The utility of phenomics in diagnosis of inherited metabolic disorders

Authors: Joyeeta Rahman[A] and Shamima Rahman[B]

ABSTRACT

Inherited metabolic disorders (IMDs) are debilitating inherited diseases, with phenotypic, biochemical and genetic heterogeneity, frequently leading to prolonged diagnostic odysseys. Mitochondrial disorders represent one of the most severe classes of IMDs, wherein defects in >350 genes lead to multisystem disease. Diagnostic rates have improved considerably following the adoption of next-generation sequencing (NGS) technologies, but are still far from perfect. Phenomic annotation is an emerging concept which is being utilised to enhance interpretation of NGS results. To test whether phenomic correlations have utility in mitochondrial disease and IMDs, we created a gene-to-phenotype interaction network with searchable elements, for Leigh syndrome, a frequently observed paediatric mitochondrial disorder. The Leigh Map comprises data on 92 genes and 275 phenotypes standardised in human phenotype ontology terms, with 80% predictive accuracy. This commentary highlights the usefulness of the Leigh Map and similar resources and the challenges associated with integrating phenomic technologies into clinical practice.

KEYWORDS: Mitochondrial disease, inherited metabolic disorders, rare disease, phenomics, Leigh syndrome

Introduction

Inherited metabolic disorders (IMDs) collectively represent one of the most common groups of Mendelian diseases, despite being individually rare (<1:2000) or ultra-rare disorders (<1:2,000,000).[1] Many genetic defects underpinning IMDs are enzyme deficiencies within vital biosynthetic and metabolic pathways. However, IMDs are increasingly recognised to include defects of structural proteins, trafficking proteins, chaperones and other regulatory proteins pertaining to a metabolic process.[2] IMDs may be associated with debilitating and often fatal phenotypic presentations including epilepsy, intellectual disability, cardiomyopathy, liver failure, myopathy, and immunodeficiency.[3,4] A crucial step in the clinical management of IMDs is obtaining a definitive genetic diagnosis, which has clinical, therapeutic,

and psychological implications for affected patients and their families (Fig 1). Despite their severity, several IMDs are amenable to clinical intervention such as dietary modification, administration of metabolites, or enzyme replacement therapy, which reduce mortality and morbidity and improve quality of life.[4] Unfortunately, patients with IMDs often undergo prolonged diagnostic odysseys because of the complexity of the clinical presentations and the underlying pathobiology typically associated with these disorders. A lack of distinguishing features, non-standardised newborn screening programmes, and a knowledge gap between front-line clinicians and metabolic experts are all obstacles to obtaining specific diagnoses. The genetic heterogeneity of IMDs is another challenging factor in that the most common groups of IMDs may be causally linked to several hundred gene defects. These include mitochondrial disorders, congenital disorders of glycosylation, and lysosomal storage disorders.[3,5,6]

In recent years the use of next-generation sequencing (NGS) technologies has helped to improve the diagnostic rate of IMDs and has exponentially increased the discovery rate of novel causative gene mutations.[7] As NGS is increasingly becoming less expensive and more rapid, it is unsurprising that it is emerging as the preferred method of genetic diagnosis in many centres around the world. Furthermore, the use of NGS can in many cases avoid the need for invasive diagnostic tests such as muscle or liver biopsy, limiting the exposure of vulnerable patients, especially children, to general anaesthesia, which in some cases may trigger metabolic decompensation. NGS, however, is not without its challenges, particularly when it comes to prioritising and analysing variants of unknown significance (VUS) or candidate disease genes for potential differential diagnosis. Variant annotation challenges are amplified in those disorders in which there is significant phenotypic heterogeneity. Furthermore, it may be labour intensive and time consuming to comb through knowledge bases or the compendium of literature to determine whether or not a variant is pathogenic. The use of phenomics, that is a complete characterisation of (preferably standardised) phenotypic manifestations associated with a particular gene defect or IMD could have utility in improving the diagnostic rate of difficult-to-diagnose IMDs.[8–10] Comprehensive knowledge bases with searchable elements and user-friendly interfaces are cost-effective and potentially efficacious additions to enhance the interpretation of NGS findings and can have secondary benefits as clinical management tools.

