Received: 24 February 2022 | Accepted: 25 March 2022

DOI: 10.1111/bjh.18191

BJHaem
BRITISH JOURNAL OF HAEMATOLOGY

GUIDELINE

The use of next-generation sequencing in the diagnosis of rare inherited anaemias: A Joint BSH/EHA Good Practice Paper*

Noémi B. A. Roy[1,2] | Lydie Da Costa[3] | Roberta Russo[4,5] | Paola Bianchi[6] | Maria del Mar Mañú-Pereira[7] | Elisa Fermo[6] | Immacolata Andolfo[4,5] | Barnaby Clark[13] | Melanie Proven[8] | Mayka Sanchez[9,10] | Richard van Wijk[11] | Bert van der Zwaag[11] | Mark Layton[12] | David Rees[13] | Achille Iolascon[4,5] | British Society for Haematology/ European Hematology Association

[1]Department of Haematology, Oxford University Hospitals, NHS Foundation Trust, Oxford, UK

[2]NIHR BRC Blood Theme, Oxford, UK

[3]Hôpital Universitaire Robert Debré, Paris, France

[4]Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Naples, Italy

[5]CEINGE Biotecnologie Avanzate, Naples, Italy

[6]UOS Fisiopatologia delle Anemie, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico Milano, Milan, Italy

[7]Translational research in Rare Anaemia Disorders, Vall d'Hebron Research Institute, Barcelona, Spain

[8]Genetics Laboratories, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[9]Department of Basic Sciences, Iron metabolism: Regulation and Diseases, Universitat Internacional de Catalunya (UIC), Barcelona, Spain

[10]BloodGenetics S.L. Diagnostics in Inherited Blood Diseases, Barcelona, Spain

[11]Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[12]Imperial College London, Hammersmith Hospital, London, UK

[13]King's College Hospital, King's College London, UK

Correspondence
BSH Administrator, British Society for Haematology, 100 White Lion Street, London, N1 9PF, UK.
Email: bshguidelines@b-s-h.org.uk

EHA Executive Office, Koninginnegracht 12b, 2514 AA - The Hague, The Netherlands
Email: guidelines@ehaweb.org

METHODOLOGY

The British Society for Haematology (BSH) produces Good Practice Papers to recommend good practice in areas where there is a limited evidence base but for which a degree of consensus or uniformity is likely to be beneficial to patient care. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) nomenclature was used to evaluate levels of evidence and to assess the strength of recommendations. The GRADE criteria can be found at http://www.gradeworkinggroup.org.

This Good Practice Paper was produced as a collaboration with the European Hematology Association (EHA) compiled according to the (BSH) process at http://scanmail.trustwave.com/?c=8248&d=68DV3b1jbPPsVn-8nm3kGS2D_-Hms9YWMWrrk5K8Eg&u=http%3a%2f%2fb-s-h%2eorg.uk. This guideline group included UK-based medical experts representing the BSH and members of the Red Cell and Iron Scientific Working Group (SWG) of EHA

*This article is being published concurrently in British Journal of Haematology and HemaSphere. The articles are identical except for minor stylistic and spelling differences in keeping with each journal's style. Citation from either journal can be used when citing this article.

Literature review details

MEDLINE, EMBASE and PubMED were searched systematically for publications in English from 2000 to 2019 using the following key words. 'NGS' and 'next-generation sequencing' or 'high throughput sequencing' AND 'haemolytic anaemia' or 'DBA or 'Diamond Blackfan anaemia' or 'CDA' or 'congenital dyserythropoietic anaemia' or 'sideroblastic anaemia' or 'HS' or 'hereditary spherocytosis' or 'red cell membrane disorders' or 'red cell enzyme disorders' or 'PK deficiency' or 'PKD'. References from relevant publications were also searched. Conference abstracts were included if deemed to be of particular relevance.

Review of the manuscript

Review of the manuscript was performed by the BSH Guidelines Committee General Haematology Task Force, the BSH Guidelines Committee and the General Haematology sounding board of the BSH. It was also on the members section of the BSH website for comment. It has also been reviewed by members of the EHA Red Cell and Iron SWG and the EHA Guidelines Executive Committee.

INTRODUCTION

The use of next-generation sequencing (NGS) in the diagnosis of rare inherited anaemias is increasingly common, as evidenced by a growing number of publications describing its clinical utility.[1–6] Excluding disorders of globin synthesis, rare anaemias include Diamond–Blackfan anaemia (DBA), congenital dyserythropoietic anaemias (CDA), congenital sideroblastic anaemias (CSA) and disorders of red cell membrane and enzymes. Other forms of genetic anaemias can also be considered while establishing NGS panels, in particular genetic syndromes where anaemia comprises one of the constellation of symptoms.

Table 1 briefly summarises the key aspects of these conditions.

The advantages of using NGS over single-gene testing, in addition to the cost effectiveness, is that clinical and laboratory features are often not specific for a particular condition, and a large number of large candidate genes might need to be analysed before making a diagnosis. A proportion of the patients also present with overlapping phenotypes, and it has been shown that in 10%–40% of cases there is a degree of misdiagnosis or no diagnosis when this is based purely on phenotype and traditional non-NGS testing.[1,6] This can result in incorrect or inadequate treatment, causing anxiety and adversely affecting quality of life and potentially cost.

The term 'NGS' refers to all types of high-throughput sequencing, and for the purposes of this good practice guideline will include targeted resequencing (t-NGS), whole exome sequencing (WES) and whole genome sequencing (WGS). Table 2 shows the advantages and disadvantages of each type of NGS. A detailed description of NGS techniques is beyond the scope of this paper; however, this is summarised in Figure 1. In t-NGS, only the genes selected are sequenced, while in WES ~30 000 genes are sequenced and in WGS all genes and intergenic regions are sequenced. However, in WES and WGS, the coding sequences of

TABLE 1 Key aspects of the rare anaemias not due to disorders of haemoglobin synthesis

	Diamond–Blackfan Anaemia (DBA)	Congenital dyserythropoietic anaemia (CDA)	Sideroblastic anaemia	Red cell membrane/cation leaking and enzyme disorders
Age at presentation	Usually 2–3 months of age or <first year of life	Usually child/young adult	Usually child/young adult	Fetal/neonate/child/young adult
Associated features	Craniofacial Skeletal Cardiac Urogenital tract	Distal limb Iron overload	Ring sideroblasts on bone marrow aspiration	Jaundice Hepatosplenomegaly Gallstones Iron overload Progressive myopathy and neurocognitive impairment[a] Lymphoedema[b]
Severity	Moderate to severe	Usually mild to moderate	Mild to severe	Mild to severe
Treatment	Corticosteroids Transfusions and chelation BMT	Interferon Transfusions and chelation Often none needed	Transfusions and chelation Often none needed	Often none needed Splenectomy Transfusions and chelation New agents
Genetics	Autosomal dominant (45%) or de novo (other inheritance for DBA-like disease) Ribosomal proteins or other genes affecting ribosome biogenesis (other genes for DBA-like disease)	Autosomal recessive or dominant, X-linked Vesicle trafficking, heterochromatin assembly, nuclear proteins, transcription factors	X-linked; autosomal recessive Haem synthesis	Autosomal dominant or recessive; X- linked RBC membrane cytoskeleton, RBC transporters and RBC enzymes

Abbreviation: RBC, red blood cell.

[a]Associated with some rare enzymopathies or rare form of glucose transporter type 1 (GLUT1) variants.

[b]Associated with some severe form of hereditary stomatocytosis.

TABLE 2 Comparison of different types of next-generation sequencing

	Targeted resequencing panels (t-NGS)	Whole exome sequencing (WES)	Whole genome sequencing (WGS)
Target of sequencing; size (base pairs [bp])	Exons of 20–200 genes with some intron/exon boundaries for splice site mutations; 500 000 bp	The 'exome'- ~30 000 exons of known coding genes (~1.5% of genome but 80%–90% of known disease-causing mutations) with some intron/exon boundaries for splice mutations; 2×10^7 bp	The whole genome (coding and non-coding space) 3×10^9 bp
Method	Capture of chosen exons, amplification steps and sequencing or amplification of chosen exons and sequencing	Capture all the exons, amplification step and sequencing	DNA is fragmented randomly, ligation of adaptors and direct sequencing (no capture or amplification)
Advantages	Cost, relative ease of interpretation, few unsolicited findings, more challenging to identify CNVs	Cost lower than WGS	Entire genome interrogated including non-coding region; more potential to identify CNVs. Can add genes to virtual panel. Relatively even coverage
Disadvantages	Will only identify mutations in targeted regions, coverage is often uneven, so mutations may be missed. Harder to detect some CNVs	Interpretation can be challenging, high chance of unsolicited findings, will only find mutations in coding regions, coverage is often uneven, may not detect CNVs. Ethical issues of incidental findings in genes that predispose to serious illness.	Interpretation challenging unless there is a trio, non-coding region cannot easily be interpreted. Ethical issues of incidental findings in genes that predispose to serious illness. Cost.

