

## **Target Enrichment and High-Throughput Sequencing of 80 Ribosomal Protein Genes to Identify Mutations Associated with Diamond-Blackfan Anemia**

Gareth Gerrard<sup>1,3</sup>, Mikel Valgañón<sup>1,3</sup>, Hui En Foong<sup>1,3</sup>, Dalia Kasperaviciute<sup>2</sup>, Deena Iskander<sup>3</sup>, Laurence Game<sup>4</sup>, Michael Müller<sup>2</sup>, Timothy J Aitman<sup>2,4</sup>, Irene Roberts<sup>3,5</sup>, Josu de la Fuente<sup>1,5</sup>, Letizia Foroni<sup>1,3</sup>, Anastasios Karadimitris<sup>3</sup>

<sup>1</sup>*Imperial Molecular Pathology Laboratory, Imperial College Healthcare NHS Trust and Academic Health Sciences Centre, Hammersmith Hospital, London W12 0HS, UK;*

<sup>2</sup>*Imperial NIHR Biomedical Research Centre, Imperial College London, London W12 0NN, UK;*

<sup>3</sup>*Centre for Haematology, Faculty of Medicine, Imperial College London, Hammersmith Hospital, London W12 0NN, UK*

<sup>4</sup>*Genomics Laboratory, MRC Clinical Sciences Centre, Imperial College London, London W12 0NN, UK;*

<sup>1,5</sup>*Pediatric Haematology and Bone Marrow Transplantation Unit, Department of Paediatrics, Imperial College Healthcare NHS Trust, London W2 1NY, UK*

**Short Title:** Next generation sequencing diagnostics in DBA

**Scientific Category:** Red Cells, Iron and Erythropoiesis

**Abstract Word Count:** 149

**Text Word Count:** 1202

**Keywords:** Diamond-Blackfan Anemia, Next Generation Sequencing

### **Correspondence:**

Gareth Gerrard, Imperial Molecular Pathology, 2<sup>nd</sup> Floor G-Block, Hammersmith Hospital, Du Cane Road, London, W12 0HS, UK; tel: +44 20 8383 2177; E-mail: g.gerrard@imperial.ac.uk

Anastasios Karadimitris, Centre for Haematology, Imperial College London, 4<sup>th</sup> Floor Commonwealth Building, Hammersmith Hospital, Du Cane Road, London, W12 0NN, UK; tel: +44 20 3313 8438 ; fax: +44 20 8742 9335; E-mail: a.karadimitris@imperial.ac.uk

### ***Key Points***

- Screening for mutations in all 80 ribosomal protein (RP) genes in DBA is a major hurdle for accurate genetic diagnosis
- We developed and validated, a practice-changing, deep sequencing-based method to rapidly and accurately screen for mutations in DBA

### ***Abstract***

The bone marrow failure syndrome Diamond-Blackfan anemia (DBA) is caused by monoallelic, inactivating mutations in ribosomal protein (RP) genes. Of the 80 RP genes, *RPS19* is mutated in up to 25% of cases and another 13 RP genes in a further 25-35% of patients. The molecular basis of almost all of the remaining 40-50% of cases is unknown. Since such cases may harbor mutations in one or more of the remaining RP genes and may occur at very low frequencies, genetic screening by conventional approaches is challenging. We therefore developed custom target enrichment technology combined with a high-throughput sequencing platform to screen all 80 RP genes. Using this approach, we identified and validated both known and novel inactivating mutations in 15/17 (88%) DBA patients. All mutations affected known DBA RP genes. Target enrichment combined with high-throughput sequencing is a robust and improved methodology for the genetic diagnosis of DBA.

## **Introduction**

Diamond-Blackfan anemia (DBA) is an heritable red cell aplasia caused by a selective defect in erythroid progenitor and precursor cell development<sup>1</sup>. Patients usually present with anemia in the first year of life, may manifest other developmental defects and have a lifetime modest risk of developing leukemia and sarcomas<sup>2</sup>. DBA is associated with monoallelic, inactivating mutations in ribosomal protein (RP) genes. The majority of patients have *de novo* mutations but in ~20% of cases the disorder is familial (predominantly autosomal dominant) and there is often clinical heterogeneity due to variable penetrance<sup>3</sup>.