Mitochondrial disorders are among the most severe metabolic disorders wherein patients suffer from multi-systemic phenotypes,

Authors: [A]PhD student, UCL Great Ormond Street Institute of Child Health, London, UK; [B]professor of paediatric metabolic medicine, UCL Great Ormond Street Institute of Child Health, London, UK

Patient with clinical suspicion of an inherited metabolic disorder

HP: 0002415
HP: 0001252
HP: 0001263

HP: 0001639
HP: 0004415

HP: 0001399
HP: 0002240

Phenomic resources

Next-generation sequencing

Phenotyping

HP: 0002415
HP: 0001263
HP: 0004415
HP: 0001639
→
SURF1
POLG
PDHX
RRM2B
ALG9
RTX1
GLA
NPC1
BTD

OMIM   ClinVar

Predictive tools

Knowledge bases

Genetic diagnosis

Clinical management

Benefit for families

Treatment

Dietary restriction

Pharmacovigilance

Further testing

Preventative screening

Genetic counselling

Carrier screening
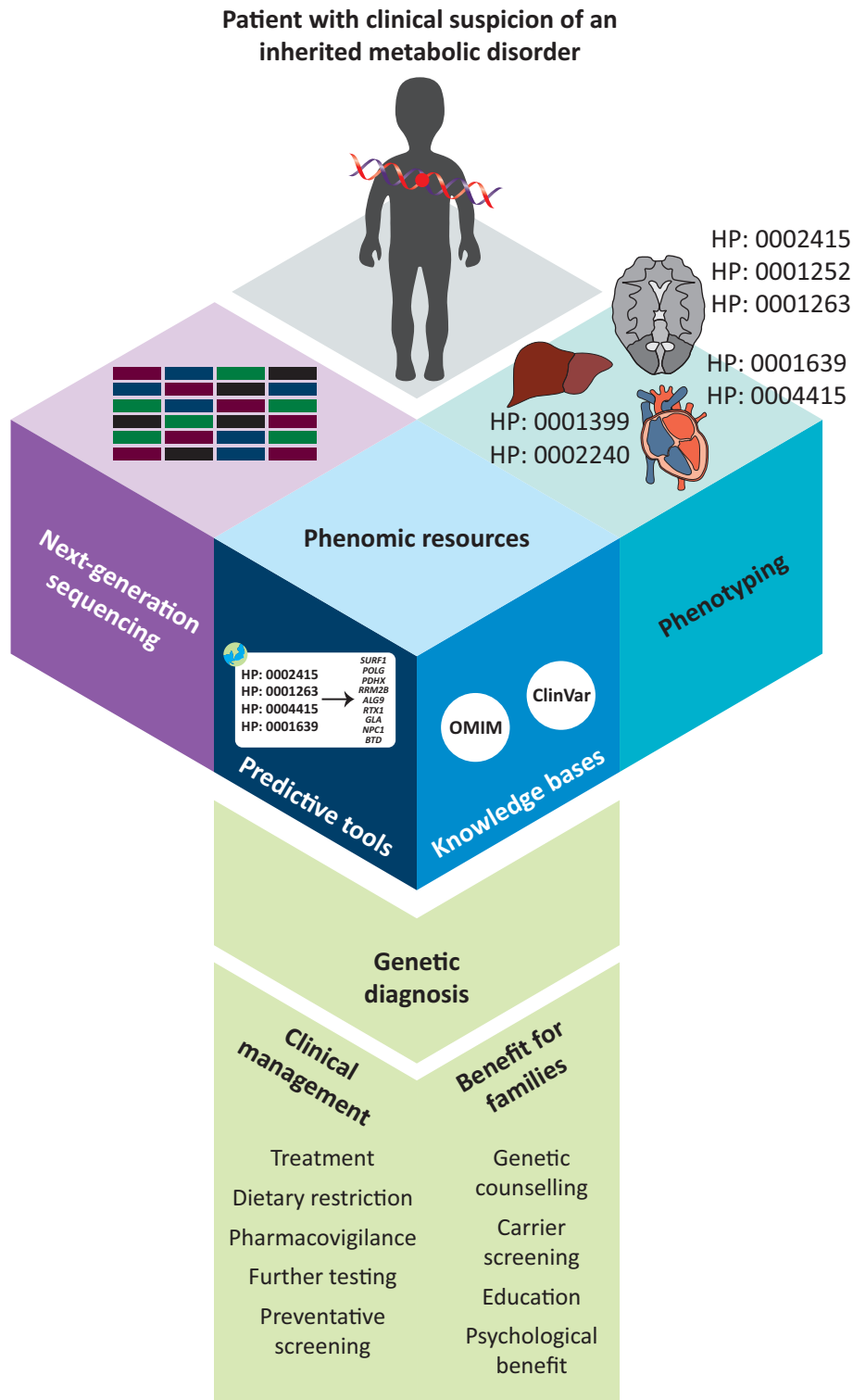
Education

Psychological benefit

Fig 1. The utility of phenomics in the inherited metabolic disease clinic. The use of phenomics resources, including predictive tools and rare disease knowledge bases can assist in obtaining a definitive genetic diagnosis when used in conjunction with next-generation sequencing tests such as whole exome or whole genome sequencing. Obtaining a genetic diagnosis improves patient care by providing treatment options, improved pharmacovigilance, and preventative screening for at-risk disorders such as cancer and neurodegenerative diseases. In addition, genetic diagnoses provide benefits to the family by facilitating accurate genetic counselling, as well as providing psychological benefit to parents.

often resulting in early death. Primary mitochondrial disorders are described as those which primarily or secondarily cause dysfunction of oxidative phosphorylation (OXPHOS) or other elements of mitochondria such as defects in vitamin and cofactor metabolism, aberrant mitochondrial dynamics, and abnormal mitochondrial lipid homeostasis.[3] Large-scale NGS gene panel studies currently place the diagnostic rate of mitochondrial disorders between ~35–60%.[11,12] The genetic heterogeneity of mitochondrial disorders and ongoing discovery of novel disease genes are caveats which can deem whole exome sequence (WES)

data unsuitable to provide clinicians with enough certainty for a definitive diagnosis in all cases.

We hypothesised that the use of phenomic annotation in mitochondrial disorders could be of diagnostic utility. Mitochondrial disorders are characterised by extreme phenotypic heterogeneity, affecting virtually any organ or system, seemingly in any combination. However, by documenting the phenotypes of reported patients with mitochondrial disease and mapping the associations to causative genes, any emerging patterns can enhance the interpretation of NGS to expedite genetic diagnoses for patients and elucidate novel gene-to-phenotype correlations. To test this strategy, we constructed a novel gene-to-phenotype interaction network using Leigh syndrome as a prototype.

Leigh Map: a case for phenomics-based tools in inherited metabolic disorders

Layout and validation of the Leigh Map