Abbreviation: CNV, copy number variant.

only a subset of genes are analysed, what is frequently referred to as a 'virtual panel'. In addition, coverage of genes is best in WGS where no DNA amplification step is required. Large duplications and deletions, involving one or more whole genes, known as copy number variants (CNVs), are more difficult to identify, but can be detected using appropriate analysis, particularly using WGS, but also WES and targeted resequencing.

It is important to note that, depending on the size of the panel, a number of variants will always be identified after all of the filters are applied, even in normal individuals. This number will depend on the number of genes, the inherent polymorphic potential of the gene and the ethnic origin of the individual tested. All variants identified post-filtering need to be assessed against strict criteria to determine their pathogenicity, based on the guidelines of American College of Medical Genetics (ACMG).[7] It is good practice to assess all variants even after a pathogenic variant has been found, to help with interpretation if this variant is identified again in the future.

Excellent comprehensive guidelines exist for the preparation of samples and the quality control that should be followed.[8] Likewise, the ACMG and Association of Clinical Genomic Science (ACGS) guidelines detail the interpretation of variants, and all laboratories should follow these criteria to determine pathogenicity of all variants identified.[7,9] The ACGS guidelines are less stringent in their assessment of evidence for pathogenicity. The ACMG system therefore scores more variants as variants of uncertain/unknown significance (VUS)/class 3 than the ACGS guidelines, increasing specificity at the expense of sensitivity.

The purpose of this paper is to give guidance on the uses of NGS that are specific to the diagnosis of rare inherited anaemias. This may be useful to laboratories wanting to set up NGS or for ones that have set this up for research and are planning to use it for clinical diagnosis. The type of NGS used, the conditions for which it can be used and the timing of it in the diagnostic pathway will partly depend on each country's healthcare system and funding arrangements. However, we aim to issue general guidance. Most of the guidance below is best suited to t-NGS as this is currently most commonly used, but the principles are equally applicable to the other technologies.

This good practice paper will address the following questions:

1. When is NGS necessary or of additional value in the diagnosis of rare anaemias?
2. At which point in the diagnostic pathway should NGS be used?
3. What are the important considerations in choosing the most appropriate NGS method and which quality criteria must be met?
4. What criteria should be used for reporting NGS variants identified?
5. How should variants identified be stored and shared between laboratories?
6. What criteria are essential for a laboratory to be able to offer clinical-grade NGS?

Question 1: When is NGS necessary and/or of additional value in the diagnosis of rare anaemias?

Most current NGS approaches include the genes involved in the pathology of DBA, CDA, CSA and disorders of red

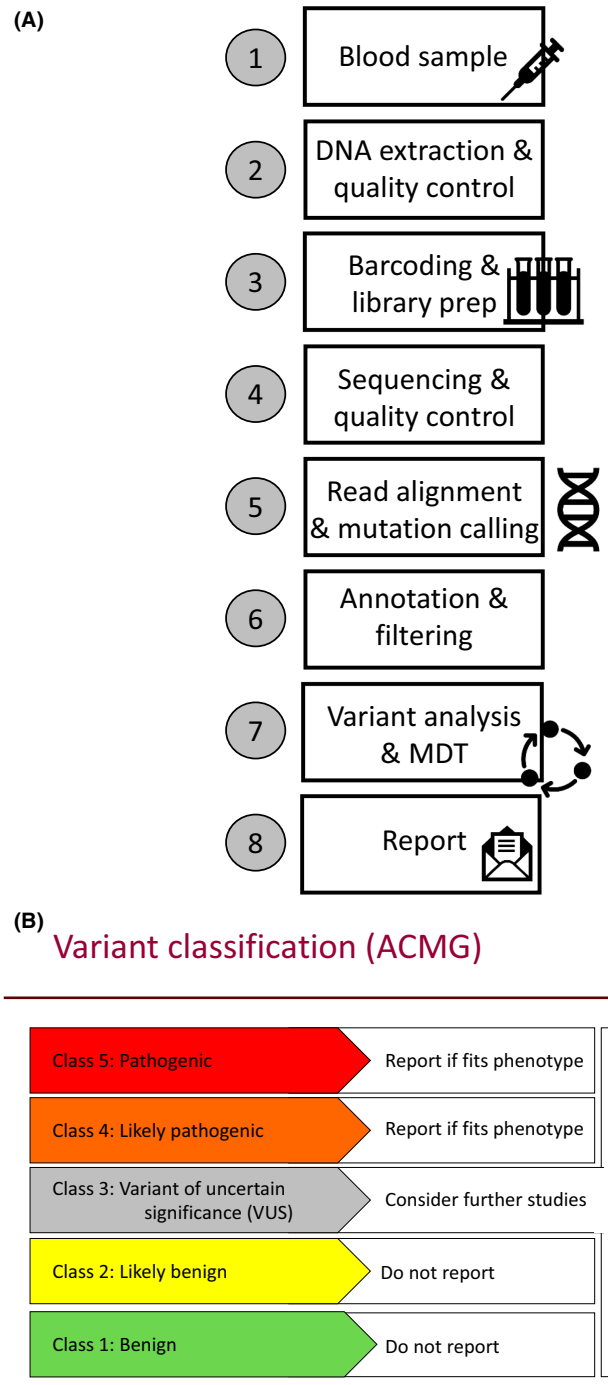

FIGURE 1 (A) Cartoon of the process of creating an NGS report from arrival of sample in the laboratory. (MDT = multidisciplinary team. Usually includes clinical scientists and clinicians.) (B) American College of Medical Genetics variant classification, with examples of further studies that can be carried out to determine the pathogenicity of Class 3 variants of uncertain significance. This includes family studies to investigate segregation, as well as functional assays such as red cell enzyme activities, EMA (eosin-5'-maleimide) dye binding for hereditary spherocytosis, and osmotic gradient ektacytometry (Osmoscan), which investigates red cell deformability for membrane disorders. This list is not exhaustive and includes other functional assays (e.g. electron microscopy for CDA, ribosomal profiling or northern blots for DBA). EMA (eosin-5'-maleimide) dye binding; ekta- ektacytometry.

cell membranes and enzymes.[1,6] The globin genes are frequently but not always included. Firstly, much of globin gene testing required for pre- and neonatal diagnosis requires a rapid turnaround time and analysis of a small number of genes, making it unwieldy and unnecessary to be testing all of the genes on a panel. In most cases, the clinical and laboratory presentation is clear and only a minority have a differential diagnosis of other haemolytic anaemias. Secondly, these are regions of very high sequence homology, potentially resulting in poor specificity and high levels of artefacts and false-positive results on NGS testing, depending on the specific technology selected. In addition, many of the pathogenic genetic abnormalities leading to haemoglobinopathies are CNVs (insertions or deletions), which can be more difficult to detect by t-NGS. In particular, some common alpha globin variants such as the 3.7 kb deletion and triplicated alpha globin gene, are especially challenging as the breakpoint sequences are not unique. However, robust validation of the panel can ensure the reliable detection of most globin variants and some panels have been designed specifically to detect CNVs in globin genes, enabling the option of using NGS for haemoglobinopathy diagnosis.

There are circumstances when globin gene sequencing is of particular importance, including in the assessment of microcytic or haemolytic anaemias. Haemoglobin subtype

analysis, including the quantitation of haemoglobin A_2, can identify or exclude most globin gene variants, but does not reliably identify many cases of unstable haemoglobin, dominant thalassaemia[10,11] or individuals with beta thalassaemia intermedia resulting from heterozygous of beta thalassaemia in the presence of triplicated alpha gene.[12] In the case of unstable haemoglobins, the patients may have a mild to severe haemolytic anaemia, including transfusion dependence.[13,14] The unstable haemoglobin is often not detectable using haemoglobin analysis, and the presence of transfused blood also makes phenotypic diagnosis more difficult, particularly if started neonatally. Globin gene variants are the commonest cause of inherited anaemia, and all patients should be formally assessed for their presence, using a combination of haemoglobin analysis and specific genetic tests for suspected variants, and by inclusion on NGS panels, depending on local practice. Particular consideration needs to be given to excluding CNVs of the alpha globin genes, which may require specific assays using a gap polymerase chain reaction (Gap-PCR) or multiplex ligation-dependent probe amplification (MLPA).

