FEBS Letters

FEBS PRESS science publishing by scientists

REVIEW ARTICLE

Genetic basis of mitochondrial diseases

Mirjana Gusic[1,2,3] and Holger Prokisch[1,2]

1 Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany
2 Institute of Human Genetics, Technical University of Munich, Germany
3 DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Germany

Correspondence
H. Prokisch, Institute of Neurogenomics, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764, Neuherberg, Germany
Tel: +49-89-3187-2890
E-mail:prokisch@helmholtzmuenchen.de

(Received 25 January 2021, revised 17 February 2021, accepted 18 February 2021)

doi:10.1002/1873-3468.14068

Edited by Peter Rehling

Mitochondrial disorders are monogenic disorders characterized by a defect in oxidative phosphorylation and caused by pathogenic variants in one of over 340 different genes. The implementation of whole-exome sequencing has led to a revolution in their diagnosis, duplicated the number of associated disease genes, and significantly increased the diagnosed fraction. However, the genetic etiology of a substantial fraction of patients exhibiting mitochondrial disorders remains unknown, highlighting limitations in variant detection and interpretation, which calls for improved computational and DNA sequencing methods, as well as the addition of OMICS tools. More intriguingly, this also suggests that some pathogenic variants lie outside of the protein-coding genes and that the mechanisms beyond the Mendelian inheritance and the mtDNA are of relevance. This review covers the current status of the genetic basis of mitochondrial diseases, discusses current challenges and perspectives, and explores the contribution of factors beyond the protein-coding regions and monogenic inheritance in the expansion of the genetic spectrum of disease.

Keywords: genetics; diagnostics; mitochondrial disease; variants; multiomics; noncoding; oxidative phosphorylation; metabolic disorders

Mitochondrial diseases are characterized by dysfunctional mitochondria, caused primarily by defects in oxidative phosphorylation (OXPHOS) or other essential mitochondrial functions [1,2]. However, it is noteworthy that there is still no common ground concerning the biochemical criteria employed to define such diseases. The prevalence of mitochondrial disease is estimated at 5–20 in 100 000 [1], but these values are based on clinical diagnosis and can be biased toward patients with clear symptoms, thus overlooking patients with atypical or unspecific phenotype. A lifetime risk for a nuclear-encoded recessive mitochondrial disease has been calculated at almost 1 in 2000 [2]. This makes it one of the most common and diverse groups of metabolic disorders. In line with the lack of a clear definition, mitochondrial diseases show extreme heterogeneity in both clinical presentation and molecular cause and can arise with 'any symptom, in any organ or tissue, at any age, with any mode of inheritance' [3]. Such heterogeneity hampers both their diagnosis and management, and many patients remain in a diagnostic odyssey, visiting different clinicians, repeating sometimes invasive tests, and even receiving a false or conflicting diagnosis. Besides cofactor deficiencies [4], there are still no curative treatments for mitochondrial disorders, and the treatment strategies are usually

Abbreviations
AD, autosomal dominant; AR, autosomal recessive; CNV, copy number variation; GWAS, genome-wide association study; LHON, Leber's hereditary optic neuropathy; lncRNA, long noncoding RNA; LRS, long-read sequencing; miRNA, microRNA; ncRNA, noncoding RNA; NGS, next-generation sequencing technologies; ORF, open reading frame; OXPHOS, oxidative phosphorylation; RNA-seq, RNA-sequencing; scRNA-seq, single-cell RNA-seq; SV, structural variant; uORF, upstream open reading frame; WES, whole exome sequencing; WGS, whole genome sequencing.

based on easing the clinical symptoms, and reducing morbidity and mortality [5]. In recent years, we have witnessed advent of personalized medicine and targeted therapies, which go hand-in-hand with molecular diagnostics. Thus, knowing the genetic cause of the disease is becoming not only a matter of diagnostic but also therapeutical relevance. The classic diagnostics approaches in mitochondrial disorders are based on an extensive characterization of clinical presentations, as well as the biochemical and histochemical evaluation of the OXPHOS enzymes, followed by the sequencing of candidate genes. However, slowly but surely, such approaches are becoming replaced with more systematic, 'genetics-first' ones [6]. Indeed, in the last two decades, we have witnessed considerable improvements in understanding of the molecular basis of mitochondrial disorders. The biggest contribution to these advances belongs to the next-generation sequencing technologies (NGS), the application of which in clinical genetics increased the diagnostic yield and accelerated discoveries of novel disease genes. Despite such improvements, a large fraction of pediatric patients with a suspected mitochondrial disease remains undiagnosed after DNA sequencing [7]. This diagnostic gap calls for not only further improvements in sequencing methods and analysis, but also a shift in focus outside the protein-coding genes and monogenic inheritance.

Here, we cover the genetic status of mitochondrial disorders, from the history to the current standards in diagnostics. We then discuss the limitations and challenges in the field and explore current methodological and computational efforts that are being developed to address them. Next, we move away from the conventional protein-centric approach and delve into the non-coding DNA regions and hidden micropeptides. Finally, we touch on factors beyond the monogenic inheritance that could have significant effects on the phenotypic presentation. Considering all these factors and the application of novel methodologies will create a more complete genetic architecture of mitochondrial disorders, and pave the way towards the improved diagnostics and the development of targeted therapies.

Diagnostics of mitochondrial disorders: current status

Suspecting a mitochondrial disease: a multidisciplinary challenge

Traditional diagnosis of a mitochondrial disorder is based on extensive characterization of clinical presentations and laboratory evaluations in which the biochemical and histochemical examinations of mitochondrial function often play a central role [8]. These findings are then followed by a genetic testing [6]. Concerning the clinical diagnosis, some patients may present with unique combinations of features that can be grouped into a discrete clinical syndrome [1], in turn guiding the genetic analysis (Fig. 1 left). As this applies only for a minority of disorders, it has been suggested to clinically suspect a mitochondrial disorder when seemingly unrelated organs are simultaneously affected [9]. For further elucidation of the disease, the biochemical and histochemical evaluations of tissue biopsies have conventionally been the first steps in the diagnostic algorithm [6,8]. These include measurements of OXPHOS enzyme activities and immunohistochemical and histoenzymatic assays following muscle biopsy, as well as measurement of lactate, pyruvate, amino acids, and organic acids levels in blood, urine, and spinal fluid [6]. However, these measurements often show low specificity and sensitivity, and are without standardized guidelines [8]. Thus, although they can indicate a mitochondrial disease, they cannot rule it

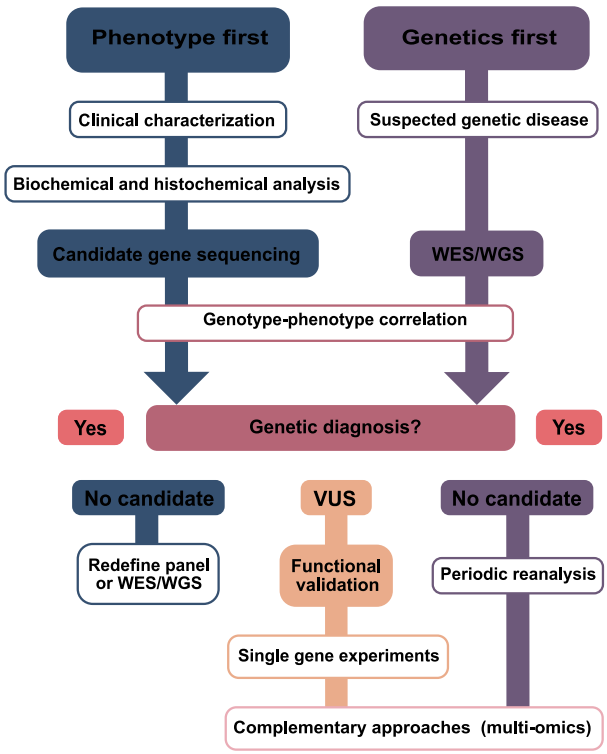

Fig. 1. Diagnostic approaches for mitochondrial disorders. Schematic showing two approaches to the molecular diagnostics of mitochondrial disorders ('the phenotype first' on the left and 'the genetics first' on the right) with the suggested follow-up steps upon the negative or inconclusive genetic finding.

out, especially in the bodily tissues that are not investigated. Finally, observed defects can be due to secondary involvement of mitochondria in other diseases [10]. For the purpose of facilitating diagnosis, mitochondrial disease criteria have been developed [11,12]. These clinical scoring systems indicate the probability of mitochondrial disease but cannot differentiate between primary and secondary OXPHOS impairment or pinpoint any precise genetic causes [12,13]. Altogether, it is clear that this approach, although thorough, is costly, time-consuming, and sadly often not reliable and conclusive. For these reasons, the 'genetics-first' approach, coupled with both clinical and biochemical investigations, is becoming more widely adopted (Fig. 1 right): This approach not only helps diagnosing mitochondrial disorders that may have been overlooked clinically and/or biochemically, but also enables the identification of mitochondrial disease phenocopies, other genetic disease, and multiple genetic disorders in a given patient [14].

Mitochondrial disease genes

In the emerging era of precision medicine, definite genetic diagnosis is becoming a cornerstone of safe medical practice [15]. For mitochondrial diseases, molecular diagnosis is challenging due to the heterogeneity of disease on both clinical and genetic levels. To start with, the phenotype–genotype association is often not straightforward. Variants in different genes can cause the same phenotype. A striking example is Leigh syndrome, with which variants in more than 75 genes are associated and this number is still in expansion [16]. Vice versa, variants in the same gene, even the same variant, can cause different clinical symptoms. This is most pronounced not only in pathogenic mtDNA point mutations [17], but also in nuclear genes encoding for proteins involved in the mtDNA synthesis and maintenance [18]. For example, numerous mutations in *POLG* can result in a spectrum of clinical manifestations, recessive or dominant mode of inheritance, and age of onset from the neonatal period to late adulthood [19].

As outlined in Schlieben and Prokisch [20], the number of genes associated with mitochondrial disease is dynamic and varies dependent on the criteria applied, which encompass clinical, genetic, biochemical, functional, and protein localization. The numbers vary between 270 disease genes, which fulfill all those criteria, and more than 400 fulfilling some of the criteria. Within the GENOMIT project (http://genomit.eu/), we compiled a list currently encompassing 343 genes, for which pathogenic variants have been associated with a defective OXPHOS, involved in a variety of mitochondrial functions (Fig. 2) [20]. For a better understanding of disease genetics, several distinct properties should be considered. First, due to the dual genetic origin of the mitochondrial proteome, pathogenic variants can reside in both nuclear DNA and mitochondrial DNA. Indeed, out of ~ 1200 mitochondrial proteins, only 13 are encoded by the mtDNA, the rest is encoded by the nuclear genome and imported [21,22]. Pathogenic variants in the mtDNA are responsible for ~ 80% of adult-onset cases of mitochondrial disease, and ~ 20% of pediatric ones [1]. Second, mitochondrial diseases exhibit any mode of inheritance: maternal, X-linked, autosomal recessive (AR), autosomal dominant (AD), and *de novo*. As the vast majority (89%) of associated disease genes are nuclear, most of them are inherited in an AR fashion, but interestingly, for variants in 23 genes, diseases are inherited as both dominant and recessive, arguing for distinct molecular pathomechanisms [18] (Fig. 2). For example, dominant and often *de novo* missense variants in *SSBP1*, encoding for a mitochondrial single-strand binding protein that binds and stabilizes the mtDNA during replication, lead to severe early-onset optic atrophy, whereas a homozygous variant extended the phenotype with cardiomyopathy, nephropathy, ataxia, and growth retardation with recessive inheritance [23]. Although evaluating the recessive type of variants is straightforward, distinguishing between benign and dominant pathogenic variants is much more challenging. Finally, the assessment of the mtDNA variants is complicated due to characteristics of the mitochondrial genome such as variable copy number in cell and high mutation rate [24], and because a variant can be present in all (homoplasmy) or a proportion of mtDNA molecules (heteroplasmy). For the biochemical defect to occur, a certain percentage of pathogenic mtDNA molecules needs to be exceeded in a cell. This threshold is often tissue- and variant-specific but generally considered to be above 60% [25]. Moreover, heteroplasmy levels vary across cells and tissues, and change over time, making the phenotype–genotype association more challenging [26]. To date, hundreds of mtDNA variants have been reported to be associated with a disease, but only 95 have a confirmed pathogenic status, defined by strong evidence of the pathogenicity coming from several laboratories [27] (Table 1). This reflects the difficulties in interpretation of mtDNA variants (reviewed in Ref. [28]). Nevertheless, pathogenic mtDNA variants can be classified as point mutations and single large-scale rearrangements (deletions and insertions) [29,30]. It is noteworthy that multiple mtDNA deletions and

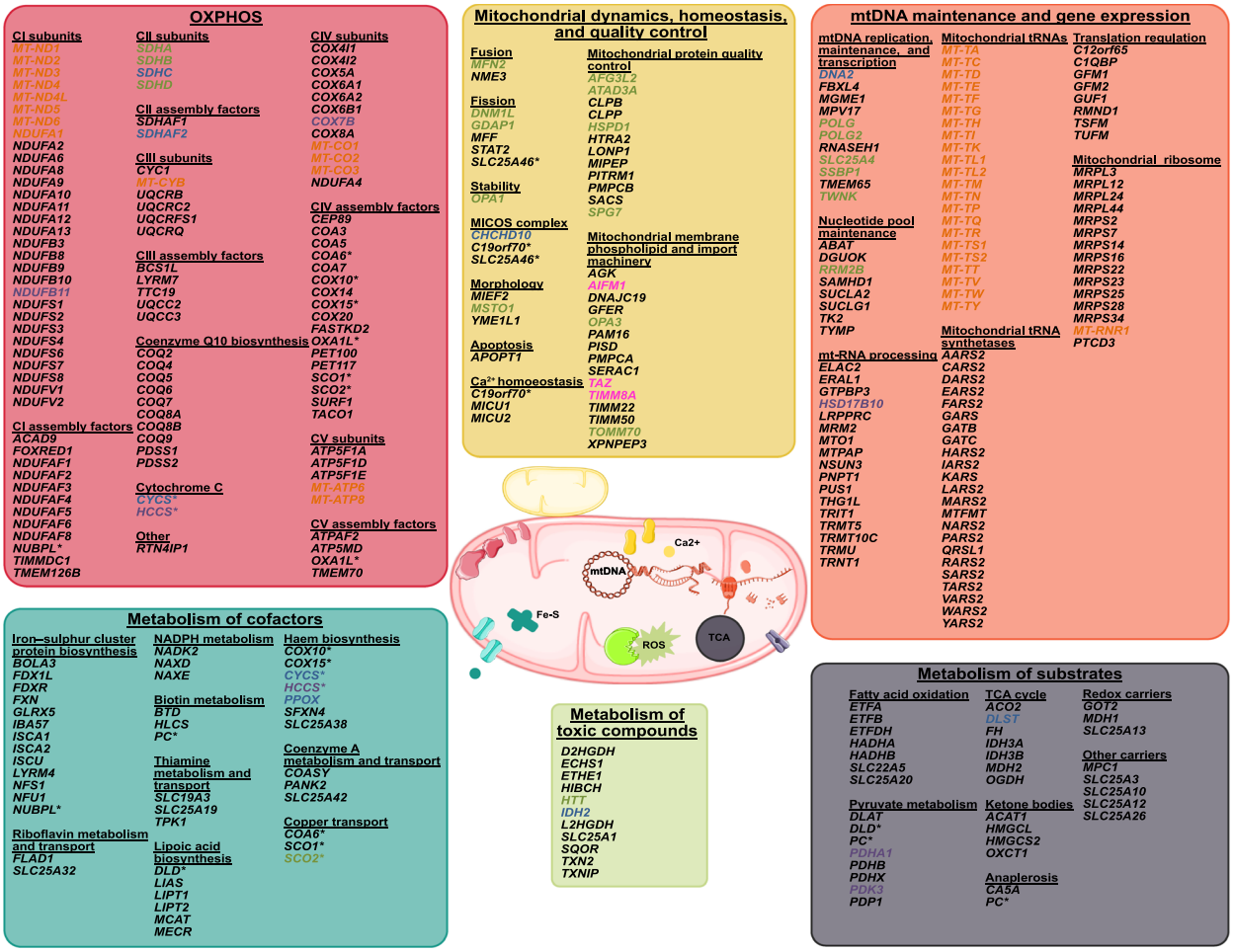

Fig. 2. Mitochondrial disease genes classified by their function. The 343 disease genes compiled in the GENOMIT consortium (http://genomit.eu/) are divided into six categories based on the functional roles of their encoding proteins. Genes that belong in more categories are indicated by an asterisk (*). Colors of letters indicate the mode of inheritance the genes were described with: AR in black (n = 266), maternal in orange (n = 36), AD in blue (n = 8), X-linked dominant in purple (n = 6), X-linked recessive in pink (n = 4), and a combination of AR and AD inheritance in green (n = 23).

depletions usually occur secondary, due to a primary defect in nuclear genes, typically with a role in the mtDNA replication and synthesis [31].

