# Molecular Pathology of Diamond Blackfan Anaemia

**By**

**Qais Al-Oqaily**

**A thesis submitted for the degree of Doctor of Philosophy**

**August 2018**

**Centre of Haematology**

**Imperial College London**

**CID:00897084**

# COPYRIGHT

# DECLARATION

In this thesis all the experiments described were designed and performed by myself unless otherwise stated. Contributions from other people are acknowledged in the appropriate sections. Experiments design, data analysis, graphs and figures were done with the guidance of my supervisors, Prof Anastasios Karadimitris and Prof Letizia Foroni and Dr. Josu de la Fuente

Qais Al-Oqaily

August 2018

# ABSTRACT

Diamond-Blackfan anaemia (DBA) is one of a rare group of genetic disorders known as 'inherited bone marrow failure (IBMF) syndromes with bone marrow failure, birth defects and higher propensity to cancer. It is a rare autosomal dominant disorder with an incidence of 7 in $10^6$ newborns, DBA is characterized by a defect in erythroid lineage development and a quantitative as well as a qualitative defect in erythroid progenitors. This disorder is inherited in 45% of cases and emerges as de novo in the remaining cases.

Diagnosis of DBA is challenging as it is based on several clinical features shared by other IBMF syndromes.

Diagnosis of this disorder has remained for many years a challenge. In recent years, the use of Next generation sequencing (NGS) of the 83 ribosomal protein (RP) genes has greatly improved the diagnosis and the future prospect of managing the disease. This new diagnostic approach provides several advantages over other conventional and classic approaches, last but not least that now over 65-80% of DBA cases have been identified to carry a genetic mutation of one of the 83 RP genes, while only few could be analyzed until recently. It also facilitates the analysis of DBA family members to facilitate both the selection of possible donors for transplant and to distinguish inherited from *de novo* mutations. RP gene mutations leading to reduced amounts of RPs (affecting both the 40S and 60S ribosomal subunits) interfere with the processing of rRNA. Until recently the different rRNA species in DBA patients could only be analyzed using Northern blot technique, with several limitations mainly

due to the poor yield of RNA in these patients.

To resolve some of these problems, in this study we designed primers specific for 32s rRNA intermediate, 18s, 28s, and 5s rRNA. We studied the relative expression of these genes in DBA compared to healthy controls.

We studied rRNA defect in a cohort of forty-eight DBA patients from resting and stimulated T cells. rRNA profile study showed a significant difference in the resting and stimulated T cells from DBA patients compared to controls.

We then applied CRISPR CAS 9 technology, electroporation, flow-cytometry, and timed cell-sorting to validate two novel heterozygous mutations discovered in our Lab, the *RPS17* c.3G>C and *RPL11* c.475_476 del AA.

We optimized a method to successfully introduce a heterozygous knock in mutation using CRISPR/Cas9 technology in K562 erythroleukemic cell line. Firstly, K562 cells were transfected with LentiCRISPRv2 plasmid that has a specific guide RNA (gRNA) ligated to it using the Amaxa nucleofection system. Seventy-two hours later, single green fluorescent protein (GFP) expressing K562 cells were sorted into 96 wells plates. After three weeks of growing these single clones, they were analyzed and the viability was shown to be 50%. Sanger sequencing was carried out to confirm the presence of heterozygous knock in of the specific mutation.

Quantitative real time PCR studies using our designed primers (18S, 28S, 32S and 5.8S rRNA) revealed that both mutations resulted in rRNA processing defect in K562 cell line compared to control.

In conclusion, our study lead to the characterization of newly established mutations and we demonstrated that such mutations were responsible and causative of a defect in the production of rRNA in an in vitro model.

# AKNOWLEDGMENT

I am greatly indebted to my funding body, The Ministry of Higher Education/Iraq without their input to my work over the last four years I would not have been able to fulfil my PhD

I would also like to thank my parents, wife and three little angels who made my PhD possible with their ultimate support

I would like to thank my colleagues at Karadimitris and Foroni's lab who had a generous input and support with protocols and techniques during my PhD:

Mai Lasa for her guidance in CRISPR technology.