We chose Leigh syndrome as our prototype disorder because it is one of the most common presentations of paediatric mitochondrial disease. Patients frequently die within 2 years of initial presentation.[13] Leigh syndrome was initially defined neuropathologically by the presence of bilateral hyperintense lesions in the basal ganglia and/or brainstem.[14] Common clinical manifestations include psychomotor retardation, with regression, and progressive neurological abnormalities related to basal ganglia and/or brainstem dysfunction such as movement disorders, speech abnormalities, and defective muscle tone. Furthermore, the majority of patients also present with multisystemic (eg cardiac, hepatic, renal or haematological) phenotypes.[13] To date 92 genes are known to cause subtypes of Leigh and Leigh-like syndromes, the majority of which are difficult to distinguish from each other, either clinically or biochemically. We hypothesised that the multisystemic features may help to characterise different genetic subtypes of Leigh syndrome.

The Leigh Map is a freely available online diagnostic resource for Leigh syndrome (www.vmh.life/#leighmap).[10] It is a gene-to-phenotype interaction network encompassing all reported defective genes known to cause Leigh syndrome, and their associated phenotypic abnormalities. The Leigh Map can be utilised to enhance the interpretation of NGS to help prioritise variants, thereby enabling faster and more accurate genetic diagnoses for patients. The network was created by data mining an extensive knowledge base of more than 900 publications to extract genetic and phenotypic data pertaining to Leigh syndrome. In addition, the Leigh Map features additional clinically relevant data. The Leigh Map is built on the Molecular Interaction NEtwoRks VisuAlization (MINERVA) framework, a molecular modelling platform that mimics the user interface of Google Maps to navigate cellular, metabolic, and in our case, phenotypic networks.[15] Navigation through the Leigh Map is therefore just as one would navigate Google Maps in that zooming in on compartments reveals progressively more detailed information. To date, 92 causative Leigh syndrome genes are included in the Leigh Map, arranged according to submitochondrial location (inner and outer mitochondrial membranes, intermembrane space, or mitochondrial matrix) and function (OXPHOS, mitochondrial deoxyribonucleic acid maintenance, and mitochondrial dynamics, among others). Clicking on a gene will display relevant information about the gene in the left-hand panel including modes of inheritance, patient demographics, magnetic resonance imaging (MRI) findings, and links to external genetic and scientific databases. Furthermore, each genetic element in the Leigh Map has an accompanying sub-map which documents all phenotypes associated with that particular gene defect. Phenotypes are annotated in human phenotypic ontology (HPO) terms, a system for standardising clinical features that are synonyms.[16] For example, 'decreased muscle tone' or 'floppiness' can be expressed as one HPO term 'muscular hypotonia' (HP: 0001252). In total, the Leigh Map currently comprises 236 HPO phenotypes (Fig 2).

