Title

Leigh syndrome: a study of 209 patients at the Beijing Children's Hospital

Running head

Leigh syndrome: 209 pediatric patients from China

Authors

Sarah L. Stenton[1,2,*], Ying Zou[3,*], Hua Cheng[4,*], Zhimei Liu[3], Junling Wang[3], Danmin Shen[3], Hong Jin[3], Changhong Ding[3], Xiaolu Tang[4], Suzhen Sun[5], Hong Han[6], Yanli Ma[7], Weihua Zhang[3,7], Ruifeng Jin[8], Hua Wang[9], Dan Sun[10], Junlan Lv[3,†], Holger Prokisch[1,2,†], Fang Fang[3,†]

*co-first authors †co-corresponding authors

Affiliations

1. Institute of Human Genetics, School of Medicine, Technical University of Munich, Munich, Germany.
2. Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany.
3. Department of Neurology, National Center for Children's Health, Beijing Children's Hospital, Capital Medical University, Beijing, China.
4. Image Center, National Center for Children's Health, Beijing Children's Hospital, Capital Medical University, Beijing, China.

5. Department of Neurology, Children's Hospital of Hebei Province, Shijiazhuang, China.
6. Department of Neurology, Children's Hospital of Shanxi Province, Taiyuan, China.
7. Department of Neurology, Children's Hospital Affiliated to Zhengzhou University, Henan Children's Hospital, Zhengzhou, China.
8. Department of Neurology, Qilu Children's Hospital of Shandong University, Jinan, China.
9. Department of Pediatrics, Shengjing Hospital of China Medical University, Shenyang, China.
10. Department of Neurology, Wuhan Children's Hospital, Wuhan, China.

Correspondence to:

Fang Fang, Department of Neurology, National Center for Children's Health, Beijing Children's Hospital, Capital Medical University, Beijing, China, No. 56 Nanlishi Road, Xicheng District, Beijing, China 100045, E-mail: fangfang@bch.com.cn.

Correspondence may also be addressed to:

Holger Prokisch, Institute of Neurogenomics (ING), Building 3537, room 8128 Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH) Ingolstaedter Landstr. 1 D-85764 Neuherberg, Germany, E-mail: prokisch@helmholtz-muenchen.de.

Junlan Lv, Department of Neurology, National Center for Children's Health, Beijing Children's Hospital, Capital Medical University, Beijing, China, No. 56 Nanlishi Road, Xicheng District, Beijing, China 100045, E-mail: lujunlan@aliyun.com.

ABSTRACT

Objective: Leigh syndrome (LS) is a heterogeneous neurodegenerative disease and the most frequent pediatric manifestation of mitochondrial disease. In the largest patient collection to date, this study aimed to provide new insights into the clinical and genetic spectrum of LS, defect-specific associations, and predictors of disease course and survival.

Methods: Clinical, metabolic, neuroimaging, onset, and survival data were collected from the medical records of 209 patients referred to the Beijing Children's Hospital with symmetrical basal ganglia and/or brainstem neuroimaging changes indicative of LS by 30 centers from the Chinese network of mitochondrial disease (mitoC-NET) between January 2013 and July 2021 for exploratory analysis.

Results: Pathogenic variants were identified in 52 genes, most frequently *MT-ATP6*, *SURF1*, and *PDHA1*. Maternally inherited variants accounted for 42% (heteroplasmy level ≥90% in 64%). Phenotypes spanned 92 Human Phenotype Ontology terms. Elevated serum lactate (144/195), global developmental delay (142/209), and developmental regression (103/209) were most frequent. Discriminating neuroimaging and/or clinical features were identified for *MT-ATP6* (m.9176T>C), *MT-ND5*, ~~*NDUFAF6,*~~ *PDHA1*, *SUCLG1*, and *SURF1*. Poorest survival was associated with *MT-ND5*, *MT-ATP6* (m.8993T>C and m.9176T>C), *SURF1*, and *ALDH5A1* (≤50% 3 year survival), in contrast to milder defects with specific treatment (*ECHS1* and *SLC19A3*, 100% 3 year survival).

Interpretation: Our data define phenotype, onset, and survival of LS in a defect-specific manner, identifying features discriminating between genetic defects and predictive of disease outcome. These

findings are essential to early diagnosis, in optimizing family counselling, and to the design and monitoring of future clinical trials, the next frontier of LS research.

INTRODUCTION

Mitochondrial disease is a heterogeneous collection of rare diseases due to defects in over 400 genes across the nuclear and mitochondrial genome (Stenton and Prokisch 2020, Schlieben and Prokisch 2020). Defects in approximately 100 of these genes can, in 35-55% of patients, manifest with Leigh syndrome (LS; subacute necrotizing encephalomyelopathy, MIM 25600). LS is the most frequent clinical presentation of mitochondrial disease in children with an estimated prevalence of 1 in 40,000 (Darin et al., 2001, Rahman et al., 1996, Skladal et al., 2003).

Despite vast clinical and genetic heterogeneity, LS patients are united by bilateral symmetrical lesions within the basal ganglia and/or brainstem structures, observed as regions of focal hyperintensity on brain magnetic resonance imaging (MRI) (Lake et al., 2015). These neurodegenerative features arise in association with rapid deterioration of cognitive and motor function, in many cases resulting in respiratory failure and death. The underlying disease pathomechanism is impaired mitochondrial energy metabolism (Rahman et al., 1996), often demonstrated by elevated serum and/or cerebrospinal fluid (CSF) lactate, reduced oxidative phosphorylation (OXPHOS) and/or pyruvate dehydrogenase (PDH) complex activity in patient-derived tissues, or indicated by the identification of disease-causing variants in a gene required for mitochondrial energy generation (Lake et al., 2016).

Though all patients with these features are gathered under the clinical diagnosis of LS irrespective of their underlying genetic diagnosis, there is increasing recognition of genetic defect-specific patterns of onset and survival, in addition to clinical, metabolic, and brain MRI features (Alves et al., 2020, Stenton et al., 2020, Sofou et al., 2018). However, defect-specific onset and survival data in LS remains limited. This limitation arises due to the rarity and broad genetic underpinning of the disease, with individual cohorts to date reaching approximately 100 genetically confirmed LS patients from European and East Asian populations (Lim et al., 2021, Ardissone et al., 2021, Lee et al., 2020, Ogawa et al., 2020, Sofou et al., 2018).

Here, we pave the way to overcome this shortfall by the characterization of 209 genetically confirmed LS patients at the Beijing Children's Hospital, a national center specialized in the diagnosis of pediatric mitochondrial disease. We thereby resolve the clinical and genetic spectrum of the largest cohort of systematically evaluated LS patients to date and uncover associations between genotype, phenotype, disease onset, and survival, essential in the design of future clinical trials for LS (Yung Tiet et al., 2021) by determining the most frequent genetic etiologies of LS and by enabling defect-specific definition of therapeutic effect on survival.

SUBJECTS/MATERIALS AND METHODS

Participants

All patients recruited to the study demonstrated symmetrical basal ganglia and/or brainstem neuroimaging changes indicative of LS, had received a molecular genetic diagnosis, and had disease

onset <18 years of age. Patients were assessed against the diagnostic criteria for LS (Rahman et al., 1996, Lake et al., 2016), as follows: i) clinical history suggestive of neurodegeneration (e.g., neurodevelopmental delay and/or regression), ii) symmetrical basal ganglia and/or brainstem neuroimaging changes, and iii) abnormal metabolism characterized by elevated lactate in serum (>2.2 mmol/L) and/or CSF (>2.78 mmol/L), a defect in OXPHOS and/or PDH complex activity, and/or a molecular genetic diagnosis in a gene related to mitochondrial dysfunction. Patients fulfilling all three criteria were classified as having LS. Patients fulfilling the first two criteria with a molecular genetic diagnosis in a gene not known to be related to mitochondrial dysfunction were classified as having LS-mimicking disease.