Mutations in *RPS19* gene account for up to 25% of all DBA cases<sup>4</sup>. Single nucleotide variations (SNV), small inversions/deletions (indels) and copy number variations (CNV) have been found in 13 other RP genes (*RPS7*, *RPS10*, *RPS17*, *RPS24*, *RPS26*, *RPS27*, *RPS29*, *RPL5*, *RPL9*, *RPL11*, *RPL19*, *RPL26*, *RPL27* and *RPL35a*) in a further 25-35% of patients<sup>5-10</sup>, meaning that up to 40% of patients have no identifiable mutations by conventional screening techniques. Given that nearly all mutations characterized so far affect RP genes, it is likely that mutations in one of the 80 RP genes will eventually be identified in the great majority of DBA patients. Current screening diagnostic methods typically employ Sanger sequencing on a per-exon/per-gene basis, with the associated time, labor and cost restrictions that this approach entails. We therefore aimed to evaluate high-throughput sequencing technology, including a bespoke target enrichment platform, to screen all 80 known RP genes, to facilitate rapid, cost-effective identification of DBA-associated mutations.

## **Methods**

### *Patients*

DNA was extracted from peripheral blood samples from 20 individuals (with appropriate parental consent), including 3 pairs of family members: affected mother and daughter; 2 affected siblings; another sibling pair, one of whom was unaffected (no defining clinical symptoms, except for a high erythrocyte adenine deaminase - eADA). Two patients with known mutations (*RPS19* c.280C>T; *RPL11* c.203delT) were included as controls. The patients' clinical details are detailed in **Table 1**.

### *Bioinformatics*

The sequencing reads were aligned to the build 37 (hg19) reference genome using BWA<sup>11</sup> and the variant calls made using GATK<sup>12</sup>; ANNOVAR was used for functional annotations of the variants<sup>13</sup>. Pipelines for both SNVs/indels and large deletions/insertions were implemented. Post-run sequencing quality was assessed by FastQC (Babraham Bioinformatics).

Target Enrichment (SureSelect XP, Agilent), Sequencing (Illumina MiSeq) and Bioinformatics analysis of *RPS17* are discussed in Supplemental Methods.

### **Results & Discussion**

We sought to establish a genomics-based approach to detect causative mutations in patients with DBA. For this purpose, we tested 20 individuals (19 DBA patients and one unaffected sibling) . Two patients were previously found to have inactivating mutations in *RPS19* and *RPL11* were used as positive controls (S07 and S18) and 18 had not previously undergone genetic testing and therefore lacked genetic diagnosis (**Table 2**).

Putative loss-of-function mutations were detected in RP genes in 15 of the 17 (88%) uncharacterised patients. No mutation was identified in the unaffected sibling S04 of S01 and the previously identified mutations were confirmed in the 2 control patients; in only two patients (S11 and S20) was no mutation identified. All mutations were in RP genes previously described in DBA, although 7 affected novel mutations (not reported on the LOVD database; <http://www.dbagenes.unito.it>) were identified. The FastQC software indicated good quality sequencing metrics for the high-throughput results and all variations were subsequently validated by Sanger sequencing of the affected gene regions.

The most commonly mutated gene in our cohort was *RPL5*, with 4 mutations (2 stop-gain SNV and 2 frameshift deletions) in 5 patients (in the mother-daughter pair S05 & S06; in S13, S10 and S17). The next most commonly mutated genes, with 3 mutations each, were *RPS26* (1 frameshift insertion, 1 start-loss and 1 stop-gain SNV) in 3 individuals (S08, S15 and S16) and *RPL11* (2 frameshift deletions and 1 frameshift insertion) in 3 individuals (S09, S18 – control and S19). A single stop-gain *RPS17* mutation was identified in the affected sibling pair S02 & S03 while 4 genes

were found to be mutated in single individuals (3 stop-gains and one splice site defect): *RPS10* in S01; *RPS19* in S07 (control); *RPS24* in S14 and *RPS7* in S12. In four cases (S01, S13, S18 and S19) samples from both parents were available for subsequent targeted screening. No mutations were detected in any of the parental samples, indicating that these cases are likely to be the result of *de novo* mutations. Ongoing pedigree studies will continue to determine the inheritance pattern for each case/mutation found.

