MD[1,2]

Purpose: Dioxygenases are oxidoreductase enzymes with roles in metabolic pathways necessary for aerobic life. 4-hydroxyphenylpyruvate dioxygenase-like protein (HPDL), encoded by *HPDL*, is an orphan paralogue of 4-hydroxyphenylpyruvate dioxygenase (HPD), an iron-dependent dioxygenase involved in tyrosine catabolism. The function and association of HPDL with human diseases remain unknown.

Methods: We applied exome sequencing in a cohort of over 10,000 individuals with neurodevelopmental diseases. Effects of HPDL loss were investigated in vitro and in vivo, and through mass spectrometry analysis. Evolutionary analysis was performed to investigate the potential functional separation of HPDL from HPD.

Results: We identified biallelic variants in *HPDL* in eight families displaying recessive inheritance. Knockout mice closely phenocopied humans and showed evidence of apoptosis in multiple cellular lineages within the cerebral cortex. *HPDL* is a single-exonic gene that likely arose from a retrotransposition event at the base of the tetrapod lineage, and unlike HPD, HPDL is mitochondria-localized. Metabolic profiling of *HPDL* mutant cells and mice showed no evidence of altered tyrosine metabolites, but rather notable accumulations in other metabolic pathways.

Conclusion: The mitochondrial localization, along with its disrupted metabolic profile, suggests *HPDL* loss in humans links to a unique neurometabolic mitochondrial infantile neurodegenerative condition.

Genetics in Medicine (2021) 23:524–533; https://doi.org/10.1038/s41436-020-01010-y

Keywords: HPDL; HPD; 4-hydroxyphenylpyruvate dioxygenase-like protein; oxidoreductase; neurodegenerative disease

INTRODUCTION

Disorders involving neurometabolism can lead to both structural and functional disturbances of the nervous system through multiple mechanisms that include abnormal accumulation of toxic substrates, depletion of key energy or metabolic intermediates, or cell death.[1] Although pediatric-onset brain diseases are often associated with genetic abnormalities, the link between metabolic impairments and

[1]Department of Neurosciences, University of California–San Diego, La Jolla, CA, USA; [2]Rady Children's Institute for Genomic Medicine, San Diego, CA, USA; [3]Biology Department, San Diego State University, San Diego, CA, USA; [4]Clinical Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; [5]Department of Pediatric Neurology, Institute of Child Health, Children Hospital Lahore, Lahore, Pakistan; [6]Division of Genetic and Genomic Medicine, Sidra Medicine and Hamad Medical Corporation, Doha, Qatar; [7]Departments of Medicine and Pharmacology, University of California–San Diego, La Jolla, CA, USA; [8]Centre for Integrative Signalling Analysis (CISA), University of Freiburg, Freiburg, Germany; [9]Institute of Biochemistry and Molecular Biology, ZBMZ, Faculty of Medicine, University of Freiburg, Freiburg, Germany; [10]CIBSS Centre for Integrative Biological Signalling Studies, University of Freiburg, Freiburg, Germany; [11]Section Metabolic Diagnostics, Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; [12]Department of Clinical and Movement Neurosciences, UCL Institute of Neurology, Queen Square, London, UK; [13]Pediatric Neurology Department, Ghaem Hospital, Mashhad University of Medical Sciences, Mashhad, Iran; [14]Department of Pediatrics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; [15]Department of Pediatric Neurology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; [16]Molecular and Clinical Sciences Institute, St. George's, University of London, Cranmer Terrace, London, UK; [17]Innovative Medical Research Center, Mashhad Branch, Islamic Azad University, Mashhad, Iran; [18]Division of Pediatric Neurology, Sidra Medicine, Doha, Qatar; [19]Genetics Research Centre, Molecular and Clinical Sciences Institute, St. George's University, London, UK; [20]Department of Neuromuscular Disorders, UCL Institute of Neurology, Queen Square, London, UK. Correspondence: Joseph G. Gleeson (jogleeson@health.ucsd.edu)
These authors contributed equally: Shereen G. Ghosh, Sangmoon Lee.

Submitted 24 July 2020; revised 6 October 2020; accepted: 6 October 2020
Published online: 14 November 2020

brain disease has been associated with relatively common disorders of the nervous system.[2] Especially in highly consanguineous populations, congenital malformations of the nervous system and neurometabolic disorders are often linked.[3] Genes involved in inborn errors of metabolism often encode enzymes that catalyze specific biochemical reactions or required cofactors.[4] Absent or abnormal functioning of such proteins leads to an accumulation of upstream substrates or deficiency of downstream products, with detrimental disease-causing effects.

Dioxygenases are oxidoreductase enzymes that utilize dioxygen (i.e., O_2), most often complexed with iron, in oxidation reactions involved in metabolic pathways or as oxygen sensors.[5] To date, 15 dioxygenases have been linked to diseases that are mostly autosomal recessive including encephalopathy, alkaptonuria (MIM 203500), and intellectual disability.[6–8] Dioxygenases can oxidize either small molecules, as in tryptophan dioxygenase,[9] or residues on macromolecules, as in EGLN1, a HIF proline dioxygenase acting as an oxygen sensor.[10]

4-hydroxyphenylpyruvate dioxygenase (HPD, MIM 609695) is a nonheme Fe(II)-dependent dioxygenase that catalyzes the second of four steps in L-tyrosine catabolism in nearly all aerobic cells: the conversion of 4-hydroxyphenylpyruvate into homogentisate.[11] Biallelic loss associates with recessive tyrosinemia type III (MIM 276710), the least frequent form of three genetic tyrosinemias.[12] Clinical features are milder than those of other genetic forms, and can include seizures, intellectual disability, and intermittent ataxia.[13,14] Serum tyrosine level is variably elevated by 5- to 20-fold (148–1769 μmol/L, reference 29–86 μmol/L).[14–17]

In addition to biallelic variants, a particular gain-of-function *HPD* variant (p.Ala33Thr) associates with the dominant disease hawkinsinuria (MIM 140350),[15] as mutant HPD can convert the normal substrate into a reactive epoxide intermediate but cannot rearrange the intermediate to homogentisate. As a result, the reactive epoxide is detoxified to either hawkinsin or hydroxycyclohexyl acetate by glutathione transferase and epoxide hydrolase, respectively. Hawkinsinuria manifests clinically with metabolic acidosis in early childhood, but most patients have normal long-term outcome.

HPDL encodes 4-hydroxyphenylpyruvate dioxygenase-like protein (HPDL), the only mammalian paralogue of HPD, and shares 44% sequence similarity. The origin and function of *HPDL* is not known, and the gene has not been linked to human disease previously. Here we describe that *HPDL* biallelic loss leads to a unique neurometabolic, infantile neurodegenerative condition.