Devising a list of conditions and genes to include in the t-NGS or virtual panel

The number of genes to include in a panel must balance inclusivity, to reduce false-negative rates, with increasing workload from needing to review and critically assess a large number of variants.

For any laboratories wishing to set up t-NGS for rare inherited anaemias, Appendix 1 contains our suggested list of genes. Any published list is rapidly out of date as new evidence accumulates. However, the majority of known genes will be valid for some time. In England, genetic testing has been harmonised nationally and all the genes on each panel offered are available on PanelApp: https://panelapp.genomicsengland.co.uk/. This list of genes has been determined and curated by specialists in the field and is updated yearly to ensure that newly published genes are included.

It is worth considering if there are conditions in which NGS is of no added value and whether the reluctance to use NGS in some cases is purely due to its cost. There are rare anaemias that are often straightforward to diagnose without recourse to DNA analysis, e.g., hereditary spherocytosis (HS). Nevertheless, for such cases the advantage of carrying out molecular analysis is that it facilitates genetic counselling. This can be especially helpful in some HS cases without a clear family history, to distinguish between recessive inheritance and a de novo variant. Conversely, laboratory tests for HS reach a sensitivity/specificity of >98%/90%, which is higher than for t-NGS. Although these are often mild conditions, they can result in significant morbidity including fetal anaemia, kernicterus and transfusion dependence, and genetic counselling is useful, particularly in families who wish to avoid further affected pregnancies.

It is particularly important to be certain of the precise diagnosis before performing splenectomy for presumed HS, to avoid ill-advised splenectomy in dehydrated hereditary stomatocytosis as this procedure is accompanied by a greatly increased risk of thromboembolic disease.[15] Phenotypically these conditions can be very similar unless some assessment of red cell hydration is performed, such as osmotic gradient ektacytometry or osmotic fragility measurement. In general, genetic diagnosis should be confirmed before recommending splenectomy in HS, and this will typically involve analysis using an NGS panel. Additionally, documenting genetic variants will eventually lead to some genotype–phenotype correlations.[16,17] This is the case with pyruvate kinase (PK) deficiency, where response to the new drug AG-348 depends on whether the mutations are missense or not.[18]

For some conditions, NGS is far superior to Sanger sequencing of specific genes, due to the phenotypic variability and the unreliability of phenotypic tests such as enzyme assays for rare enzymopathies, making it difficult to target genes precisely, particularly when the patient is transfusion dependent. Because of frequent misdiagnosis of 'dyserythropoiesis' in some haemolytic anaemias,[1,6] genetic analysis should always be used to confirm a 'CDA'.

One condition where genetic analysis is particularly useful is dehydrated hereditary stomatocytosis (xerocytosis) due to autosomal dominant mutations in the gene Piezotype mechanosensitive ion channel component 1 (*PIEZO1*), a mechanosensitive calcium channel. Patients with this condition are probably at high risk of developing post-splenectomy thrombosis and splenectomy in these cases is generally contra-indicated.[15,19] This condition is difficult to diagnose, and can be associated with only occasional stomatocytes on the blood film; genetic diagnosis should usually be performed before splenectomy when there is a possibility that the diagnosis could be dehydrated hereditary stomatocytosis; this will include most cases of presumed HS.

Finally, NGS-based genetic testing is useful for the identification of complex modes of inheritance that are recognised to account for at least 4% of diagnosed Mendelian conditions.[20]

Recommendations

- NGS should only be used in cases where acquired causes are thought to be very unlikely (IA)
- Appropriate consent should be obtained (IA)
- Globin gene abnormalities should be considered and investigated appropriately before NGS is carried out, including haemoglobin analysis and sequencing of individual globin genes, depending on the genetic distribution that is already known in the local population. Specific consideration should be given to globin gene CNVs, with use of Gap-PCR and MLPA as appropriate (IIB)

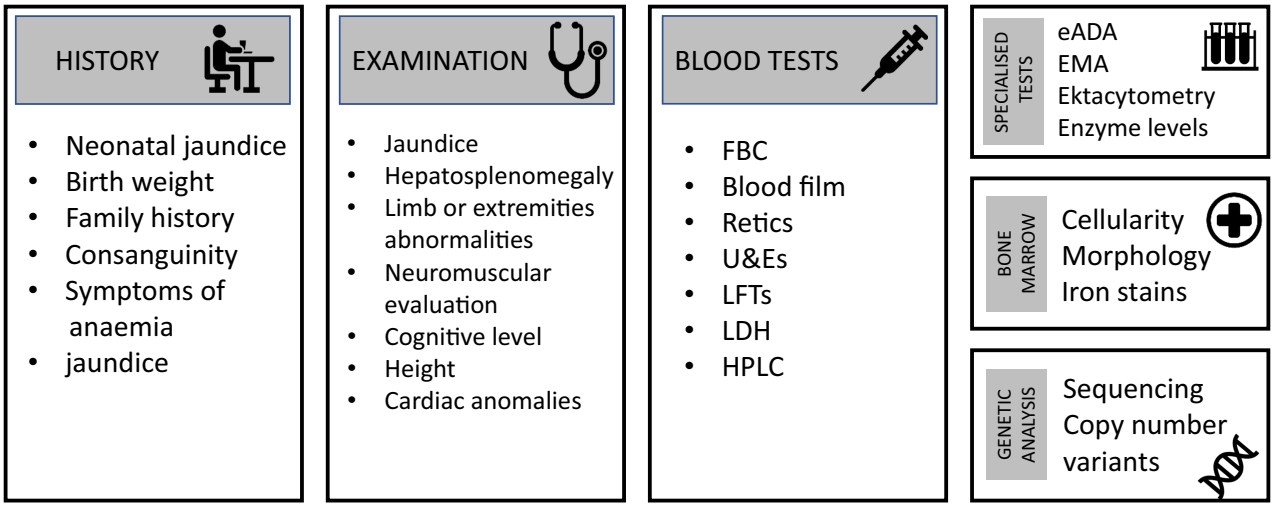

FIGURE 2 Clinical and laboratory assessment of the patient with a suspected diagnosis of inherited anaemia. These are indicative only and not exhaustive. (FBC: full blood count, retics: reticulocytes, U&Es: urea and electrolytes, LFTs: liver function tests, LDH: lactate dehydrogenase, HPLC: high performance liquid chromatography, eADA: erythrocyte adenosine deaminase, EMA: eosin-5'- maleimide test).

- Conditions that should be tested on the panel include DBA, CDA, CSA, suspected red cell enzyme deficiencies and red cell membrane disorders (IIB)
- Genetic analysis should be used to confirm conditions when there is diagnostic uncertainty (IIB)
- Genetic analysis should be performed before undertaking splenectomy for inherited haemolytic anaemias or other irreversible procedures such as bone marrow transplantation, where the genetic variant should be excluded from a potential stem cell sibling donor (IIB)

Question 2: At which point in the diagnostic pathway should NGS be used?

The use of NGS will partly depend on each country or hospital system's technical and reimbursement characteristics. The traditional investigative pathway is to take a history and examination, full blood count (FBC), reticulocyte count, and haemolytic markers, before selecting specialised tests (enzyme assays, osmotic gradient ektacytometry, eosin-5-maleimide [EMA] test, erythrocyte adenosine deaminase [eADA] etc.). In some cases, this may lead to a bone marrow biopsy or aspiration, with genetic analysis being kept at the end of the pathway. In other places, genetic analysis may occur much earlier in the pathway.[21] The advantages are that this may lead to a more rapid diagnosis, may be cost effective in reducing delay in diagnosis (at the expense of a higher cost upfront) and may (in some conditions) preclude the need for a bone marrow biopsy. Figure 2 shows examples of aspects of the history and examination that should be sought when evaluating the patient, as well as standard blood tests. The requirement for specialised tests, bone marrow aspiration and biopsy, and genetic analysis and the order in which they are requested, will differ between services, but in time, genetic analysis is likely to be carried out earlier in the pathway, with specialised functional analysis used to confirm the genetic diagnosis.