Discovery of pathogenic variants

In the last two decades, we have witnessed rapid developments of technologies for assessing genetic variation and gene regulation, even on a cellular level. The utility of these technologies was very quickly tested in clinical practice [32], including also diagnostics of mitochondrial disorders. To further support application of these technologies in clinical and basic research, publically available databases have been created, providing catalogs of functional annotation of the genome and genetic variation across populations, as well as clinical associations of genes and variants, and even patient records. These milestones are shown in Fig. 3.

Since the first discovery of mitochondrial disease-causing variant in the mtDNA in 1988 [61], technologies for genetic testing have evolved from the targeted mtDNA and candidate gene Sanger sequencing, to the more unbiased and systematic technologies based on the NGS [62] (summarized in Fig. 4). Although candidate gene and mtDNA sequencing remain fast and cost-effective methods for genetically and phenotypically well-defined syndromes, such as the Leber's hereditary optic neuropathy (LHON)* [62] (Box 1), the genetic heterogeneity of mitochondrial disorders, together with often unspecific biochemical and metabolic findings, makes the choice of feasible number of

BOX 1. LHON

LHON is a mitochondrial disorder characterized by degeneration of retinal ganglion cells (RGCs) and their axons that usually begins in adolescence or young adulthood, and ultimately leads to an acute or subacute loss of central vision. In more than 95% of nonsporadic cases, it is caused by three distinct point mtDNA variants at positions m. 11778, 3460, or 14484 [62].

Table 1. Pathogenic mtDNA variants with a confirmed status (as of January 2021). Table created in accordance with Ref. [27].

Index	Gene	Position and variant
1	*MT-ATP6*	m.8851T>C
2	*MT-ATP6*	m.8969G>A
3	*MT-ATP6*	m.8993T>C
4	*MT-ATP6*	m.8993T>G
5	*MT-ATP6*	m.9035T>C
6	*MT-ATP6*	m.9155A>G
7	*MT-ATP6*	m.9176T>C
8	*MT-ATP6*	m.9176T>G
9	*MT-ATP6*	m.9185T>C
10	*MT-ATP6*	m.9205_9206delTA
11	*MT-ATP8/6*	m.8528T>C
12	*MT-CO1*	m.7445A>G
13	*MT-CYB*	m.14849T>C
14	*MT-CYB*	m.15579A>G
15	*MT-ND1*	m.3376G>A
16	*MT-ND1*	m.3460G>A
17	*MT-ND1*	m.3635G>A
18	*MT-ND1*	m.3697G>A
19	*MT-ND1*	m.3700G>A
20	*MT-ND1*	m.3733G>A
21	*MT-ND1*	m.3890G>A
22	*MT-ND1*	m.3902_3908ACCTTGCinv
23	*MT-ND1*	m.4171C>A
24	*MT-ND3*	m.10158T>C
25	*MT-ND3*	m.10191T>C
26	*MT-ND3*	m.10197G>A
27	*MT-ND4L*	m.10663T>C
28	*MT-ND4*	m.11777C>A
29	*MT-ND4*	m.11778G>A
30	*MT-ND5*	m.12706T>C
31	*MT-ND5*	m.13042G>A
32	*MT-ND5*	m.13051G>A
33	*MT-ND5*	m.13094T>C
34	*MT-ND5*	m.13379A>C
35	*MT-ND5*	m.13513G>A
36	*MT-ND5*	m.13514A>G
37	*MT-ND6*	m.14459G>A
38	*MT-ND6*	m.14482C>A
39	*MT-ND6*	m.14482C>G
40	*MT-ND6*	m.14484T>C
41	*MT-ND6*	m.14487T>C
42	*MT-ND6*	m.14495A>G
43	*MT-ND6*	m.14568C>T
44	*MT-RNR1*	m.1494C>T
45	*MT-RNR1*	m.1555A>G
46	*MT-TA*	m.5650G>A
47	*MT-TE*	m.14674T>C
48	*MT-TE*	m.14709T>C
49	*MT-TE*	m.14710G>A
50	*MT-TF*	m.583G>A
51	*MT-TF*	m.616T>C
52	*MT-TG*	m.10010T>C
53	*MT-TH*	m.12147G>A
54	*MT-TH*	m.12201T>C
55	*MT-TI*	m.4298G>A
56	*MT-TI*	m.4300A>G
57	*MT-TI*	m.4308G>A
58	*MT-TK*	m.8306T>C
59	*MT-TK*	m.8313G>A
60	*MT-TK*	m.8340G>A
61	*MT-TK*	m.8344A>G
62	*MT-TK*	m.8356T>C
63	*MT-TK*	m.8363G>A
64	*MT-TL1*	m.3243A>G
65	*MT-TL1*	m.3243A>T
66	*MT-TL1*	m.3256C>T
67	*MT-TL1*	m.3258T>C
68	*MT-TL1*	m.3260A>G
69	*MT-TL1*	m.3271delT
70	*MT-TL1*	m.3271T>C
71	*MT-TL1*	m.3280A>G
72	*MT-TL1*	m.3291T>C
73	*MT-TL1*	m.3302A>G
74	*MT-TL1*	m.3303C>T
75	*MT-TL2*	m.12276G>A
76	*MT-TL2*	m.12294G>A
77	*MT-TL2*	m.12315G>A
78	*MT-TL2*	m.12316G>A
79	*MT-TM*	m.4450G>A
80	*MT-TN*	m.5690A>G
81	*MT-TN*	m.5703G>A
82	*MT-TN*	m.5728T>C
83	*MT-TP*	m.15990C>T
84	*MT-TQ*	m.4332G>A
85	*MT-TS1 precursor*	m.7445A>G
86	*MT-TS1*	m.7471_7472insC
87	*MT-TS1*	m.7497G>A
88	*MT-TS1*	m.7510T>C
89	*MT-TS1*	m.7511T>C
90	*MT-TS2*	m.12258C>A
91	*MT-TV*	m.1606G>A
92	*MT-TV*	m.1630A>G
93	*MT-TV*	m.1644G>A
94	*MT-TW*	m.5521G>A
95	*MT-TW*	m.5537_5538insT

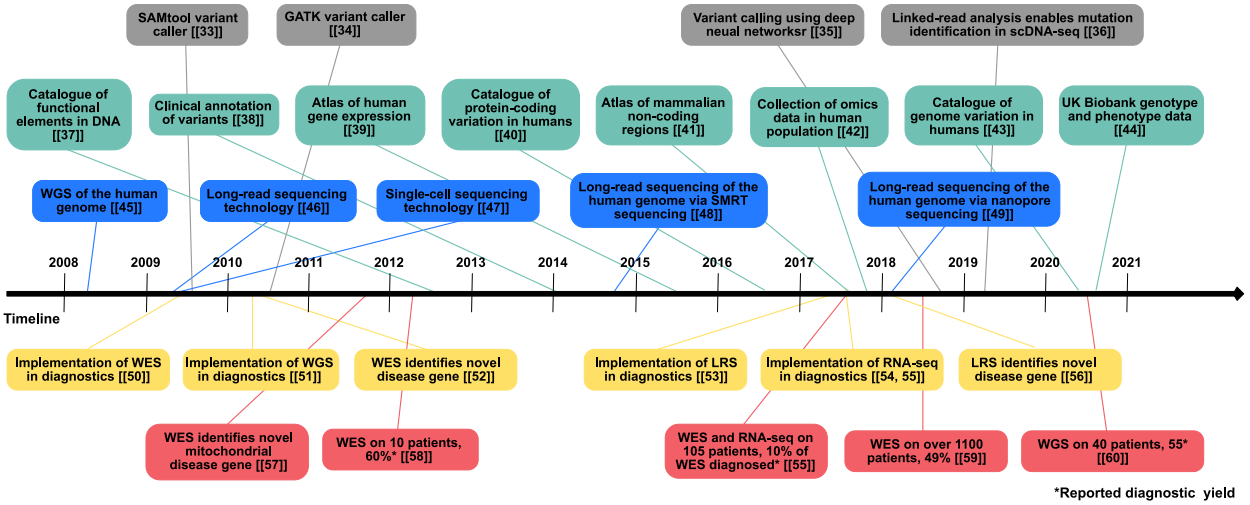

Fig. 3. Milestones in genetics and genetic diagnosis of mitochondrial disorders. Timeline of major computational, data, technological, and diagnostic milestones, as well as their application to the diagnostics of mitochondrial disorders.

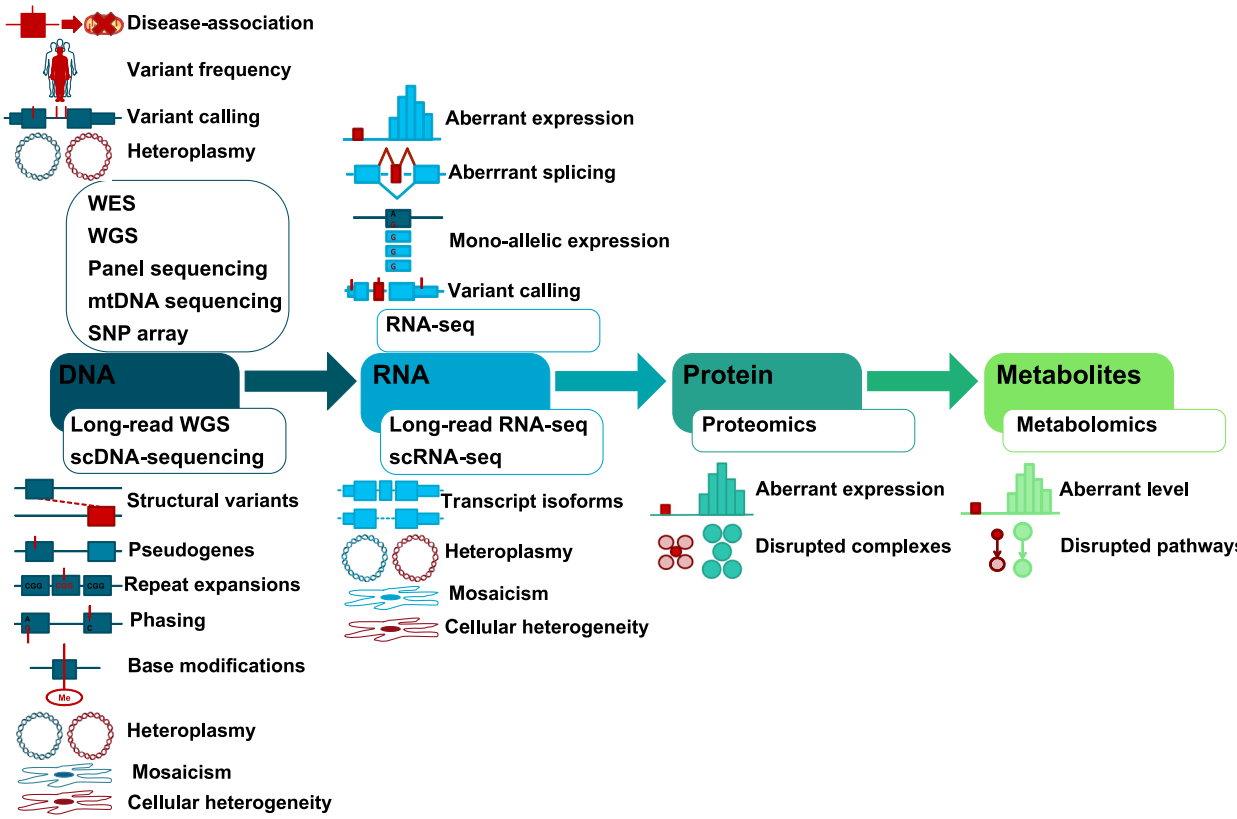

Fig. 4. Present and future OMICS approach for discovery of pathogenic variants. Based on the central dogma of molecular biology, the figure depicts the flow of genetic information and the OMICS technologies together with the observations that can be provided by each.

candidate genes difficult. Indeed, screening of 64 candidate genes established a diagnosis in just 11% of cases [63]. On the other hand, the NGS era, beginning with *ACAD9* in 2010 [57], caused a revolution in genetics of mitochondrial disease. Apart from increasing diagnostic rates and expanding the genotype–phenotype association, it accelerated discovery of novel disease genes, which is over 20 per year since 2012 [64]. Not to be overlooked, such impact comes also thanks to developments of bioinformatics tools

for sequence alignment, annotation, variant calling, and prioritization [65], as well as shared databases for variant curation and variant frequencies in population [38,43].

Starting with the more targeted approaches, application of NGS to sequence mtDNA is a routine first step in many diagnostic centers, especially for the cases with the adult onset and where phenotype is highly evocative of an mtDNA etiology. Apart from providing variant discovery, it also allows exact measurement of heteroplasmy levels [66]. Analysis is usually performed in blood and urine in pediatric-onset cases [67], but more commonly in muscle in adult-onset ones, as the affected tissue is most informative and causative variants may be undetected in blood due to tissue-specific heteroplasmy. This especially refers to progressive external ophthalmoplegia (PEO) and Kearns–Sayre syndrome, in which causative single large-scale mtDNA deletions are mostly affecting the postmitotic skeletal muscle [68]. Nevertheless, one must be cautious in interpreting negative results, as many pathogenic mtDNA variants are restricted to the affected tissues.

Expanding the diagnostic focus to the nuclear genes, NGS gene panels provide a targeted, deep sequencing of the predefined sets of mitochondrial disease genes, as well as candidate genes encoding for the proteins involved in essential mitochondrial function, whose disruption is thus likely to cause a disease. Available panels range from 100 genes associated with complex I efficiency to the 'MitoExome', targeting the predicted mitochondrial proteome [69]. The success rate of panels varies from 7% to 31% [69–74]. Although panels do offer advantages in the higher coverage of targeted regions, as well as easier data interpretation, the constant updates of reported disease genes, often low phenotype–genotype correlation, the inability to surly define a mitochondrial disease by clinical symptoms, and lower diagnostic yield compared to the whole exome sequencing (WES) have made the latter the more preferable choice.