Antonella Rotolo for her guide in tissue culture and flowcytometry.

Valentina Caputo for her guide in cloning and molecular techniques

Nikolaos Trasanidis for his help in designing rRNA primers in addition to molecular techniques

David Pitcher, Katerina Goudevenou and Kanagaraju Ponnusamy for their guidance in tissue culture and viral transduction

Gareth Gerrard and Mary Alikian for their supervision in NGS

Sarmad Toma and Foong Hui En for explaining PCR and sanger sequencing.

Molecular diagnostic Lab at Hammersmith Hospital for hosting me to do all the sequencing especially during my first year of PhD.

I would also like to sincerely thank my supervisors Anastasios Karadimitris, Letizia Foroni and Josu De La Fuente.

# ABBREVIATIONS

| | |
|---|---|
| ACTB | Actin beta |
| Cas9 | CRISPR-associated protein 9 |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| Ct | Cycle threshold |
| DBA | Diamond Blackfan Anaemia |
| DC | Yskeratosis Congenita |
| DSB | double-stranded break |
| DTT | dithiothreitol |
| eADA | erythrocyte adenosine deaminase |
| FastAP | Thermosensitive Alkaline Phosphatase |
| FBS | inactivated Foetal Bovine Serum |
| GFP | Green fluorescent protein |
| GvHD | Graft Versus Host Disease |
| HSCT | Haematopoietic stem cell transplant |
| EDTA | Ethylenediaminetetraacetic acid |
| gRNA | guide RNA |
| HCST | Haematopoietic stem cell transplant |
| HDM | Human double minute |
| HDR | Homology directed repair |
| IBMF | Inherited bone marrow failure |
| IL-2 | Interleukin 2 |
| IPO8 | Importin 8 |
| LB | lysogeny broth |
| MDS | Myelodysplastic syndrome |
| NGS | Next Generation Sequencing |
| NHEJ | Non homologous end joining repair |

| | |
|---|---|
| NAIC | North American Indian childhood cirrhosis |
| PAM | Protospacer adjacent motifs |
| PNK | polynucleotide kinase |
| RNAP | RNA polymerase |
| RP | Ribosomal protein |
| RPLP0 | Ribosomal protein lateral stalk subunit P0 |
| SBDS | Shwachman-Bodian-Diamond syndrome |
| SDS | Shwachman Diamond syndrome |
| SOC | Stable Outgrowth Medium |
| TALEN | transcription activator-like effector nucleases |
| TAE | Tris base, acetic acid and EDTA. |
| TBP | Tata box binding |
| TCS | Treacher Collins syndrome |
| TEC | transient erythroblastopenia |
| TFRC | Transferrin receptor |
| ZFN | Zinc-finger nucleases |

# OBJECTIVES

RP gene mutations and deletions are responsible for at least 65% of DBA cases(1–6). However, our lab has found ribosomal protein mutations in up to 88% of DBA patients (7) using next generation sequencing, though some of the mutations still require functional validation (a relevant component of this project). In 12% of patients, who on clinical and laboratory grounds have been diagnosed with DBA, mutation screening has failed to identify pathogenic mutations. It is possible that in these patients, DBA is associated with mutations in novel, non-RP genes.

As ribosomal biogenesis is a complex process consisting of the coordinated synthesis and assembly of four ribosomal RNAs (rRNA) which are 18S, 28S, 5.8S and 5S rRNAs with about 83 ribosomal proteins together with more than 150 non-ribosomal proteins and other factors (8), DBA may result from mutations in any of these non-RP genes.

Currently the RP genes found to be mutated in DBA are: *RPS19* (in 25%, the highest incidence), *RPL5, RPL11, RPL26, RPL35A, RPS7, RPS10, RPS17, RPS24* and *RPS26* (1–4,6,9–12). Other genetic abnormalities of less certain significance have been identified in isolated patients or families. These include: *RPL3, RPL7, RPL9, RPL14, RPL15, RPL19, RPL23A, RPL25, RPL35, RPL36, RPS8, RPS15* and *RPS27A*(1–3).