Recommendations

- NGS should primarily be used once the phenotype has been characterised. In particular, it should be established whether the patient has haemolysis, ineffective erythropoiesis, dyserythropoiesis or bone marrow failure, as this may direct the analysis of the variants identified (IC)
- Clinical grade NGS should ensure that variants are reported with reference to the phenotype of the patient (a sample request form detailing minimal phenotypic information can be found in Appendix 2) (IC)
- If further investigations are required to confirm the diagnosis (e.g., family studies, RNA studies, specialist haematological tests directed by the variant identified), these can be recommended on the genetics report (IC)

Question 3: What are the important considerations in choosing the most appropriate NGS method and which quality criteria must be met?

Most panels are currently carried out as t-NGS, although some diagnostic laboratories carry out target enrichment across thousands of regions, then analyse the variants among genes that have been grouped together into virtual panels. As some countries move towards conducting all genetic analysis in the form of WGS, virtual panels will be increasingly used. The choice of using t-NGS over virtual panels is mostly due to availability, cost and turnaround time. While cost-per-base may be lower for WGS, this requires a capital investment beyond the scope of most diagnostic laboratories. However, a

major disadvantage of using t-NGS is that if any new genes are found to be associated with a known phenotype, adding a gene to the panel requires complete redesign and revalidation. This time-consuming and expensive process limits updating t-NGS panels to about once a year. WGS is also better suited for determination of CNVs, a common genetic cause of a number of inherited anaemias, with alpha globin gene deletions remaining a particular challenge for all technologies. New bioinformatic protocols to improve CNV assessment from targeted panels are improving their detection across modalities. Bait capture and unique molecular indexed amplicon methods may be combined with bioinformatic algorithms to determine the breakpoint mapping from short reads.[22] As the selected method will depend on many factors, it is critical that a laboratory is aware of the limitations of the technique, and that additional steps are taken to either overcome some of these limitations (e.g., gap-filling by Sanger Sequencing) or that the report produced is explicitly clear on the limitations of the analysis. This may require suggesting alternative methods (e.g., MLPA) to address CNVs that may not be detected reliably by t-NGS.

The availability of complementary diagnostic tools such as erythrocyte morphology, red cell and reticulocyte indices, EMA dye binding or osmotic gradient ektacytometry for red cell membrane disorders, may allow a phenotypic confirmation of the diagnosis in the absence of a definitive genetic diagnosis.

Recommendations

- The NGS method should be chosen based on local resources and required turnaround time (IC)
- Depending on the method chosen, the laboratory should be aware of the limitations and either reduce these (MLPA, gap-fill) or make it clear in the report what has not been tested (IC)

Question 4: What criteria should be used for reporting NGS variants identified?

Once variants have been identified and graded for pathogenicity, a multidisciplinary team meeting (MDT) is carried out, where variants are discussed in the context of the clinical presentation and a final report is written. In cases of an established pathogenic variant that fits with the phenotype, a report can be issued by the clinical scientists in the absence of an MDT meeting.

The ACMG guidelines must be followed for pathogenicity of single-nucleotide variants (five classes) - pathogenic (class 5) and likely pathogenic (class 4) variants related to the clinical suspicion should be included in the report.

Recommendations

Variant types to include in the final report (IIB):

1. Pathogenic/likely pathogenic variants related to the clinical suspicion

Variants to which a pathogenic role can be attributed with certainty, including:

- known variants in genes already associated with phenotype/disease
- novel variants in genes already associated with the phenotype/disease that have a clear causative role (e.g., loss-of-function of a known gene that is associated with disease with a mechanism of haploinsufficiency), and fits with the pattern of inheritance, if available.

2. Pathogenic variants unrelated to the clinical suspicion

Variants with a well-known pathogenic role not related to clinical suspicion, including:

- causative variants in genes already associated with a phenotype but different from the suspected disease (*reverse phenotyping*)
- incidental findings (e.g., carrier state for other condition), which should be reported only if consent explicitly signed for this as per the ACMG guidelines.

3. Variants with unknown clinical and functional role (*VUSs*) that could provide a diagnosis pending further investigation or evidence

These variants can be identified in:

- genes related to the suspected phenotype, which can be included in the final report. However, it should be made clear that the variant is a VUS and that without functional and/or family studies one cannot be sure that this variant is involved in the pathogenicity of the condition.
- genes not related to the clinical suspicion, which should not be included in the report.

In general, it is not recommended that intronic/splice (non-canonical) and 5′ and 3′ variants are reported unless substantial functional data is available. Some laboratories may report a recessive disorder where one pathogenic mutation (class 4 or 5) has been found together with a VUS. Family studies are strongly recommended, and the report must make clear that there is no definite pathogenicity associated with the second variant.

This also includes circumstances where two very rare VUSs are identified in a gene(s) implicated in the phenotype, and family studies indicate they are in *trans* and functional data supports this gene as being causative. Variants of uncertain significance or variants that would suggest a novel complex mode of inheritance can form the basis of research studies, with the caveat that this almost universally requires a different form of consent to that obtained for diagnostic testing.

Question 5: How should variants be stored and shared between laboratories?

The sharing of variants between laboratories plays a very important role in ensuring high-quality data, high-diagnostic

rates and cost efficiency. However, this is often much more difficult to achieve than might be imagined, with issues such as data storage and the practicalities of sharing variants being significant obstacles.

One of the pre-requisites for variant sharing is that participating laboratories use the same system for variant classification. Sharing of variants is difficult because ideally the information to be shared includes the clinical phenotype, how the pathogenicity was assessed including individual components of the overall score, and knowledge of the other variants found in the same patient. A potentially pathogenic variant where a different definitive genetic cause has also been found in the patient means the first one is less likely to be pathogenic.

However, the more variants are shared, the more identifiable the data are, raising the possibility that individuals may be identified according to specific haplotypes. Other obstacles to routine variant sharing across laboratories include practical technical reasons (not everyone shares and stores data in the same way) and time (the need to keep the database up to date and curated, with someone to take responsibility for any discrepancies).

Variant sharing is also predicated upon using the same nomenclature (e.g., Human Genome Variation Society [HGVS]) and reporting against the same transcripts. In cases of multiple transcripts, the specific 'disease transcript' must be used, but this is not always known. Laboratories should make reasonable efforts to ensure that the transcript they are using is expressed in erythroid cells. LRG (Locus Reference Genomic) may be useful in this assessment https://www.lrg-sequence.org/.

Recommendation

- Laboratories should share variants with other laboratories analysing the same genes (IIC)
- Laboratories should ensure they are using commonly used transcripts which have been shown to be expressed in erythroid cells (IIC)
- Ethical and legal issues in sharing variants between laboratories, often located in different countries should be clearly reported and discussed (IIC)

Question 6: What criteria are essential for a laboratory to be able to offer clinical-grade NGS?

For a laboratory to offer clinical-grade NGS, a number of parameters must be met, which relate to the laboratory itself, the panel design, the analytical pathway and the report.

Patient consent

This will depend on the legal framework of each country. However, some general guidance can be issued. Each laboratory should have a clear, written policy on consent, which is available to referring clinicians, and takes into account the following considerations:

- the risk of identifying non-paternity
- incidental findings (this will depend on the gene content of the panel, the way it is analysed, and national regulations)
- identification of carrier status, which could have reproductive implications or implications for family members
- the test may not find mutations that explain the phenotype
- whether the DNA will be re-tested as new genes are published
- whether residual DNA in the laboratory could be used as a control
- how long the DNA sample will be stored for

Finally, it may be useful to begin a discussion with patients at the stage of consent regarding the concept of 'variants of uncertain significance' and that they may be asked to provide further samples to help validate some of the findings. In addition they can be made aware that other family members may be investigated to clarify co-segregation of a variant of interest.

Laboratory characteristics

The laboratory must be involved in an accreditation scheme (e.g., Clinical Pathology Association [CPA], International Organisation for Standardisation [ISO], United Kingdom Accreditation Service [UKAS]) and must participate in external quality assessment (EQA) for NGS. Where none exist, the laboratory should make a reasonable effort to carry out some sample sharing with other similar laboratories to ensure there is an element of external assessment. Where the laboratory carrying out the NGS does so on DNA samples it receives from external laboratories, these laboratories should also be accredited.

Pre-testing

Clinical grade NGS will ensure that variants are reported against the phenotype of the patient and where this is unavailable, the laboratory should make reasonable attempts to obtain it from clinicians or issue a limited report if there is no phenotypic information.

Assessing pathogenicity

Assessing pathogenicity in the laboratory according to the ACMG guidelines as stated above is critical as numerous publications assigning pathogenicity in the literature predate the ACGM recommendations. In some cases, the variants described are accompanied by sufficient functional evidence to confirm pathogenicity. As this is not universally the case, it is our recommendation that original publications are reviewed to assess this.