In modern diagnostics, WES has become a desired first-tier tool of investigation, especially in the cases of early-onset mitochondrial disease, where the cause of disease likely lies in the nuclear DNA [6] and because it also allows the analysis of mtDNA in the given tissue [75]. Although WES captures only 2% of the genome that encodes proteins, this region includes up to 99% of reported pathogenic variants [38]. By revealing almost 9000 protein-affecting variants per individual [44], WES enables not only discovery of pathogenic variants, but also novel disease genes, making it an effective sequencing tool in diagnostics [6,76]. Within rare disease-diagnostic cohorts, mitochondrial diseases sit at the upper end of the WES diagnostic rate [11], ranging from 35% to 70% across nine published cohorts [6,58,73,77–82].

Finally, limitations of WES regarding the genome coverage can be overcome with whole genome sequencing (WGS). It enables the identification of structural variants (SVs), deep intronic variants, and some coding regions in which WES shows a sequencing bias [83,84], revealing in summary 4–5 million variants per individual [85]. Based on PCR-free protocols and requiring less starting material, WGS has been proven effective in neonatal and pediatric intensive care units, where a rapid diagnosis can significantly influence the outcome [86]. So far, WGS was implemented in a single mitochondrial disease cohort, reaching diagnosis in 55% of 40 patients [60]. In singleton approaches, it led to the discovery of pathogenic variants elusive to WES in three novel disease genes [55,87,88].

Genotype–phenotype correlations: (inter)national resources/approaches

Research and clinical trials on rare disease are hampered by the small sample size. For this reason, patient registries and collaborative networks present key tools to gather data and promote collaborations between clinicians and researchers, altogether providing better natural disease history data, and facilitating genetic diagnosis and clinical trials. Mitochondrial disease research has benefited from GENOMIT, a global network of eight mitochondrial disorders research centers (http://genomit.eu/), and a number of national networks for mitochondrial disorders: Mitocon (Italy, https://www.mitocon.it/), Mitochondrial Patient Cohort (UK, https://www.newcastle-mitochondria.com/clinical-professional-home-page/mitochondrial-cohort/), mitoNET (Germany, http://mitonet.org/), NAMDC (USA, https://www.rarediseasesnetwork.org/cms/namdc), and Mito Foundation (Australia, https://www.mito.org.au/), and J-MO-Bank (Japan, http://mo-bank.com/).

Addressing the diagnostic gap: new tools in diagnostics

Limitations of current genetic analysis

The rise and implementation of the sequencing technologies in human genetics have improved our diagnostic capabilities dramatically, and these advances are predicted to continue. NGS offers us a high-throughput, genome-wide, automatic approaches with simplified protocols and statistical analysis that apply to all genetic disorders. As outlined above, WES has become

a widely accepted tool for the molecular diagnosis of genetic diseases and novel disease gene discoveries and tremendously improved our understanding of Mendelian disorders [89]. Yet, the present diagnostic gap also calls for assessing the bottlenecks: the issue of data sharing concerning the genetic privacy, the need for better bioinformatics tools for variant calling and interpretation, and overall computational infrastructure. Indeed, a significant fraction of the mitochondrial disease patients remains undiagnosed, about half in WES studies [59]. Several reasons could lie behind the negative results. Inconclusive WES can be attributed to its limitation in variant detection, but a similar diagnostic yield after WGS [60] suggests that the reasons behind it rather lie in variant prioritization and interpretation. Although bioinformatics tools for the sequence alignment, variant calling, functional annotation, and *in silico* prediction are constantly being improved [90], there are still no recommended tools for clinical assessments of noncoding, splicing, synonymous, and UTR variants [91,92]. Additional challenges arise in the mtDNA, where NGS raised the issues of interpretation of very low mtDNA mutational loads and the coexistence of mtDNA variant [28]. Indeed, it has been reported that even low-level heteroplasmy mtDNA variants can cause a disease [93,94]. Without functional validation, these variants remain variants of uncertain significance (VUS) (Fig. 1). As VUS compose 47% of all reported variants [38], it is clear that additional functional evidence apart from sequence information predicted consequence on a transcript/protein level, and frequency is needed to annotate variant´s pathogenicity. Trio sequencing, where the index and parents are sequenced simultaneously, has been shown to be more effective than singleton one despite higher costs and longer data analysis [95], and is thus increasingly applied in diagnostics [86,95–97]. Such analysis presents a powerful tool not only for the interpretation of heterozygous variants in dominant disease genes but also for the detection of *de novo* variants and variant phasing. *De novo* variants can be more harmful than inherited ones and are recognized as a major cause of severe early-onset genetic disorders [98]. Interestingly, an increasing number of studies in recent years have reported *de novo* pathogenic variants across mitochondrial disease genes in cases with a suspected recessive etiology [99–104]. Moreover, in case of *SLC25A4*, the phenotype associated with *de novo* variant did not resemble previously reported dominant or recessive ones [99]. Having this in mind, we recommend the sequencing of entire family whenever possible. Next, complementary tools to the DNA analysis could provide an additional level of functional evidence and thus a better understanding of variant pathogenicity [105] (Fig. 1). Finally, shortcomings of short-read sequencings concerning the variant detection, as well as sequencing of bulk population, can be overcome by long-read and single-cell sequencing technologies (Box 2). A summary of the currently implemented and developed tools is provided in Fig. 4.

BOX 2. Long-read sequencing technologies

Commercially available LRS methods include Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) and Oxford Nanopore Technologies' (ONT) nanopore sequencing [151]. SMRT [152] sequencing is currently more developed and therefore diagnostically applicable. The main reason is that its errors are relatively randomly distributed [152] and can be overcome with increased read depths [153]. Library preparation requires larger amounts of DNA (at least 5 mg) and includes the generation of circularized target DNA molecules by the hairpin sequencing adaptors (the SMRTbell) [154]. The method is based on special flow cells with wells, termed zero-mode waveguides, whose bottom has a fixed single polymerase. Polymerases bind single DNA molecules and incorporate labeled bases, upon which the fluorescent signal is detected in real time by a camera. On the other side, nanopore sequencing is based on a membrane that contains a nanopore linked to a polymerase or helicase [46]. As the single DNA molecule passes through the nanopore, the ionic current fluctuations that occur across the membrane get detected and amplified [46,155]. As opposed to SMRT, its errors are more systematic and related to the features of the DNA fragments, and thus more difficult to overcome [156].

Single-cell sequencing refers to next-generation sequencing of the genome or transcriptome of individual cells, providing a high-resolution overview of single-cell heterogeneity on a global scale. Single-cell genome sequencing workflow is based on single-cell isolation, DNA extraction, amplification of DNA [178,179], and subsequent library preparation and sequencing using the standard NGS protocols [180]. In addition, currently the most employed single-cell technology is the single-cell RNA-seq, first described in 2009 [47,181]. It is based on sequencing of cDNA of a previously isolated viable cell that was exponentially amplified by PCR or *in vitro* transcription [47,182].

Complementary tools to DNA analysis

Transcriptomics

Following the central dogma of molecular biology, RNA analysis, revealing the effect of the variant on a transcript, can facilitate variant interpretation and (re) prioritization, especially for the noncoding variants [106]. Although the diagnostic focus lies in the protein as a product of gene expression, up to 30% of pathogenic variants impact RNA and fall within the noncoding regions [107,108], and up to 10% of pathogenic exonic variants affect splicing [109], suggesting that transcript disruption is a more occurring pathomechanism than often considered. Although RT-PCR is frequently used to assess the effect of splice variants, RNA-sequencing (RNA-seq) offers the systematic analysis of the cellular transcriptome in the form of both sequence and transcript levels [106]. It is becoming a companion to DNA sequencing in diagnostics. To date, six studies reported its diagnostic utility in monogenic diseases, reaching a diagnostic yield of between 8% and 36% over DNA sequencing alone in cohorts of 63–115 patients [54,55,110–113]. With RNA-seq detecting up to 15 000 expressed genes, these studies implemented different computational tools to systematically detect three aberrant transcript events that can be disease causative: aberrant expression level, aberrant splicing, and monoallelic expression [114] (Fig. 4). In addition, calling variants on RNA-seq can provide information about variants in regions not well covered by WES, such as the UTRs and introns, exemplified by detection of a pathogenic heterozygous 5′ UTR variant in *GMPPB* [111]. The published RNA-seq studies based their analysis on in-house databases, where patients serve as controls to each other [55], as well as publically available ones, such as GTEx [110]. Both approaches proved as successful. Diagnostic utility of RNA-seq has been shown on mitochondrial diseases, where a study on skin-derived fibroblasts led to a diagnosis in 10% of previously WES-undiagnosed cases [55]. There, a manageable number of aberrant transcript events were detected: a median one aberrantly expressed gene, five aberrant splicing events, and six monoallelically expressed rare variants. Such approach also enabled discovery of novel disease gene, *TIMMDC1*, encoding a complex I assembly factor. Here, RNA-seq captured aberrant expression and splicing in the transcript in two unrelated individuals, caused by a deep intronic variant that was not detected by WES [55]. Despite encouraging results regarding the reported diagnostic yield, important considerations for the application of RNA-seq in diagnostics should be taken. As tissues are characterized by unique gene expression patterns, ideally the RNA source material should provide an adequate proxy of gene expression and splicing in the disease-relevant tissue. In a diagnostic setting, this tissue specificity could potentially hamper gene detection as clinicians′ choice is usually restricted to easily accessible blood and skin biopsies [115]. More invasive skeletal muscle biopsies are also sometimes available from patients with mitochondrial and muscular disorders [6,54,55,111]. Alternative approaches could be to reprogram the available cells into induced pluripotent stem cells [116] or transdifferentiate them into ones representing the disease-relevant tissues [111,116], but such approaches are more laborious, time-consuming, and therefore only applicable to research. For mitochondrial disease investigation, we are in a more privileged position due to the ubiquitous role of mitochondria throughout the organism. Indeed, over 80% of mitochondrial disease genes are detected across a range of tissues [18]. A study of Kremer *et al.* [55] was performed on skin-derived fibroblasts, which are, apart from being easily obtainable, also an established biological model for the functional validation of potential pathogenic variants in mitochondrial disorders [117]. Although not representing the affected tissue, their usage may enable easier distinguishing between causal gene and its downstream effects, as the consequence of the causal gene defect on the other genes is less prevalent [118]. Regarding RNA-seq data analysis, it has been simplified by the recent developments of specialized methods that detect aberrant transcript defects: OUTRIDER [119] for aberrant expression, LeafCutterMD [120], SPOT [121], and FRASER [122] for splicing outliers, and a method based on a negative binominal test [55] and ANEVA-DOT [123] for monoallelic expression. The computational pipeline DROP offers an all-in-one solution, integrating preprocessing steps, quality control, and outlier detection [124]. As after the pioneer studies in 2017 more studies emerge every year, it is expected that RNA-seq will become a routine part of genetic diagnostics in foreseeable future.

Proteomics

Transcriptome analyses remain inconclusive for missense variants and underestimate the effects of post-transcriptional and post-translational regulation. The effect of the variant may not be detected by RNA-seq but may change the protein levels. Thus, determining protein levels, as well as the protein/RNA ratios, is important for the understanding human physiology and disease, as well as gene regulation [125,126]. For

the systematic analysis of the cellular proteome, proteomics can be implemented. This mass spectrometry analysis of protein lysates provides information on protein abundance. Systematic proteomics promises detection of aberrantly expressed proteins [55], as well as quantification of protein complexes [127] (Fig. 4). In doing so, it can replace the laborious and targeted western blot analyses. This could be of considerable benefit in mitochondrial diseases that are characterized by OXPHOS defect and where missense variants present the most common type of pathogenic variants [59]. To date, proteomics has been reported once as a diagnostic tool [128], where analysis of 150 fibroblasts cell lines from undiagnosed patients with suspected mitochondrial disorders enabled validation of 44% missense VUS, and novel diagnosis in 12% of cases. Previously, singleton proteomics provided functional validation of a VUS in three cases, as well as assessed their effect on complex I and mitoribosome, respectively [55,129,130]. For example, it showed lack of detectable TIMMDC1 and decreased levels of complex I subunits, confirming the loss-of-function effect in *TIMMDC1* detected by RNA-seq, and going in line with the role of this protein in complex I assembly [55]. It also provided functional validation for the biallelic variants in *MRPS34* in a patient with Leigh syndrome and combined OXPHOS defects, showing that they lead to not only depletion of MRPS34, but also reduction in all other small mitoribosome protein subunits, as well as complex I and IV protein subunits [129]. Despite such encouraging results, wider application of proteomics is limited by the still incomplete coverage of cellular proteins and tissue specificity. For example, it detects ~ 8000 proteins in fibroblast cell lines [128]. In addition, it detects the effect of variants on the protein stability and abundance of protein complexes [128] but does not provide information on protein activity and only limited information on physical interactions, which was reported as the main effect of genetic variant [131,132].

Metabolomics

Another layer of information on the top of the DNA sequence may be provided by metabolomics. Based on mass spectrometry of small molecules, this OMICS technology enables the detection and quantification of thousands of small molecule metabolites in a single assay [133]. Although in a clinical setting, metabolomics has been applied for biomarker discovery [134], developments in its analytics and methods for data analysis expanded its application to the studies of disturbed metabolic pathways [135], as well as the variant interpretation [136] (Fig. 4). For mitochondrial disorders, metabolomics has been rarely used as a study tool, probably due to the small number of well-characterized disease samples (< 100) and large disease heterogeneity. Metabolic investigation of very few metabolites (lactate, pyruvate, alanine, TCA cycle intermediates, ethylmalonic acid, 3-methylglutaconic acid, dicarbonic acids, and acylcarnitines) belongs to the routine diagnostic [18]. These metabolites may be employed as biomarkers, but still have low sensitivity and specificity [137]. Fibroblast growth factor 21 (FGF21) and growth/differentiation factor 15 (GDF15) have been introduced as valuable serum biomarkers in patients with mitochondrial diseases, showing high sensitivity and specificity [138,139]. As their value is yet to be shown in larger cohorts, current studies rather argue for the usage of a combination of biomarkers ('metabolic biosignatures') as a screening method [140–142]. For diagnostic purposes, metabolomics was used to confirm defects in proline metabolism due to *ALDH18A1* variants [55], as well as metabolic interpretation of *OPA1* variants [143]. However, a large-scale study with more than 100 samples and measured metabolites has not been published yet for mitochondrial disorders, although it proved of great value for the diagnostics of inborn errors of metabolism, where it showed potential to provide key diagnostic metabolites [144]. Like all technologies, metabolomics has limitations. Measured metabolite levels are results of not only the primary genetic defect, but also the impact of genetic background and environmental exposures [145]. In addition, differences in metabolite properties influence their detection, and our knowledge of metabolic pathways is still far from complete [133]. This altogether makes the metabolomics data more difficult to interpret.