All frameshift events introduced a premature stop codon in either the same or an upstream exon, which along with the explicit stop-gains and the splice-site defect meant that all the associated transcripts would most likely be flagged for nonsense mediated decay and lead to haploinsufficiency<sup>14,10</sup>. The *RPS26* loss of methionine translation start signal would predict that no protein would be generated from that allele, ultimately leading to the DBA phenotype displayed by the affected individual<sup>11</sup>. Although mutations in *RPS19* have been reported in up to 25% of patients, there was only a single *RPS19* mutation in our cohort corresponding to the known positive control (S07). However, the frequency of *RPS19* mutations varies between different ethnic groups, being as low as 11% in the Japanese cohort<sup>15</sup>. Indeed, the overall frequency of *RPS19* mutations in our entire cohort of patients, in which at least 20% are non-Caucasian, including the patients described here, is 5/48 (10.4%; Table 1 and data not shown).

Of the 3 individuals in whom no definitive mutation was identified, a diagnosis of DBA was excluded in one (S04, sibling of patient S01) because of lack of clinical and laboratory diagnostic criteria for DBA, except for a single measurement of increased eADA. However, in patients S11 and S20, both with a confirmed diagnosis of DBA, lack of mutations would be consistent with the presence of either large RP gene deletions or lesions in non-RP genes. Large RP gene deletions were recently described in up to 20% of patients with DBA<sup>6,16,17</sup> by a variety of methods, including SNP arrays. In principle, our high throughput sequencing approach, by indentifying exonic and intronic SNP in all 80 RP genes, could also be employed as a screening approach for large deletions.

Comprehensive analysis of large cohorts of patients with DBA suggests that non-RP genes may be causative in a minority of patients. Indeed, *GATA1* mutations were identified by whole exome sequencing in 2 cases with apparent X-linked inheritance<sup>18</sup>. In our cohort, parental consanguinity for S20 is suggestive of an

autosomal recessive trait, thus raising the possibility of a non-RP gene mutation in this family.

In summary, we showed here that target-gene enrichment followed by multiplexing and sequencing on a bench-top class high-throughput sequencing platform provides several advantages over conventional and even more global approaches. In particular, it allows rapid and accurate screening of all 80 RP genes, an attribute which is crucial for selecting family members as donors for allogeneic stem cell transplants and also for purposes of pre-natal diagnosis and counselling. Furthermore, it lacks the ethical implications of whole exome sequencing resulting from identification of off-target disease-associated mutations. Therefore high throughput sequencing technology combined with a bespoke target enrichment platform for RP genes is a feasible, efficient and relatively rapid diagnostic tool for the detection of causative mutations in DBA.

***Acknowledgements***

The authors would like to acknowledge the clinical staff at St Mary's Hospital Paediatric Haematology and Stem Cell Transplantation Unit and all the patients and their families. This work was funded by the Imperial Healthcare Comprehensive Biomedical Research Centre with infrastructure support from an NIHR Capital Investment award.

***Authorship Contributions***

GG designed the target enrichment baits, performed the library preparation, helped analyse the data and write the manuscript; MV & HF performed the DNA extraction and Sanger sequencing and commented on the manuscript; DI extracted DNA, compiled clinical data and commented on the paper; DK & MM performed the bioinformatics and commented on the manuscript; LG ran the high-throughput sequencing facility and commented on the manuscript; TJA advised on study design, contributed infrastructure support and commented on the manuscript; IR helped design the study and commented on the manuscript; JdlF provided clinical patient care, helped design the study and commented on the manuscript; LF & AK designed the study and helped write the manuscript.