In total, 209 patients of Chinese ethic background were included in the study (35 previously reported, Fang et al., 2017) referred by one of 30 recruiting centers in the Chinese mitochondrial disease network (mitoC-NET, founded in 2018). A standardized proforma was used to capture molecular genetic, clinical, metabolic, and neuroradiological data, in addition to family history, treatment, disease onset, follow-up time, and survival status, by a designated investigator. Phenotype data were standardized as Human Phenotype Ontology (HPO) terms. This study was approved by the local ethical review board and was performed under the ethical guidelines of the Declaration of Helsinki. Written informed consent was obtained for all participants.

Molecular genetic investigations and analyses

DNA was extracted from blood and urine for analysis. Variants were detected by fluorescent restriction fragment length polymorphism analysis (n=8), Sanger sequencing (n=3), various next-

generation sequencing-based gene panels (n=111), or exome sequencing (n=87). For mtDNA variants, heteroplasmy level was measured in blood and/or urine, and carrier testing was offered to all maternal family members. For autosomal variants, segregation was confirmed in parental samples.

Disease genes reported to be associated with LS were identified from the Leigh Map (vmh.life/#leighmap, Rahman et al., 2017) and literature reports (Chakravarty et al., 2003, Fang et al., 2017, Kamps et al., 2018, Vögtle et al., 2018, Borna et al., 2019, Rubegni et al., 2019, Hu et al., 2020, Alves et al., 2020, Tian et al., 2021). Disease genes reported to be associated with mitochondrial energy metabolism were taken from Stenton and Prokisch 2020 and Schlieben and Prokisch 2020. Designation of variant pathogenicity was in accordance with the American College of Medical Genetics and Genomics and the Association for Molecular Pathology guidelines (Richards et al., 2015).

Statistical analysis

Due to the rarity of the disease, the statistical analyses (performed using R version 4.1.0) were exploratory, and no sample size calculation was performed in advance to power the study for specific hypotheses. Fisher exact tests were performed to compare categorical variables between two groups. Survival analysis used the Log rank test. In the analyses, the following features were investigated for their association to each of 18 molecular genetic diagnoses with pathogenic variants identified in ≥3 patients: age at onset below versus above 6 months (early-onset LS) and below versus above 2 years (late-onset LS), and 58 clinical, metabolic, and neuroimaging HPO terms reported in ≥10 patients. All statistical tests were two-sided, and statistical significance was determined as ≤0.05. The ~~adjusted~~ *p*

values and confidence intervals ~~was~~ reported ~~based~~ underwent ~~on~~ Bonferroni correction for multiple testing. Principal component analysis (PCA) was utilized to visualize high dimensional HPO data.

Data availability

The molecular genetic (variant-level) and HPO phenotype (patient-level) data included in this study are freely available on GENOMITexplorer (Beijing Leigh Group Project, https://prokischlab.github.io/GENOMITexplorer/) and are found in the **Online Supplementary Tables (Tables S1-S3)**.

RESULTS

Study population

A total of 209 genetically diagnosed patients were enrolled into the study, inclusive of three sibling pairs (203 pedigrees) (134 males, 64%), 207 patients with disease meeting the criteria for LS and two patients with LS-mimicking disease (see **Supplementary Online Table 1** for patient-level data). The median age of onset was 1.1 years (range = 0-10.8 years, interquartile range [IQR] = 2.1 years), with 66% of patients (n = 137) presenting before 2 years of age. At last contact, 155 patients were alive (median age = 6 years, range = 0.5-18.3 years, IQR = 4.8 years, median follow-up time 48 months) and 54 patients were deceased (median age = 3 years, range = 0.3-14 years, IQR = 4.4 years, median follow-up time 13 months). 85% of patients with sufficient follow-up length were alive at or beyond 3 years of age (153/180 patients at risk). A maternal family history of disease was reported in 20 families, 17 due to a pathogenic mtDNA variant and three due to an X-linked dominant (XLD) variant

in either *PDHA* or *HSD17B10*. Most pregnancies were uneventful (174, 89%). Of those with a complication, oligohydramnios (n = 5), preterm birth (n = 5, <37 completed weeks of gestation), decreased fetal movement (n = 5), and intrauterine growth restriction (n = 4) were most frequent. In 69 patients, a disease-triggering factor was reported (infection, n = 66; vaccination, n = 2; or operation, n = 1).

Mutational spectrum

Across the 209 patients, 87 (42%) had disease-causing variants in the mtDNA and 122 (58%) in the nuclear DNA. P/LP variants were identified in 46 LS-associated genes (nine mtDNA-encoded, 37 nuclear-encoded), four nuclear-encoded mitochondrial disease genes, so far not linked to LS (*COA6*, n = 1; *HSD17B10*, n = 2; *L2HGDH*, n = 1; *TARS2*, n = 2) and two inborn errors of metabolism (*ALG1*, n = 1; *GAMT*, n = 1) presenting with the typical clinical and neuroradiological features of LS in the absence of mitochondrial dysfunction (**Figure 1 a**). The nuclear diagnoses stratify into 104 recessively inherited defects (89 compound heterozygous and 15 homozygous) and 18 XLD inherited defects, of which 7/18 arose *de novo* (*PDHA1* n = 6 and *HSD17B10* n = 1). Across both genomes, most frequently, disease-causing variant were identified in *MT-ATP6* (n = 28) and *SURF1* (n = 22), in-keeping with other large LS cohorts reported in the literature (Lim et al., 2021, Ardissone et al., 2021, Ogawa et al., 2020, Lee et al., 2020, Sofou et al., 2018) (**Figure 1 b**-**c**).

In total, here we report 203 unique P/LP variants (see **Supplementary Online Table 2** for patient-level variant data), 20 unique mtDNA variants (all reported in MITOMAP) and 183 unique nuclear variants (38 reported in as P/LP ClinVar). Seventeen of the nuclear variants were detected in ≥2