MATERIALS AND METHODS

Patient recruitment

The institutional review board (IRB) at the University of California–San Diego (UCSD) approved this study. All study participants signed informed consent documents, and the study was performed in accordance with the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rules. The procedures followed for recruitment and data collection were in accordance with the ethical standards of the responsible committee on human experimentation at the respective participating institute, and proper informed consent was obtained.

DNA sequencing

Samples of the probands were subjected to exome sequencing. Sanger sequencing of polymerase chain reaction (PCR) products was performed for validation and segregation. Further details are in the Supplementary data.

Cell culture

HEK293T cells (ATCC) were transfected with expression vectors using Lipofectamine 2000 according to manufacturer's protocol (100 ng of DNA and 0.5 μl of Lipofectamine 2000 per well) and immunostained 48 hours after transfection. Methods used for cloning mammalian expression vectors, transfection, and immunostaining are described in Supplementary Methods.

Methods used to generate HPDL knockout (KO) CACO-2 cell line are described in Supplementary Methods. Immunostaining and imaging of wild-type and KO CACO-2 cells details are described in Supplementary Methods.

Animal experiments

Animal use followed National Institutes of Health (NIH) guidelines and was approved by the Institutional Animal Care and Use Committee (IACUC) at UCSD. The frameshift variant in *Hpdl* (NM_146256.3:c.59_65del) was introduced to C57BL/6 mouse using CRISPR/Cas9 (single-guide RNA [sgRNA] sequence: CUU CCA GCC CCU GGC GGU GA) according to standard protocols at the UCSD Transgenic Mouse Core. Mice were genotyped by PCR followed by Sanger sequencing (Hpdl_KO_F/R primers; Table S1), and those with correct genotypes were backcrossed for at least five generations prior to experimental crosses of heterozygous animals. Timed pregnant animals were obtained by plug checks, where the day of the observed vaginal plug was determined to be E0.5. All mice were group-housed under a 12-hour light/dark cycle with access to chow and water. Breeding was performed until 30 homozygous pups were born to analyze survival, growth, histology, and immunofluorescence staining. The investigators were not blind to the analysis, since phenotypes of KO animals were very apparent. Methods used for histological analysis, immunofluorescence staining, and imaging are described in Supplementary Methods.

Direct infusion–based metabolomics

Collection of cell lysates and mouse brain for metabolomics experiments are described in Supplementary Methods.

A nonquantitative direct infusion high-resolution mass spectrometry metabolomics method was used as previously

described with minor adaptations for cell and tissue extracts.[18] In brief, 70 µl of cell lysate in methanol was diluted with 60 µl 0.3% formic acid (Emsure, Darmstadt, Germany) and filtered using a methanol preconditioned 96-well filter plate (Acro prep, 0.2 um GHP, NTRL, 1-ml well; Pall Corporation, Ann Arbor, MI, USA) and a vacuum manifold. The sample filtrate was collected in a 96-well plate (Advion, Ithaca, NY, USA).

Samples were analyzed using a TriVersa NanoMate system (Advion, Ithaca, NY, USA) controlled by Chipsoft software (version 8.3.3, Advion). Data acquisition was performed using Xcalibur software (version 3.0, Thermo Scientific, Waltham, MA, USA). A peak calling pipeline, developed in R programming language, annotated the raw mass spectrometry data according to the Human Metabolome Database (HMDB).

Phylogenetic tree reconstruction

Amino acid sequences were aligned using PROMALs3d.[19] Three alignments were computed: (1) all sequences, (2) HPD sequences, and (3) HPDL sequences. We used a maximum likelihood method using RAxML[20] on XSEDE via CIPRES to obtain the gene tree topology using the JTT model and 1000 bootstrap replicates. The gene tree was visualized using iTol,[21] and the final rooted topology was obtained by rerooting the tree with a midpoint root.

Estimates of evolutionary rates

Branch lengths were obtained directly from the phylogenetic trees. To calculate d*N*/d*S* ratios over the tree and subtrees (HPD and HPDL), we used CODEML in the PAML package (version 4.8).[22] To test multiple rates on the tree on CODEML, we used the model = 2 parameter to estimate an HPD, HPDL, HPD tetrapod, and an HPDL tetrapod rate. Pairwise d*N* estimates were used using the runmode = −2 parameter. To perform the CODEML analysis, the CDS alignment was performed using TranslatorX.[23] The dist. alignment function was called in seqinr to compute the pairwise distance matrix using the amino acid identity and similarity.[24]

RESULTS

Identification of an infantile-onset neurodegenerative condition in eight families with biallelic variants in *HPDL*

We identified eight independent consanguineous families carrying *HPDL* biallelic variants leading to a range of neurological phenotypes, which included spastic tetraplegia, microcephaly, brain atrophy, epilepsy, and severe intellectual and motor disability (Fig. **1a** and Table S2). After obtaining informed consent from all participating individuals in accordance with the ethical standards set by the UCSD IRB, we identified a total of eight distinct variants in *HPDL* in the eight families by exome sequencing. We recruited family 1 with three affected siblings showing limb spasticity and mild motor and intellectual disability. Although parental consanguinity suggested the presence of a homozygous variant as the cause, no homozygous variants in any gene passed filter criteria. Instead, we identified compound heterozygous p.Gly50Asp; p.Tyr118* variants in *HPDL*, which was the only candidate identified by exome sequencing and segregated in the family. Family 2 was recruited with documented parental consanguinity with one affected girl showing a nearly identical clinical pattern as family 1. We identified a homozygous p.Gly50Asp variant in the affected individual as the most likely cause of disease, one of the same variants identified in family 1 (Table S3).

Through collaboration with colleagues, we identified four additional families (3, 4, 5, and 8) with homozygous missense variants and two additional families (6 and 7) with homozygous truncating variants in *HPDL* independently identified as likely most pathogenic based upon inheritance pattern, computational prediction, population allele frequency, and segregation (Tables S1 and S2). These variants included homozygous p.Ala78Thr (family 3), p.Gly126Ser (family 4), p.Leu164Pro (family 5), p.Gly319Argfs*15 (family 6), p.Gln32* (family 7), and p.Gly301Val (family 8). Thus, we identified a total of eight families comprising 17 individuals with biallelic *HPDL* damaging variants with variable phenotypes. Because HPDL is encoded by a single exon, and thus not subject to alternative splicing, the presence of homozygous early truncating variants suggests *HPDL* loss of function as the mechanism of this infantile-onset neurodegenerative disease.

Affected children were born full-term without complications during pregnancy or delivery (Table S2). Where data were available, birth weight and head circumference were not remarkable. However, most affected individuals showed a trend toward smaller head circumference by 5 years of age; 7 of 14 affected individuals, where data were available, met criteria for microcephaly, defined as a head circumference <3 SD below mean (Table S2). Six of 14 affected individuals also displayed nonspecific facial dysmorphisms. Most individuals presented with minimal to absent psychomotor development, including gross motor skills, language, and social skills. Most displayed spontaneous epileptic seizures starting by 1 year. Seizures were typically myoclonic, focal, or tonic and occurred daily to weekly. Neurological findings included hypertonia, hyperreflexia, spastic gait, and positive Babinski sign—features that became more pronounced over time, suggesting a progressive course. Many individuals were unable to walk and displayed absent language skills. In six severely affected individuals, brain magnetic resonance images (MRIs) were available for review, demonstrating cortical atrophy, white matter hyperintensity, corpus callosum thinning, reduced size of the cerebellum, ventriculomegaly, and increased extra-axial fluid (Fig. **1b**), suggesting loss of brain parenchyma. Brain atrophy and the progressive course were consistent with neurodegeneration.