New sequencing technologies

Despite the Human Genome Project being completed in 2003, the complexity and repetitiveness of the human genome still present a major in the analysis of the human genome. Both WES and WGS belong to the second-generation DNA, massively parallel, short-read sequencing methods. These methods generate short reads (~ 150–300 bp) and require template amplification, resulting in copying errors, amplification biases, and loss of base modification information, which ultimately brings certain limitations to the analyses. Even with the constant improvements of technologies and bioinformatics algorithms, these methods are not able to accurately map, or even assemble, all reads originating from regions harboring structural

variation [146], pseudogenes, or homologous regions, highly polymorphic regions, and repetitive sequences. As a consequence, some genetic variants are not accurately called [147]. In addition, short-read sequencing provides limited variant phasing information, which is crucial for the demonstration of compound heterozygosity in cases of the AR disease, especially if parental samples are not available for the segregation analysis. The development of third-generation DNA sequencing methods or long-read sequencing (LRS)* promises to overcome these limitations. Based on PCR-free, real-time, single-molecule sequencing, these technologies generate fast (within 2–10 h) reads > 10 kb in average [148], lack amplification biases and can detect base modifications, such as methylation patterns [149,150] (Box 2). These features promise high-resolution *de novo* genome assembly [157], mapping, identification of transcript isoforms [158], and detection of epigenetic variation and complex regions of the genome [159] (Fig. 4). Although LRS comes with higher error rates, the base calling accuracy has improved in recent years, with current error rates reported to be below 1% for SMRT [160] and < 5% for nanopore sequencers [46]. LRS is still not a part of routine diagnostics; however, recent studies highlighted its role in resolving clinically relevant genomic regions that were previously problematic to sequence [161]. Indeed, in several studies it led to the identification of disease-causing locus [56] and SVs [53], enabled sequencing of pathogenic repeat expansions [162] and mapping of highly polymorphic regions [163], resolved variant phasing to determine the parental origin of variants or study mosaicism [164,165], and helped discriminate genes of interest from its homolog pseudogenes [166]. For genetic diagnosis of mitochondrial disease, LRS could prove useful for assessing the *ATAD3* gene cluster, composed of three highly homologous genes formed *via* tandem segmental duplication: *ATAD3C*, *ATAD3B*, and *ATAD3A*. Due to the high sequence homology among the genes, the *ATAD3* region exhibits structural instability and is prone to nonallelic homologous recombination (NAHR), which can result in copy number variations (CNVs) [167]. Indeed, a variety of pathogenic variants has been reported in this locus, including the (*de novo*) dominant and recessive SNVs and CNVs [100,168–171]. LRS was recently employed to validate the presence of chimeric *ATAD3A*/*ATAD3C* gene in one patient with a *de novo* *ATAD3* duplication [172]. As emerging studies argue the higher incidence of *ATAD3* as a cause of pediatric-onset disease, LRS may help overcome the challenges in assessing its sequence. In addition, LRS could replace PCR and Southern blotting in testing for mtDNA deletions. Single study on mitochondrial disease cohort reports that LRS is able to detect mtDNA deletions, but has issues with determination of breakpoints and false-positive rates [173]. Thus, further improvements in the method and bioinformatic tools are needed to confidently detect pathogenic mtDNA SVs. Overall, the higher costs, more demanding library preparation, and still immature tools for data analysis are standing in the way of more routine clinical implementation [174,175], and LRS has been mainly used in combination with short-read sequencing or as a follow-up tool for the undiagnosed cases. Yet, constant improvements in all aspects suggest that LRS will move forward from the current targeted approaches in foreseeable future.

Present diagnostic studies focus on germline variants, frequently overlooking cell-to-cell variability. However, with the rise of single-cell sequencing* in the last decade, we have a possibility of a comprehensive analysis of individual cells [176] on both genome and transcriptome levels [47,177] (Box 2, Fig. 4). This helps resolving heterogeneous cell populations and discovering rare cell subpopulations [183,184]. Single-cell genomic sequencing enables the detection of cell-specific somatic DNA mutations. Studies employing the method revealed surprisingly large genomic diversity and *de novo* mutations arising from germ cells [185,186], as well as the presence of private and clonal somatic CNVs in neurons [187]. Recently, Ludwig *et al.* reported single-cell mtDNA sequencing [188] to detect mtDNA mutations and heteroplasmy on the level of an individual cell, and use them as genetic barcodes for clonal and lineage tracing. Such methods will improve recording of heteroplasmy and its impact on mitochondrial disorders caused by the mtDNA variants. On the other hand, the single-cell RNA-seq (scRNA-seq) enables investigations of cellular heterogeneity and transcriptomic changes in individual cells and has been successfully employed in several medical fields. In basic research, it enabled the identification of specific immune cell subtypes [189] and insights into neural development [190,191]. In clinical research, scRNA-seq helped pinpoint disease-related cell populations or track the course of a disease [192–196]. Compared with the standard bulk sequencing, application of single-cell sequencing is more complicated due to higher costs, more demanding sample preparation, and challenging variant calling [197,198], although a recent method promises accurate identification of mutations [36]. An important limitation is also the sequencing depth and detecting power, well below the standard bulk transcriptome [199]. Yet, considering previous encouraging results [200], for mitochondrial disease

these methods could be an option to improve the understanding of phenotypic variability and tissue-specific expression of defects.

Broadening the diagnostic focus: beyond the protein-coding genes

Noncoding regions

In the search for missing heritability, it is also important to note that the primary focus in diagnostics of Mendelian disorders remains on genetic variants in protein-coding genes that ultimately affect the protein product. Although easier to interpret and functionally validate, such focus is limited to a small portion of the genome. Still poorly understood, the noncoding region is gaining more attention, and accumulated evidence argues for its role in disease causality. First, the development of high-throughput OMICS technologies has accelerated the identification of gene regulatory elements on a genomic scale. Large-scale projects such as ENCODE [37], FANTOM [41], and GTEx [39] are enabling a better understanding of the complexity behind gene regulation, as well as improving functional annotation of the genome. Indeed, ENCODE project reported functionality of 80% of the human genome, defined by a gene product (protein or ncRNA), protein and RNA binding, and specific chromatin structure [37]. Second, more than 90% of disease-associated SNPs identified by genome-wide association studies (GWAS) fall within the noncoding region [201,202]. Finally, the disease risk and variable clinical presentation can be attributed to common coding and regulatory variants influencing the gene expression [203,204]. One could hypothesize that at least a fraction of undiagnosed cases could be due to the disruption of the functional elements in the noncoding regions: regulatory regions, such as promoters, enhancers, structural elements on the one side, and ncRNAs, as well as micropeptides hidden within ncRNAs, on the other (Fig. 5).

Functional interpretation of variants in regulatory regions is challenging, as they may be located far from the gene they control. Slowly but surely, annotation of such variants is improving from correlation with a disease to a disease causality [205]. Pathogenic regulatory SNPs, as well as SVs, have been described in a relatively small number of Mendelian disorders [206,207]. They can act *via* loss of function, resulting in downregulation of gene expression, or a gain of function, resulting in mis- or overexpression, frequently in a dominant mode of inheritance. For example, point mutations in *ZRS*, an enhancer located 1 Mb from its target gene, *SHH*, cause polydactyly [208]. Biallelic variants in the 25-kb region downstream of *PTF1A* abolish regions′ enhancer activity, disrupt the expression of *PTF1A*, and cause isolated pancreatic agenesis [209]. SVs in regulatory regions can also disturb normal chromatin folding. For example, rearrangements of the *WNT6/IHH/EPHA4/PAX3* region results in limb malformations due to disruption of protein-coding topologies and consequent inappropriate

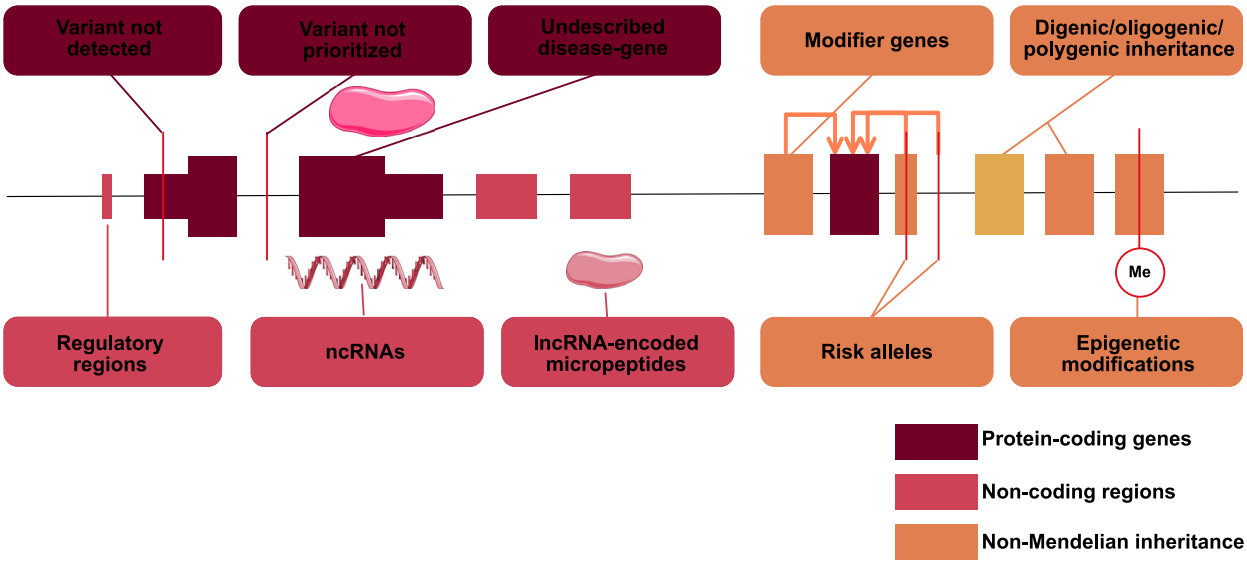

Fig. 5. Reasons leading to an inconclusive genetic diagnosis. Each box provides a potential cause or a contributor to the disease across protein-coding genes, noncoding regions, and the non-Mendelian inheritance.

expressions and interactions [210]. Finally, *de novo* mutations are enriched in neurodevelopmental disorders and could provide a diagnosis for up to 3% of currently undiagnosed patients [211]. Increasing genome sequencing in mitochondrial disorders such as in the UK or mitoNET is considering such modes of disease. Although the prioritization and interpretation of variants in regulatory regions are still underdeveloped, the implementation of complementary OMICS tools to the DNA analysis could provide additional information to pinpoint the affected transcript and protein.

Apart from the regulatory regions, ncRNAs* are also emerging as important factors in pathogenesis (Box 3). Indeed, transcriptome analyses revealed that up to 75% of the genome is pervasively transcribed [215]. It is clear that apart from mRNAs, other various classes of ncRNAs contribute to the cellular transcriptome and have important rules in gene regulation [216]. Among ncRNAs, microRNAs (miRNAs) and their role in transcriptional and translational regulation have been well studied [212]. Moreover, long noncoding RNA (lncRNAs) and circRNAs are emerging as important players in gene regulation and chromatin states [214,217]. Large-scale genome-wide screenings reported that 2–8% of targeted lncRNAs are important for cellular growth, implying their significance [218–220]. Although miRNAs and lncRNAs have been usually associated with the human disease just through changes in expression levels or as part of syndromic deletions [221,222], a handful of Mendelian disorders has been caused by variants in miRNAs [223–225]. In addition, Ang *et al.* (2019) reported mutations in lncRNAs in patients with neurodevelopmental disorders, identifying *lnc-NR2F1* as an important regulator of neurogenesis and a potential cause of disease [226]. Still, little is known about the variants affecting the genomic loci that encode ncRNAs as their effect may not be as straightforward as for protein-coding genes. Focusing on mitochondria, numerous miRNAs and lncRNAs have been described to target variety of mitochondrial functions by binding not only mRNAs, but also miRNAs and proteins, and even to be encoded by the mtDNA itself [216]. For example, lncRNA *Cerox1* has been described as a modulator of OXPHOS, by binding miR-488-3p in the cytoplasm and preventing its inhibitory effect on the complex I transcripts [227]. However, the field of ncRNAs in mitochondria is heavily understudied, with disappointingly small study overlap and rare replications of findings, and direct association between ncRNAs and mitochondrial disease is yet to be reported [216]

In the last decade, the ribosome profiling data and proteomic analyses identified a surprisingly large number of micropeptides* derived from upstream open reading frames (uORFs*), as well as ORFs corresponding to the lncRNAs [125,228–231], arguing for the coding potential of noncoding regions and faults in current genome annotation (Box 3). More than twenty ncRNA-encoded proteins have been characterized in depth [235], including the mtDNA-encoded humanin, MOTS-c, and SHLPs [236–238], with roles in mitochondrial bioenergetics and metabolism, and mitoregulin, involved in the formation of OXPHOS supercomplexes, fatty acid oxidation, and Ca^{2+} dynamics [239]. Interestingly, 5% of mitochondrial proteome is represented with micropeptides [21], and the newly reported ncRNA-encoded micropeptides are enriched in the organelle [230,231]. With their correct genome annotations, these discoveries will increase the sequence spectrum while searching for causes of mitochondrial disease, and potentially lead to identification of a novel disease gene. This emerging coding, but also disease-causing, potential of noncoding regions presents an additional argument for implementation of WGS over WES, where these regions are invisible.

BOX 3. ncRNAs

ncRNAs present a heterogeneous class of transcripts that are not translated into proteins. miRNAs are a class of short ncRNAs. Single-stranded 19–23 nucleotides in length regulate gene expression by binding to the mRNAs, leading to cleavage, degradation, or translational repression [212]. Long noncoding RNAs (lncRNAs) are defined as non-protein-coding transcripts longer than 200 nucleotides. They include intergenic transcripts, enhancer RNAs (eRNAs), and overlapping (sense or antisense) transcripts that overlap other genes [213]. circRNAs are a class of ncRNAs formed during alternative splicing of pre-mRNAs, with the linked 3′ and 5′ ends [214].

Micropeptides are biologically active peptides smaller than 100 amino acids, resulting in them often being overlooked by the traditional ORF cutoffs [232].

uORF presents an ORF within the 5'-untranslated region (5'UTR) of an mRNA. These short coding sequences usually act as *cis*-acting repressors of translation [233,234].

Non-Mendelian inheritance

Implementation of systematic approaches, for instance, WES, in diagnostics has enabled finding novel

disease-associated genes at a fast pace [18]. Yet, these discoveries will likely reach the plateau in the future, paving a way to studying more complex pathomechanisms that have been barely reported in mitochondrial disease. Currently, the search for a genetic cause is made on the assumption that the pathogenic variant is rare and fully penetrant. However, incomplete penetrance, especially for the mtDNA-encoded pathogenic variants [240,241] and variable phenotype expressivity, presents well-known features and challenges of mitochondrial disorders. The contribution of the other (epi)genetic factors, including genetic modifiers, digenic, oligogenic, and polygenic inheritance, as well as tissue specificity and environmental influences, is neglected during genetic diagnosis, but slowly gaining the attention of researchers (Fig. 5).

In addition to the mtDNA variants, two examples of reduced penetrance associated with nuclear-encoded disease gene have been reported. Homozygous *SDHB* variant was shared by a patient with leukoencephalopathy and complex II deficiency and her healthy sister [242]. Across 29 families, 30 of 31 homozygous male carriers and just three of seven homozygous male carriers of rare *DNAJC30* variants expressed a LHON phenotype, resembling the similar observations as in the maternally inherited LHON [130]. This story, revealing a novel mitochondrial complex I repair mechanism, also represents one more example how the discovery of novel mitochondrial disease genes improves our knowledge of mitochondrial physiology and pathomechanism. Detection and interpretation of such variants are exceptionally difficult due to the absence of conclusive segregation analysis, higher allele frequency in the population, and usually few mitochondrial disease cases with the same clinical presentation.