unrelated patients, indicating potential founder events in the Chinese population, of which five were confirmed to be present in multiple individuals across Asian populations in gnomad (*NDUFAF6*, 8:96047755T>C, n = 3; *SURF1*, 9:136218994delCTCT, n = 3; *ALDH5A1*, 6:24505104detT, n = 2; *ECHS1*, 10:135182478G>A, n = 2; *GTPBP3*, 19:17449956A>C, n = 2). Frequently arising mtDNA variants included those in *MT-ATP6* (m.9176T>C n = 12, m.8993T>C n = 7, m.8993T>G n = 5), *MT-ND1* (m.3697G>A n = 6), *MT-ND3* (m.10197G>A n = 7, m.10191T>C n = 5, m.10158T>C n = 5), *MT-ND5* (m.13513G>A n = 9), *MT-ND6* (m.14487T>C n = 9), and *MT-TK* (m.8344A>G n = 5). Across all pathogenic mtDNA variants, the median degree of heteroplasmy was 97% (range = 37-100%, IQR = 14%) in blood (n = 84) and median 100% (range = 65-100%, IQR = 9%) in urine (n = 28). In 56 of the 87 patients with mtDNA variants (64%), the variant was detected with a high degree of heteroplasmy (≥90% in at least one tissue), in 36 patients (41%) even homoplasmic. The lowest patient-level clinical expression thresholds were associated with m.13513G>A and m.13094T>C in *MT-ND5* (37% and 46% in blood, respectively). The degree of variant heteroplasmy in the patients' mother was investigated in all pedigrees where maternal DNA from blood and/or urine was available. Of the 20 unrelated patients with a maternal family history of disease, maternal DNA was available for all and heteroplasmy in the blood and/or urine of the mother ranged from undetectable to 90% (median 80%). Of the 67 unrelated patients with no reported maternal family history of disease, maternal DNA was available for 61 and heteroplasmy level in blood and/or urine ranged from non-detectable (23 mothers), indicating *de novo* occurrence of the variant, to ≥90% (median 4%). The variants carried by asymptomatic mothers at ≥90% heteroplasmy level were as follows: m.8993T>C, m.9176T>C, and m.9185T>C in *MT-ATP6*, m.3697G>A in *MT-ND1*, m.14487T>C in *MT-ND6*. Across all measured samples there was a high correlation of heteroplasmy degrees between blood and urine ($R = 0.94$, $p = 6.3\times10^{-14}$, Spearman's rank test).

Clinical, metabolic, and neuroimaging feature spectrum

Clinical and metabolic data were available for all 209 patients. A median of 12 HPO terms (range = 3-23) spanning a median of four systems (range = 1-8) were recorded per patient, resulting in a total of 92 unique HPO terms across the cohort (see **Supplementary Online Table 3** for patient-level HPO terms). In 56% of patients, at least three individual organ systems (not including metabolic derangement) were affected during the course of the disease, most frequently neurological, musculoskeletal, and ophthalmological (**Figure 2 a**). Summative analysis of the clinical data revealed the most common clinical examination findings to be global developmental delay (142/209), followed by developmental regression (103/209), failure to thrive (101/209), and muscle weakness (91/209) (**Figure 2 b**). Developmental regression was the most frequent initial presenting symptom (47/209), at a median age of 0.7 years. The clinical features at onset in relation to age at onset are depicted in **Figure 2 c**. All other clinical features were less frequent at onset or developed later in the disease course. The key demographic and clinical features from this and previous studies are displayed in **Table 1**.

Serum lactate was measured in 195 patients (in 154 patients on one visit, in 41 patients on two visits) (mean = 3.9 mmol/L, standard deviation [s.d.] = 2.8 mmol/L, reference range = 0.5-2.2 mmol/L) and was elevated on at least one visit in 144/195 patients (74%). In seven patients, serum lactate was elevated on only one of two visits. CSF lactate was measured in 65 patients (mean = 3.2 mmol/L, s.d. = 1.4 mmol/L, = reference range 1-2.78 mmol/L) and was elevated in 43 (66%). Overall, correlation was mild between serum and CSF lactate (R = 0.31, p-value = 0.006).

In accordance with our inclusion criteria, neuroimaging data (brain MRI) revealed symmetrical basal ganglia and/or brainstem signal abnormalities in all patients (basal ganglia n = 160; brainstem n = 152, both involved n = 103). Lesions were further stratified by discrete location into those affecting the lenticular nucleus (150/160) and/or caudate nucleus (87/160) in the basal ganglia, and those affecting the medulla oblongata (81/152) in the brainstem. The neuroimaging from patients with defects in genes newly reported to be associated with LS or reported to result in LS-mimicking disease are depicted in **Figure 3**.

The profile of clinical, metabolic, and MRI features (58 reported HPO terms in ≥10 patients) was compared between the most frequent genetic diagnoses (18 genetic diagnoses made in ≥3 patients) (see **Methods**). Principal component analysis (PCA) demonstrated potential clustering of genetic defects based on HPO terms (**Figure 4 a**), with basal ganglia/brainstem lesions on MRI explaining the maximal amount of variance (**Figure 4 b**). To explore this further, we compared each of these 18 diagnoses in turn to all remaining patients to look for significant defect-specific associations (**Table 2**). Stringently correcting the confidence intervals for multiple-testing, the analysis revealed significant MRI feature associations in *MT-ND5* and *SURF1* defect. *MT-ND5* was ~~enriched for thalamus involvement and~~ depleted for basal ganglia lesions. *SURF1* was enriched for medulla oblongata and cerebellar involvement and depleted for basal ganglia lesions. Moreover, the analysis uncovered several significant clinical feature associations. *MT-ATP6* (m.9176T>C) was enriched for ptosis, ~~*MT-ND5* for cardiomyopathy, *NDUFAF6* for dysarthria,~~ *PDHA1* for peripheral neuropathy, *SUCLG1* for methylmalonic aciduria, and *SURF1* for hypertrichosis, peripheral neuropathy, and nystagmus. The

frequency of HPO terms by molecular diagnosis and the results of the complete association analysis are displayed in **Supplementary Online Table 4**.

Clinical, metabolic, neuroimaging, and genetic factors associated with disease onset, course, and survival

Early-onset LS (<6 months, n = 55) in comparison to late-onset LS (>2 years, n = 72) was enriched for global developmental delay (49/55 vs. 32/72, $p = 6.6\times10^{-6}$), feeding difficulties (22/55 vs. 8/72, $p = 1.5\times10^{-2}$), and hypotonia (25/55 vs. 10/72, $p = 6.6\times10^{-3}$), and depleted for abnormality of movement (8/55 vs. 45/72, $p = 5.3\times10^{-7}$), ataxia (4/55 vs. 34/72, $p = 3.1\times10^{-5}$), dysarthria (0/55 vs. 21/72, $p = 1.7\times10^{-4}$), and dystonia (3/55 vs. 27/72, $p = 8.8\times10^{-4}$).

Follow-up clinical data was available for analysis in a total of 160 LS patients (median follow-up 48 months, range 0.2-184 months), at which point parents were asked to give a subjective judgement on their child's disease course by the selection of one term from: "improvement", "stable", or "progression" by consideration of the child's gain/loss of developmental milestones. This indicated clinical improvement in 36 patients (27%). Of the remaining 97 patients, 48 (36%) were stable, 11 (8%) demonstrated clinical progression, and 38 (29%) were deceased at the point of last follow-up.

Disease onset and survival demonstrated defect-specific patterns (**Figure 4 c**). The earliest age of onset was associated with *SUCLG1* (median 0.1 years, range = 0-0.3 years, IQR = 0.3 years) and the *MT-ATP6* variant m.8993T>G (median 0.5 years, range = 0.2-0.6 years, IQR = 0.2 years), and the latest with the *MT-ATP6* variant m.9176T>C (median 4.1 years, range = 0-9.5 years, IQR = 6.1 years) and *NDUFAF6* (median 3.5 years, range = 0-4.7 years, IQR = 2 years). The lowest 1 year and 3 year survival rates were associated with *MT-ND5*, *MT-ATP6* variants m.8993T>C and m.9176T>C, *SURF1*, and *ALDH5A1* defects (range 33%-60% 1 year survival, range 30-50% 3 year survival). In association with *ECHS1* and *SLC19A3* defects, all patients with follow-up exceeding 3 years were alive.

The survival analysis for the study population as a whole is shown in **Figure 5 a**, demonstrating 3 year survival in excess of 75%. Disease onset, clinical, metabolic, neuroimaging, and genetic factors were analyzed for their association with patient survival (see **Methods**). Following multiple-testing correction, dyspnea was found to be associated with significantly poorer survival (**Figure 5 b**) and dysarthria, a phenotype associated with late-onset LS, with significantly better survival (**Figure 5 c**).