***Disclosure of Conflicts of Interest***

The authors confirm that there are no relevant conflicts of interest to disclose.

## References

1. Ball S. Diamond Blackfan anemia. *Hematology Am Soc Hematol Educ Program*. 2011;2011:487-491.
2. Vlachos A, Ball S, Dahl N, et al. Diagnosing and treating Diamond Blackfan anaemia: results of an international clinical consensus conference. *Br J Haematol*. 2008;142(6):859-876.
3. Farrar JE, Dahl N. Untangling the phenotypic heterogeneity of Diamond Blackfan anemia. *Semin Hematol*. 2011;48(2):124-135.
4. Draptchinskaia N, Gustavsson P, Andersson B, et al. The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat Genet*. 1999;21(2):169-175.
5. Gazda HT, Preti M, Sheen MR, et al. Frameshift mutation in p53 regulator RPL26 is associated with multiple physical abnormalities and a specific pre-ribosomal RNA processing defect in diamond-blackfan anemia. *Hum Mutat*. 2012;33(7):1037-1044.
6. Farrar JE, Vlachos A, Atsidaftos E, et al. Ribosomal protein gene deletions in Diamond-Blackfan anemia. *Blood*. 2011;118(26):6943-6951.
7. Lipton JM, Ellis SR. Diamond-Blackfan anemia: diagnosis, treatment, and molecular pathogenesis. *HematolOncol Clin North Am*. 2009;23(2):261-282.
8. Doherty L, Sheen MR, Vlachos A, et al. Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. *Am J Hum Genet*. 2010;86(2):222-228.
9. Ito E, Yoshida Y, Yusuke Okuno Y, et al. Identification of Two New DBA Genes, RPS27 and RPL27, by Whole-Exome Sequencing in Diamond-Blackfan Anemia Patients. *54th American Society of Hematology Annual Meeting and Exposition*. 2012. Abstract 984.
10. Mirabello L, Ballew B, Giri N, et al. RPS29 is Mutated in a Multi-Case Diamond Blackfan Anemia Family. *54th American Society of Hematology Annual Meeting and Exposition*. 2012. Abstract 511.
11. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595.
12. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.
13. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
14. Gazda HT, Zhong R, Long L, et al. RNA and protein evidence for haplo-insufficiency in Diamond-Blackfan anaemia patients with RPS19 mutations. *Br J Haematol*. 2004;127(1):105-113.
15. Konno Y, Toki T, Tandai S, et al. Mutations in the ribosomal protein genes in Japanese patients with Diamond-Blackfan anemia. *Haematologica*. 2010;95(8):1293-1299.
16. Kuramitsu M, Sato-Otsubo A, Morio T, et al. Extensive gene deletions in Japanese patients with Diamond-Blackfan anemia. *Blood*. 2012;119(10):2376-2384.
17. Quarello P, Garelli E, Brusco A, et al. High frequency of ribosomal protein gene deletions in Italian Diamond-Blackfan anemia patients detected by multiplex ligation-dependent probe amplification assay. *Haematologica*. 2012;97(12):1813-1817.
18. Sankaran VG, Ghazvinian R, Do R, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J Clin Invest*. 2012.