HPDL encodes 371 amino acids containing a mitochondrial targeting sequence (MTS), two vicinal oxygen chelate (VOC) domains, and three iron-binding sites. The locations of all the variants occurred within the two VOC domains (Fig. **1c**), thus predicted to impact enzymatic function. Of note,

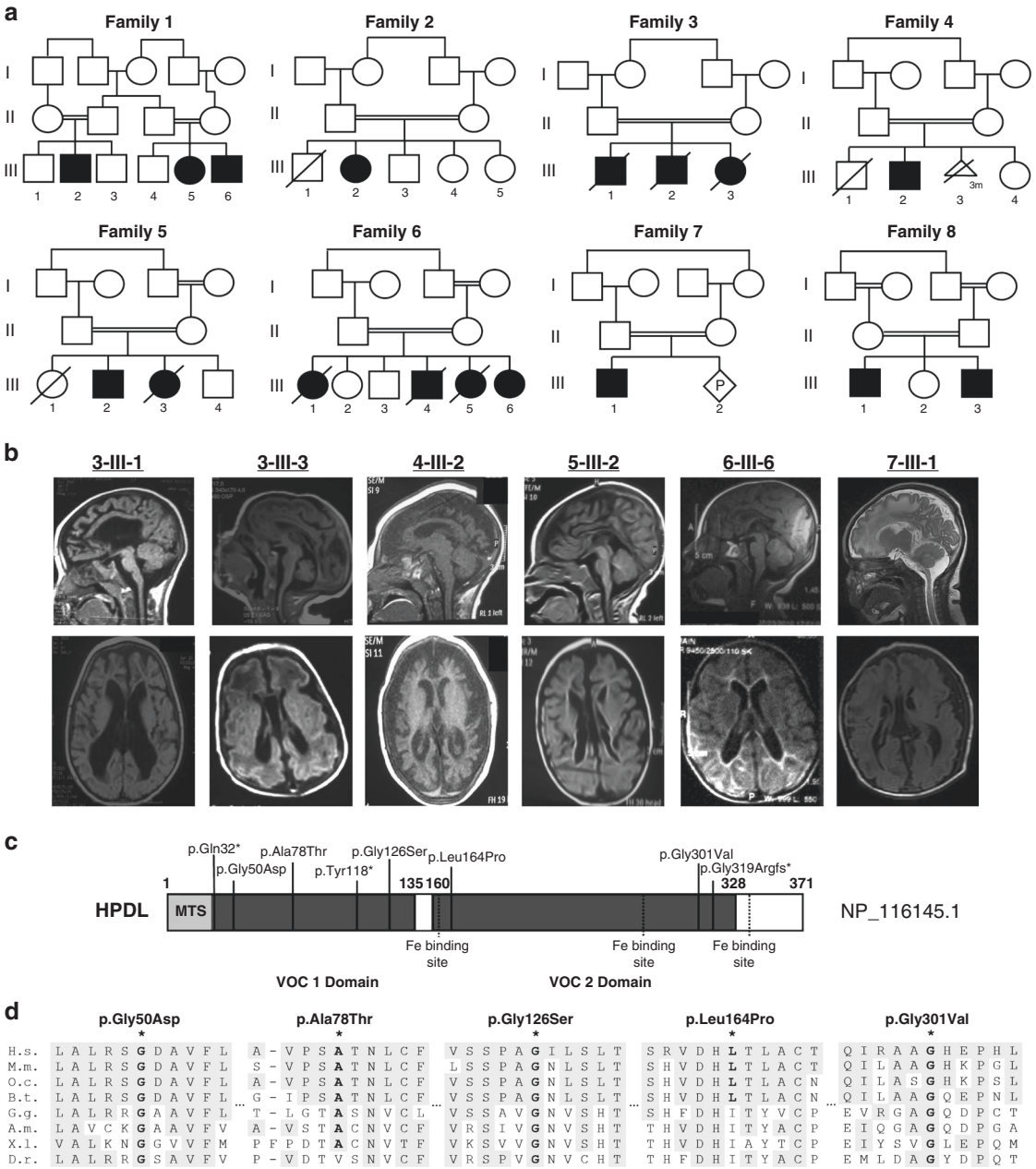

Fig. 1 Variants in *HPDL* in eight independent consanguineous families lead to microcephaly and brain atrophy. (**a**) Pedigrees of families 1–8 showing consanguineous marriages (double line) with a total of 17 affected children. All unfilled members are without neurological disease, epilepsy, or neurodegeneration. Dashed line: deceased. (**b**) Panels show magnetic resonance imaging (MRI) scans for six affected individuals from five different families. Shown are sagittal and axial images demonstrating enlarged ventricles, thin corpus callosum, and severe cortical and cerebellar atrophy. (**c**) Schematic of the HPDL protein structure depicting a predicted mitochondrial targeting sequence (MTS), two vicinal oxygen chelate (VOC) domains from amino acids 7–135 and 160–328 (shown in dark gray) and three iron (Fe) binding sites (dashed lines). Solid black lines indicate the locations of the variants. (**d**) Amino acid alignments for all identified missense variants across different vertebrate species. Amino acids highlighted in gray indicate conserved residues. A.m. *Alligator mississippiensis*, B.t. *Bos taurus*, D.r. *Danio rerio*, G.g. *Gallus gallus*, H.s. *Homo sapiens*, M.m. *Mus musculus,* O.c. *Oryctolagus cuniculus*, X.l. *Xenopus laevis*.

the phenotypes between homozygous truncating variants (p.Gln32* and p.Gly319Argfs*15), destroying at least one iron-binding site, and three of the missense variants (p.Ala78Thr, p.Gly126Ser, and p.Leu164Pro), were indistinguishable, suggesting that the missense variants cause loss of function of HPDL. The p.Gly50Asp variant was found only in families 1 and 2, and was associated with milder phenotypes compared with the more severely affected individuals in other families with brain atrophy and epilepsy, suggesting that this variant has a milder impact on HPDL function. The variants were unique in our data set of >5000 exomes from individuals with neurodevelopmental phenotypes, were predicted to be disease-causing, and were not represented (p.Ala78Thr, p.Leu164Pro, p.Gly301Val, and p.Gly319Argfs*) or very rare