Content query is another Google Maps feature which is embedded within MINERVA. The Leigh Map features more than 1700 gene-to-phenotype interactions, all of which can be queried by the user using the search function. Querying the name or HPO number of a particular phenotype will result in a list of genes with which that phenotype has been associated. The Leigh Map supports the query of multiple phenotypes simultaneously, therefore this feature can be used in a diagnostic setting to elucidate causative disease genes. Conversely, genes can also be queried to browse phenotypes, which can be used as a clinical surveillance tool.[10]

Two non-clinical independent testers validated the Leigh Map, using unpublished test cases consisting of clinical and biochemical vignettes (excluding genetic information) of patients from a national mitochondrial disease clinic who had a prior established genetic diagnosis of Leigh syndrome. The testers were able to elucidate the correct gene as a candidate in 80% of the test cases. We defined a gene as a candidate if it contained >50% of the query phenotypes. Candidate gene lists typically comprised 8–10 Leigh Map genes, thereby eliminating 85–90% of the genes in the network. The failure to deduce the correct candidate genes for a minority of our test cases was due to the predominant presence of common Leigh syndrome phenotypes such as hypotonia, lactic acidosis, and developmental delay without more specific organ involvement. We found more success in 'diagnosing' patients with discriminating phenotypes, usually involving more than one system including cardiomyopathy, hepatopathy, and renal tubulopathy.

Limitations and future prospects

Given its success rate in predicting causative genes by non-clinical individuals, the Leigh Map is an efficacious diagnostic resource that, in combination with WES or whole genome sequence (WGS) data and metabolic testing in blood, urine and/or cerebrospinal fluid, can be used by clinicians to provide patients with accurate diagnoses or to guide further biochemical investigation.

Currently the most significant limitation of the Leigh Map is the lack of a multiple advanced search facility. Future work aims to implement this feature to the network, in addition to a scoring system so that candidate genes can be ranked more easily. Another issue is that the outputs from the Leigh Map are affected by the data inputs, ie the breadth of literature available for individual genes. For example, well-established or common gene defects such as *SURF1, POLG,* and *MT-TL1* are the subject of numerous case reports/publications while newer genes or genes with unclear functions have considerably fewer reports and associated phenotypes.