Across the study cohort, 36 patients received a defect-specific treatment: 12 *PDHA1* patients (12/15) and one *TPK1* patient (1/1) received thiamine 4-30mg/kg/day, four *SLC19A3* patients (4/6) received thiamine 5-20mg/kg/day in addition to biotin 3-45mg/kg/day, one *COQ7* patient (1/1) received coenzyme Q10 10-50mg/kg/day, and 18 *ECHS1*/*HIBCH* (18/18) patients received a valine-restricted diet. Collectively, patients receiving a defect-specific treatment demonstrated overall better survival in comparison to their untreated counterparts ($p = 0.0069$, Log rank test) (**Figure 5 d**), with all patients

with sufficient follow-up duration being alive at 3 years of age. There was no significant difference in gender nor in age of onset to provide an alternative explanation for this difference (treated, median age of onset 1 year, range 0-4.8 years, IQR 1.2 years; untreated, median age of onset 1.2 years, range 0-10.8 years, IQR 2.6 years).

Follow-up neuroimaging (brain MRI and/or DWI) data were available for analysis in 67 patients (median follow-up time = 10 months, range 0.5-102 months, IQR = 19 months). Lesion changes were analyzed in accordance with Bonfante et al., 2016 and classified as either: i) stable – no change on T2 MRI or DWI, ii) progression – new lesions present or the number and/or extension of previously visualized lesions had increased, iii) regression – complete resolution on both T2 MRI and DWI of lesions previously visualized, or iv) evolution – normalization of DWI with persistent T2 MRI changes, or decreased size of the T2 MRI signal changes as a result of encephalomalacia. In total, regression of MRI abnormalities was reported in four patients (4/67, 6%), seven remained stable (7/67, 9%), 16 demonstrated evolution (16/67, 24%), and 40 progressed by the point of last follow-up (40/67, 58%) (**Figure 6**). Comparing disease course in 59 patients with both clinical and imaging follow-up data available for analysis demonstrated all patients with regression of MRI abnormalities to have a corresponding clinical improvement, resulting in an enrichment for MRI recovery in the patients with clinical improvement (3/14 vs. 0/45, $p = 0.04$, Fisher exact test). Conversely, all patients who were deceased at last follow-up demonstrated enrichment for progression of the MRI abnormality (17/19 vs. 20/40, $p = 0.02$, Fisher exact test).

DISCUSSION

LS is a rare and heterogeneous disease, demonstrated in this large cohort of 209 patients by its underpinning across 52 different disease genes and presentation with over 90 different phenotypes. LS is considered to be a disease of infancy and early childhood, typically reported to present within the first two years of life, and characteristically presenting during a period of metabolic stress following a period of normal development (Rahman et al., 1996, Lee et al., 2009, Naess et al., 2009). In-keeping with this, the median age of onset in our cohort was 1.1 years and 66% of patients presented before 2 years of age. In 33% of our patients, disease was triggered by infection, vaccination, or an operation, and we report almost one quarter of patients (22%) to present with developmental regression after a period of normal development at a median of 0.7 years. However, we also report over one third of patients to present with late-onset LS (34%), defined as onset over the age of two years (Sofou et al., 2014). This value is higher than the approximate 20% previously reported (Sofou et al., 2014) and may reflect the genetic underpinning of the cohort with the majority of late-onset cases accounted for by *MT-ATP6* (m.9176T>C and m.8993T>C) and *MT-ND1, -ND3, -ND5,* and *-ND6* defects. This highlights the defect-specific patterns of the age of onset, with earliest onset in *SUCLG1* and *MT-ATP6* (m.8993T>G) defect, latest onset in *MT-ATP6* (m.9176T>C) and *NDUFAF6* defect, and a number of defects such as *MT-ND3*, *-ND5,* and *-ND6* demonstrating wider distributions. Moreover, we found age of onset to have important implications for disease manifestation, with late-onset LS associated with disorders of movement and coordination, including dystonia, ataxia, and dysarthria, as previously reported (Sofou et al., 2014). We did not, however, find a significant association between age of onset and overall survival to support previous reports of poorer survival in patients with disease onset <6 months (Ogawa et al., 2020, Sofou et al., 2014, Naess et al., 2009).

We report 42% of LS to be maternally inherited, this figure is in-keeping with other large European and East Asian cohorts of LS patients published in the last 5 years, collectively describing 514 genetically confirmed LS patients, and reporting figures ranging from 27-52% (Lim et al., 2021, Ardissone et el., 2021, Ogawa et al., 2020, Lee et al., 2020, Sofou et al., 2018). On the individual gene-level, combing these cohorts, *MT-ATP6* accounts for the largest proportion of LS cases (18%), followed by *SURF1* (15%), with all other defects accounting for <10% (see **Figure 1 b**). The variable distribution of defects across East Asian and European cohorts suggests that genetic defects may vary based on ethnicity or between populations, such as due to founder mutations as previously described for *ECHS1* and *GTPBP3* in Chinese patients (Sun et al., 2020, Yan et al., 2021). We, however, found only limited evidence for the contribution of founder variants to disease, with only five variants of potential founder status identified in our Chinese LS population. LS was primarily clinical defined in the pre-genome area, that, as we know today, comprises a large heterogeneous group of approximately 100 distinct rare genetic disorders. Accordingly, studies summarizing LS-associated clinical features, disease onset, and survival data on the cohort level show high variability given high dependency on the underlying genetic composition of the cohort and provide limited information on single gene defects. Thus, emphasizing the importance to define such features on the genetic defect-specific level. Moreover, knowledge of the prevalence of LS on the defect-level aids in directing the need for treatment development and in the appropriate design and selection of primary end points for clinical trials. Interestingly, amongst the defects reported in our cohort were four mitochondrial disease genes across six unrelated patients (*COA6, HSD17B10, L2HGDH,* and *TARS2*) not currently reported to be associated with LS, in addition to two inborn errors of metabolism (*ALG1* and *GAMT*) each reported in one patient presenting with the typical clinical and neuroradiological features of LS in the absence of mitochondrial dysfunction (see **Figure 3**). In all instances, the patient(s) manifested with the

phenotype typically reported for the genetic etiology of the disease, e.g., developmental delay, seizures, and liver dysfunction in *ALG1* (MIM 608540) and cardiomyopathy plus a mitochondrial complex IV defect in *COA6* (MIM 616501), in combination with clinical features of neurodegeneration and LS characteristic bilateral basal ganglia and/or brainstem lesions on MRI. Further, in 75% (6/8) of these patients, biochemical evidence of mitochondrial dysfunction was reported (elevated serum and/or CSF lactate). Given that we report *HSD17B10* and *TARS2* each in two unrelated patients, evidence for their associated with LS is strongest, demonstrating the ever-increasing phenotypic spectrum of mitochondrial diseases with the application of untargeted genomic sequencing approaches.