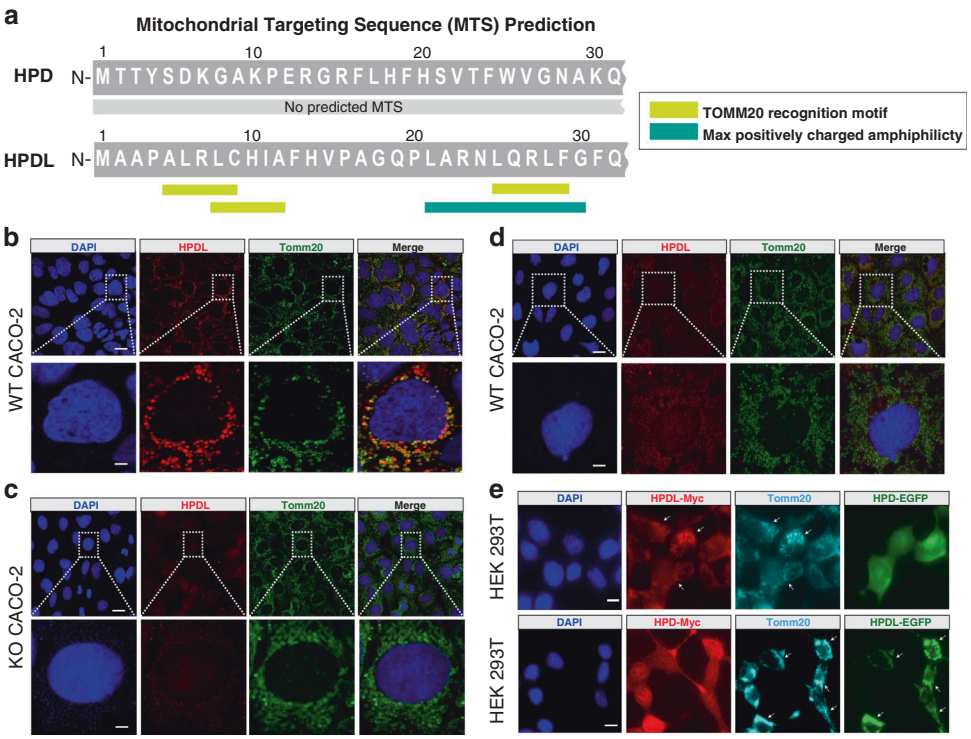

Fig. 2 HPD and HPDL display different subcellular localization patterns. (**a**) Schematic of HPD and HPDL N-terminal amino acid sequences highlighting the presence or absence of a predicted mitochondrial targeting sequence (MTS) as predicted by Mitofates. HPD has no predicted MTS, while HPDL has three different TOMM20 recognition motifs (yellow) and a region of max positively charged amphiphilicity (green). (**b**) Immunofluorescent staining of HPDL wild-type (WT) CACO-2 cells with DAPI (blue), HPDL (red), and TOMM20 (green). Scale bars: 10 µm (upper panels) and 2 µm (lower panels). (**c**) Immunofluorescent staining of HPDL knockout (KO) CACO-2 cells with DAPI (blue), HPDL (red), and TOMM20 (green). Scale bars: 10 µm (upper panels) and 2 µm (lower panels). (**d**) Immunofluorescent staining of HPDL WT CACO-2 cells with DAPI (blue), HPD (red), and TOMM20 (green). Scale bars: 10 µm (upper panels) and 2 µm (lower panels). (**e**) Immunostaining for TOMM20 (cyan) and DAPI (blue) of HEK293T cells transfected with both HPDL-Myc (red) and HPD-EGFP (green) or HPD-Myc (red) and HPDL-EGFP (green) to determine the subcellular localization of HPD and HPDL. White arrows: cells with mitochondrial localization of HPDL. Scale bar: 10 µm.

(p.Gln32* [allele frequency; AF = 8.46e-6], p.Gly50Asp [AF = 1.02e-5], p.Tyr118* [AF = 3.19e-5], and p.Gly126Ser [AF = 8.67e-6]), in the gnomAD database.[25] No instances of biallelic damaging variants in *HPDL* among over 100,000 sequenced individuals occur in gnomAD (Fisher's exact test $P < 0.00001$). Additionally, all missense variants are highly conserved among vertebrates (Fig. **1d**). All variants were confirmed by Sanger sequencing and segregated according to a recessive mode of inheritance, with heterozygotes showing no apparent clinical features (Fig. S1).

HPD and *HPDL* have differing tissue-wide expression and subcellular localization

To functionally distinguish between *HPDL* and the paralogous *HPD*, we sought differences in tissue expression and subcellular localization patterns. The GTEx database indicates that *HPD* is almost exclusively expressed in the liver and kidney, whereas *HPDL* is widely expressed in most organs (Fig. S2).[26] Importantly, the expression level of *HPDL* is much higher than *HPD*, specifically in the brain (Fig. S3).[26] Publicly available human and mouse brain single-cell RNA-seq data suggests that *HPDL* transcripts are encountered ~5× more commonly than *HPD* in the brain, with brain astrocytes as the major source of *HPDL* transcripts.[27–29]

Protein databases suggest HPD localizes to the cytosol while HPDL localizes to mitochondria.[30] We used Mitofates, an online MTS prediction tool, to assess for an N-terminal mitochondrial presequence or cleavable localization signal.[31] While HPD had no predicted MTS, HPDL contained three different recognition motifs for TOMM20, a mitochondrial import receptor subunit responsible for the recognition and translocation of cytosolically synthesized mitochondrial preproteins. In addition, there was a region of maximum positively charged amphiphilicity, another essential feature for mitochondrial presequence function (Fig. **2a**).[32] To confirm this experimentally, we utilized CACO-2 cells, a human colon adenocarcinoma cell line expressing levels of *HPDL* higher than most other cell lines.[33] To validate the specificity of the anti-HPDL antibody, we generated *HPDL* KO CACO-2 cells using CRISPR/Cas9, and confirmed correct biallelic targeting (Fig. S4A, B).

Immunofluorescence in WT cells confirmed the colocalization of TOMM20 and HPDL within mitochondria (Fig. **2b**), while staining in the KO cells confirmed absent HPDL signal (Fig. **2c**). HPD showed no colocalization with TOMM20, but

instead was localized diffusely to the cytoplasm (Fig. 2d). To assess localization of tagged versions of the proteins, we cotransfected HEK293T cells with C-terminally Myc-tagged HPDL and C-terminally EGFP-tagged HPD and costained with anti-TOMM20. We used C-terminally tagged vectors to ensure that any N-terminal MTS would not be disturbed. Again, TOMM20 overlapped in localization with HPDL, but not HPD. We also flipped the vectors and repeated the experiment (C-terminally Myc-tagged HPD and C-terminally EGFP-tagged HPDL) to ensure that the tag did not affect localization, which provided further confirmation (Fig. 2e). We utilized a Seahorse assay to compare the oxygen consumption rate in these cells, which illustrated slightly lower oxygen consumption rate in the HPDL KO cells compared with WT (Fig. S5). Together, these results demonstrate that HPDL, but not HPD, localizes to mitochondria, suggesting that these enzymes act in different locations on either the same or on different substrates. The slightly reduced oxygen consumption rate in the KO cells further suggests that the absence of HPDL may have an effect on mitochondrial metabolic function.