Work in our lab has established a next generation sequencing (NGS)-based platform entailing RP gene capture and amplicon deep sequencing. This is the most comprehensive approach for the simultaneous screening for mutations and deletions of 83 genes (7). The only RP gene not included in the DBA diagnostic panel is *RPS17*. This is because of the homology between *RPS17* and its pseudogene that made designing of primers for this gene too complex, however, Sanger sequencing was

performed on all DBA patients to screen for *RPS17* mutation. Furthermore, our diagnostic panel also includes *GATA1* gene which is reported to be involved in DBA (13,14).

Using this approach, heterozygous mutations or deletions in RP genes have been identified in 88% of patients with DBA in our cohort of referrals compared to less than 65% in the era prior to the implementation of next generation sequencing. NGS approach has generated a wealth of genetic data and improved substantially the genetic diagnosis of DBA.

However, several of the identified variants, and missense mutations, although by genetic and computational criteria are predicted to be pathogenic, warrant functional validation. In this project we are planning to validate some of these mutations.

In **Objective 1** of my project, I will aim to develop and validate an alternative methodology to northern blotting to assess rRNA processing in DBA cells.

In **Objective 2**, I will address the functional impact of these missense mutations on rRNA processing and erythropoiesis using the clustered regularly interspaced short palindromic repeats (CRISPR/CAS9) technology.

**In Objective 3**, I plan to employ exome sequencing on patients with no clear RP mutation, with the aim to identify mutations in other non-RP genes

# 1. INTRODUCTION

## 1.1. Diamond Blackfan Anaemia (DBA)

Diamond-Blackfan Anaemia (DBA) is a rare autosomal dominant disorder with an incidence of 7 in $10^6$ and is characterized by pure red cell aplasia resulting from early death of erythroid progenitors (15). DBA is inherited in 45% of cases and emerges as de novo in the rest of cases (16)

DBA is one of a rare group of genetic disorders known as 'inherited bone marrow failure syndromes' (IBMFS) with bone marrow failure, birth defects and higher propensity to cancer (17). DBA is very heterogeneous genetically and clinically. It is characterized by an intrinsic defect in erythroid lineage development (15), and specifically by a quantitative as well as a qualitative defect in erythroid progenitors(18). As a result, DBA patients usually present in the first year of life with severe macrocytic anaemia with reticulocytopenia and a normocellular bone marrow with a paucity of erythroid precursors (19). DBA patients generally exhibit increased levels of fetal hemoglobin and the activity of erythrocyte adenosine deaminase (eADA) is elevated in 80–85% of all patients (20,21). The cancer incidence in DBA is significantly elevated especially for acute myeloid leukemia, colon carcinoma, osteogenic sarcoma and female genital cancers (17).

Various physical malformations are seen in about 50% of all DBA patients usually affecting the face, hands, heart, or urogenital region. DBA patients can also present with congenital anomalies without anaemia (22). Atypical DBA patients usually manifest later in life (17).

Patients with DBA has an equal 1:1 male to female ratio and it is virtually seen in all ethnic groups (17).

Anaemia can be controlled with corticosteroids in 30-40% of patients, the rest of the patients require regular blood transfusion. Some patients who respond to steroids end up with an every-other-day dose while others (20%) end up in complete remission during the first decade of life (17). A lesser percentage (9%) of DBA patients will require haematopoietic stem cell transplant (HSCT) which cures the haematological manifestations of DBA (23).