Genetic modifiers, alleles that modulate the effect of the disease-causing one, are becoming more recognized as contributors to Mendelian diseases [243]. In mitochondrial diseases, several studies reported the contribution of a variant in nuclear-encoded genes to the phenotypic effect of pathogenic mtDNA variants. For example, a heterozygous start-loss variant in *SSBP1* cosegregates with hearing loss in carriers of the homoplasmic m.1555A>G variant [244]. Two studies reported that the presence of X-linked c.157C>T in *PRICKLE3* and c.572G>T in *YARS2* in LHON patients with m.11778G>A leads to a visual failure [245,246], offering a possible explanation for the variations in penetrance and sex well known in LHON. The same variant in *YARS2* leads to deafness in combination with the m.7511A>G variant [247]. It is noteworthy that these studies were based on a genetic and segregation analysis of a small number of large multigenerational families, and a more systematic, singleton study is missing.

Another consideration could be that the suspected disease is caused by digenic or oligogenic factors. Such pathomechanisms are seldom considered during standard DNA analysis as they are difficult to interpret. One example was an adult patient with PEO and severe neurological presentations, where the combined effect of *de novo ANT1* and homozygous *POLG1* may explain such a complex clinical picture [248]. Implementations of complementary OMICS tools may help pinpoint the contribution of more than one gene to the disease. Indeed, such approach was recently utilized to decipher the molecular mechanism behind complete spontaneous recovery observed in patients with reversible infantile respiratory chain deficiency. Here, digenic inheritance of homoplasmic m.14674T>C variant with heterozygous pathogenic variants in nuclear genes, *EARS2* and *TRMU* variants, respectively, leads to the clinical manifestation of the disease, activating the three-phase metabolic events that eventually lead to recovery [249]. Concerning the mtDNA, it has been postulated that the phenotypic variability of mtDNA homoplasmic variants could be explained by additional low-level heteroplasmic variants [250].

Variable expressivity in mitochondrial disease contributes a lot to the heterogeneous clinical presentations of patients. It may be attributable to 'weaker' even common variants, and their interactions. However, these interactions, better covered in the research of complex diseases, are difficult to study in a rare disease setting, as they require a large number of patients with comprehensive genotypic and phenotypic data. Such catalogs could be generated in more common mitochondrial disease, such as LHON. Indeed, the effect of pathogenic mtDNA variants on LHON phenotype was shown to be influenced by the mtDNA haplogroups, as reported in a study of over 3600 patients [251]. However, it is rather unlikely to generate such a cohort for the majority of mitochondrial diseases, but should nevertheless serve as an encouragement for mitochondrial researchers to expand their collaborations and share their data. The GENOMIT network stands as a positive example, with the collection of ~ 3000 WES data and developing a global registry of mitochondrial disease patients. To overcome the issue of statistical power, the latest studies were based on data from large GWAS of complex traits. They demonstrated the involvement of common polygenic factors in the variable expressivity of rare diseases [252–254]. Mitochondrial function and genetic variation have been previously associated with metabolic diseases such as obesity, metabolic syndrome,

insulin resistance, and diabetes, as well as cardiovascular disease, Parkinson´s disease, and immune response [255–262]. Also, recent studies focused on regulatory links between the mitochondrial and nuclear genome. They have identified 11 significant associations between mitochondrial genome mutations and nuclear gene expression [263], associations between variants in mitochondrial RNA-binding proteins and haplogroup-specific mtDNA gene expression [264], and 64 nuclear loci associated with expression levels of 14 mtDNA genes [265]. These results provide evidence for the *in trans* effects between two genomes regarding the gene expression. Finally, by performing meta-analysis across 45 cohorts with up to 170 000 individuals, Kraja *et al.* (2018) reported that seven mtDNA SNVs and 109 nuclear genes predicted to have a role in mitochondria have a significant association with at least one of seven inspected metabolic traits [266]. This suggests that the common genetic variants have downstream effects that may influence disease risk and manifestation.

Finally, the cause of the disease may be outside the four base pair genetic code. Human disease can also be caused by rare epigenetic variation (epivariation, epimutations) [267], defined as an alteration in DNA methylation, irrespective of their underlying etiology (sporadic events such as imprinting anomalies [268] or sequence variation [269]). Although the contribution of epivariations to mitochondrial disorders remains unexplored, recent studies reported enrichment of *de novo* epivariation in congenital disorders and neurodevelopmental disorders, followed by altered gene expression [270,271]. Epivariation has been reported as a cause of the inborn error of metabolism, where the cause of the disease, *MMACHC*, harbored a secondary epimutation in its promoter on one allele and a frameshift variant on the other. The secondary epimutation, marked by promoter hypermethylation, was caused by the antisense transcription of adjacent mutated *PRDX1* [272]. These studies indicate that adding methylation profiling to the sequencing approaches may reveal pathogenic epivariation and fill in the diagnostic gap.

Conclusions and Perspectives

The last decade has greatly advanced our understanding of the genetics behind mitochondrial disorders: Of almost 1200 currently described mitochondrial proteins, more than 500 are associated with disease, out of which over 340 as a cause of mitochondrial disease [20]. Following the developments in genetic tools, genetic diagnostics has shifted from candidate gene screens into the unbiased analysis of the complete genetic variation landscape. NGS methods, inseparable from computational tools, have allowed generating large genome databases, an increase in the discovery of pathogenic variants, and exponential growth in the number of newly discovered disease-associated genes. These patient datasets now provide big data that can be used not only for diagnostics but also for unraveling mitochondrial biology. The newly diagnosed cases help both disease management and unraveling its underlying etiology. On the one side, the established genetic cause may offer vast opportunities for clinical trials and the development of therapeutic approaches such as gene replacement, antisense oligonucleotides, and gene editing [273]. On the other side, the undiagnosed cases call for the implementation of better methods, but also require us to go beyond the protein-coding regions when formulating a diagnosis. Complementary OMICS tools, as well as long-read and single-cell sequencing, offer improvements in the variant prioritization, interpretation, and detection. Moving away from the protein-centric approach, the roles of noncoding regions in biological functions and human diseases are gaining more attention. This could not only lead to discoveries of disease-related noncoding genes and variants, but also unravel new pathomechanisms. Finally, the non-Mendelian factors are emerging as modulators of mitochondrial disorders. Recent examples of digenic nuclear–mitochondrial interactions suggest that such co-occurrence should be evaluated in mtDNA diseases. Moreover, the influence of genetic background and common variants is yet to be elucidated and will provide a better understanding of the disease risk and clinical manifestation. Such expanding views and hopefully consequent findings will further explain the mechanisms behind the complexity of the mitochondrial disease. To conclude, the development of technologies and increased publicly available datasets will steadily improve diagnostics and enable discoveries of novel pathomechanisms behind, thus paving the way for developing gene-targeted therapies for mitochondrial diseases.

Acknowledgements

The authors acknowledge the support by a German Federal Ministry of Education and Research (BMBF) Grant to the German Network for Mitochondrial Disorders (mitoNET, 01GM1906B), and by the German BMBF and Horizon2020 through the E-Rare Project GENOMIT European Network for Mitochondrial Disease (01GM1920A) and the ERA PerMed Project PerMiM (01KU2016A). MG is grateful for the support of the DZHG (German Centre for

Cardiovascular Research) and the BMBF (German Federal Ministry for Education and Research).

Author contributions

MG and HP designed the review. MG wrote the review and generated the table and figures. HP supervised the writing process, provided the critical feedback, and contributed to the final version of the review.

References

1 Gorman GS, Chinnery PF, DiMauro S, Hirano M, Koga Y, McFarland R, Suomalainen A, Thorburn DR, Zeviani M and Turnbull DM (2016) Mitochondrial diseases. *Nat Rev Dis Primers* **2**, 16080.

2 Tan J, Wagner M, Stenton SL, Strom TM, Wortmann SB, Prokisch H, Meitinger T, Oexle K and Klopstock T (2020) Lifetime risk of autosomal recessive mitochondrial disorders calculated from genetic databases. *EBioMedicine* **54**, 102730.

3 Munnich A and Rustin P (2001) Clinical spectrum and diagnosis of mitochondrial disorders. *Am J Med Genet* **106**, 4–17.

4 Zhou J, Li J, Stenton SL, Ren X, Gong S, Fang F and Prokisch H (2020) NAD(P)HX dehydratase (NAXD) deficiency: a novel neurodegenerative disorder exacerbated by febrile illnesses. *Brain* **143**, e8.

5 El-Hattab AW, Zarante AM, Almannai M and Scaglia F (2017) Therapies for mitochondrial diseases and current clinical trials. *Mol Genet Metab* **122**, 1–9.

6 Wortmann SB, Mayr JA, Nuoffer JM, Prokisch H and Sperl W (2017) A guideline for the diagnosis of pediatric mitochondrial disease: the value of muscle and skin biopsies in the genetics era. *Neuropediatrics* **48**, 309–314.

7 Wortmann SB, Koolen DA, Smeitink JA, van den Heuvel L and Rodenburg RJ (2015) Whole exome sequencing of suspected mitochondrial patients in clinical practice. *J Inherit Metab Dis* **38**, 437–443.

8 Rodenburg RJ (2011) Biochemical diagnosis of mitochondrial disorders. *J Inherit Metab Dis* **34**, 283–292.

9 Munnich A, Rotig A, Chretien D, Cormier V, Bourgeron T, Bonnefont JP, Saudubray JM and Rustin P (1996) Clinical presentation of mitochondrial disorders in childhood. *J Inherit Metab Dis* **19**, 521–527.

10 Niyazov DM, Kahler SG and Frye RE (2016) Primary mitochondrial disease and secondary mitochondrial dysfunction: importance of distinction for diagnosis and treatment. *Mol Syndromol* **7**, 122–137.

11 Wolf NI and Smeitink JA (2002) Mitochondrial disorders: a proposal for consensus diagnostic criteria in infants and children. *Neurology* **59**, 1402–1405.

12 Morava E, van den Heuvel L, Hol F, de Vries MC, Hogeveen M, Rodenburg RJ and Smeitink JA (2006) Mitochondrial disease criteria: diagnostic applications in children. *Neurology* **67**, 1823–1826.

13 Witters P, Saada A, Honzik T, Tesarova M, Kleinle S, Horvath R, Goldstein A and Morava E (2018) Revisiting mitochondrial diagnostic criteria in the new era of genomics. *Genet Med* **20**, 444–451.

14 Raymond FL, Horvath R and Chinnery PF (2018) First-line genomic diagnosis of mitochondrial disorders. *Nat Rev Genet* **19**, 399–400.

15 Wright CF, FitzPatrick DR and Firth HV (2018) Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet* **19**, 325.

16 Rahman J, Noronha A, Thiele I and Rahman S (2017) Leigh map: a novel computational diagnostic resource for mitochondrial disease. *Ann Neurol* **81**, 9–16.

17 Schon EA, DiMauro S and Hirano M (2012) Human mitochondrial DNA: roles of inherited and somatic mutations. *Nat Rev Genet* **13**, 878–890.

18 Stenton SL and Prokisch H (2020) Genetics of mitochondrial diseases: Identifying mutations to help diagnosis. *EBioMedicine* **56**, 102784.

19 Milone M and Massie R (2010) Polymerase gamma 1 mutations: clinical correlations. *Neurologist* **16**, 84–91.

20 Schlieben LD and Prokisch H (2021) The dimensions of primary mitochondrial disorders. *Front Cell Dev Biol* **8**, 600079.

21 Calvo SE, Clauser KR and Mootha VK (2016) MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res* **44**, D1251–D1257.

22 Harbauer AB, Zahedi RP, Sickmann A, Pfanner N and Meisinger C (2014) The protein import machinery of mitochondria-a regulatory hub in metabolism, stress, and disease. *Cell Metab* **19**, 357–372.

23 Del Dotto V, Ullah F, Di Meo I, Magini P, Gusic M, Maresca A, Caporali L, Palombo F, Tagliavini F, Baugh EH *et al.* (2020) SSBP1 mutations cause mtDNA depletion underlying a complex optic atrophy disorder. *J Clin Invest* **130**, 108–125.

24 Marcelino LA and Thilly WG (1999) Mitochondrial mutagenesis in human cells and tissues. *Mutat Res* **434**, 177–203.

25 Rossignol R, Faustin B, Rocher C, Malgat M, Mazat JP and Letellier T (2003) Mitochondrial threshold effects. *Biochem J* **370**, 751–762.

26 de Laat P, Koene S, van den Heuvel LP, Rodenburg RJ, Janssen MC and Smeitink JA (2012) Clinical features and heteroplasmy in blood, urine and saliva in 34 Dutch families carrying the m.3243A > G mutation. *J Inherit Metab Dis* **35**, 1059–1069.

27 MITOMAP: A Human Mitochondrial Genome Database (2019) http://www.mitomap.org.

28 Bris C, Goudenege D, Desquiret-Dumas V, Charif M, Colin E, Bonneau D, Amati-Bonneau P, Lenaers G, Reynier P and Procaccio V (2018) Bioinformatics tools

and databases to assess the pathogenicity of mitochondrial DNA variants in the field of next generation sequencing. *Front Genet* **9**, 632.
29 Sallevelt SC, de Die-Smulders CE, Hendrickx AT, Hellebrekers DM, de Coo IF, Alston CL, Knowles C, Taylor RW, McFarland R and Smeets HJ (2017) *De novo* mtDNA point mutations are common and have a low recurrence risk. *J Med Genet* **54**, 73–83.
30 Chinnery PF, DiMauro S, Shanske S, Schon EA, Zeviani M, Mariotti C, Carrara F, Lombes A, Laforet P, Ogier H *et al.* (2004) Risk of developing a mitochondrial DNA deletion disorder. *Lancet* **364**, 592–596.
31 Viscomi C and Zeviani M (2017) MtDNA-maintenance defects: syndromes and genes. *J Inherit Metab Dis* **40**, 587–599.
32 Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurles ME, Kathiresan S, Kenny EE, Lindgren CM, MacArthur DG *et al.* (2020) A brief history of human disease genetics. *Nature* **577**, 179–189.
33 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R & 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2089.
34 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al.* (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303.
35 Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT *et al.* (2018) A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* **36**, 983–987.
36 Bohrson CL, Barton AR, Lodato MA, Rodin RE, Luquette LJ, Viswanadham VV, Gulhan DC, Cortes-Ciriano I, Sherman MA, Kwon M *et al.* (2019) Linked-read analysis identifies mutations in single-cell DNA-sequencing data. *Nat Genet* **51**, 749–754.
37 The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
38 Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM and Maglott DR (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980–D985.
39 Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ *et al.* (2015) Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660–665.
40 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291.
41 Kawaji H, Kasukawa T, Forrest A, Carninci P and Hayashizaki Y (2017) The FANTOM5 collection, a data series underpinning mammalian transcriptome atlases in diverse cell types. *Sci Data* **4**, 170113.
42 Brody JA, Morrison AC, Bis JC, O'Connell JR, Brown MR, Huffman JE, Ames DC, Carroll A, Conomos MP, Gabriel S *et al.* (2017) Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat Genet* **49**, 1560–1563.
43 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443.
44 Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, Gonzaga-Jauregui C, Khalid S, Ye B, Banerjee N *et al.* (2020) Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756.
45 Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876.
46 Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S and Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* **4**, 265–270.
47 Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–382.
48 Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M *et al.* (2015) Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608–611.
49 Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT *et al.* (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* **36**, 338–345.
50 Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE *et al.* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276.
51 Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H,