Hpdl KO mice display epilepsy, early lethality, smaller brain sizes, and cellular apoptosis in the brain

To investigate the physiological role of HPDL, we created a KO mouse model with a homozygous frameshift deletion of *Hpdl* generated using CRISPR/Cas9 (*Hpdl* KO, Fig. S6). KOs were born healthy and at Mendelian ratios (WT:heterozygote:homozygote = 30:68:26, test for Hardy–Weinberg equilibrium $P = 0.368$). Although KOs were indistinguishable at birth, by postnatal day 5 (P5), KOs exhibit lethargy and spontaneous movements suggesting of epileptic seizures (Supplemental video S1). Almost all KOs died by P10, and none survived to P15 (Fig. 3a). Although body weights and brain size were similar at birth, by P9–10 KO mice weighed approximately half that of littermates (Fig. 3b, c). Immunostaining of P9 brain tissue sections for DAPI and Cleaved Caspase 3 (CC3) demonstrated significant cellular apoptosis (CC3+ cells) in the KO brains compared with WT pups (23 vs. <1 CC3+ cell per 500-μm column), which correlated with smaller brain size (Fig. 3c–e). To determine the cellular identity of CC3+ cortical cells, we performed coimmunostaining with CTIP2 (upper layer cortical neurons), OLIG2 (oligodendrocytes), and GFAP (astrocytes). CTIP2/CC3 staining revealed the most colocalization compared with OLIG2/CC3 and GFAP/CC3 (5.9% vs. 1.9% and 3.8%, respectively, Fig. 3f, g). Taken together, these data suggest cellular apoptosis in the brain occurs in the absence of HPDL function.

Metabolomic analysis reveals disrupted metabolic signature but normal tyrosine catabolism

Deficiencies of specific enzymes of the tyrosine catabolism pathway are known to give rise to a number of severe metabolic disorders in humans. In the oxidative degradation of tyrosine by vertebrates, the first conversion step is to 4-hydroxyphenylpyruvate in a reaction catalyzed by tyrosine transaminase (Fig. S7A).[34] HPD then catalyzes the oxidation of 4-hydroxyphenylpyruvate to homogentisate.[12] Defects in these two steps in tyrosine catabolism lead to tyrosinemia type II and type III, respectively. Normally, homogentisate is further catabolized by the next enzyme in the pathway, homogentisate 1,2-dioxygenase.[8] In alkaptonuria, however, the enzyme is deficient, leading to this rare hereditary metabolic disorder. In tyrosinemia type I, fumarylacetoacetate hydrolase is mutated, causing the subsequent buildup of byproducts.[35] Because both HPD and HPDL have highly similar amino acid homology, we speculated that HPD and HPDL may have similar substrates. To test this, we performed mass spectrometry analysis on *HPDL* WT and KO CACO-2 cell extracts and conditioned cell culture media. The analysis revealed several significantly downregulated and upregulated metabolites in both the cell extracts and culture media of KO cells compared with WT; however, we did not detect an abnormal accumulation of any metabolites involved in tyrosine catabolism (Fig. S7B, C, Table S4). We then performed a similar analysis in both *Hpdl* WT and KO mouse brain extracts and serum at P5, which again showed no abnormal accumulation of tyrosine catabolites (Fig. S7D, E, Table S4). Taken together, these data suggest that HPDL likely acts on a different substrate than HPD, and perhaps on a different cellular pathway than tyrosine catabolism.

Independent *HPD* duplication events gave rise to *HPDL* in vertebrates

HPD, conserved in all eukaryotes, is encoded on human chromosome 12, with its major isoform consisting of 14 coding exons (Fig. 4a). One alternative splice isoform is described; however: the major isoform encodes a protein 393 amino acids long. *HPDL*, on the other hand, is a single-exon gene located on chromosome 1, with no alternative splice isoforms reported, encoding 371 amino acids, apparent in *Xenopus tropicalis* (frog) but not in lower species like *Danio rerio* (zebrafish) (Fig. 4b). There are no other *HPD* paralogues in mammals, suggesting that *HPDL* arose as a result of a retrotransposition event at the base of the tetrapod lineage (Fig. 4c). However, there is an *HPDL* paralogue in other lower deuterostomes, including *Ciona intestinalis* (sea vase) and zebrafish, but this *HPDL* contains introns. Thus, other models are possible by which *HPDL* arose from *HPD* during evolution.

Several lines of evidence suggest that mammalian *HPDL* arose as a separate gene duplication event from the event that yielded *HPDL* in lower deuterostomes. First, only mammalian HPDL contains a predicted MTS. Second, mammalian *HPDL* shows evidence of stronger positive selective pressure than fish *hpdl*, as evidenced by faster accumulation of nonsynonymous to synonymous base changes (Figs. 4d and S8A). Third, mammalian *HPDL* is in a region of conserved synteny in frog but not in zebrafish (Fig. S8B). Finally, and most convincingly, it is more likely that a retrotransposition event yielded an intronless gene than removal of introns from an