While there are no current curative therapies for mitochondrial disease, there are numerous compounds which are aimed at
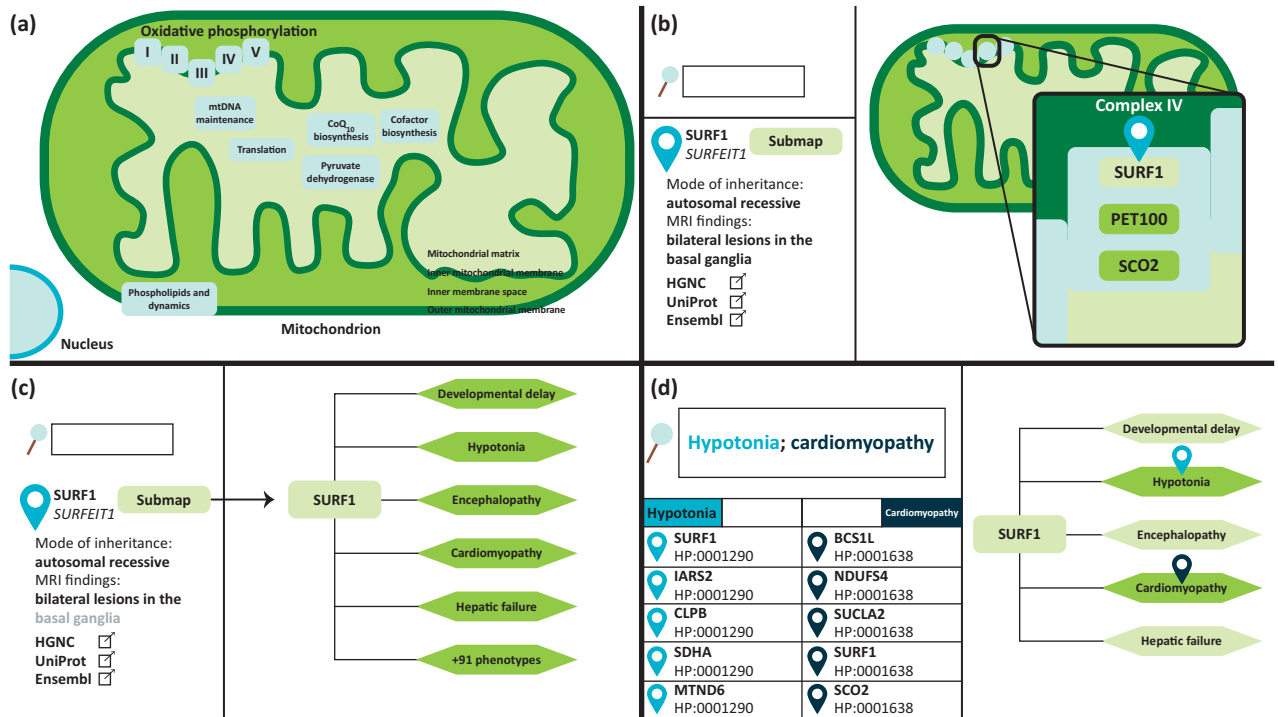
Fig 2. The Leigh Map. The Leigh Map is an interactive gene-to-phenotype interaction network which can be used as a diagnostic resource for Leigh syndrome. It is built on a platform which mimics the user interface of Google Maps. **(a)** Causative Leigh syndrome genes are arranged according to their mitochondrial location and function. **(b)** Zooming in and clicking on a gene will display further information, such as modes of inheritance, MRI findings, patient demographics, and external annotations in the left-hand panel. **(c)** In addition, each gene has an associated sub-map which displays all phenotypes ever reported in patients with a particular gene defect. **(d)** The genetic and phenotypic elements of the Leigh map can be queried. Multiple phenotypes can be queried when separated by a semicolon. Results will be displayed on the left-hand panel and clicking on the sub-map of a query result will provide the user with a visual representation of the query phenotypes.

symptomatic management, including anti-convulsant drugs and cofactor and vitamin supplements, such as coenzyme Q_{10}, thiamine, and biotin, used to treat corresponding deficiencies.[3] The addition of drug targets (a current feature of the MINERVA platform) to the Leigh Map could potentially provide insight into the effectiveness of various agents in treating mitochondrial disease in specific genetic contexts. This is useful for clinical management as certain mitochondrial diseases are potentially treatable.[17]

The computational nature of the Leigh Map allows for the addition of novel disease genes or phenotypes with relative ease, therefore clinicians have access to a database of all current causative genes, which can enhance the interpretation of WES data. This is especially beneficial within the context of mitochondrial diseases since novel genes are constantly being identified. In Leigh syndrome alone, one third of the causative genes were identified in the last five years.[18] Ideally, we will update both the phenotypic and genetic components of the Leigh Map concurrently with the literature. We are also exploring crowdsourcing options from clinicians and machine learning methods to keep the Leigh Map as up-to-date as possible.

Within the domain of mitochondrial disorders there are several phenotypes which have considerable genetic heterogeneity which would also benefit from the creation of specific diagnostic resources, as we have done with the Leigh Map. Our ongoing efforts are to create additional resources for these phenotype

groups, including mitochondrial epilepsy, cardiomyopathy, and liver disease. Ultimately, we also aim to create a large comprehensive diagnostic network to encompass the entirety of mitochondrial disease.

The utility of phenomics in IMD diagnosis

Benefits of phenomic databases

The adoption of phenomics into biomedical knowledge bases is beneficial and one can argue essential in the era of NGS. A caveat of WGS or WES outputs are VUS, which can number in the thousands. It is often difficult to disentangle pathogenic variants from variants caused by genetic variation in the population and naturally occurring individual mutations (an average of ~100 naturally occurring loss-of-function variants per genome).[8] The annotation of clinical abnormalities and their integration into the diagnostic pipeline therefore can identify functional groups of genes, which can help to pinpoint truly pathogenic variants and obtain definitive genetic diagnoses.