**Table 1. Summary of clinical characteristics of individuals included in the study**

ID	Age (yr)/ Gender/ Ethnicity	Birth details	Age at diagnosis - presentation	eADA (nmol/hr/mgHb)	Additional features	Family History	Steroid response	Current treatment
S01	4/F/ Caucasian	Normal birth at term	17 mo Hb 1.7g/dl	U	Growth retardation; vitamin D deficiency; Cathie facies	1 unaffected sib (S04)	Steroid responsive	None- in remission
S02	4/F/ Caucasian	Normal birth at term	11 mo anemia	U	Growth retardation; ostium secundum ASD	Sib with DBA (S03)	Steroid responsive	Prednisolone (0.09mg/kg alt days)
S03	11/M/ Caucasian	Normal birth at term	Birth high eADA in cord blood, sib with DBA	U	Hernia, neutropenia	Sib with DBA (S02)	Steroid responsive	Prednisolone (0.08mg/kg alt days)
S04	2/M/ Caucasian	U	NA	136	NA	Sib with DBA (S01)	NA	NA
S05	10/F/ Caucasian	EMLSCS at 35/40 -IUGR and fetal distress	Birth Anemia, high eADA, mother with DBA	U	VSD (resolved), cleft palate; clinodactyly; Cathie facies; growth retardation; vitamin D deficiency; severe hepatic iron overload	Mother of S06	Secondary resistance	Prednisolone (0.9mg/kg alt days) + transfusions
S06	39/F/ Caucasian	Normal birth at term	5 yr Hb 5.2g/dl	109	Growth retardation; osteoporosis, thumb abnormalities; severe hepatic iron overload	Daughter of S05	Steroid responsive	Prednisolone (0.12mg/kg od)
S07	1/M/ Caucasian	U	U	U	U	U	U	U
S08	38/M/ Caucasian	U	U	U	U	U	U	U
S09	7/M/ Caucasian	EMLSCS at 36/40 -fetal distress	Birth Hb 8.2g/dl	305	Patent Ductus Arteriosus with spontaneous closure; hypoplastic thumbs; recurrent chest infections; vitamin D deficiency	None	Steroid responsive	Prednisolone (0.24mg/kg od)

S10	16 <sup>c</sup> /M/ Caucasian	U	U	U	Cleft palate; osteoporosis; poor feeding requiring gastrostomy tube	9 unaffected sibs	Secondary resistance	Sibling bone marrow transplant aged 10yr
S11	31/F/ Caucasian	Normal birth at 34/40	3 yr anemia DBA in sib	628	Pyloric stenosis at 6m; recurrent chest infections; vitamin D deficiency; osteopenia; eczema; nasal polyps	Sib with DBA	Never had steroid trial	None- Hb adequate and stable off treatment
S12	10/F/ Caucasian	EMLSCS at term- poor progression in labour	2 mo Hb 3g/dl	U	Growth retardation; Cathie facies; neutropenia; mild hepatic iron overload	2 unaffected sibs	Secondary resistance	Sibling bone marrow transplant aged 11yr
S13	4/M/ Caucasian	Normal birth at term	7 wks Hb 3g/dl	U	Growth retardation; cleft palate; oesophageal strictures; eosinophilic oesophagitis; triphalangeal thumbs; obstructive sleep apnoea; severe hepatic iron overload	1 unaffected sib	Primary resistance	TD
S14	1.5/M/ Other Asian	EMLSCS at 37/40- premature rupture of membranes	Birth Hb 7.8g/dl	U	Growth retardation; factor XI deficiency; delayed linguistic development	1 unaffected sib	Secondary resistance	TD
S15	8/M/ Caucasian	Normal birth at 38/40	6 wks Hb1.9g/dl	U	Winged scapula; vitamin D deficiency; moderate iron overload	2 unaffected sibs	Primary resistance	Sibling bone marrow transplant aged 8 yr
S16	6/F/ Arab	U	2 wks Hb 3.9g/dl	U	Growth retardation; recurrent pneumonia in infancy; VSD; vitamin D deficiency; choanal atresia; bilateral auditory canal atresia; Klippel- Feil syndrome; Sprengel shoulder deformity	None	Primary resistance	TD