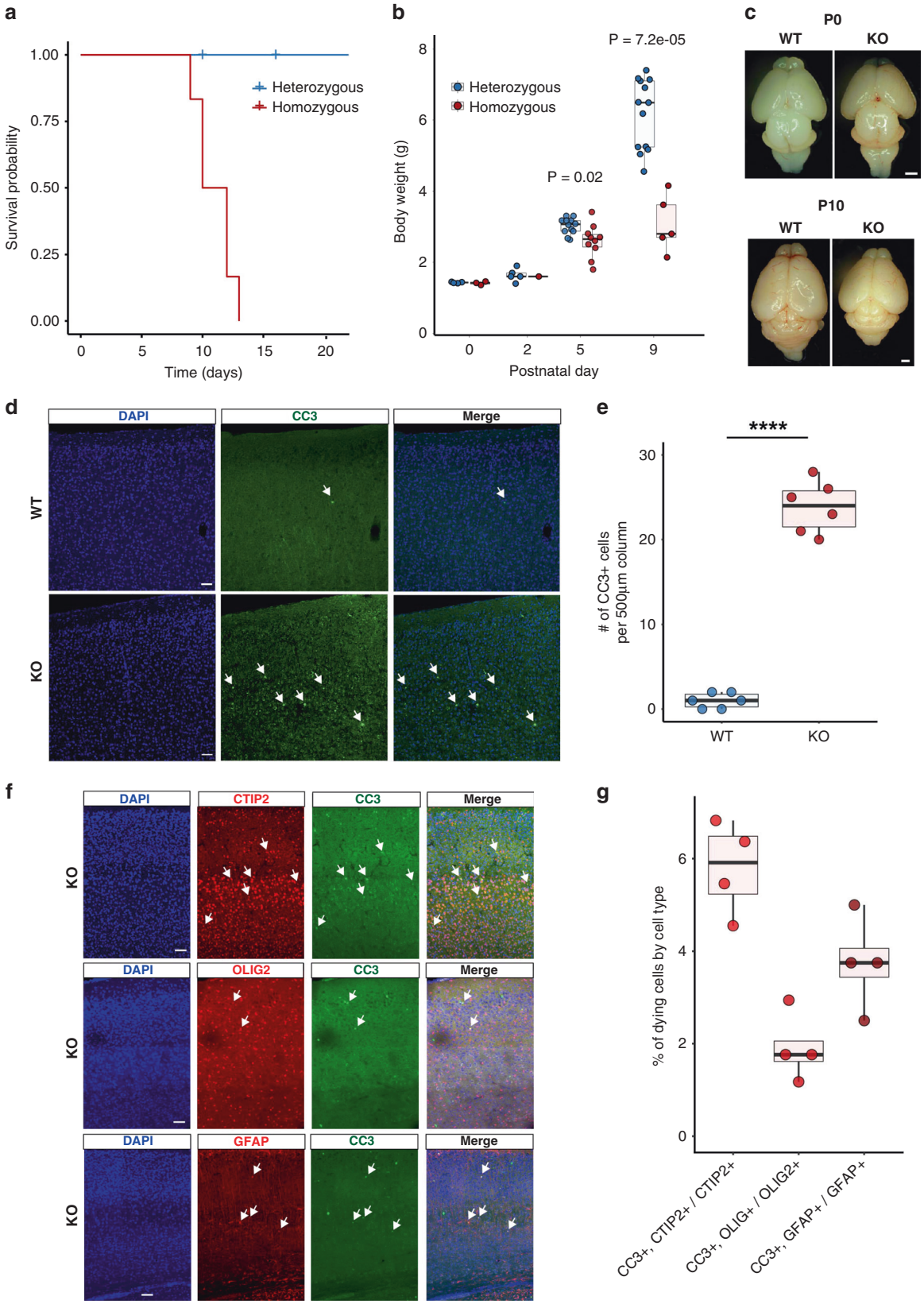
a
Survival probability
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
Time (days)
Heterozygous
Homozygous
b
Body weight (g)
8
6
4
2
0
0
2
5
9
Postnatal day
P = 7.2e-05
P = 0.02
Heterozygous
Homozygous
c
P0
WT
KO
P10
WT
KO
d
DAPI
CC3
Merge
WT
KO
e
****
# of CC3+ cells per 500µm column
30
20
10
0
WT
KO
f
DAPI
CTIP2
CC3
Merge
KO
DAPI
OLIG2
CC3
Merge
KO
DAPI
GFAP
CC3
Merge
KO
g
% of dying cells by cell type
6
4
2
0
CC3+, CTIP2+ / CTIP2+
CC3+, OLIG+ / OLIG2+
CC3+, GFAP+ / GFAP+

Fig. 3 *Hpdl* knockout (KO) mice display early lethality, smaller brain sizes, and cellular apoptosis. (**a**) Kaplan–Meier curve displaying the survival probability of *Hpdl* heterozygous mice ($n = 10$, blue) versus *Hpdl* KO mice ($n = 6$, red). (**b**) Graph depicting body weight measurements of *Hpdl* heterozygous mice versus *Hpdl* KO mice at postnatal days 0, 2, 5, and 9. *P* values were obtained by an independent *t*-test. (**c**) Whole-brain images of wild-type (WT) versus *Hpdl* KO mice at P0 and P10. Scale bar: 1 mm (**d**) Coronal sections of P9 WT and KO mouse brain were stained for DAPI (blue) and Cleaved Caspase 3 (CC3, green). The P9 KO shows strong CC3+ immunoreactivity. Note the absence of CC3 immunoreactivity in the P9 WT. Scale bar: 50 µm (**e**) Quantification of CC3+ cells in *Hpdl* WT vs KO mouse brain per 500-µm column. Statistical analysis (two-way analysis of variance [ANOVA]) was performed between the categories. $N = 3$ independent experiments with two sections per mouse quantified (****$P < 0.0001$). (**f**) Coronal sections of P9 KO mouse brain were stained for DAPI (blue) and CC3 (green) in combination with either CTIP2 (red), OLIG2 (red), or GFAP (red). White arrows depict colocalization of the cell identity markers with CC3. Scale bar: 50 µm. (**g**) Quantification of the number of CTIP2+, OLIG2+, or GFAP+ cells which were also CC3+. Analysis shows CTIP2+ cells to be most affected, followed by GFAP+ cells, followed by OLIG2+ cells. $N = 2$ independent experiments. Two sections per mouse were used for quantification.