In recent years several gene-to-phenotype databases and/ or knowledge bases have emerged, including large-scale international expert-curated knowledge bases such as ClinVar and ClinGen.[19,20] However, there are also several niche expert phenomic resources similar to the Leigh Map, which were developed by experts within a specific disease field to enhance

Table 1. Current live phenomic knowledge bases and tools

Type of Resource	Resource Name	URL	Disease Area	Reference
Phenotype-genotype database	ClinGen	clinicalgenome.org	GD	[20]
	ClinVar	ncbi.nlm.nih.gov/clinvar	GD	[19]
	Online Mendelian Inheritance In Man	omim.org	GD	[34]
	Orphanet Rare Disease Ontology	orphadata.org/cgi-bin/inc/ordo_orphanet.inc.php	RGD	[35]
Niche phenomic knowledge base	Amyotrophic Lateral Sclerosis Online Genetics Database	alsod.iop.kcl.ac.uk	ALS	[36]
	Fanconi Anemia Mutation Database	www2.rockefeller.edu/fanconi	Fanconi anaemia	[22]
	IEMBase	iembase.org	IMDs	[23]
	Metagene	metagene.de	IMDs	NA
	Face2Gene	face2gene.com	IMDs	[33]
	Leigh Map	www.vmh.life/#leighmap	Leigh syndrome	[10]
	Osteogenesis Imperfecta Variant Database	oi.gene.le.ac.uk/home.php	Osteogenesis imperfecta	[37]
	PhenoDis	mips.helmholtz-muenchen.de/phenodis	Rare cardiac disorders	[21]
Predictive phenotypic tool	Exomizer	sanger.ac.uk/science/tools/exomiser	GD	[38]
	Find Zebra	findzebra.com	RGD	[29]
	FunSimMat	funsimmat.de	GD	[39]
	Gene Cards	genecards.org	GD	[40]
	GeneIO	bioinformatics.ibioba-mpsp-conicet.gov.ar/GenIO	RGD	[41]
	Phen IX	compbio.charite.de/PhenIX	GD	[42]
	PhenGen	phen-gen.org	RGD	[25]
	Phenolyzer	phenolyzer.wglab.org	GD	[27]
	Phenominer	phenominer.mml.cam.ac.uk	GD	[43]
	Phenomizer	compbio.charite.de/phenomizer	GD	[44]
	Phenotate	phenotate.org	GD	NA
	Phenotips	phenotips.org	GD	[45]
	PHEVOR	yandell-lab.org/software/phevor.html	GD	[26]

ALS = amyotrophic lateral sclerosis; GD = genetic disorders; IMDs = inherited metabolic disorders; NA = not available; RGD = rare genetic disorders

diagnosis and/or clinical management. Such resources have been created for IMDs, Fanconi anaemia, and rare cardiac disorders.[21–23] Other tools can be integrated within the bioinformatics pipeline of analysis of WES/WGS which use machine learning and data mining of several clinical and biological knowledge bases to enhance predictive capability.[24–27] Resources such as 'Phenolyzer' and 'PHEVOR', for example, automatically consult several rare disease knowledge bases and, based on text inputs from the user, provide a candidate gene list based on both known gene-phenotype associations and alternative genes within the same gene family.[26,27]

In the era of digitised medicine, online tools are increasingly being utilised by experts and non-experts alike. Clinicians often use online tools when faced with atypical phenotypic presentations, which is not an infrequent scenario in the context of rare genetic disorders. Patients and their families may also use the internet to access information pertaining to their diagnosis.[28] Typical search engines such as Google, while user-friendly, may not be equipped to handle medically specific queries.[29] Therefore the increasing presence of high-quality curated user-friendly phenomic tools will help to expedite diagnostic odysseys and also provide education to clinicians, patients and families.

Remaining challenges

The increasing availability of broad and narrow phenomic repositories is an exciting addition to the field of omics. As with all novel technologies, there are a number of emerging

challenges. It is important to consider the quality and the number of outputs that phenomic resources provide. Phenomic resources are supplements to and not replacements for clinical and genetic expertise. If a particular disease or genetic defect is associated with a pathognomonic clinical presentation well known to clinicians, it is unlikely that there will be a need to consult a phenomic resource. Conversely, some resources generate hundreds of outputs that may be too large to be manageable or interpreted meaningfully in a short amount of time. A manageable number of outputs is something that is more easily achieved in niche resources such as the Leigh Map, but also needs to be considered in larger resources.