Jing C, Wheeler DA *et al*. (2010) Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* **362**, 1181–1191.
52 Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G *et al*. (2010) *De novo* mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* **42**, 483–485.
53 Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, Waggott D, Utiramerur S, Hou Y, Smith KS *et al*. (2018) Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med* **20**, 159–163.
54 Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V, Waddell LB, Sandaradura SA, O'Grady GL *et al*. (2017) Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* **9**, eaal5209.
55 Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, Haack TB, Graf E, Schwarzmayr T, Terrile C *et al*. (2017) Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* **8**, 15824.
56 Aneichyk T, Hendriks WT, Yadav R, Shin D, Gao D, Vaine CA, Collins RL, Domingo A, Currall B, Stortchevoi A *et al*. (2018) Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. *Cell* **172**, 897–909.e21.
57 Haack TB, Danhauser K, Haberberger B, Hoser J, Strecker V, Boehm D, Uziel G, Lamantea E, Invernizzi F, Poulton J *et al*. (2010) Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat Genet* **42**, 1131–1141.
58 Haack TB, Haberberger B, Frisch EM, Wieland T, Iuso A, Gorza M, Strecker V, Graf E, Mayr JA, Herberg U *et al*. (2012) Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *J Med Genet*. **49**, 277–283.
59 Stenton SL, Alhaddad B, Chang C, Haack T, Wortmann S, Mayr JA, Büchner B, Hempel M, Distelmaier F, Freisinger P *et al*. (2019) The genetic landscape of mitochondrial disease: a study of 1116 exomes. 51st European Society of Human Genetics Conference. Milan, Italy.
60 Riley LG, Cowley MJ, Gayevskiy V, Minoche AE, Puttick C, Thorburn DR, Rius R, Compton AG, Menezes MJ, Bhattacharya K *et al*. (2020) The diagnostic utility of genome sequencing in a pediatric cohort with suspected mitochondrial disease. *Genet Med* **22**, 1254–1261.
61 Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, Elsas LJ 2nd and Nikoskelainen EK (1988) Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* **242**, 1427–1430.
62 Yu-Wai-Man P, Turnbull DM and Chinnery PF (2002) Leber hereditary optic neuropathy. *J Med Genet* **39**, 162–169.
63 Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, Rodenburg RJ, Yntema HG, Spruijt L, Vermeer S *et al*. (2013) A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat* **34**, 1721–1726.
64 Frazier AE, Thorburn DR and Compton AG (2019) Mitochondrial energy generation disorders: genes, mechanisms, and clues to pathology. *J Biol Chem* **294**, 5386–5395.
65 Pereira R, Oliveira J and Sousa M (2020) Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. *J Clin Med* **9**, 132.
66 Tang S, Wang J, Zhang VW, Li FY, Landsverk M, Cui H, Truong CK, Wang G, Chen LC, Graham B *et al*. (2013) Transition to next generation analysis of the whole mitochondrial genome: a summary of molecular defects. *Hum Mutat* **34**, 882–893.
67 Paediatric single mitochondrial DNA deletion disorders: an overlapping spectrum of disease
68 Gorman G, Ng YS, Sommerville E, Schaefer A, Taylor R, McFarland R and Turnbull D (2016) Diagnostic algorithm in adult-onset mendelian PEO associated with mitochondrial disease. *Neurology* **86**, P5.070.
69 Calvo SE, Compton AG, Hershman SG, Lim SC, Lieber DS, Tucker EJ, Laskowski A, Garone C, Liu ST, Jaffe DB *et al*. (2012) Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med* **4**, 118ra10.
70 Calvo SE, Tucker EJ, Compton AG, Kirby DM, Crawford G, Burtt NP, Rivas M, Guiducci C, Bruno DL, Goldberger OA *et al*. (2010) High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat Genet* **42**, 851–858.
71 Vasta V, Ng SB, Turner EH, Shendure J and Hahn SH (2009) Next generation sequence analysis for mitochondrial disorders. *Genome Med* **1**, 100.
72 DaRe JT, Vasta V, Penn J, Tran NTB and Hahn SH (2013) Targeted exome sequencing for mitochondrial disorders reveals high genetic heterogeneity. *Bmc Med Genet* **14**, 118.
73 Legati A, Reyes A, Nasca A, Invernizzi F, Lamantea E, Tiranti V, Garavaglia B, Lamperti C, Ardissone A, Moroni I *et al*. (2016) New genes and pathomechanisms in mitochondrial disorders unraveled by NGS technologies. *Bba-Bioenergetics* **1857**, 1326–1335.

74 Lieber DS, Calvo SE, Shanahan K, Slate NG, Liu ST, Hershman SG, Gold NB, Chapman BA, Thorburn DR, Berry GT *et al*. (2013) Targeted exome sequencing of suspected mitochondrial disorders. *Neurology* **80**, 1762–1770.
75 Wagner M, Berutti R, Lorenz-Depiereux B, Graf E, Eckstein G, Mayr JA, Meitinger T, Ahting U, Prokisch H, Strom TM *et al*. (2019) Mitochondrial DNA mutation analysis from exome sequencing-A more holistic approach in diagnostics of suspected mitochondrial disease. *J Inherit Metab Dis* **42**, 909–917.
76 van Dijk EL, Auger H, Jaszczyszyn Y and Thermes C (2014) Ten years of next-generation sequencing technology. *Trends Genet* **30**, 418–426.
77 Taylor RW, Pyle A, Griffin H, Blakely EL, Duff J, He LP, Smertenko T, Alston CL, Neeve VC, Best A *et al*. (2014) Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. *JAMA* **312**, 68–77.
78 Ohtake A, Murayama K, Mori M, Harashima H, Yamazaki T, Tamaru S, Yamashita Y, Kishita Y, Nakachi Y, Kohda M *et al*. (2014) Diagnosis and molecular basis of mitochondrial respiratory chain disorders: exome sequencing for disease gene identification. *Bba-Gen Subjects* **1840**, 1355–1359.
79 Kohda M, Tokuzawa Y, Kishita Y, Nyuzuki H, Moriyama Y, Mizuno Y, Hirata T, Yatsuka Y, Yamashita-Sugahara Y, Nakachi Y *et al*. (2016) A comprehensive genomic analysis reveals the genetic landscape of mitochondrial respiratory chain complex deficiencies. *Plos Genet* **12**, e1005679.
80 Pronicka E, Piekutowska-Abramczuk D, Ciara E, Trubicka J, Rokicki D, Karkucinska-Wieckowska A, Pajdowska M, Jurkiewicz E, Halat P, Kosinska J *et al*. (2016) New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre. *J Transl Med* **14**, 174.
81 Puusepp S, Reinson K, Pajusalu S, Murumets U, Oiglane-Shlik E, Rein R, Talvik I, Rodenburg RJ and Ounap K (2018) Effectiveness of whole exome sequencing in unsolved patients with a clinical suspicion of a mitochondrial disorder in Estonia. *Mol Genet Metab Rep* **15**, 80–89.
82 Theunissen TEJ, Nguyen M, Kamps R, Hendrickx AT, Sallevelt S, Gottschalk RWH, Calis CM, Stassen APM, de Koning B, Mulder-Den Hartog ENM *et al*. (2018) Whole exome sequencing is the preferred strategy to identify the genetic defect in patients with a probable or possible mitochondrial cause. *Front Genet* **9**, 400.
83 Alfares A, Aloraini T, Al Subaie L, Alissa A, Al Qudsi A, Alahmad A, Al Mutairi F, Alswaid A, Alothaim A, Eyaid W *et al*. (2018) Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genet Med* **20**, 1328–1333.
84 Meienberg J, Bruggmann R, Oexle K and Matyas G (2016) Clinical sequencing: is WGS the better WES? *Hum Genet* **135**, 359–362.
85 Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, Khera AV, Zhou W, Bloom JM, Engreitz JM *et al*. (2018) Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat Commun* **9**, 3391.
86 Clark MM, Hildreth A, Batalov S, Ding Y, Chowdhury S, Watkins K, Ellsworth K, Camp B, Kint CI, Yacoubian C *et al*. (2019) Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci Transl Med* **11**, eaat6177.
87 Tamiya G, Makino S, Hayashi M, Abe A, Numakura C, Ueki M, Tanaka A, Ito C, Toshimori K, Ogawa N *et al*. (2014) A mutation of COX6A1 causes a recessive axonal or mixed form of charcot-marie-tooth disease. *Am J Hum Genet* **95**, 294–300.
88 Malicdan MCV, Vilboux T, Ben-Zeev B, Guo J, Eliyahu A, Pode-Shakked B, Dori A, Kakani S, Chandrasekharappa SC, Ferreira CR *et al*. (2018) A novel inborn error of the coenzyme Q10 biosynthesis pathway: cerebellar ataxia and static encephalomyopathy due to COQ5 C-methyltransferase deficiency. *Hum Mutat* **39**, 69–79.
89 Smith HS, Swint JM, Lalani SR, Yamal JM, Otto MCD, Castellanos S, Taylor A, Lee BH and Russell HV (2019) Clinical application of genome and exome sequencing as a diagnostic tool for pediatric patients: a scoping review of the literature. *Genet Med* **21**, 3–16.
90 Dias R and Torkamani A (2019) Artificial intelligence in clinical and genomic diagnostics. *Genome Med* **11**, 70.
91 Boycott KM and Ardigo D (2018) Addressing challenges in the diagnosis and treatment of rare genetic diseases. *Nat Rev Drug Discov* **17**, 151–152.
92 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E *et al*. (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. *Genet Med* **17**, 405–424.
93 Ng YS, Lax NZ, Maddison P, Alston CL, Blakely EL, Hepplewhite PD, Riordan G, Meldau S, Chinnery PF, Pierre G *et al*. (2018) MT-ND5 mutation exhibits highly variable neurological manifestations at low mutant load. *EBioMedicine* **30**, 86–93.
94 Leung DG, Cohen JS, Michelle EH, Bai R, Mammen AL and Christopher-Stine L (2018) Mitochondrial DNA deletions with low-level heteroplasmy in adult-onset myopathy. *J Clin Neuromuscul Dis* **19**, 117–123.

95 Tan TY, Lunke S, Chong B, Phelan D, Fanjul-Fernandez M, Marum J, Kumar VS, Stark Z, Yeung A, Brown NJ *et al.* (2019) A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. *Eur J Hum Genet* **27**, 1632–1633.
96 Lelieveld SH, Reijnders MRF, Pfundt R, Yntema HG, Kamsteeg EJ, de Vries P, de Vries BBA, Willemsen MH, Kleefstra T, Lohner K *et al.* (2016) Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci* **19**, 1194–1196.
97 Wang HJ, Qian YY, Lu YL, Qin Q, Lu GP, Cheng GQ, Zhang P, Yang L, Wu BB and Zhou WH (2020) Clinical utility of 24-h rapid trio-exome sequencing for critically ill infants. *Npj Genom Med* **5**, 20.
98 Acuna-Hidalgo R, Veltman JA and Hoischen A (2016) New insights into the generation and role of *de novo* mutations in health and disease. *Genome Biol* **17**, 241.
99 Harel T, Yoon WH, Garone C, Gu S, Coban-Akdemir Z, Eldomery MK, Posey JE, Jhangiani SN, Rosenfeld JA, Cho MT *et al.* (2016) Recurrent *de novo* and biallelic variation of ATAD3A, encoding a mitochondrial membrane protein, results in distinct neurological syndromes. *Am J Hum Genet* **99**, 831–845.
100 Thompson K, Majd H, Dallabona C, Reinson K, King MS, Alston CL, He L, Lodi T, Jones SA, Fattal-Valevski A *et al.* (2016) Recurrent *de novo* dominant mutations in SLC25A4 cause severe early-onset mitochondrial disease and loss of mitochondrial DNA copy number. *Am J Hum Genet* **99**, 860–876.
101 Ehmke N, Graul-Neumann L, Smorag L, Koenig R, Segebrecht L, Magoulas P, Scaglia F, Kilic E, Hennig AF, Adolphs N *et al.* (2017) *De novo* mutations in SLC25A24 cause a craniosynostosis syndrome with hypertrichosis, progeroid appearance, and mitochondrial dysfunction. *Am J Hum Genet* **101**, 833–843.
102 Fahrner JA, Liu R, Perry MS, Klein J and Chan DC (2016) A novel *de novo* dominant negative mutation in DNM1L impairs mitochondrial fission and presents as childhood epileptic encephalopathy. *Am J Med Genet A* **170**, 2002–2011.
103 Beck DB, Cho MT, Millan F, Yates C, Hannibal M, O'Connor B, Shinawi M, Connolly AM, Waggoner D, Halbach S *et al.* (2016) A recurrent *de novo* CTBP1 mutation is associated with developmental delay, hypotonia, ataxia, and tooth enamel defects. *Neurogenetics* **17**, 173–178.
104 Legati A, Reyes A, Ceccatelli Berti C, Stehling O, Marchet S, Lamperti C, Ferrari A, Robinson AJ, Muhlenhoff U, Lill R *et al.* (2017) A novel *de novo* dominant mutation in ISCU associated with mitochondrial myopathy. *J Med Genet* **54**, 815–824.
105 Karczewski KJ and Snyder MP (2018) Integrative omics for health and disease. *Nat Rev Genet* **19**, 299–310.
106 Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD and Craig DW (2016) Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* **17**, 257–271.
107 Ma M, Ru Y, Chuang LS, Hsu NY, Shi LS, Hakenberg J, Cheng WY, Uzilov A, Ding W, Glicksberg BS *et al.* (2015) Disease-associated variants in different categories of disease located in distinct regulatory elements. *BMC Genom* **16** (Suppl 8), S3.
108 Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD and Cooper DN (2017) The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* **136**, 665–677.
109 Soemedi R, Cygan KJ, Rhine CL, Wang J, Bulacan C, Yang J, Bayrak-Toydemir P, McDonald J and Fairbrother WG (2017) Pathogenic variants that alter protein code often disrupt splicing. *Nat Genet* **49**, 848–855.
110 Fresard L, Smail C, Ferraro NM, Teran NA, Li X, Smith KS, Bonner D, Kernohan KD, Marwaha S, Zappala Z *et al.* (2019) Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med* **25**, 911–919.
111 Gonorazky HD, Naumenko S, Ramani AK, Nelakuditi V, Mashouri P, Wang P, Kao D, Ohri K, Viththiyapaskaran S, Tarnopolsky MA *et al.* (2019) Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am J Hum Genet* **104**, 1007.
112 Lee H, Huang AY, Wang LK, Yoon AJ, Renteria G, Eskin A, Signer RH, Dorrani N, Nieves-Rodriguez S, Wan J *et al.* (2019) Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med* **22**, 490–499.
113 Murdock DR, Dai H, Burrage LC, Rosenfeld JA, Ketkar S, Muller MF, Yepez VA, Gagneur J, Liu P, Chen S *et al.* (2020) Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J Clin Invest*.
114 Kremer LS, Wortmann SB and Prokisch H (2018) "Transcriptomics": molecular diagnosis of inborn errors of metabolism via RNA-sequencing. *J Inherit Metab Dis* **41**, 525–532.
115 Aicher JK, Jewell P, Vaquero-Garcia J, Barash Y and Bhoj EJ (2020) Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet Med* **22**, 1181–1190.
116 Sterneckert JL, Reinhardt P and Scholer HR (2014) Investigating human disease using stem cell models. *Nat Rev Genet* **15**, 625–639.
117 Danhauser K, Iuso A, Haack TB, Freisinger P, Brockmann K, Mayr JA, Meitinger T and Prokisch H