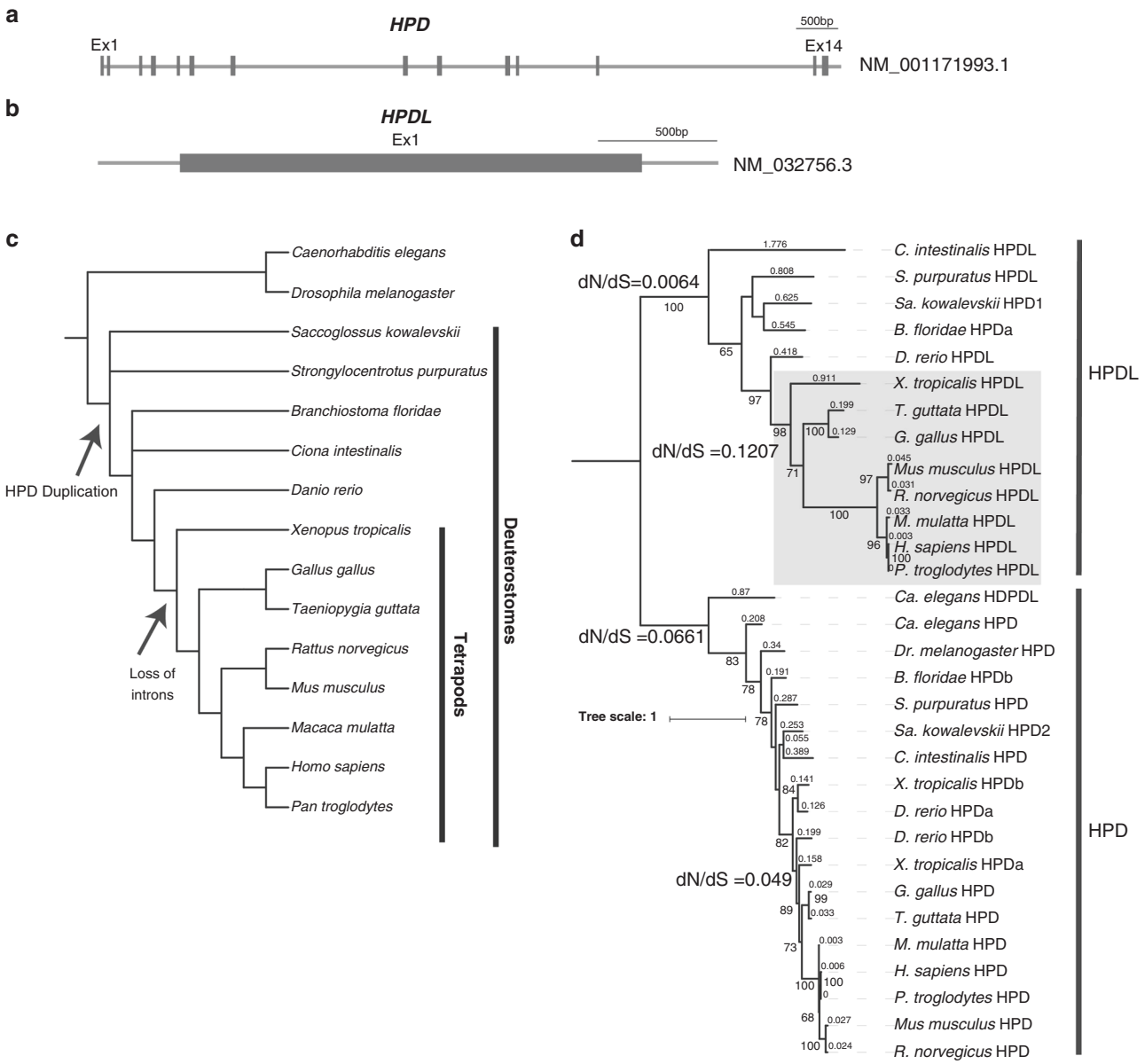

Fig. 4 HPD and HPDL evolutionary analysis reveals two separate, yet significant, evolutionary events leading to the emergence of vertebrate HPDL with novel function. (**a**) Schematic of the major human *HPD* isoform depicting the coding sequence spanning 14 exons using the University of California–Santa Cruz (UCSC) Genome Browser as a reference. (**b**) Schematic of human *HPDL* depicting the coding sequence spanning a single exon and the 5′ and 3′ UTRs using the UCSC Genome Browser as a reference. Scale bars are shown. (**c**) Species tree of genomes used in this analysis. Arrows indicate *HPD* gene duplication and a possible retrotransposition event leading to the loss of exons in *HPDL*. (**d**) Gene tree for the *HPD* and *HPDL* genes constructed with RAxML with the JTT model. Branch lengths are on top of each branch, bootstrap support of 65 and above are shown below branches. Gray box indicates a loss of *HPDL* exons in tetrapods. d*N*/d*S* ratios are estimated using codeml and the model showing four different d*N*/d*S* ratios is shown (Fig. S8).

intron-containing gene. Human HPDL is more similar to zebrafish hpdl than to human HPD (40% vs. 25% amino acid sequence identity). The data suggest that lower deuterostome *HPDL* was lost contemporaneously with its retrotransposition to intronless *HPDL*, which subsequently acquired an MTS and amino acid changes that provided a new essential function.

DISCUSSION

Here, we implicate *HPDL* in a recessive infantile-onset neurodegenerative disease, characterized by intellectual and motor disability, epilepsy, spastic tetraplegia, and brain atrophy. Radiographic phenotypes include severe cortical and cerebellar atrophy, thinning of corpus callosum, and ventriculomegaly. There are likely a range of developmental brain phenotypes resulting from loss of *HPDL*, although our subjects are likely to be at the most severe end of the spectrum given the nature of the alleles. HPDL is not the first mitochondrial dioxygenase to be implicated in a neurometabolic disorder. ETHE1, another mitochondrial dioxygenase, is linked to ethylmalonic encephalopathy, an inherited disorder defined by elevated excretion of ethylmalonic acid, leading to developmental delay, seizures, and hypertonia.[36] Future studies will be required to elucidate the full range of phenotypes, genotype–phenotype correlations, and mechanisms of pathogenicity.

We demonstrate that HPD and HPDL not only have different tissue-wide expression patterns, but also show different subcellular localizations. Localization of HPDL to mitochondria, together with the disrupted metabolomes in *HPDL* KO cells and mice, implicates this disease not only as a mitochondrial neurodegenerative condition, but also as a likely neurometabolic condition. In a direct comparison between *Hpdl* WT and KO mice, we highlight the pathophysiological importance of the HPDL enzyme for neurodevelopment and survival. Interestingly, the fact that the mice are born healthy and begin regressing soon after birth suggests that the mice lacking *Hpdl* may depend on an environmental or some other factor provided by their mother to stay well. Further, although astrocytes have been shown to express HPDL at the highest levels in the brain,[27–29] in *HPDL* KO brain we found more neurons than glia with evidence of apoptosis, suggesting that neurons are ultimately more sensitive to *HPDL* loss.

There are several possible reasons why we were unable to identify a specific substrate for HPDL despite comprehensive metabolic profiling. First, the metabolite may be unidentifiable by the current approach, where we focused on small molecules that have been preannotated in public databases. Second, the amount of the potential substrate may have been too small to detect due to large sample volume. Therefore, future work should aim toward more targeted approaches (i.e., analyzing specific brain regions rather than the whole brain, or analyzing mitochondrial-specific extracts rather than whole-cell extracts). Finally, it is still possible that HPDL is not an enzyme, since we did not confirm its enzymatic activity. However, given its close sequence similarity to the enzyme HPD, presence of iron-binding motifs, disease caused by biallelic null variants, and metabolic dysregulation seen in the cell and mouse models, this hypothesis is less likely.

This finding was further supported by our evolutionary analysis, suggesting that HPDL and HPD are functionally different. We have shown that the *HPDL* found in humans likely emerged through a series of *HPD* gene duplication events, the last of which probably involved a retrotransposition event at the transition from deuterostomes to tetrapods. This suggests that at some point in time, *HPDL* existed as a multiexonic gene before losing its introns and gaining an MTS.

Taken together, we describe a novel, early-onset neurodegenerative disorder caused by variants in *HPDL*, and also illustrate the importance of this protein for neurodevelopment and survival. Through evolutionary modeling and analysis, we illustrate the route by which human *HPDL* emerged as a paralogue of *HPD*, with proposed novel biochemical functions. Although further work is needed to identify the specific function of mitochondrial HPDL, we believe our work is a significant cornerstone to elucidate the downstream mechanisms of HPDL, and to potentially lead to a treatment for this disease.

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-020-01010-y) contains supplementary material, which is available to authorized users.

DATA AVAILABILITY

Data are available in a public, open access repository. The exome sequencing from individuals from the UCSD study site have been deposited in the Database of Genotypes and Phenotypes under accession numbers phs001272.v1.p1 and phs000744.v4.p2.