S17	12/F/ Caucasian	EMLSCS at 35/40 – fetal distress, IUGR and polyhydramnios	11 yr anemia (also transfused perinatally)	191	Growth retardation; ASD; gastroesophageal reflux disease; growth retardation; hearing difficulty; asthma; cleft palate; frequent otitis media; neutropenia	1 unaffected sib	Steroid trial ongoing	TD
S18	5/F/ Indian	EMLSCS at term - fetal distress and IUGR	3 mo Hb 4g/dl	U	Growth retardation, recurrent infections	None	Secondary resistance	TD
S19	7/F/ Caucasian	EMLSCS at 38/40 - fetal distress	1yr Hb 5g/dl	322	Growth retardation; recurrent otitis media; mouth ulcers; neutropenia; hypoplastic thumbs; scoliosis; Cathie facies; Sprenger shoulder deformity	None	Monitoring	Monitoring
S20 <sup>a</sup>	1.5/M/ Pakistani	Normal birth at term	6 wks Hb1.8g/dl	150	ASD; VSD	None	Steroid trial ongoing	TD

eADA: erythrocyte adenosine deaminase; ASD: atrial septal defect; BM: bone marrow; EMLSCS: Emergency lower segment Caesarian Section; IUGR: intrauterine growth restriction; NA: not applicable; TD: transfusion dependent; U: unknown; VSD: ventricular septal defect.

<sup>a</sup> Parents first cousins, no parental consanguinity or status unknown for the rest of patients

<sup>b</sup> Normal range: 40-100 nmol/hr/mgHb

<sup>c</sup> Diagnosis of DBA made on pre-transplant sample

**Table 2. Gene variations flagged as loss of function and validated by Sanger sequencing**

ID	Gene	Type of variation	Depth	Exon	Base	Codon	Chr	Start/End	Ref	Seen	Previously Reported?	Notes
S01	RPS10	Stopgain SNV	215	4	C337T	R113*	6	34389570/ 34389570	G	A	Yes	Sibling of S04
S02	RPS17	Stopgain SNV	44	3	T159G	Y53*	15	82823390/ 82823390	T	G	No	Sibling of S03
S03	RPS17	Stopgain SNV	55	3	T159G	Y53*	15	82823390/ 82823390	T	G	No	Sibling of S02
S04	-											Non-affected sibling of S01
S05	RPL5	Stopgain SNV	226	4	G244T	E82*	1	93300390/ 93300390	G	T	No	Daughter of S06
S06	RPL5	Stopgain SNV	225	4	G244T	E82*	1	93300390/ 93300390	G	T	No	Mother of S05
S07	RPS19 <sup>a</sup>	Stopgain SNV	48	4	C280T	R94*	19	42373208/ 42373208	C	T	Yes (rs61762293)	
S08	RPS26	Frameshift insertion	230	3	212- 213ins	L71fs	12	56437177/ 56437177	-	A	No	Premature stop in exon 3
S09	RPL11	Frameshift deletion	110	5	475_476d el	159_159d el	1	24022366/ 24022367	A A	-	No	Premature stop in exon 6
S10	RPL5	Frameshift deletion	42	3	166_169d el	56_57del	1	93299194/ 93299197	A C A	-	Yes	Premature stop in exon 4
S11	-											

S1 2	RPS7	Splicing	95	3	147+1G> T	-	2	3623479/ 3623479	G	T	Yes (G>A)	Splice site defect
S1 3	RPL5	Stopgain SNV	183	6	C664T	Q222*	1	93303149/ 93303149	C	T	No	
S1 4	RPS24	Stopgain SNV	226	2	C46T	R16*	10	79795152/ 79795152	C	T	Yes (rs104894189)	
S1 5	RPS26	Start Loss SNV	108	1	A1G	M1V	12	56435951/ 56435951	A	G	Yes (rs143951267)	Loss of translation start
S1 6	RPS26	Stopgain SNV	164	3	C259T	R87*	12	56437224/ 56437224	C	T	Yes (rs148942765)	
S1 7	RPL5	Frameshift deletion	144	3	172_173d el	58_58del	1	93299200	A G	-	Yes	Same seq as 175_176delGA
S1 8	RPL11 <sup>a</sup>	Frameshift deletion	210	3	203del	I68del	1	24020342/ 93299201	T	-	No	Premature stop in exon 3
S1 9	RPL11	Frameshift insertion	196	3	202_203i ns	I67fs	1	24020341/ 24020341	A T	AGT	No	Frameshift Substitution
S2 0	-											

<sup>a</sup> Control sample previously characterized by conventional Sanger sequencing screen