At present, there are few efforts to standardise the collection, annotation, and storage of phenotypic data. This decreases the interoperability between differen t resources and geographical regions. Perhaps in future phenomic resources can benefit from similar standardisation to the American College of Medical Genetics and Genomics variant prioritisation guidelines.[30] The annotation of phenotypes in HPO is helpful as there are several thousand phenotypes documented so far. However, there are still several phenotypes, for example biochemical phenotypes, that do not yet have an HPO number and cannot be integrated in resources linked with HPO. This is problematic since these biochemical phenotypes, which are particularly relevant to IMDs, cannot be queried in the aforementioned tools, thus diminishing the utility of these tools for this disease group.[31] However, several ongoing projects aim to improve standardisation and interoperability in the phenomics field. One such initiative, harmonising phenomics information for better interoperability in rare disease, aims to create a rare-disease-specific bioinformatics system which integrates expert curated annotations, automated concept recognitions, and standardisation via HPO to maximise the clinically relevant potential of phenomics. Notably the project emphasises the need for standardised data collection and annotation from patient repositories and biobanks worldwide, but these data can be difficult to access and make publically available, especially with the implementation of stringent data protection laws.[9,32]

There is also the issue of keeping phenomic resources up-to-date. Resources such as the Leigh Map, which are literature dependent, face the challenge of being as current as possible given that novel gene defects are constantly being characterised. Possible solutions that are being explored include expert curation, crowd-sourcing, and automated machine learning methods.[9,19,27,29,33]

Concluding remarks

In recent years, there has been an increase in the utilisation of phenomics to help to diagnose and manage rare diseases, especially those with complex genetic and phenotypic heterogeneity, as observed for many IMDs. The benefits of interactive user-friendly platforms in the interpretation of NGS results are becoming increasingly apparent. While several challenges still need to be mitigated, the advent of phenomics in the IMD field will certainly increase the number of genetic diagnoses, facilitating timely and appropriate clinical intervention, and potentially improving the overall prognosis for affected patients. ■

References

1   Hennekam RCM. Care for patients with ultra-rare disorders. *Eur J Med Genet* 2011;54:220–4.

2   Morava E, Rahman S, Peters V, Baumgartner MR, Patterson M, Zschocke J. *Quo vadis*: the redefinition of 'inborn metabolic diseases'. *J Inherit Metab Dis* 2015;38:1003–6.

3   Rahman J, Rahman S. Mitochondrial medicine in the omics era. *Lancet* 2018;391:2560–74.

4   Tarailo-Graovac M, Shyr C, Ross CJ *et al*. Exome sequencing and the management of neurometabolic disorders. *N Engl J Med* 2016;374:2246–55.

5   Ng BG, Freeze HH. Perspectives on glycosylation and its congenital disorders. *Trends Genet* 2018;34:466–76.

6   Boustany R-MN. Lysosomal storage diseases – the horizon expands. *Nat Rev Neurol* 2013;9:583–98.

7   Koboldt Daniel C, Steinberg Karyn M, Larson David E, Wilson Richard K, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell* 2013;155:27–38.

8   Brookes AJ, Robinson PN. Human genotype–phenotype databases: aims, challenges and opportunities. *Nat Rev Genet* 2015;16:702–15.

9   Maiella S, Olry A, Hanauer M *et al*. Harmonising phenomics information for a better interoperability in the rare disease field. *Eur J Med Genet* 2018;61:706–14.

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

11   Kohda M, Tokuzawa Y, Kishita Y *et al*. A comprehensive genomic analysis reveals the genetic landscape of mitochondrial respiratory chain complex deficiencies. *PLoS Genetics* 2016;12:e1005679.

12   Pronicka E, Piekutowska-Abramczuk D, Ciara E *et al*. New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre. *J Transl Med* 2016;14:174.

13   Sofou K, De Coo IF, Isohanni P *et al*. A multicenter study on Leigh syndrome: disease course and predictors of survival. *Orphanet J Rare Dis* 2014;9:52.

14   Leigh D. Subacute necrotizing encephalomyelopathy in an infant. *J Neurol Neurosurg Psychiatry* 1951;14:216–21.

15   Noronha A, Daníelsdóttir AD, Gawron P *et al*. ReconMap: an interactive visualization of human metabolism. *Bioinformatics* 2017;33:605–7.

16   Köhler S, Vasilevsky NA, Engelstad M *et al*. The human phenotype ontology in 2017. *Nucleic Acids Res* 2017;45:D865–76.

17   Fassone E, Wedatilake Y, DeVile CJ, Chong WK, Carr LJ, Rahman S. Treatable Leigh-like encephalopathy presenting in adolescence. *BMJ case rep* 2013;2013:bcr2013200838.