Remarkably, across the detected pathogenic mtDNA variants, the heteroplasmy level threshold for clinical manifestation in the LS patients was reported as low as 37% in blood. The most frequent mtDNA encoded gene implicated in LS in our cohort was *MT-ATP6*. In keeping with other large studies of *MT-ATP6* defects (Stendel et al., 2020), a >50% contribution of the m.9176T>C variant to the LS phenotype in *MT-ATP6* was found. Among these patients, we identified a number with overlapping features of NARP (neuropathy, ataxia, and retinitis pigmentosa), a second mitochondrial disease syndrome presentation commonly associated with *MT-ATP6* defects, by neuropathy (1/28), ataxia (10/28), and retinitis pigmentosa (1/28, in association with m.8993T>G), as reported by Ng et al., 2019. In general, LS is expected to manifest at a heteroplasmy level ≥90% for variants in *MT-ATP6*, however, we report four patients with heteroplasmy levels ranging from 66-88% in blood, indicating LS to also manifest at moderate heteroplasmy levels. Conversely, it is known from the literature that there is no threshold for expression of a phenotype with m.8993T>C, m.9176T>C, and m.9185T>C variants in *MT-ATP6,* and homoplasmic asymptomatic carriers are reported (Stendel et

al., 2020, Ng et al., 2019, Ganetzky et al., 2019). In keeping with this, we report five asymptomatic maternal carriers of pathogenic mtDNA variants at heteroplasmy levels ≥90%.

One of the key diagnostic criteria for LS is an indicator of impaired energy metabolism (Rahman et al., 1996). Elevated serum lactate is a readily measurable indicator in clinical practice. However, like previously described cohorts with reportedly normal serum/CSF lactate in 25%-30% of cases (Lim et al., 2021, Ardissone et al., 2021, Sofou et al 2014), in just 74% of investigated patients in our LS cohort the serum lactate was elevated, and when measured on more than one visit in the same patient, discrepancies arose in almost 20%. This underlines the difficulty to appropriately interpret serum lactate, given the risk of missing a transient elevation in LS, and reporting a nonspecific elevation in an acutely unwell, a struggling child, or due to firm tourniquet use. This renders identification of the molecular diagnosis in combination with characteristic neuroimaging and clinical features the most accurate method of LS definition, as reflected by an amendment to the LS diagnostic criteria to allow the identification of disease-causing variants in a gene related to mitochondrial energy generation as an alternative to biochemical measurement of dysfunction by Lake et al., 2016. Importantly, defects in a number of genes linked to LS, such as *MT-ATP6* and *ECHS1* (Stendel et al., 2020, Marti-Sanchez et al., 2021), can present with a different clinical picture to LS and future clinical trials for defect-specific treatments may need to be stratified by the molecular basis and the clinical presentation.

To date, few large cohort studies of genotype-phenotype correlations have been conducted (Sofou et al., 2018, Alves et al., 2020) and this is, to our knowledge, the largest phenotype-genotype correlation study in LS to be performed on systematically evaluated patients. Consequently, characteristic disease-

associated clinical features have only been described for a small number of LS-associated disease genes in the past, such as hypertrichosis in association with pathogenic variants in *SURF1* (Lake et al., 2016), also recapitulated in our cohort, and methylmalonic aciduria in *SUCLA2* (Sofou et al., 2018, Carrozzo et al., 2016, Ostergaard et al., 2007), as reported for *SUCLG1* in our cohort, two defects resulting in Succinate-CoA ligase deficiency. Facilitated by the size of our molecularly defined cohort, we were able to expand the phenotype associations of *SURF1*, and report further significant associations for *MT-ATP6*, *MT-ND5*, ~~*NDUFAF6,*~~ and *PDHA1* (see **Table 2**), important for recognition of disease. ~~This is exemplified by cardiomyopathy, reported in six of 11 patients with *MT-ND5* defects, previously only reported in a single patient amongst a cohort of 20 patients with *MT-ND5* defects (Ng et al., 2018) and a number of case reports (Wesół-Kucharska et al., 2021, Zhou et al., 2019). This knowledge is important for implementing preventative measures and anticipatory care in routine clinical practice (e.g., echocardiogram surveillance) (Quadir et al., 2019).~~

As required for the diagnosis of LS, all patients displayed characteristic bilateral lesions in the basal ganglia (77%) or brainstem (73%), or both (49%) on MRI and/or DWI, in addition to less frequently reported neuroradiological abnormalities such as cerebral atrophy (34%), white matter involvement (19%), and cerebellar lesions (17%). In search of brain MRI features indicative of specific molecular genetic diagnoses, we found *MT-ND5* to be ~~associated with involvement of the thalamus and to be~~ depleted for basal ganglia lesions, and *SURF1* to be associated with medulla oblongata lesions and cerebellar involvement, and depleted for basal ganglia lesions, associations we have previously described in a subset of 139 patients from our cohort (Stenton et al., 2020) and partially supported in an independent cohort of 53 LS brain MRI datasets (Alves et al., 2020), amongst other smaller studies (Ng et al., 2018). Importantly, we found regression of MRI abnormalities to invariably result in a

corresponding clinical improvement in the patient, and enrichment for progression of the MRI lesions in patients with clinical deterioration resulting in death, indicating radiological features to have the potential to serve as therapeutic biomarkers in future clinical trials.

The present study reports the survival data of Chinese patients with LS for the first time. LS survival is generally considered to be poor, with death reportedly occurring in the first few years of life (Rahman et al., 1996, Lake et al., 2015). Encouragingly, the overall survival time in our cohort was higher than expected, with 109 patients alive at 5 years of age or older, 85% of patients at risk alive at or beyond 3 years of age, and 63% of patients reported as clinically improving or stable at a median of 48 months follow-up from disease onset. These higher-than-expected survival rates are supported by two further studies collectively analyzing almost 300 patients and reporting survival rates of >75% at or beyond 3 years of age (Sofou et al., 2014, Ogawa et al., 2020). Survival was found to be defect-specific, with poorest survival associated with defects in *MT-ND5, MT-ATP6* (m.8993T>C and m.9176T>C), *SURF1*, and *ALDH5A1*. These findings are largely in agreement with earlier studies (Ogawa et al., 2020), with the exception of *SURF1* for which reports are conflicting. *SURF1* has generally been considered to have a more favorable survival in comparison to other causes of LS. A pooled analysis of 142 *SURF1* patients by Wedatilake et al., 2014 reported a median survival of 5.4 years and approximate 75% 3 year survival, figures also supported by smaller studies (Ogawa et al., 2020). We, however, report a 3 year survival of just 44% and found *SURF1* to be amongst the most unfavorable defects. This finding is supported by the most recent LS cohort studies by Ardissone et al., 2021 and Lim et al., 2021, reporting *SURF1* to be the most severe defect across LS patients in the Italian Collaborative Network of Mitochondrial Diseases and the Mitochondrial Disease Patient Cohort (MitoCohort) UK, respectively. One potential explanation for the discrepancy in *SURF1*

survival across cohorts is the near absence of overlap in the reported underlying variants, with just three of over 100 unique *SURF1* variants found in common between our cohort and the Wedatilake et al., 2014 pooled study. Overall, in our study, the most favorable survival was found in association with defects in *ECHS1* and *SLC19A3*. Investigating predictors of poor survival, we found dyspnea to be significant after stringent multiple testing correction. Dyspnea may be a sign of impending respiratory failure, a typical cause of death in LS brainstem dysfunction in combination with muscle weakness (Sofou et al., 2014, Finsterer 2008). Cardiomyopathy has also previously been associated with poor survival more generally in mitochondrial disease (Holmgren et al., 2003, Imai-Okazaki et al., 2019). Cardiomyopathy was reported in 25 patients of our LS patient (12%) and demonstrated a significant association to *MT-ND5* defects, notably, the molecular diagnosis with the poorest survival rate (see **Figure 4 c**).