Panel characteristics

All clinically relevant regions of known genes should be included and at least 90% of the relevant/coding regions must be covered by the sequencing. The coverage at individual genes may be included in the report, or merely retained and stored in the laboratory and available upon request. Sufficient sequencing into the introns is needed to identify splice variants and should cover ±50 bp to ensure branch point mutations are covered. All types of known causative variants should be detected by the method- for example this would include variants localised in the promoter region (e.g., 5'-aminolevulinate synthase 2 [*ALAS2*] GATA1 binding site in promoter[23] or for PKLR)[23].

Depth of coverage should be sufficient to detect a heterozygous variant, including single-nucleotide variants, indels and structural variants including CNVs.

Very common pathogenic variants should be reliably detected. Panels have to be large enough to detect phenotypic overlap.[1,6] In cases of syndromes, where anaemia may be the presenting feature, these genes can also be included in the panel (e.g., TRNA nucleotidyl transferase 1 [*TNRT1*], carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase [*CAD*] or ATP-binding cassette subfamily B member 7 [*ABCB7*]). Genes that are not known to be pathogenic should not be included unless specific research consent has been sought.[24]

Validation of the test should be carried out

As with the introduction of any new test, robust validation should be carried out before introduction into the diagnostic laboratory. Validation should be based on the Rehm 2013 criteria and are summarised in Appendix 3. Depending on the country, some laboratories only need to validate the technology, while others need to validate each individual panel.

Quality control (QC) metrics in the report

The report should provide also the following data: average sequencing depth, coverage (target regions)_20×, coverage (target regions)_50×.

Filtering

Filtering is generally set to exclude variants occurring at an allele frequency of >1% in the general population, although this may vary depending on the patient's ethnic group, which should ideally be available for each patient. Any variant with a ClinVar annotation of pathogenicity should be included irrespective of the allele frequency. In HS, common pathogenic modifiers are present above the 1% cut-off (e.g., alpha-LeLY/LePRA) and should be specifically sought in patients where HS/hereditary elliptocytosis is a possible diagnosis, especially in the presence of other spectrin alpha, erythrocytic 1 (*SPTA1*) variants. It should be noted that while alpha-LeLY is usually included in most t-NGS panels, alpha-LePRA (c.4339-99C>T) is not usually covered with standard t-NGS platforms. However, alpha-LePRA is frequently co-inherited with the α-Bughill allele[25] (missense variant p.[Ala970Asp]) that is covered by standard t-NGS platforms.

Confirmation of variants

Sanger sequencing is not needed routinely to confirm individual variants if the panel has been validated and the whole pathway is assessed as part of accreditation. However, in some cases Sanger sequencing is necessary before a report can be issued. This includes any unusual variant allele frequency (VAF) or low-coverage reads (e.g., 30 reads); any novel CNVs should be confirmed where possible, as should any variants that are present in only forward or reverse reads, as this may result from allele dropout and erroneously suggest homozygosity. Sanger sequencing can also be used to confirm that the correct sample has been tested and therefore the report is being issued to the recipient intended. Confirmation of patient identity may alternatively be provided by DNA analysis of a new sample and/or analysis of polymorphisms (snapshot technique).

Phenotypic/pedigree confirmation

Pedigree confirmation or additional functional tests may increase or decrease pathogenicity of individual variants as per the ACMG guidelines. It is best practice to request samples from relevant family members upfront so that family studies may be carried out, including testing for de novo inheritance. When a clearly pathogenic variant is identified in only one allele of a recessive gene that is associated with the clinical phenotype, it is common practice to report that the pathogenic variant may be *contributing* to the phenotype. The report should be clear on the limitations of the technique and further work can be suggested where this would be helpful.

Assessment for false-negative rate

The genes that have been assessed, the coverage of those genes and the types of genetic abnormalities must be stated. This is to ensure that if a report states that no pathogenic variants were detected, the clinician is clear about what possible causes of the condition have not been excluded. This includes common pathogenic variants not detected by the current method. Other ways of picking up deletions include, e.g., MLPA, array comparative genomic hybridisation and single-nucleotide polymorphism arrays. The false-negative rate will be different for each phenotype that is being assessed by the panel, depending on what proportion of the phenotype is known to be explained at the molecular level. This is likely to improve in time as novel genes are identified.

Report writing

The clinical report should be easy to read and be understood by clinicians. The technical characteristics of the test should be made available upon request. Details regarding which transcript was used should be clear.

Recommendations

- Consent for DNA analysis should be obtained (IIB)
- Phenotypic information, and functional test results, if available, should be provided and the genetic variants assessed against this (IIB)
- It should be clear which genes have been tested, the coverage of these genes, and whether or not CNVs are detectable by this method (IIB)
- Filtering of variants against general population should be set between 1% and 5% (IIB)
- Variants with a ClinVar entry and with known modifier effect (e.g., alpha-LeLY and LePRA) should be sought and included in the report (IIB)
- Ideally a false-negative rate should be provided, and it should be made clear that a lack of genetic diagnosis does not exclude an appropriate clinical diagnosis (IC)
- Class 4 and 5 variants should be reported if they correspond to the phenotype. In occasional circumstances, some class 3 variants may be reported with careful wording of why they were included (IC)
- The whole panel must be fully validated before use (IA)

ACKNOWLEDGEMENTS

The BSH General Haematology Task Force members at the time of writing this good practice paper were Mamta Garg, Charlotte Bradbury, Emmy Dickens, Savio Fernandes, John Grainger, Ciaran Mooney, Noémi Roy, Barbara Delasalle, Nicola Svenson, Sara Stuart-Smith and Suzanne Docherty. The authors would like to thank them, the BSH sounding board, and the BSH guidelines committee for their support in preparing this good practice paper. The authors would also like to thank the EHA Guidelines Committee as well as EHA Executive Office for their support in preparing and publishing this good practice paper. Noémi Roy, Lydie Da Costa, Roberta Russo, Paola Bianchi, Maria del Mar Mañú-Pereira, Elisa Fermo, Immacolata Andolfo, Barnaby Clark, Melanie Proven, Mayka Sanchez, Richard van Wijk, Bert van der Zwaag, Mark Layton, David Rees and Achille Iolascon all took part in the data synthesis and writing of the guidelines.

CONFLICT OF INTERESTS

The BSH paid the expenses incurred during the writing of this good practice paper. All authors have made a declaration of interests to the BSH and Task Force Chairs which may be viewed on request. There were no conflicts of interest that impacted on the writing of this paper.

REVIEW PROCESS

Members of the writing group will inform the writing group Chair, Noemi Roy, if any new pertinent evidence becomes available that would alter the strength of the recommendations made in this document or render it obsolete. The document will be archived and removed from the BSH current guidelines website if it becomes obsolete. If new recommendations are made an addendum will be published on the BSH guidelines website (www.b-s-h.org/guidelines).

DISCLAIMER

While the advice and information in this guidance is believed to be true and accurate at the time of going to press, neither the authors, the BSH/EHA nor the publishers accept any legal responsibility for the content of this guidance.

ORCID

Noémi B. A. Roy https://orcid.org/0000-0002-0919-0790
Roberta Russo https://orcid.org/0000-0002-3624-7721
Paola Bianchi https://orcid.org/0000-0001-5976-5233
Immacolata Andolfo https://orcid.org/0000-0003-0493-812X
David Rees https://orcid.org/0000-0003-3647-1050
Achille Iolascon https://orcid.org/0000-0002-9558-0356