18   Lake NJ, Compton AG, Rahman S, Thorburn DR. Leigh syndrome: one disorder, more than 75 monogenic causes. *Ann Neurol* 2016;79:190–203.

19   Landrum MJ, Lee JM, Riley GR *et al*. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42:D980–5.

20   Rehm HL, Berg JS, Brooks LD *et al*. ClinGen – The clinical genome resource. *N Engl J Med* 2015;372:2235–42.

21   Adler A, Kirchmeier P, Reinhard J *et al*. PhenoDis: a comprehensive database for phenotypic characterization of rare cardiac diseases. *Orphanet J Rare Dis* 2018;13:22.

22   Chandrasekharappa SC, Lach FP, Kimble DC *et al*. Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. *Blood* 2013;121:e138–48.

23   Lee JJY, Wasserman WW, Hoffmann GF, van Karnebeek CDM, Blau N. Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism. *Genet Med* 2018;20:151–8.

24   Groza T, Köhler S, Doelken S *et al*. Automatic concept recognition using the human phenotype ontology reference and test suite *corpora. Database (Oxford)* 2015;2015:bav005.

25   Javed A, Agrawal S, Ng PC. Phen-Gen: combining phenotype and genotype to analyze rare disorders. *Nat Methods* 2014;11:935–7.

26 Singleton Marc V, Guthery Stephen L, Voelkerding Karl V *et al*. Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet* 2014;94:599–610.

27 Yang H, Robinson PN, Wang K. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. *Nat Methods* 2015;12:841–3.

28 Svenstrup D, Jørgensen HL, Winther O. Rare disease diagnosis: a review of web search, social media and large-scale data-mining approaches. *Rare Dis* 2015;3:e1083145.

29 Dragusin R, Petcu P, Lioma C *et al*. FindZebra: a search engine for rare diseases. *Int J Med Inform* 2013;82:528–38.

30 Richards S, Aziz N, Bale S *et al*. 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* 2015;17:405–24.

31 Lee JJY, Gottlieb MM, Lever J *et al*. Text-based phenotypic profiles incorporating biochemical phenotypes of inborn errors of metabolism improve phenomics-based diagnosis. *J Inherit Metab Dis* 2018;41:555–62.

32 McCall B. What does the GDPR mean for the medical community? *Lancet* 2018;391:1249–50.

33 Pantel JT, Zhao M, Mensah MA *et al*. Advances in computer-assisted syndrome recognition by the example of inborn errors of metabolism. *J Inherit Metab Dis* 2018;41:533–9.

34 Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: online mendelian inheritance in man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* 2015;43:D789–98.

35 Kibbe WA, Arze C, Felix V *et al*. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res* 2015;43:D1071–8.

36 Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSoD: a user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum Mutat* 2012;33:1345–51.

37 Dalgleish R. The human type I collagen mutation database. *Nucleic Acids Res* 1997;25:181–7.

38 Smedley D, Jacobsen JOB, Jager M *et al*. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc* 2015;10:2004–15.

39 Schlicker A, Lengauer T, Albrecht M. Improving disease gene prioritization using the semantic similarity of gene ontology terms. *Bioinformatics* 2010;26:i561–7.

40 Stelzer G, Rosen N, Plaschkes I *et al*. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics* 2016;54:1.30.1–33.

41 Koile D, Cordoba M, de Sousa Serro M, Kauffman MA, Yankilevich P. GenIO: a phenotype-genotype analysis web server for clinical genomics of rare diseases. *BMC Bioinformatics* 2018;19:25.

42 Zemojtel T, Köhler S, Mackenroth L *et al*. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci Transl Med* 2014;6:252ra123.

43 Collier N, Groza T, Smedley D, Robinson PN, Oellrich A, Rebholz-Schuhmann D. PhenoMiner: from text to a database of phenotypes associated with OMIM diseases. *Database (Oxford)* 2015;2015:bav104.

44 Köhler S, Schulz MH, Krawitz P *et al*. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet* 2009;85:457–64.

45 Girdea M, Dumitriu S, Fiume M *et al*. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat* 2013;34:1057–65.

Address for correspondence: Professor Shamima Rahman, UCL Great Ormond Street Institute of Child Health, Genetics and Genomic Medicine, 30 Guilford Street, London, WC1N 1EH, UK.
Email: shamima.rahman@ucl.ac.uk