Across the study cohort, just 36 patients received defect-specific treatment. Thereby, with the exception of patients with *PDHA1* and *SLC19A3* defects, where the vast majority were treated with thiamine, and *HIBCH* and *ECHS1* where all were treated with a valine-restricted diet, the defect-specific survival data presented are not influenced by treatment. In these patients, the provision of defect-specific treatment was shown to afford clinical improvement by a significant improvement in survival (see **Figure 5 d**). This is supported by established reports of thiamine responsivity in *PDHA1* (van Dongen et al., 2014) and *SLC19A3* (Wang et al., 2021) defect and positive response to valine-restricted diet in *ECHS1* defect (Wang et al., 2021) in the literature, though may also reflect a milder course of disease associated with these defects as we do not have sufficient numbers of untreated patients with these specific defects to directly compare to. These molecular diagnoses exemplify three of the numerous medically actionable LS-defects with defect-specific potentially disease-modifying

treatments, the development of which is only made possible by knowledge of the disease pathomechanism. Further examples include, but are not limited to, the beneficial effects of ketogenic diet for pyruvate dehydrogenase (PDH) deficiency (Sofou et al., 2017), triheptanoin for *ECHS1* defect (Engelstad et al., 2021), and the mTOR inhibitor rapamycin for specific causes of LS, such as *NDUFS4* defect (Sage-Schwaede et al., 2019, Barriocanal-Casado et al., 2019).

To conclude, our data pave the way to defining the phenotype, onset, and survival data of LS in a defect-specific manner, providing indispensable knowledge on genetic defect-level needed for assessing the efficacy of experimental treatments in clinical trials, and adding considerably to existing knowledge concerning features associated with specific gene defects or predictive of disease outcome. The emergence of such genotype-phenotype associations is only made possible by the high numbers of patients with shared genetic defects in our cohort and will continue to emerge in the future with further accumulation of patients, optimally in the setting of national (e.g., mitoC-NET in China) and international (GENOMIT, genomit.eu in Europe, Asia, and the USA) registries for detailed and standardized data collection. Prospectively, GENOMIT will provide a platform for common data acquisition for four national LS cohorts published to date, including this study (Lim et al., 2021, Ardissone et al., 2021, Ogawa et al., 2020). Diagnostic delay and limited systematic evaluation of treatments inevitably impacts on clinical outcome for patients with LS (Yung Tiet, et al., 2021), these data are thereby essential to disease recognition in expediting diagnosis, appropriate surveillance for associated complications, and targeted treatment, and in providing optimal genetic and prognostic counselling by the knowledge gained on disease course to ensure appropriate monitoring of the effect of defect-specific therapies, the next frontier of LS research.

Acknowledgement statement (including conflict of interest and funding sources)

This study was supported by the German BMBF and Horizon2020 through the ERA PerMed project PerMiM (01KU2016A), the European Joint Programme on Rare Diseases project GENOMIT (01GM1920A), the Capital Health Development Research Foundation Project (2018-2-2096), the National Natural Science Foundation of China (81541115), the 2019 High-end Foreign Expert Introduction Program (G20190101010), and the 2020 High-end Foreign Expert Introduction Program (G20200101021)

Author contributions

Conceived and supervised the study, FF, HP; analyzed and interpreted results, SLS, YZ, HC; provided essential materials, all authors; wrote the manuscript, SLS and HP; edited manuscript, all authors.

Potential conflict of interest

Nothing to report.

REFERENCES

Alves, C. A., Teixeira, S. R., Martin-Saavedra, J. S., Guimarães Gonçalves, F., Lo Russo, F., Muraresku, C., ... & Zuccoli, G. (2020). Pediatric leigh syndrome: neuroimaging features and genetic correlations. Annals of neurology, 88(2), 218-232.

Barriocanal-Casado, E., Hidalgo-Gutiérrez, A., Raimundo, N., González-García, P., Acuña-Castroviejo, D., Escames, G., & López, L. C. (2019). Rapamycin administration is not a valid therapeutic strategy for every case of mitochondrial disease. EBioMedicine, 42, 511-523.

Bonfante, E., Koenig, M. K., Adejumo, R. B., Perinjelil, V., & Riascos, R. F. (2016). The neuroimaging of Leigh syndrome: case series and review of the literature. Pediatric radiology, 46(4), 443-451.

Borna, N. N., Kishita, Y., Kohda, M., Lim, S. C., Shimura, M., Wu, Y., ... & Okazaki, Y. (2019). Mitochondrial ribosomal protein PTCD3 mutations cause oxidative phosphorylation defects with Leigh syndrome. neurogenetics, 20(1), 9-25.

Carrozzo, R., Verrigni, D., Rasmussen, M., De Coo, R., Amartino, H., Bianchi, M., ... & Ostergaard, E. (2016). Succinate-CoA ligase deficiency due to mutations in SUCLA2 and SUCLG1: phenotype and genotype correlations in 71 patients. Journal of inherited metabolic disease, 39(2), 243-252.

Chakravarty, A., Mukherjee, A., & Sen, A. (2003). Familial pediatric rapidly progressive extrapyramidal syndrome: is it Hallervorden-Spatz disease?. Pediatric neurology, 29(2), 170-172.

Darin, N., Oldfors, A., Moslemi, A. R., Holme, E., & Tulinius, M. (2001). The incidence of mitochondrial encephalomyopathies in childhood: clinical features and morphological, biochemical, and DNA abnormalities. Annals of neurology, 49(3), 377-383.

Engelstad, K., Salazar, R., Koenigsberger, D., Stackowtiz, E., Brodlie, S., Brandabur, M., & De Vivo, D. C. (2021). Exploring triheptanoin as treatment for short chain enoyl CoA hydratase deficiency. Annals of Clinical and Translational Neurology, 8(5), 1151-1157.

Fang, F., Shen, Y., Shen, D. M., Liu, Z. M., Ding, C. H., Zhang, W. C., ... & Wu, H. S. (2017). Clinical and genetic characteristics of children with Leigh syndrome. Zhonghua er ke za zhi= Chinese Journal of Pediatrics, 55(3), 205-209.

Finsterer, J. (2008). Leigh and Leigh-like syndrome in children and adults. Pediatric neurology, 39(4), 223-235.

Ganetzky, R. D., Stendel, C., McCormick, E. M., Zolkipli-Cunningham, Z., Goldstein, A. C., Klopstock, T., & Falk, M. J. (2019). MT-ATP6 mitochondrial disease variants: phenotypic and biochemical features analysis in 218 published cases and cohort of 14 new cases. Human mutation, 40(5), 499-515.

Holmgren, D., Wahlander, H., Eriksson, B. O., Oldfors, A., Holme, E., & Tulinius, M. (2003). Cardiomyopathy in children with mitochondrial disease: clinical course and cardiological findings. European heart journal, 24(3), 280-288.

Hu, C., Li, X., Zhao, L., Shi, Y., Zhou, S., Wu, B., & Wang, Y. (2020). Clinical and molecular characterization of pediatric mitochondrial disorders in south of China. European journal of medical genetics, 63(8), 103898.

Imai-Okazaki, A., Kishita, Y., Kohda, M., Mizuno, Y., Fushimi, T., Matsunaga, A., ... & Okazaki, Y. (2019). Cardiomyopathy in children with mitochondrial disease: prognosis and genetic background. International journal of cardiology, 279, 115-121.