REFERENCES

1. Roy NB, Wilson EA, Henderson S, Wray K, Babbs C, Okoli S, et al. A novel 33-gene targeted resequencing panel provides accurate, clinical-grade diagnosis and improves patient management for rare inherited anaemias. Br J Haematol. 2016;175(2):318–30. https://doi.org/10.1111/bjh.14221
2. Agarwal AM, Nussenzveig RH, Reading NS, Patel JL, Sangle N, Salama ME, et al. Clinical utility of next-generation sequencing in the diagnosis of hereditary haemolytic anaemias. Br J Haematol. 2016;174(5):806–14. https://doi.org/10.1111/bjh.14131
3. He Y, Luo J, Lei Y, Jia S, Liao N. A novel PKLR gene mutation identified using advanced molecular techniques. Pediatr Transplant. 2018;22(2). https://doi.org/10.1111/petr.13143
4. Mirabello L, Khincha PP, Ellis SR, Giri N, Brodie S, Chandrasekharappa SC, et al. Novel and known ribosomal causes of diamond-Blackfan anaemia identified through comprehensive genomic characterisation. J Med Genet. 2017;54(6):417–25. https://doi.org/10.1136/jmedgenet-2016-104346
5. Hamada M, Doisaki S, Okuno Y, Muramatsu H, Hama A, Kawashima N, et al. Whole-exome analysis to detect congenital hemolytic anemia mimicking congenital dyserythropoietic anemia. Int J Hematol. 2018;108(3):306–11. https://doi.org/10.1007/s12185-018-2482-7
6. Russo R, Andolfo I, Manna F, Gambale A, Marra R, Rosato BE, et al. Multi-gene panel testing improves diagnosis and management of patients with hereditary anemias. Am J Hematol. 2018;93(5):672–82. https://doi.org/10.1002/ajh.25058
7. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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(5):405–24. https://doi.org/10.1038/gim.2015.30
8. Cree IA, Deans Z, Ligtenberg MJ, Normanno N, Edsjö A, Rouleau E, et al. Guidance for laboratories performing molecular pathology for cancer patients. J Clin Pathol. 2014;67(11):923–31. https://doi.org/10.1136/jclinpath-2014-202404

9. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013;15(9):733–47. https://doi.org/10.1038/gim.2013.92
10. Vetter B, Neu-Yilik G, Kohne E, Arnold R, Sinha P, Gaedicke G, et al. Dominant beta-thalassaemia: a highly unstable haemoglobin is caused by a novel 6 bp deletion of the beta-globin gene. Br J Haematol. 2000;108(1):176–81. https://doi.org/10.1046/j.1365-2141.2000.01820.x
11. Luo HY, Chui DH. Diverse hematological phenotypes of β-thalassemia carriers. Ann N Y Acad Sci. 2016;1368(1):49–55. https://doi.org/10.1111/nyas.13056
12. Traeger-Synodinos J, Kanavakis E, Vrettou C, Maragoudaki E, Michael T, Metaxotou-Mavromati A, et al. The triplicated alpha-globin gene locus in beta-thalassaemia heterozygotes: clinical, haematological, biosynthetic and molecular studies. Br J Haematol. 1996;95(3):467–71. https://doi.org/10.1046/j.1365-2141.1996.d01-1939.x
13. Williamson D. The unstable haemoglobins. Blood Rev. 1993;7(3):146–63. https://doi.org/10.1016/0268-960x(93)90002-l
14. Yates AM, Mortier NA, Hyde KS, Hankins JS, Ware RE. The diagnostic dilemma of congenital unstable hemoglobinopathies. Pediatr Blood Cancer. 2010;55(7):1393–5. https://doi.org/10.1002/pbc.22702
15. Iolascon A, Andolfo I, Barcellini W, Corcione F, Garçon L, De Franceschi L, et al. Recommendations regarding splenectomy in hereditary hemolytic anemias. Haematologica. 2017;102(8):1304–13. https://doi.org/10.3324/haematol.2016.161166
16. van Vuren A, van derZwaag B, Huisjes R, Lak N, Bierings M, Gerritsen E, et al. The complexity of genotype-phenotype correlations in hereditary spherocytosis: a cohort of 95 patients: genotype-phenotype correlation in hereditary spherocytosis. Hema. 2019;3(4):e276. https://doi.org/10.1097/HS9.0000000000000276
17. Bianchi P, Fermo E. Molecular heterogeneity of pyruvate kinase deficiency. Haematologica. 2020;105(9):2218–28. https://doi.org/10.3324/haematol.2019.241141
18. Grace RF, Rose C, Layton DM, Galactéros F, Barcellini W, Morton DH, et al. Safety and efficacy of Mitapivat in pyruvate kinase deficiency. N Engl J Med. 2019;381(10):933–44. https://doi.org/10.1056/NEJMoa1902678
19. Andolfo I, Russo R, Rosato BE, Manna F, Gambale A, Brugnara C, et al. Genotype-phenotype correlation and risk stratification in a cohort of 123 hereditary stomatocytosis patients. Am J Hematol. 2018;93(12):1509–17. https://doi.org/10.1002/ajh.25276
20. Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, et al. Resolution of disease phenotypes resulting from multilocus genomic variation. N Engl J Med. 2017;376(1):21–31.
21. Roy NBA, Babbs C. The pathogenesis, diagnosis and management of congenital dyserythropoietic anaemia type I. Br J Haematol. 2019;185(3):436–49. https://doi.org/10.1111/bjh.15817
22. Nord AS, Lee M, King MC, Walsh T. Accurate and exact CNV identification from targeted high-throughput sequence data. BMC Genomics. 2011;12:184. https://doi.org/10.1186/1471-2164-12-184
23. Kaneko K, Furuyama K, Fujiwara T, Kobayashi R, Ishida H, Harigae H, et al. Identification of a novel erythroid-specific enhancer for the ALAS2 gene and its loss-of-function mutation which is associated with congenital sideroblastic anemia. Haematologica. 2014;99(2):252–61. https://doi.org/10.3324/haematol.2013.085449
24. Gerrard G, Valgañón M, Foong HE, Kasperaviciute D, Iskander D, Game L, et al. Target enrichment and high-throughput sequencing of 80 ribosomal protein genes to identify mutations associated with diamond-Blackfan anaemia. Br J Haematol. 2013;162(4):530–6. https://doi.org/10.1111/bjh.12397
25. Gallagher PG, Maksimova Y, Lezon-Geyda K, Newburger PE, Medeiros D, Hanson RD, et al. Aberrant splicing contributes to severe α-spectrin-linked congenital hemolytic anemia. J Clin Invest. 2019;129(7):2878–87. https://doi.org/10.1172/JCI127195

How to cite this article: Roy NB, Da Costa L, Russo R, Bianchi P, Mañú-Pereira MdM, Fermo E, et al. The use of next-generation sequencing in the diagnosis of rare inherited anaemias: A Joint BSH/EHA Good Practice Paper. Br J Haematol. 2022;198:459–477. https://doi.org/10.1111/bjh.18191

SUPPLEMENTAL DATA I

Example list of genes implicated in rare inherited anaemias[§]

	Approved symbol	Location	Gene/locus name	Transcript RefSeq	Phenotype	Inheritance
Congenital Dyserythropoietic Anaemia (CDA)	*CDAN1*	15q15.2	Codanin 1	NM_138477.4	Dyserythropoietic anaemia, congenital, type Ia	AR
	CDIN1	15q14	CDAN1 Interacting Nuclease 1	NM_001321759.2	Dyserythropoietic anaemia, congenital, type Ib	AR
	GATA1	Xp11.23	GATA-binding protein-1 (globin transcription factor-1)	NM_002049.4	Anaemia, X-linked, with/ without neutropenia and/ or platelet abnormalities	XL
					Thrombocytopenia with beta-thalassemia, X-linked	XL
					Thrombocytopenia, X-linked, with or without dyserythropoietic anaemia	XL
					Diamond–Blackfan anaemia	XL
	KIF23	15q23	Kinesin family member 23	NM_138555.4	Dyserythropoietic anaemia, congenital, type III	AD
	KLF1	19p13.13	Kruppel-like factor 1, erythroid	NM_006563.5	[Hereditary persistence of fetal haemoglobin]	AD; AR
					Blood group--Lutheran inhibitor	AD
					Dyserythropoietic anaemia, congenital, type IV	AD
	LPIN2	18p11.31	Lipin 2	NM_014646.2	Majeed syndrome	AR
	SEC23B	20p11.23	SEC23 homologue B, coat complex II component	NM_006363.6	Dyserythropoietic anaemia, congenital, type II	AR
Diamond–Blackfan Anaemia (DBA)	*RPL5*	1p22.1	Ribosomal protein L5	NM_000969.5	Diamond–Blackfan anaemia 6	AD
	RPL9	4p14	Ribosomal protein L9	NM_000661.5	?Diamond–Blackfan anaemia	AD
	RPL11	1p36.11	Ribosomal protein L11	NM_000975.5	Diamond–Blackfan anaemia 7	AD
	RPL15	3p24.2	Ribosomal protein L15	NM_002948.5	?Diamond–Blackfan anaemia 12	AD
	RPL18	19q13.33	Ribosomal protein L18	NM_000979.4	?Diamond–Blackfan anaemia 18	AD
	RPL26	17p13.1	Ribosomal protein L26	NM_000987.5	?Diamond–Blackfan anaemia 11	AD
	RPL27	17q21.31	Ribosomal protein L27	NM_000988.5	?Diamond–Blackfan anaemia 16	AD
	RPL31	2q11.2	Ribosomal protein L31	NM_001098577.3	?Diamond–Blackfan anaemia	AD
	RPL35	9q33.3	Ribosomal protein L35	NM_007209.4	?Diamond–Blackfan anaemia 19	AD
	RPL35A	3q29	Ribosomal protein L35a	NM_000996.4	Diamond–Blackfan anaemia 5	AD
	RPS7	2p25.3	Ribosomal protein S7	NM_001011.4	Diamond–Blackfan anaemia 8	AD
	RPS10	6p21.31	Ribosomal protein S10	NM_001014.5	Diamond–Blackfan anaemia 9	AD
	RPS15A	16p12.3	Ribosomal protein S15a	NM_001019.5	?Diamond–Blackfan anaemia 20	AD
	RPS17	15q25.2	Ribosomal protein S17	NM_001021.6	Diamond–Blackfan anaemia 4	AD
	RPS19	19q13.2	Ribosomal protein S19	NM_001022.4	Diamond–Blackfan anaemia 1	AD
	RPS24	10q22.3	Ribosomal protein S24	NM_033022.4	Diamond–Blackfan anaemia 3	AD
	RPS26	12q13.2	Ribosomal protein S26	NM_001029.5	Diamond–Blackfan anaemia 10	AD
	RPS27	1q21.3	Ribosomal protein S27	NM_001030.6	Diamond–Blackfan anaemia 17	AD
	RPS27A	2p16.1	Ribosomal protein S27a	NM_002954.6	?Diamond–Blackfan anaemia	AD