ACKNOWLEDGEMENTS

The authors thank the patients and their families for participation in this study. This work was supported by NIH U01 MH108898, R01NS048453, R01 NS098004, the Simons Foundation Autism Research Initiative (SFARI), and Qatar National Research Foundation 6–1463 (J.G.G.); NIH S10OD020025 and R01ES027595 (M.J.); the Deutsche Forschungsgemeinschaft (DFG) under Germany´s Excellence Strategy (EXC-2189 project ID: 390939984) and a European Research Council (ERC) Consolidator Grant (648235, N.W.); and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre (D.M.). S.G.G. is supported by the Ruth L. Kirschstein Institutional National Research Service Award (T32 GM008666) from the National Institute on Deafness and Other Communication Disorders and by award F31HD095602 from the NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development. We thank the Rady Children's Institute for Genomic Medicine, Broad Institute (U54HG003067 to E. Lander and UM1HG008900 to D. MacArthur), the Yale Center for Mendelian Disorders (U54HG006504 to R. Lifton and M. Gunel) for sequencing

support, and the Matchmaker Exchange. We also thank UCSD Mouse Transgenic Core and UCSD Neuroscience Microscopy Core (P30 NS047101). We acknowledge M. Gerstein, S. Mane, A.B. Ekici, S. Uebe, E.S. Cauley, and UCSD IGM Genetics Center for sequencing support and analysis; the Yale Biomedical High-Performance Computing Center for data analysis and storage; the Yale Program on Neurogenetics; and the Yale Center for Human Genetics. We thank Pradipta Ghosh for sharing CACO-2 cells.

DISCLOSURE

The authors declare no conflicts of interest.

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

REFERENCES

1. Brassier A, Ottolenghi C, Boddaert N, et al. Prenatal symptoms and diagnosis of inherited metabolic diseases. Arch Pediatr. 2012;19:959–969.
2. Tarailo-Graovac M, Shyr C, Ross CJ, et al. Exome sequencing and the management of neurometabolic disorders. N Engl J Med. 2016;374: 2246–2255.
3. BoAli AY, Alfadhel M, Tabarki B. Neurometabolic disorders and congenital malformations of the central nervous system. Neurosciences (Riyadh). 2018;23:97–103.
4. Willemsen MA, Harting I, Wevers RA. Neurometabolic disorders: five new things. Neurol Clin Pract. 2016;6:348–357.
5. Abu-Omar MM, Loaiza A, Hontzeas N. Reaction mechanisms of mononuclear non-heme iron oxygenases. Chem Rev. 2005;105:2227–2252.
6. Burlina A, Zacchello F, Dionisi-Vici C, et al. New clinical phenotype of branched-chain acyl-CoA oxidation defect. Lancet. 1991;338:1522–1523.
7. Boissel S, Reish O, Proulx K, et al. Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. Am J Hum Genet. 2009;85:106–111.
8. Fernández-Cañón JM, Granadino B, Beltrán-Valero de Bernabé D, et al. The molecular basis of alkaptonuria. Nat Genet. 1996;14:19–24.
9. Ferreira P, Shin I, Sosova I, et al. Hypertryptophanemia due to tryptophan 2,3-dioxygenase deficiency. Mol Genet Metab. 2017;120:317–324.
10. Bruick RK, McKnight SL. A conserved family of prolyl-4-hydroxylases that modify HIF. Science. 2001;294:1337–1340.
11. Moran GR. 4-Hydroxyphenylpyruvate dioxygenase. Arch Biochem Biophys. 2005;433:117–128.
12. Rüetschi U, Cerone R, Pérez-Cerda C, et al. Mutations in the 4-hydroxyphenylpyruvate dioxygenase gene (HPD) in patients with tyrosinemia type III. Hum Genet. 2000;106:654–662.
13. Kubo S, Kiwaki K, Awata H, et al. In vivo correction with recombinant adenovirus of 4-hydroxyphenylpyruvic acid dioxygenase deficiencies in strain III mice. Hum Gene Ther. 1997;8:65–71.
14. Szymanska E, Sredzinska M, Ciara E, et al. Tyrosinemia type III in an asymptomatic girl. Mol Genet Metab Rep. 2015;5:48–50.
15. Tomoeda K, Awata H, Matsuura T, et al. Mutations in the 4-hydroxyphenylpyruvic acid dioxygenase gene are responsible for tyrosinemia type III and hawkinsinuria. Mol Genet Metab. 2000;71:506–510.
16. Item CB, Mihalek I, Lichtarge O, et al. Manifestation of hawkinsinuria in a patient compound heterozygous for hawkinsinuria and tyrosinemia III. Mol Genet Metab. 2007;91:379–383.
17. Barroso F, Correia J, Bandeira A, et al. Tyrosinemia type III: a case report of siblings and literature review. Rev Paul Pediatr. 2020;38:e2018158.
18. Haijes HA, Willemsen M, Van der Ham M, et al. Direct infusion based metabolomics identifies metabolic disease in patients' dried blood spots and plasma. Metabolites. 2019;9:12.
19. Pei J, Grishin NV. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol Biol. 2014;1079:263–271.
20. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–1313.
21. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–245.
22. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–1591.
23. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010;38:W7–13.
24. Charif D, Thioulouse J, Lobry JR, Perriere G. Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics. 2005;21:545–547.
25. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020; 581:434–443.
26. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–660.
27. Zhang Y, Sloan SA, Clarke LE, et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. Neuron. 2016;89:37–53.
28. Saunders A, Macosko EZ, Wysoker A, et al. Molecular diversity and specializations among the cells of the adult mouse brain. Cell. 2018;174: 1015–.e1016.
29. Vanlandewijck M, He L, Mäe MA, et al. A molecular atlas of cell types and zonation in the brain vasculature. Nature. 2018;554:475–480.
30. Thul PJ, Akesson L, Wiking M, et al. A subcellular map of the human proteome. Science. 2017;356:eaal3321.
31. Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Mol Cell Proteomics. 2015;14:1113–1126.
32. Roise D, Theiler F, Horvath SJ, et al. Amphiphilicity is essential for mitochondrial presequence function. Embo j. 1988;7:649–653.
33. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347:1260419.
34. Natt E, Kida K, Odievre M, Di Rocco M, Scherer G. Point mutations in the tyrosine aminotransferase gene in tyrosinemia type II. Proc Natl Acad Sci U S A. 1992;89:9297–9301.
35. Phaneuf D, Lambert M, Laframboise R, Mitchell G, Lettre F, Tanguay RM. Type 1 hereditary tyrosinemia. Evidence for molecular heterogeneity and identification of a causal mutation in a French Canadian patient. J Clin Invest. 1992;90:1185–1192.
36. Tiranti V, Viscomi C, Hildebrandt T, et al. Loss of ETHE1, a mitochondrial dioxygenase, causes fatal sulfide toxicity in ethylmalonic encephalopathy. Nat Med. 2009;15:200–205.