(2011) Cellular rescue-assay aids verification of causative DNA-variants in mitochondrial complex I deficiency. *Mol Genet Metab* **103**, 161–166.
118 Gagneur J, Stegle O, Zhu CC, Jakob P, Tekkedil MM, Aiyar RS, Schuon AK, Pe'er D and Steinmetz LM (2013) Genotype-environment interactions reveal causal pathways that mediate genetic effects on phenotype. *Plos Genet* **9**, e1003803.
119 Brechtmann F, Mertes C, Matuseviciute A, Yepez VA, Avsec Z, Herzog M, Bader DM, Prokisch H and Gagneur J (2018) OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA sequencing data. *Am J Hum Genet* **103**, 907–917.
120 Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK and Pritchard JK (2018) Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**, 151–158.Fraser
121 Ferraro NM, Strober BJ, Einson J, Abell NS, Aguet F, Barbeira AN, Brandt M, Bucan M, Castel SE, Davis JR *et al*. (2020) Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science* **369**, eaaz5900.
122 Mertes C, Scheller I, Yepez V, Çelik M, Liang Y, Kremer L, Gusic M, Prokisch H and Gagneur J (2021) Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun* **12**, 529.
123 Mohammadi P, Castel SE, Cummings BB, Einson J, Sousa C, Hoffman P, Donkervoort S, Jiang Z, Mohassel P, Foley AR *et al*. (2019) Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science* **366**, 351–356.
124 Yépez VA, Mertes C, Müller M, Andrade D, Wachutka L, Fresard L, Gusic M, Scheller I, Goldberg P, Prokisch H *et al*. (2021) Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc* **16**, 1276–1296.
125 Wang D, Eraslan B, Wieland T, Hallstrom B, Hopf T, Zolg DP, Zecha J, Asplund A, Li LH, Meng C *et al*. (2019) A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol* **15**, e8503.
126 Jiang L, Wang M, Lin S, Jian R, Li X, Chan J, Dong G, Fang H, Robinson AE, GTEx Consortium and Snyder MP (2020) A quantitative proteome map of the human body. *Cell* **183**, 269–283.e19.
127 Stroud DA, Surgenor EE, Formosa LE, Reljic B, Frazier AE, Dibley MG, Osellame LD, Stait T, Beilharz TH, Thorburn DR *et al*. (2016) Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature* **538**, 123–126.
128 Kopajtich R, Smirnov D, Loipfinger S, Meng C, Ghezzi D, Murayama K, Mayr JA, Freisinger P, Metodiev MD, Rötig A *et al*. (2019) Multi-Omics integration for molecular diagnostics of mendelian disorders. ASHG Meeting 2019, Houston, USA.
129 Lake NJ, Webb BD, Stroud DA, Richman TR, Ruzzenente B, Compton AG, Mountford HS, Pulman J, Zangarelli C, Rio M *et al*. (2019) Biallelic mutations in MRPS34 lead to instability of the small mitoribosomal subunit and Leigh syndrome. *Eur J Hum Genet* **27**, 860–861.
130 Stenton SL, Sheremet NL, Catarino CB, Andreeva NA, Assouline Z, Barboni P, Barel O, Berutti R, Bychkov I, Caporali L *et al*. (2021) Impaired complex I repair causes recessive Leber's hereditary optic neuropathy. *J Clin Invest* 131, e138267.
131 Sahni N, Yi S, Taipale M, Bass JIF, Coulombe-Huntington J, Yang F, Peng J, Weile J, Karras GI, Wang Y *et al*. (2015) Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647–660.
132 Fragoza R, Das J, Wierbowski SD, Liang J, Tran TN, Liang SQ, Beltran JF, Rivera-Erick CA, Ye KX, Wang TY *et al*. (2019) Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations. *Nat Commun* **10**, 4141.
133 Johnson CH, Ivanisevic J and Siuzdak G (2016) Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* **17**, 451–459.
134 Tolstikov V, Moser AJ, Sarangarajan R, Narain NR and Kiebish MA (2020) Current status of metabolomic biomarker discovery: impact of study design and demographic characteristics. *Metabolites* **10**, 224.
135 Johnson CH, Ivanisevic J, Benton HP and Siuzdak G (2015) Bioinformatics: the next frontier of metabolomics. *Anal Chem* **87**, 147–156.
136 Alaimo JT, Glinton KE, Liu N, Xiao J, Yang Y, Sutton VR and Elsea SH (2020) Integrated analysis of metabolomic profiling and exome data supplements sequence variant interpretation, classification, and diagnosis. *Genet Med* **22**, 1560–1566.
137 Esterhuizen K, van der Westhuizen FH and Louw R (2017) Metabolomics of mitochondrial disease. *Mitochondrion* **35**, 97–110.
138 Lehtonen JM, Forsstrom S, Bottani E, Viscomi C, Baris OR, Isoniemi H, Hockerstedt K, Osterlund P, Hurme M, Jylhava J *et al*. (2016) FGF21 is a biomarker for mitochondrial translation and mtDNA maintenance disorders. *Neurology* **87**, 2290–2299.
139 Montero R, Yubero D, Villarroya J, Henares D, Jou C, Rodriguez MA, Ramos F, Nascimento A, Ortez CI, Campistol J *et al*. (2016) GDF-15 is elevated in children with mitochondrial diseases and is induced by mitochondrial dysfunction. *PLoS One* **11**, e0148709.
140 Reinecke CJ, Koekemoer G, van der Westhuizen FH, Louw R, Lindeque JZ, Mienie LJ and Smuts I (2012) Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics* **8**, 264–283.

141 Venter L, Lindeque Z, van Rensburg PJ, van der Westhuizen F, Smuts I and Louw R (2015) Untargeted urine metabolomics reveals a biosignature for muscle respiratory chain deficiencies. *Metabolomics* **11**, 111–121.
142 Buzkova J, Nikkanen J, Ahola S, Hakonen AH, Sevastianova K, Hovinen T, Yki-Jarvinen H, Pietilainen KH, Lonnqvist T, Velagapudi V *et al.* (2018) Metabolomes of mitochondrial diseases and inclusion body myositis patients: treatment targets and biomarkers. *Embo Mol Med* **10**, e9091.
143 de la Barca JMC, Fogazza M, Rugolo M, Chupin S, Del Dotto V, Ghelli AM, Carelli V, Simard G, Procaccio V, Bonneau D *et al.* (2020) Metabolomics hallmarks OPA1 variants correlating with their *in vitro* phenotype and predicting clinical severity. *Hum Mol Genet* **29**, 1319–1329.
144 Coene KLM, Kluijtmans LAJ, van der Heeft E, Engelke UFH, de Boer S, Hoegen B, Kwast HJT, van de Vorst M, Huigen MCDG, Keularts IMLW *et al.* (2018) Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients. *J Inherit Metab Dis* **41**, 337–353.
145 Pinu FR, Goldansaz SA and Jaine J (2019) Translational metabolomics: current challenges and future opportunities. *Metabolites* **9**, 108.
146 Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK *et al.* (2011) Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65.
147 Ashley EA (2016) Towards precision medicine. *Nat Rev Genet* **17**, 507–522.
148 van Dijk EL, Jaszczyszyn Y, Naquin D and Thermes C (2018) The third revolution in sequencing technology. *Trends Genet* **34**, 666–681.
149 Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM, Bartlett JJ, Pavlenok M and Gundlach JH (2013) Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci USA* **110**, 18904–18909.
150 Yang Y, Sebra R, Pullman BS, Qiao WQ, Peter I, Desnick RJ, Geyer CR, DeCoteau JF and Scott SA (2015) Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). *BMC Genom* **16**, 350.
151 Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME and Gouil Q (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**, 30.
152 Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138.
153 Hebert PDN, Braukmann TWA, Prosser SWJ, Ratnasingham S, deWaard JR, Ivanova NV, Janzen DH, Hallwachs W, Naik S, Sones JE *et al.* (2018) A Sequel to Sanger: amplicon sequencing that scales. *BMC Genom* **19**, 219.
154 Travers KJ, Chin CS, Rank DR, Eid JS and Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* **38**, e159.
155 Jain M, Olsen HE, Paten B and Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* **17**, 239.
156 Krishnakumar R, Sinha A, Bird SW, Jayamohan H, Edwards HS, Schoeniger JS, Patel KD, Branda SS and Bartsch MS (2018) Systematic and stochastic influences on the performance of the MinION nanopore sequencer across a range of nucleotide bias. *Sci Rep* **8**, 3159.
157 Madoui MA, Engelen S, Cruaud C, Belser C, Bertrand L, Alberti A, Lemainque A, Wincker P and Aury JM (2015) Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genom* **16**, 327.
158 Byrne A, Beaudin AE, Olsen HE, Jain M, Cole C, Palmer T, DuBois RM, Forsberg EC, Akeson M and Vollmers C (2017) Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat Commun* **8**, 16027.
159 Beaulaurier J, Zhang XS, Zhu S, Sebra R, Rosenbluh C, Deikus G, Shen N, Munera D, Waldor MK, Chess A *et al.* (2015) Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. *Nat Commun* **6**, 7438.
160 Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND *et al.* (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* **37**, 1155–1162.
161 Mantere T, Kersten S and Hoischen A (2019) Long-read sequencing emerging in medical genetics. *Front Genet* **10**, 426.
162 Loomis EW, Eid JS, Peluso P, Yin J, Hickey L, Rank D, McCalmon S, Hagerman RJ, Tassone F and Hagerman PJ (2013) Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome Res* **23**, 121–128.
163 Albrecht V, Zweiniger C, Surendranath V, Lang K, Schofl G, Dahl A, Winkler S, Lange V, Bohme I and Schmidt AH (2017) Dual redundant sequencing strategy: full-length gene characterisation of 1056 novel and confirmatory HLA alleles. *HLA* **90**, 79–87.
164 Wilbe M, Gudmundsson S, Johansson J, Ameur A, Stattin EL, Anneren G, Malmgren H, Frykholm C and Bondeson ML (2017) A novel approach using

long-read sequencing and ddPCR to investigate gonadal mosaicism and estimate recurrence risk in two families with developmental disorders. *Prenatal Diag* **37**, 1146–1154.
165 Gudmundsson S, Wilbe M, Ekvall S, Ameur A, Cahill N, Alexandrov LB, Virtanen M, Pigg MH, Vahlquist A, Torma H *et al.* (2017) Revertant mosaicism repairs skin lesions in a patient with keratitis-ichthyosis-deafness syndrome by second-site mutations in connexin 26. *Hum Mol Genet* **26**, 1070–1077.
166 Borras DM, Vossen RHAM, Liem M, Buermans HPJ, Dauwerse H, van Heusden D, Gansevoort RT, den Dunnen JT, Janssen B, Peters DJM *et al.* (2017) Detecting PKD1 variants in polycystic kidney disease patients by single-molecule long-read sequencing. *Hum Mutat* **38**, 870–879.
167 Carvalho CMB and Lupski JR (2016) Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* **17**, 224–238.
168 Cooper HM, Yang Y, Ylikallio E, Khairullin R, Woldegebriel R, Lin KL, Euro L, Palin E, Wolf A, Trokovic R *et al.* (2017) ATPase-deficient mitochondrial inner membrane protein ATAD3A disturbs mitochondrial dynamics in dominant hereditary spastic paraplegia. *Hum Mol Genet* **26**, 1432–1443.
169 Peralta S, Gonzalez-Quintana A, Ybarra M, Delmiro A, Perez-Perez R, Docampo J, Arenas J, Blazquez A, Ugalde C and Martin MA (2019) Novel ATAD3A recessive mutation associated to fatal cerebellar hypoplasia with multiorgan involvement and mitochondrial structural abnormalities. *Mol Genet Metab* **128**, 452–462.
170 Desai R, Frazier AE, Durigon R, Patel H, Jones AW, Rosa ID, Lake NJ, Compton AG, Mountford HS, Tucker EJ *et al.* (2017) ATAD3 gene cluster deletions cause cerebellar dysfunction associated with altered mitochondrial DNA and cholesterol metabolism. *Brain* **140**, 1595–1610.
171 Gunning AC, Strucinska K, Oreja MM, Parrish A, Caswell R, Stals KL, Durigon R, Durlacher-Betzer K, Cunningham MH, Grochowski CM *et al.* (2020) Recurrent *de novo* NAHR reciprocal duplications in the ATAD3 gene cluster cause a neurogenetic trait with perturbed cholesterol and mitochondrial metabolism. *Am J Hum Genet* **106**, 272–279.
172 Frazier AE, Compton AG, Kishita Y, Hock DH, Welch AE, Amarasekera SSC, Rius R, Formosa LE, Imai-Okazaki A, Francis D *et al.* (2020) Fatal perinatal mitochondrial cardiac failure caused by recurrent *de novo* duplications in the ATAD locus. *Med* **2**, 49.
173 Wood E, Parker MD, Dunning MJ, Hesketh S, Wang D, Pink R and Fratter C (2019) Clinical long-read sequencing of the human mitochondrial genome for mitochondrial disease diagnostics. *bioRxiv*. "[PREPRINT]" https://doi.org/10.1101/597187
174 Sedlazeck FJ, Lee H, Darby CA and Schatz MC (2018) Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet* **19**, 329–346.
175 Watson M and Warr A (2019) Errors in long-read assemblies can critically affect protein prediction. *Nat Biotechnol* **37**, 124–126.
176 Huang S (2009) Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853–3862.
177 Li L and Clevers H (2010) Coexistence of quiescent and active adult stem cells in mammals. *Science* **327**, 542–545.
178 Lasken RS (2007) Single-cell genomic sequencing using multiple displacement amplification. *Curr Opin Microbiol* **10**, 510–516.
179 Zong C, Lu S, Chapman AR and Xie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626.
180 Gawad C, Koh W and Quake SR (2016) Single-cell genome sequencing: current state of the science. *Nat Rev Genet* **17**, 175–188.
181 Hwang B, Lee JH and Bang D (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* **50**, 96.
182 Wen L and Tang F (2018) Boosting the power of single-cell analysis. *Nat Biotechnol* **36**, 408–409.
183 Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S *et al.* (2017) Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573.
184 Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D *et al.* (2011) Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94.
185 Wang JB, Fan HC, Behr B and Quake SR (2012) Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. *Cell* **150**, 402–412.
186 Hou Y, Fan W, Yan LY, Li R, Lian Y, Huang J, Li JS, Xu LY, Tang FC, Xie XS *et al.* (2013) Genome analyses of single human oocytes. *Cell* **155**, 1492–1506.
187 Cai XY, Evrony GD, Lehmann HS, Elhosary PC, Mehta BK, Poduri A and Walsh CA (2014) Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep* **8**, 1280–1289.
188 Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, Pelka K, Ge W, Oren Y, Brack A *et al.* (2019) Lineage tracing in humans enabled by

mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339.
189 Chen H, Ye F and Guo G (2019) Revolutionizing immunology with single-cell RNA sequencing. *Cell Mol Immunol* **16**, 242–249.
190 Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, Wildberg A, Gao D, Fung HL, Chen S *et al.* (2016) Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590.
191 Zhong S, Zhang S, Fan X, Wu Q, Yan L, Dong J, Zhang H, Li L, Sun L, Pan N *et al.* (2018) A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528.
192 Velmeshev D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, Bhaduri A, Goyal N, Rowitch DH and Kriegstein AR (2019) Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689.
193 Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, Menon M, He L, Abdurrob F, Jiang X *et al.* (2019) Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337.
194 Chaudhry F, Isherwood J, Bawa T, Patel D, Gurdziel K, Lanfear DE, Ruden DM and Levy PD (2019) Single-cell RNA sequencing of the cardiovascular system: new looks for old diseases. *Front Cardiovasc Med* **6**, 173.
195 Szabo PA, Levitin HM, Miron M, Snyder ME, Senda T, Yuan J, Cheng YL, Bush EC, Dogra P, Thapa P *et al.* (2019) Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat Commun* **10**, 4706.
196 Gonzalez-Silva L, Quevedo L and Varela I (2020) Tumor functional heterogeneity unraveled by scRNA-seq technologies. *Trends Cancer* **6**, 13–19.
197 Nguyen QH, Pervolarakis N, Nee K and Kessenbrock K (2018) Experimental considerations for single-cell RNA sequencing approaches. *Front Cell Dev Biol* **6**, 108.
198 Lahnemann D, Koster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Campbell KR, Beerenwinkel N, Mahfouz A *et al.* (2020) Eleven grand challenges in single-cell data science. *Genome Biol* **21**, 31.
199 Zhang MJ, Ntranos V and Tse D (2020) Determining sequencing depth in a single-cell RNA-seq experiment. *Nat Commun* **11**, 774.
200 Tang X, Huang Y, Lei J, Luo H and Zhu X (2019) The single-cell sequencing: new developments and medical applications. *Cell Biosci* **9**, 53.
201 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al.* (2009) Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.
202 Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195.
203 Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, Guigo R, Iossifov I, Vasileva A and Lappalainen T (2018) Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat Genet* **50**, 1327–1334.
204 Niemi MEK, Martin HC, Rice DL, Gallone G, Gordon S, Kelemen M, McAloney K, McRae J, Radford EJ, Yu S *et al.* (2018) Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* **562**, 268–271.
205 Gloss BS and Dinger ME (2018) Realizing the significance of noncoding functionality in clinical genomics. *Exp Mol Med* **50**, 1–8.
206 Mathelier A, Shi WQ and Wasserman WW (2015) Identification of altered cis-regulatory elements in human disease. *Trends Genet* **31**, 67–76.
207 Zhang F and Lupski JR (2015) Non-coding genetic variants in human disease. *Hum Mol Genet* **24**, R102–R110.
208 Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE and de Graaff E (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725–1735.
209 Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, Rodriguez-Segui SA, Shaw-Smith C, Cho CHH, Allen HL *et al.* (2014) Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* **46**, 61–64.
210 Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Hom D, Kayserili H, Opitz JM, Laxova R *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025.
211 Short PJ, Mcrae JF, Gallone G, Sifrim A, Won H, Geschwind DH, Wright CF, Firth HV, FitzPatrick DR, Barrett JC *et al.* (2018) *De novo* mutations in regulatory elements in neurodevelopmental disorders. *Nature* **555**, 611–616.
212 Gebert LFR and MacRae IJ (2019) Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol* **20**, 21–37.
213 Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their

gene structure, evolution, and expression. *Genome Res* **22**, 1775–1789.
214 Ragan C, Goodall GJ, Shirokikh NE and Preiss T (2019) Insights into the biogenesis and potential functions of exonic circular RNA. *Sci Rep* **9**, 2048.
215 Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F *et al*. (2012) Landscape of transcription in human cells. *Nature* **489**, 101–108.
216 Gusic M and Prokisch H (2020) ncRNAs: new players in mitochondrial health and disease? *Front Genet* **11**, 95.
217 Kopp F and Mendell JT (2018) Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407.
218 Ramilowski JA, Yip CW, Agrawal S, Chang JC, Ciani Y, Kulakovskiy IV, Mendez M, Ooi JLC, Ouyang JF, Parkinson N *et al*. (2020) Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome Res* **30**, 1060–1072.
219 Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, He D, Attenello FJ, Villalta JE, Cho MY, Chen YW *et al*. (2017) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, aah7111.
220 Liu Y, Cao Z, Wang Y, Guo Y, Xu P, Yuan P, Liu Z, He Y and Wei W (2018) Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat Biotechnol* **36**, 1203–1210.
221 Mendell JT and Olson EN (2012) MicroRNAs in stress signaling and human disease. *Cell* **148**, 1172–1187.
222 Cheetham SW, Gruhl F, Mattick JS and Dinger ME (2013) Long noncoding RNAs and the genetics of cancer. *Br J Cancer* **108**, 2419–2425.
223 Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, Olavarrieta L, Aguirre LA, del Castillo I, Steel KP, Dalmay T *et al*. (2009) Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet* **41**, 609–613.
224 de Pontual L, Yao E, Callier P, Faivre L, Drouin V, Cariou S, Van Haeringen A, Genevieve D, Goldenberg A, Oufadem M *et al*. (2011) Germline deletion of the miR-17 approximately 92 cluster causes skeletal and growth defects in humans. *Nat Genet* **43**, 1026–1030.
225 Hughes AE, Bradley DT, Campbell M, Lechner J, Dash DP, Simpson DA and Willoughby CE (2011) Mutation altering the miR-184 seed region causes familial keratoconus with cataract. *Am J Hum Genet* **89**, 628–633.
226 Ang CE, Ma Q, Wapinski OL, Fan S, Flynn RA, Lee QY, Coe B, Onoguchi M, Olmos VH, Do BT *et al*. (2019) The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *Elife* **8**, e41770.
227 Sirey TM, Roberts K, Haerty W, Bedoya-Reina O, Rogatti-Granados S, Tan JY, Li N, Heather LC, Carter RN, Cooper S *et al*. (2019) The long non-coding RNA Cerox1 is a post transcriptional regulator of mitochondrial complex I catalytic activity. *Elife* **8**, e45051.
228 Ingolia NT (2016) Ribosome footprint profiling of translation throughout the genome. *Cell* **165**, 22–33.
229 Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H *et al*. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587.
230 van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, Kirchner M, Maatz H, Blachut S, Sandmann CL *et al*. (2019) The translational landscape of the human heart. *Cell* **178**, 242–260.e29.
231 Chen J, Brunner AD, Cogan JZ, Nunez JK, Fields AP, Adamson B, Itzhak DN, Li JY, Mann M, Leonetti MD *et al*. (2020) Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140–1146.
232 Yeasmin F, Yada T and Akimitsu N (2018) Micropeptides encoded in transcripts previously identified as long noncoding RNAs: a new chapter in transcriptomics and proteomics. *Front Genet* **9**, 144.
233 Iacono M, Mignone F and Pesole G (2005) uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* **349**, 97–105.
234 Johnstone TG, Bazzini AA and Giraldez AJ (2016) Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J* **35**, 706–723.
235 Wang S, Mao C and Liu S (2019) Peptides encoded by noncoding genes: challenges and perspectives. *Signal Transduct Target Ther* **4**, 57.
236 Hashimoto Y, Niikura T, Tajima H, Yasukawa T, Sudo H, Ito Y, Kita Y, Kawasumi M, Kouyama K, Doyu M *et al*. (2001) A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Abeta. *Proc Natl Acad Sci USA* **98**, 6336–6341.
237 Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim SJ, Mehta H, Hevener AL, de Cabo R *et al*. (2015) The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab* **21**, 443–454.
238 Cobb LJ, Lee C, Xiao J, Yen K, Wong RG, Nakamura HK, Mehta HH, Gao Q, Ashur C, Huffman DM *et al*. (2016) Naturally occurring mitochondrial-derived peptides are age-dependent regulators of apoptosis, insulin sensitivity, and inflammatory markers. *Aging* **8**, 796–809.

239 Stein CS, Jadiya P, Zhang X, McLendon JM, Abouassaly GM, Witmer NH, Anderson EJ, Elrod JW and Boudreau RL (2018) Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency. *Cell Rep* **23**, 3710–3720.e8.
240 Grady JP, Pickett SJ, Ng YS, Alston CL, Blakely EL, Hardy SA, Feeney CL, Bright AA, Schaefer AM, Gorman GS *et al.* (2018) mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease. *EMBO Mol Med* **10**, e8262.
241 Bianco A, Martinez-Romero I, Bisceglia L, D'Agruma L, Favia P, Ruiz-Pesini E, Guerriero S, Montoya J and Petruzzella V (2016) Mitochondrial DNA copy number differentiates the Leber's hereditary optic neuropathy affected individuals from the unaffected mutation carriers. *Brain* **139**, e1.
242 Ardissone A, Invernizzi F, Nasca A, Moroni I, Farina L and Ghezzi D (2015) Mitochondrial leukoencephalopathy and complex II deficiency associated with a recessive SDHB mutation with reduced penetrance. *Mol Genet Metab Rep* **5**, 51–54.
243 Kousi M and Katsanis N (2015) Genetic modifiers and oligogenic inheritance. *Cold Spring Harb Perspect Med* **5**, a017145.
244 Kullar PJ, Gomez-Duran A, Gammage PA, Garone C, Minczuk M, Golder Z, Wilson J, Montoya J, Hakli S, Karppa M *et al.* (2018) Heterozygous SSBP1 start loss mutation co-segregates with hearing loss and the m.1555A>G mtDNA variant in a large multigenerational family. *Brain* **141**, 55–62.
245 Yu J, Liang X, Ji Y, Ai C, Liu J, Zhu L, Nie Z, Jin X, Wang C, Zhang J *et al.* (2020) PRICKLE3 linked to ATPase biogenesis manifested Leber's hereditary optic neuropathy. *J Clin Invest* **130**, 4935–4946.
246 Jiang P, Jin X, Peng Y, Wang M, Liu H, Liu X, Zhang Z, Ji Y, Zhang J, Liang M *et al.* (2016) The exome sequencing identified the mutation in YARS2 encoding the mitochondrial tyrosyl-tRNA synthetase as a nuclear modifier for the phenotypic manifestation of Leber's hereditary optic neuropathy-associated mitochondrial DNA mutation. *Hum Mol Genet* **25**, 584–596.
247 Fan W, Zheng J, Kong W, Cui L, Aishanjiang M, Yi Q, Wang M, Cang X, Tang X, Chen Y *et al.* (2019) Contribution of a mitochondrial tyrosyl-tRNA synthetase mutation to the phenotypic expression of the deafness-associated tRNA(Ser(UCN)) 7511A>G mutation. *J Biol Chem* **294**, 19292–19305.
248 Galassi G, Lamantea E, Invernizzi F, Tavani F, Pisano I, Ferrero I, Palmieri L and Zeviani M (2008) Additive effects of POLG1 and ANT1 mutations in a complex encephalomyopathy. *Neuromuscul Disord* **18**, 465–470.
249 Hathazi D, Griffin H, Jennings MJ, Giunta M, Powell C, Pearce SF, Munro B, Wei W, Boczonadi V, Poulton J *et al.* (2020) Metabolic shift underlies recovery in reversible infantile respiratory chain deficiency. *EMBO J* **39**, e105364.
250 Ballana E, Govea N, de Cid R, Garcia C, Arribas C, Rosell J and Estivill X (2008) Detection of unrecognized low-level mtDNA heteroplasmy may explain the variable phenotypic expressivity of apparently homoplasmic mtDNA mutations. *Hum Mutat* **29**, 248–257.
251 Hudson G, Carelli V, Spruijt L, Gerards M, Mowbray C, Achilli A, Pyle A, Elson J, Howell N, La Morgia C *et al.* (2007) Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background. *Am J Hum Genet* **81**, 228–233.
252 Oetjens MT, Kelly MA, Sturm AC, Martin CL and Ledbetter DH (2019) Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. *Nat Commun* **10**, 4897.
253 Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT *et al.* (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219–1224.
254 Khera AV, Chaffin M, Zekavat SM, Collins RL, Roselli C, Natarajan P, Lichtman JH, D'Onofrio G, Mattera J, Dreyer R *et al.* (2019) Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. *Circulation* **139**, 1593–1602.
255 Lee HS and Park T (2017) Pathway-driven approaches of interaction between oxidative balance and genetic polymorphism on metabolic syndrome. *Oxid Med Cell Longev* **2017**, 6873197.
256 Chow J, Rahman J, Achermann JC, Dattani MT and Rahman S (2017) Mitochondrial disease and endocrine dysfunction. *Nat Rev Endocrinol* **13**, 92–104.
257 Ridler C (2016) Obesity: inheritance via mitochondria. *Nat Rev Endocrinol* **12**, 497.
258 Szendroedi J, Phielix E and Roden M (2011) The role of mitochondria in insulin resistance and type 2 diabetes mellitus. *Nat Rev Endocrinol* **8**, 92–103.
259 West AP and Shadel GS (2017) Mitochondrial DNA in innate immune responses and inflammatory pathology. *Nat Rev Immunol* **17**, 363–375.
260 Ridge PG and Kauwe JSK (2018) Mitochondria and Alzheimer's disease: the role of mitochondrial genetic variation. *Curr Genet Med Rep* **6**, 1–10.
261 Fetterman JL, Liu C, Mitchell GF, Vasan RS, Benjamin EJ, Vita JA, Hamburg NM and Levy D (2018) Relations of mitochondrial genetic variants to measures of vascular function. *Mitochondrion* **40**, 51–57.

262 Brown DA, Perry JB, Allen ME, Sabbah HN, Stauffer BL, Shaikh SR, Cleland JG, Colucci WS, Butler J, Voors AA *et al*. (2017) Expert consensus document: mitochondrial function as a therapeutic target in heart failure. *Nat Rev Cardiol* **14**, 238–250.

263 Kassam I, Qi T, Lloyd-Jones L, Holloway A, Jan Bonder M, Henders AK, Martin NG, Powell JE, Franke L, Montgomery GW *et al*. (2016) Evidence for mitochondrial genetic control of autosomal gene expression. *Hum Mol Genet* **25**, 5332–5338.

264 Cohen T, Levin L and Mishmar D (2016) Ancient Out-of-Africa mitochondrial DNA variants associate with distinct mitochondrial gene expression patterns. *PLoS Genet* **12**, e1006407.

265 Ali AT, Boehme L, Carbajosa G, Seitan VC, Small KS and Hodgkinson A (2019) Nuclear genetic regulation of the human mitochondrial transcriptome. *Elife* **8**, e41927.

266 Kraja AT, Liu C, Fetterman JL, Graff M, Have CTGu, Yanek C, Feitosa LR, Arking MF, Chasman DE, Young DI *et al*. (2019) Associations of mitochondrial and nuclear mitochondrial variants and genes with seven metabolic traits. *Am J Hum Genet* **104**, 112–138.

267 Garg P, Jadhav B, Rodriguez OL, Patel N, Martin-Trujillo A, Jain M, Metsu S, Olsen H, Paten B, Ritz B *et al*. (2020) A Survey of rare epigenetic variation in 23,116 human genomes identifies disease-relevant epivariations and CGG expansions. *Am J Hum Genet* **107**, 654–669.

268 Buiting K, Gross S, Lich C, Gillessen-Kaesbach G, El-Maarri O and Horsthemke B (2003) Epimutations in Prader-Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. *Am J Hum Genet* **72**, 571–577.

269 Lacroix AJ, Stabley D, Sahraoui R, Adam MP, Mehaffey M, Kernan K, Myers CT, Fagerstrom C, Anadiotis G, Akkari YM *et al*. (2019) GGC repeat expansion and exon 1 methylation of XYLT1 is a common pathogenic variant in Baratela-Scott syndrome. *Am J Hum Genet* **104**, 35–44.

270 Barbosa M, Joshi RS, Garg P, Martin-Trujillo A, Patel N, Jadhav B, Watson CT, Gibson W, Chetnik K, Tessereau C *et al*. (2018) Identification of rare *de novo* epigenetic variations in congenital disorders. *Nat Commun* **9**, 2064.

271 Garg P and Sharp AJ (2019) Screening for rare epigenetic variations in autism and schizophrenia. *Hum Mutat* **40**, 952–961.

272 Gueant JL, Chery C, Oussalah A, Nadaf J, Coelho D, Josse T, Flayac J, Robert A, Koscinski I, Gastin I *et al*. (2018) A PRDX1 mutant allele causes a MMACHC secondary epimutation in cblC patients. *Nat Commun* **9**, 67.

273 Russell OM, Gorman GS, Lightowlers RN and Turnbull DM (2020) Mitochondrial diseases: hope for the future. *Cell* **181**, 168–188.