	Approved symbol	Location	Gene/locus name	Transcript RefSeq	Phenotype	Inheritance
	RPS28	19p13.2	Ribosomal protein S28	NM_001031.5	Diamond–Blackfan anaemia 15 with mandibulofacial dysostosis	AD
	RPS29	14q21.3	Ribosomal protein S29	NM_001032.5	Diamond–Blackfan anaemia 13	AD
	EPO	7q22.1			?Diamond–Blackfan anaemia-like	AR
	ADA2	22q11.1	Adenosine deaminase 2	NM_001282225.2	Vasculitis, autoinflammation, immunodeficiency, and haematological defects syndrome	AR
	TSR2	Xp11.22	20S rRNA accumulation, *S. cerevisiae*, homologue of	NM_058163.3	?Diamond–Blackfan anaemia 14 with mandibulofacial dysostosis	XL
Enzyme defects	*AK1*	9q34.11	Adenylate kinase 1	NM_000476.3	Haemolytic anaemia due to adenylate kinase deficiency	AR
	G6PD	Xq28	Glucose-6-phosphate dehydrogenase	NM_001042351.3	Haemolytic anaemia, G6PD deficient (favism)	XL
	ALDOA	16p11.2	Aldolase, Fructose-Bisphosphate A	NM_001127617.2	Glycogen storage disease due to aldolase A deficiency	AR
	GCLC	6p12.1	Glutamate-cysteine ligase, catalytic subunit	NM_001498.4	Haemolytic anaemia due to gamma-glutamylcysteine synthetase deficiency	AR
	GPI	19q13.11	Glucose phosphate isomerase; neuroleukin	NM_000175.5	Haemolytic anaemia, non-spherocytic, due to glucose phosphate isomerase deficiency	AR
	GSS	20q11.22	Glutathione synthetase	NM_000178.4	Haemolytic anaemia due to glutathione synthetase deficiency	AR
	HK1	10q22.1	Hexokinase-1	NM_000188.3	Haemolytic anaemia due to hexokinase deficiency	AR
	NT5C3A	7p14.3	5′ nucleotidase, cytosolic IIIA (uridine 5′ monophosphate hydrolase 1)	NM_016489.14	Anaemia, haemolytic, due to UMPH1 deficiency	AR
	PGK1	Xq21.1	Phosphoglycerate kinase 1	NM_000291.4	Phosphoglycerate kinase 1 deficiency	XL
	PKLR	1q22	Pyruvate kinase, liver and RBC type	NM_000298.6	Pyruvate kinase deficiency	AR
	TPI1	12p13.31	Triosephosphate isomerase 1	NM_000365.6	Haemolytic anaemia due to triosephosphate isomerase deficiency	AR
Red cell membrane/ hydration disorder	*ABCB6*	2q35	ATP-binding cassette, subfamily B, member 6	NM_005689.4	[Blood group, Langereis system]	AR
					Pseudohyperkalaemia, familial, 2, due to red cell leak	AD
	ABCG5	2p21	ATP-binding cassette, subfamily G, member 5	NM_022436.3	Sitosterolaemia 2	AR
	ABCG8	2p21	ATP-binding cassette, subfamily G, member 8	NM_022437.3	Sitosterolaemia 1	AR
	ANK1	8p11.21	Ankyrin-1, erythrocytic	NM_000037.4	Spherocytosis, type 1	AD; AR
	ATP11C	Xq27.1	ATPase, class VI, type 11C	NM_001010986.3	?Haemolytic anaemia, congenital, X-linked	XL
	EPB41	1p35.3	Erythrocyte surface protein band 4.1	NM_004437.4	Elliptocytosis 1	AD
	EPB42	15q15.2	Erythrocyte surface protein band 4.2	NM_000119.3	Spherocytosis, type 5	AR

	Approved symbol	Location	Gene/locus name	Transcript RefSeq	Phenotype	Inheritance
	KCNN4	19q13.31	Potassium intermediate/ small conductance calcium-activated channel, subfamily N, member 4	NM_002250.3	Dehydrated hereditary stomatocytosis 2/Gardos Channelopathy	AD
	PIEZO1	16q24.3	PIEZO1 ion channel	NM_001142864.4	Dehydrated hereditary stomatocytosis with or without pseudohyperkalaemia and/ or perinatal oedema	AD
	RHAG	6p12.3	Rh blood group-associated glycoprotein	NM_000324.3	Overhydrated hereditary stomatocytosis	AD
					Anaemia, haemolytic, Rh-null, regulator type	AR
	SLC2A1	1p34.2	Solute carrier family 2 (facilitated glucose transporter), member 1	NM_006516.4	Stomatin-deficient cryohydrocytosis with neurological defects	AD
	SLC4A1	17q21.31	Solute carrier family 4, anion exchanger, member 1 (erythrocyte membrane protein band 3, Diego blood group)	NM_000342.4	[Blood group, Diego – Froese – Swann – Waldner – Wright]	AD
					Cryohydrocytosis	AD
					Distal renal tubular acidosis 4 with haemolytic anaemia	AR
					Ovalocytosis, Southeast Asian type	AD
					Spherocytosis, type 4	AD
	SPTA1	1q23.1	Spectrin, alpha, erythrocytic-1	NM_003126.4	Elliptocytosis 2	AD
					Pyropoikilocytosis	AR
					Spherocytosis, type 3	AR
	SPTB	14q23.3	Spectrin, beta, erythrocytic	NM_001355436.2	Elliptocytosis-3	AD
					Spherocytosis, type 2	AD
Atypical microcytic anaemias	*SLC11A2*	12q13.12	Solute carrier family 11 (proton-coupled divalent metal ion transporter) member 2	NM_000617.3	Anaemia, hypochromic microcytic, with iron overload 1	AR
	STEAP3	2q14.2	STEAP3 metalloreductase	NM_001008410.2	?Anaemia, hypochromic microcytic, with iron overload 2	AD
	TMPRSS6	22q12.3	Transmembrane protease, serine 6 (matriptase 2)	NM_153609.4	Iron-refractory iron deficiency anaemia	AR
Sideroblastic Anaemia	*ABCB7*	Xq13.3	ATP-binding cassette-7	NM_004299.6	Anaemia, sideroblastic, with ataxia	XL
	ALAS2	Xp11.21	Aminolevulinate, delta-, synthase-2	NM_000032.5	Anaemia, sideroblastic, 1	XL
	GLRX5	14q32.13	Glutaredoxin 5	NM_016417.3	Anaemia, sideroblastic, 3, pyridoxine-refractory	AR
	HSCB	22q12.1	Mitochondrial iron–sulfur cluster co-chaperone	NM_172002.5	?Anaemia, sideroblastic, 5	AR
	HSPA9	5q31.2	Heat-shock 70 kD protein-9 (mortalin)	NM_004134.7	Anaemia, sideroblastic, 4	AD
	LARS2	3p21.31	Leucyl-tRNA synthetase, mitochondrial	NM_015340.4	Hydrops, lactic acidosis, and sideroblastic anaemia	AR
					Perrault syndrome 4	AR
	NDUFB11	Xp11.3	NADH–ubiquinone oxidoreductase 1 beta subcomplex, 11	NM_019056.7	?Mitochondrial complex I deficiency, nuclear type 30	XL