Kamps, R., Szklarczyk, R., Theunissen, T. E., Hellebrekers, D. M., Sallevelt, S. C., Boesten, I. B., ... & Smeets, H. J. (2018). Genetic defects in mtDNA-encoded protein translation cause pediatric,

mitochondrial cardiomyopathy with early-onset brain disease. European Journal of Human Genetics, 26(4), 537-551.

Lake, N. J., Bird, M. J., Isohanni, P., & Paetau, A. (2015). Leigh syndrome: neuropathology and pathogenesis. Journal of Neuropathology & Experimental Neurology, 74(6), 482-492.

Lake, N. J., Compton, A. G., Rahman, S., & Thorburn, D. R. (2016). Leigh syndrome: one disorder, more than 75 monogenic causes. Annals of neurology, 79(2), 190-203.

Lee, H. F., Tsai, C. R., Chi, C. S., Lee, H. J., & Chen, C. C. C. (2009). Leigh syndrome: clinical and neuroimaging follow-up. Pediatric neurology, 40(2), 88-93.

Lee, J. S., Yoo, T., Lee, M., Lee, Y., Jeon, E., Kim, S. Y., ... & Chae, J. H. (2020). Genetic heterogeneity in Leigh syndrome: highlighting treatable and novel genetic causes. Clinical genetics, 97(4), 586-594.

Marti-Sanchez, L., Baide-Mairena, H., Marcé-Grau, A., Pons, R., Skouma, A., López-Laso, E., ... & Pérez-Dueñas, B. (2021). Delineating the neurological phenotype in children with defects in the ECHS1 or HIBCH gene. Journal of Inherited Metabolic Disease, 44(2), 401-414.

Naess, K., Freyer, C., Bruhn, H., Wibom, R., Malm, G., Nennesmo, I., ... & Larsson, N. G. (2009). MtDNA mutations are a common cause of severe disease phenotypes in children with Leigh syndrome. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1787(5), 484-490.

Ng, Y. S., Lax, N. Z., Maddison, P., Alston, C. L., Blakely, E. L., Hepplewhite, P. D., ... & Gorman, G. S. (2018). MT-ND5 mutation exhibits highly variable neurological manifestations at low mutant load. EBioMedicine, 30, 86-93.

Ng, Y. S., Martikainen, M. H., Gorman, G. S., Blain, A., Bugiardini, E., Bunting, A., ... & McFarland, R. (2019). Pathogenic variants in MT-ATP6: A United Kingdom–based mitochondrial disease cohort study. Annals of neurology, 86(2), 310-315.

Ogawa, E., Fushimi, T., Ogawa-Tominaga, M., Shimura, M., Tajika, M., Ichimoto, K., ... & Murayama, K. (2020). Mortality of Japanese patients with Leigh syndrome: Effects of age at onset and genetic diagnosis. Journal of inherited metabolic disease, 43(4), 819-826.

Ogawa, E., Shimura, M., Fushimi, T., Tajika, M., Ichimoto, K., Matsunaga, A., ... & Murayama, K. (2017). Clinical validity of biochemical and molecular analysis in diagnosing Leigh syndrome: a study of 106 Japanese patients. Journal of Inherited Metabolic Disease: Official Journal of the Society for the Study of Inborn Errors of Metabolism, 40(5), 685-693.

Ostergaard, E., Hansen, F. J., Sorensen, N., Duno, M., Vissing, J., Larsen, P. L., ... & Schwartz, M. (2007). Mitochondrial encephalomyopathy with elevated methylmalonic acid is caused by SUCLA2 mutations. Brain, 130(3), 853-861.

~~Quadir, A., Pontifex, C. S., Robertson, H. L., Labos, C., & Pfeffer, G. (2019). Systematic review and meta-analysis of cardiac involvement in mitochondrial myopathy. Neurology Genetics, 5(4).~~

Rahman, J., Noronha, A., Thiele, I., & Rahman, S. (2017). Leigh map: a novel computational diagnostic resource for mitochondrial disease. Annals of neurology, 81(1), 9.

Rahman, S., Blok, R. B., Dahl, H. H., Danks, D. M., Kirby, D. M., Chow, C. W., ... & Thorburn, D. R. (1996). Leigh syndrome: clinical features and biochemical and DNA abnormalities. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 39(3), 343-351.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine, 17(5), 405-423.

Rubegni, A., Pasquariello, R., Dosi, C., Astrea, G., Canapicchi, R., Santorelli, F. M., & Nesti, C. (2019). Teaching NeuroImages: Leigh-like features expand the picture of PMPCA-related disorders. Neurology, 92(2), e168-e169.

Sage-Schwaede, A., Engelstad, K., Salazar, R., Curcio, A., Khandji, A., Garvin Jr, J. H., & De Vivo, D. C. (2019). Exploring mTOR inhibition as treatment for mitochondrial disease. Annals of clinical and translational neurology, 6(9), 1877-1881.

Schlieben, L. D., & Prokisch, H. (2020). The dimensions of primary mitochondrial disorders. Frontiers in Cell and Developmental Biology, 8.

Skladal, D., Halliday, J., & Thorburn, D. R. (2003). Minimum birth prevalence of mitochondrial respiratory chain disorders in children. Brain, 126(8), 1905-1912.

Sofou, K., De Coo, I. F., Isohanni, P., Ostergaard, E., Naess, K., De Meirleir, L., ... & Darin, N. (2014). A multicenter study on Leigh syndrome: disease course and predictors of survival. Orphanet journal of rare diseases, 9(1), 1-16.

Sofou, K., Dahlin, M., Hallböök, T., Lindefeldt, M., Viggedal, G., & Darin, N. (2017). Ketogenic diet in pyruvate dehydrogenase complex deficiency: short-and long-term outcomes. Journal of inherited metabolic disease, 40(2), 237-245.

Sofou, K., de Coo, I. F., Ostergaard, E., Isohanni, P., Naess, K., De Meirleir, L., ... & Darin, N. (2018). Phenotype-genotype correlations in Leigh syndrome: new insights from a multicentre study of 96 patients. Journal of medical genetics, 55(1), 21-27.

Stendel, C., Neuhofer, C., Floride, E., Yuqing, S., Ganetzky, R. D., Park, J., ... & ATP6 Study Group. (2020). Delineating MT-ATP6-associated disease: From isolated neuropathy to early onset neurodegeneration. Neurology Genetics, 6(1).

Stenton, S. L., Shimura, M., Piekutowska-Abramczuk, D., Freisinger, P., Distelmaier, F., Mayr, J., ... & Prokisch, H. (2021). Diagnosing pediatric mitochondrial disease: lessons from 2,000 exomes. medRxiv.

Stenton, S. L., Zou, Y., Cheng, H., Prokisch, H., & Fang, F. (2020). Pediatric Leigh Syndrome: Neuroimaging Features and Genetic Correlations. Annals of Neurology.

Stenton, S. L., & Prokisch, H. (2020). Genetics of mitochondrial diseases: Identifying mutations to help diagnosis. EBioMedicine, 56, 102784.

Sun, D., Liu, Z., Liu, Y., Wu, M., Fang, F., Deng, X., ... & Zhu, Y. (2020). Novel ECHS1 mutations in Leigh syndrome identified by whole-exome sequencing in five Chinese families: case report. BMC Medical Genetics, 21(1), 1-12.

Tian, X. J., Li, X., Fang, F., Liu, Z. M., Wu, W. J., Liu, K., & Sun, S. Z. (2021). Molybdenum cofactor deficiency type B manifested as Leigh-like syndrome: a case report and literature review. Zhonghua er ke za zhi= Chinese Journal of Pediatrics, 59(2), 119-124.

van Dongen, S., Brown, R. M., Brown, G. K., Thorburn, D. R., & Boneh, A. (2014). Thiamine-responsive and non-responsive patients with PDHC-E1 deficiency: a retrospective assessment. In JIMD Reports, Volume 15 (pp. 13-27). Springer, Berlin, Heidelberg.

Vögtle, F. N., Brändl, B., Larson, A., Pendziwiat, M., Friederich, M. W., White, S. M., ... & Helbig, I. (2018). Mutations in PMPCB encoding the catalytic subunit of the mitochondrial presequence protease cause neurodegeneration in early childhood. The American Journal of Human Genetics, 102(4), 557-573.

Wang, J., Wang, J., Han, X., Liu, Z., Ma, Y., Chen, G., ... & Fang, F. (2021). Report of the largest Chinese cohort with SLC19A3 gene defect and literature review. Frontiers in Genetics, 12, 1138.

Wang, J., Liu, Z., Xu, M., Han, X., Ren, C., Yang, X., ... & Fang, F. (2021). Cinical, Metabolic, and Genetic Analysis and Follow-Up of Eight Patients With HIBCH Mutations Presenting With Leigh/Leigh-Like Syndrome. Frontiers in pharmacology, 12.

Wedatilake, Y., Brown, R. M., McFarland, R., Yaplito-Lee, J., Morris, A. A., Champion, M., ... & Rahman, S. (2013). SURF1 deficiency: a multi-centre natural history study. Orphanet journal of rare diseases, 8(1), 1-13.

~~Wesół-Kucharska, D., Greczan, M., Witulska, K., Piekutowska-Abramczuk, D., Ciara, E., Kowalski, P., & Rokicki, D. (2021). Improvement of cardiomyopathy after ketogenic diet in a patient with Leigh syndrome caused by MTND5 mutation.~~

Yan, H. M., Liu, Z. M., Cao, B., Zhang, V. W., He, Y. D., Jia, Z. J., ... & Wang, H. (2021). Novel Mutations in the GTPBP3 Gene for Mitochondrial Disease and Characteristics of Related Phenotypic Spectrum: The First Three Cases From China. Frontiers in Genetics, 12, 933.

Yung Tiet, M., Lin, Z., Gao, F., Jennings, M. J., & Horvath, R. Targeted Therapies for Leigh Syndrome: Systematic Review and Steps Towards a 'Treatabolome'. Journal of Neuromuscular Diseases, (Preprint), 1-13.

~~Zhou, N., Tang, L., Jiang, Y., Qin, S., Cui, J., Wang, Y., ... & Shu, X. (2019). Whole-exome sequencing reveals a novel mutation of MT-ND5 gene in a mitochondrial cardiomyopathy pedigree: Patients who show biventricular hypertrophy, hyperlactacidemia, pulmonary hypertension, and decreased exercise tolerance. Anatolian journal of cardiology, 21(1), 18.~~

FIGURE LEGENDS

Figure 1. Genetic underpinning of LS. A, Frequency of molecular genetic diagnoses in our cohort indicating the proportion of molecular genetic diagnoses by inheritance mode. **B,** Frequency of molecular genetic diagnoses in LS patients reported by six large LS cohort studies to date across Europe and East Asia: this study (n = 209), Sofou et al., 2018 (n = 96), Ogawa et al., 2020 (n = 103), and Lee et al., 2020 (n = 41), Ardissone et al., 2021 (n = 110), and Lim et al., 2021 (n=59). Genes defects described in ≥5 patient are displayed. **C,** Frequency of molecular diagnoses displayed in **B** in the pediatric mitochondrial disease HPO-genotype association database GENOMITexplorer (prokischlab.github.io/GENOMITexplorer/) indicating the proportion of patients with disease-causing variants in these genes manifesting with LS. LS, Leigh syndrome; AR, autosomal recessive; comp het, compound heterozygous; hom, homozygous; XLD, X-linked dominant; MD, mitochondrial disease.

Figure 2. Phenotypic spectrum of LS. Frequency of **A,** system-level and **B,** individual phenotype-level HPO terms for all phenotypes reported in ≥5 patients. **C,** Distribution of age of onset for initial symptoms reported at onset in ≥5 patients. Individual patient ages are displayed with the median. The lower and upper hinges correspond to the 25th and 75th percentiles. CSF, cerebrospinal fluid; MRS, magnetic resonance spectroscopy; CK, creatine kinase.

Figure 3. Characteristic LS lesions on MRI in disease genes newly reported to be associated with LS or presenting with LS-mimicking disease. A, T2: lesions in the subthalamus. **B,** DWI: lesions in

the subthalamus. **C,** T2FLAIR: lesions in the ventral medulla (brainstem). **D,** DWI: lesions in the ventral medulla (brainstem). **E,** T2: lesions in bilateral pallidum of the lenticular nucleus (basal ganglia). **F,** DWI: lesions in the bilateral pallidum of the lenticular nucleus (basal ganglia). **G,** T2: lesions in pons of the midbrain (brainstem). **H,** DWI: lesions in the pons of the midbrain (brainstem). **I-J,** T2: lesions in the basal ganglia. **K-L,** DWI: lesions in the basal ganglia. **M,** T2: lesions in the basal ganglia. **N,** T2: long signal in the near subcortical region. **O,** DWI: lesions in the basal ganglia. **P,** T2: long signal in the dentate nucleus (cerebellum). **Q-R,** T2: lesions on the cerebral peduncles and peri-aqueduct of the midbrain (brainstem). **S-T,** T2: lesions in the basal ganglia.

Figure 4. Defect-specific phenotypes, onset, disease course, and survival. Principal component analysis for HPO terms reported in ≥10 patients (n = 58) for patients with P/LP variants in one of 18 genetic diagnoses with ≥3 patients in the cohort, colored by **A,** genetic diagnosis and **B,** location of MRI lesions. **C,** Distribution of the age of onset and age at last follow-up for all genetic diagnoses with ≥3 patients in the cohort, stratified by alive or deceased at last report. Individual patient ages are displayed with the median. The lower and upper hinges correspond to the 25th and 75th percentiles. The 1 year and 3 year survival rates are indicated with the number at risk in parentheses and by the survival curve in the top corner of each display. BG, basal ganglia; BS, brainstem. PC, principal component.

Figure 5. LS survival and significant features determining survival. A, Kaplan-Meier survival curve for the entire study cohort of 209 LS patients. Kaplan-Meier survival curve for comparing LS patients with and without **B,** dyspnea (HR = 2.77, p = 0.02, Log rank test), **C,** dysarthria (HR = 0.00,

p = 0.001, Log rank test), and **D,** receival of a defect-specific treatment (HR = 0.18, p = 0.0069, Log rank test). Patients were censored when last seen alive at follow-up.

Figure 6. LS MRI changes over time. A, Regression of T2FLAIR MRI in a *MT-ND4* patient with lesions in the cerebral peduncles of the midbrain (brainstem) completely resolving between 4 and 7 years of age. **B,** Stable T2 MRI and DWI in a *MECR* patient with lesions in the basal ganglia at 7 and 12 years of age. **C,** Evolution of T2 MRI and DWI in a *ECHS1* patient with lesions in the basal ganglia developing cystic change (encephalomalacia) 5 and 11 months of age. **D,** Progression of T2 MRI and DWI in a *SURF1* patient with lesions in the midbrain (brainstem) rapidly increasing in size in just 19 days, demonstrating how rapidly LS can aggravate.