	Approved symbol	Location	Gene/locus name	Transcript RefSeq	Phenotype	Inheritance
	PUS1	12q24.33	Pseudourine synthase 1	NM_025215.6	Myopathy, lactic acidosis, and sideroblastic anaemia 1	AR
	SLC19A2	1q24.2	Solute carrier family 19 (thiamine transporter), member 2	NM_006996.3	Thiamine-responsive megaloblastic anaemia syndrome	AR
	SLC25A38	3p22.1	Solute carrier family 25, member 38	NM_017875.4	Anaemia, sideroblastic, 2, pyridoxine-refractory	AR
	TRNT1	3p26.2	tRNA nucleotidyltransferase, CCA-adding, 1	NM_182916.3	Sideroblastic anaemia with B-cell immunodeficiency, periodic fevers, and developmental delay	AR
					Retinitis pigmentosa and erythrocytic microcytosis	AR
	YARS2	12p11.21	Tyrosyl-tRNA synthetase 2	NM_001040436.3	Myopathy, lactic acidosis, and sideroblastic anaemia 2	AR

Brackets, '[]', indicate 'non-diseases', mainly genetic variations that lead to apparently abnormal laboratory test values.

A question mark, '?', before the phenotype name indicates that the relationship between the phenotype and gene is provisional.

Abbreviations: AR, autosomal recessive; AD, autosomal dominant; XL, X-linked.

[§]Genes with high/moderate evidence of association with rare anaemias.

SUPPLEMENTAL DATA II

PATIENT

SURNAME: ..... NAME: .....
GENDER: .....
D.O.B.: .....
ADDRESS: .....

PATIENT REFERENCE NUMBER: .....

REQUESTER DETAILS

CLINICIAN NAME: .....

EMAIL ADRESS: .....
PHONE NUMBER: .....
HOSPITAL: .....
ADDRESS: .....

SUSPECTED DIAGNOSIS: .....
.....

□ **THIS IS THE PROBAND**
□ **THIS IS A FAMILY MEMBER**
PROBAND NAME: .....
RELATIONSHIP TO THE PROBAND: .....
AFFECTED/UNAFFECTED
INFORMED CONSENT TO PERFORM GENETIC INVESTIGATIONS: YES/NO

FAMILY HISTORY

Maternal Origin:.....
Paternal Origin:.....

CONSANGUINITY: NO YES

FAMILY TREE

Please identify the proband and the patient in study (if family member)

Previous pregnancies/miscarriages:.....
.....

OTHER: .....
.....

DATA AT BIRTH if available:

Pregnancy history/complications: .....
Gestational age and birth weight: .....
Neonatal jaundice: NO YES
Bilirubin level: .....
Phototherapy: NO YES
Exchange transfusion: NO YES
Hb g/L: .....
Transfusions at birth: NO YES
Other:.....

Onset of anaemia (age) .....

First available data

Age: ……

Hb: …… RBC: …… WBC: …… Plt: ……

MCV: …… MCH: …… MCHC: ……

Reticulocytes (x10^9/l): …… Unconjugated Bilirubin: ……

LDH: …… Haptoglobin: ……

Iron: …… Ferritin: …… Transferrin saturation: ……

DAT: ……

Recent data:

Age: ……

Hb: …… RBC: …… WBC: …… Plt: ……

MCV: …… MCH: …… MCHC: ……

Reticulocytes (x10^9/l): …… Unconjugated Bilirubin: ……

LDH: …… Haptoglobin: ……

Iron: …… Ferritin: …… Transferrin saturation: ……

DAT: ……

RBC morphology description:……

Other laboratory tests performed: ……

……

CLINICAL DATA:

Jaundice	NO	YES	
Splenomegaly	NO	YES	cm below costal margin……
Hepatomegaly	NO	YES	cm below costal margin……
Aplastic crises	NO	YES	(when)……..
Haemolytic crises	NO	YES	(when)……..
Infections (CMV,PVB19..)	NO	YES	(when)……..
Gallstones	NO	YES	
Malleolar ulcers	NO	YES	
Pancreatic Insufficiency	NO	YES	
Dysmorphic facies	NO	YES	
Developmental delay	NO	YES	
Learning difficulties	NO	YES	
Short stature/failure to thrive	NO	YES	
Skeletal, limb or digit abnormalities	NO	YES	

THERAPY

TRANSFUSIONS: □ NO □YES (n°/year)…… □ OCCASIONAL(when)……

SPLENECTOMY: □ NO □YES age ……..

TRANSFUSIONS after splenectomy:

□ NO □YES (n°/year)…… □ OCCASIONAL(when)…….

CHOLECYSTECTOMY: □ NO □YES (age) ……..

IRON CHELATION □ NO □ YES (from …….. to ……..)

Other therapies ……

……

SUPPLEMENTAL DATA III

Validation strategy and available databases for variant verification

Validation:

- Determine number of reads required to pass the run. Set pass/fail criteria
- Determine sensitivity and specificity of SNVs and indels separately
- Evaluate size of indels that can be detected
- Determine minimum coverage to call heterozygosity
- Ensure coverage of paralogous genes
- Gap filling required issuing a negative report in those case with strong clinical suspicion
- Determine sensitivity and specificity, ensuring that types of variants that are disease causing are reflected in the types of variants that are used for validation
- Robustness and assay precision – e.g., three runs running three of the same samples as a minimum
- e.g., ~300 variants should be tested independently (they do not have to be pathogenic variants), at least 30 samples run on same/different runs

Database name	URL	Content	Access
Alamut Visual	https://www.interactive-biosoftware.com/alamut-visual/	Software to visualise genetic variants to help in the assessment of pathogenicity	Pay for licence
ClinGen	https://clinicalgenome.org/	NIH funded resource to define the clinical relevance of genes and variants for use in precision medicine and research.	Publicly available
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/	Part of the NCBI databases – genetic variation and relationship to health	Publicly available
DBA Mutation Database	http://www.dbagenes.unito.it		[Boria et al. 2008)
Ensembl Genome Browser	https://www.ensembl.org/index.html	Helpful to visualise regions of interest and transcripts of interest, can overlay other tracks such as transcription factor binding sites. Use to find ensemble gene transcripts, also variant effect predictor.	Publicly available
gnomAD	https://gnomad.broadinstitute.org/	Genome aggregation database. Describes variants identified in whole genome and exome sequencing studies and the frequency in specific populations	Publicly available
HbVar	http://globin.cse.psu.edu/hbvar/menu.html	Database of variants specific to haemoglobinopathies.	Publicly available
HGMD	http://www.hgmd.cf.ac.uk/ac/all.php	Database of genetic variants – mostly pathogenic	There is a public and a pay to view version with different content
ithanet	https://www.ithanet.eu/db/ithagenes	Sequence variation affecting haemoglobin disorders	Publicly available
LOVD	http://databases.lovd.nl/whole_genome/variants/PKLR	Locus specific database of variants identified in the *PKLR* gene	Publicly available
NCBI	https://www.ncbi.nlm.nih.gov/	Multiple tools and databases available	Publicly available
PKLR Paper	https://doi.org/10.3324/haematol.2019.241141	Molecular heterogeneity of pyruvate kinase deficiency	Publicly available
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/	Prediction of functional effects of human SNPs	Publicly available
SIFT	https://sift.bii.a-star.edu.sg/	Predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids.	Publicly available
Splice site prediction tools	http://www.phenosystems.com/www/index.php/links-to-various-tools-and-information/splice-prediction-tools	Landing page for multiple tools such as GeneSplicer, Splice Predictor, ESEfinder, MIT splice predictor	Publicly available

Database name	URL	Content	Access
UCSC Genome Browser	http://genome.ucsc.edu/	Helpful to visualise regions of interest and transcripts of interest, can overlay other tracks such as transcription factor binding sites	Publicly available
Variant Validator	https://variantvalidator.org/	Accurate validation, mapping and formatting of sequence variants using HGVS nomenclature.	Publicly available
Varsome	https://varsome.com/	Assessing pathogenicity of genetic variants	Publicly available but can have your own account
VASA	https://vasa.rosalind.kcl.ac.uk/	Repository of scored variants with a tool to support variant scoring. Includes multiple red cell-specific genes.	Free but access by request