## 1.2. Genetics of DBA

Most of the genetic variations in DBA genes are sporadic or de novo in around 55% of cases and familial in the remaining 45%. In many cases, DBA patients inherit the mutation from a parent who will not show any overt phenotype and is considered “silent carrier”. Silent carriers may also exhibit only macrocytosis without anemia and/or an elevated eADA (24).
There is a marked heterogeneity in the phenotypic expression within the various genotypes and even within the same family. This heterogeneity involves both the hematologic and the non-hematologic manifestations of DBA (17).
The first gene responsible for DBA was *RPS19* in 1999 (25). Subsequently, the identification of other mutations were revealed in *RPS24, RPS17, RPL5, RPL11, RPS10,* and *RPS26*(1,5,6,11). Many other mutations in RP genes have been identified lately and today the list includes *RPS7, RPS10, RPS15A, RPS17, RPS19, RPS24, RPS26, RPS27, RPS28, RPS29; RPL5, RPL9, RPL11, RPL15, RPL18, RPL26, RPL27, RPL31, RPL35,* and *RPL35A*(3,26–32).

Most of the mutations (>90%) include large deletions occurring in only 6 genes (*RPS19, RPL5, RPS26, RPL11, RPL35A*, and *RPS24*). Other genes (such as *RPS29, RPS17, RPS7, RPS10, RPL15*, and others) account for less than 10% of all mutated cases (24).

To date, all the RP gene mutations identified in DBA patients are heterozygous. Homozygosity is largely suspected to be lethal. This is based on the lethality of homozygous RP gene mutations in several animal models including zebrafish and mice(33,34)

A wide range of mutation types is also evident in DBA. Most of the missense mutations have been identified in the *RPS19* gene while nonsense mutations, small deletions or insertions, and splice site mutations are predominant in *RPL5* and *RPL11* (1,10). Partial and whole-gene deletions were detected in *RPS17*, *RPL35A, RPS19* (4,35,36).

Although DBA is considered to be almost entirely linked to RP gene mutations, two non-RP genes have been reported in patients including *GATA1* and *TSR2* (30,37,38). Having said that, these genetic lesions are extremely rare. *GATA1* mutation was reported in only five DBA patients and only one with a *TSR2* mutation. The *TSR2* gene is involved in pre-rRNA processing and binds to *RPS26* protein and is related to ribosome biogenesis. The *GATA1* is not reported to be involved in ribosome biogenesis but it encodes for the major erythroid transcription factor GATA1. In approximately 30% of DBA patients, the underlying genetic defects remain to be elucidated.

## Table 1.1: Genes reported to be pathogenic in DBA. Adapted from Gazda et al(39).

| Gene | % of affected individuals | Type of mutation | References |
|---|---|---|---|
| *RPS19* | ~25 | Single-nucleotide variant (SNV) (missense, nonsense, splice site mutations), microinsertions, microdeletions, single copy deletions | Draptchinskaia et al., 1999; Farrar et al., 2011; Gazda et al., 2004 |
| *RPL5* | ~7 | SNV (missense, nonsense, splice site mutations), microinsertions, microdeletions | Cmejla et al., 2009; Gazda et al., 2008; Quarello et al., 2012 |
| *RPS26* | ~6.5 | SNV (missense and splice site mutations), microinsertions, single copy deletions | Doherty et al., 2010; Farrar et al., 2011 |
| *RPL11* | ~5 | SNV (missense, nonsense, splice site mutations), microinsertions, microdeletions | Cmejla et al., 2009; Gazda et al., 2008; Quarello et al., 2012 |
| *RPL35A* | ~3 | SNV (missense, nonsense, splice site mutations), microdeletion, single copy deletions | Farrar et al., 2008; Farrar et al., 2011 |
| *RPS10* | ~2.5 | SNV (missense and nonsense), microinsertion | Doherty et al., 2010 |
| *RPS24* | ~2 | SNV (nonsense), microinsertion/ microdeletion, large deletion | Gazda et al., 2006b; Landowski et al., 2013 |
| *RPS17* | ~1 | SNV (missense), microinsertion, single copy deletions | Cmejla et al., 2007; Farrar et al., 2011; Gazda et al., 2008 |
| *RPS7* | <1 | SNV (splice site mutation) | Gazda et al., 2008 |
| *RPL26* | <1 | Microdeletion | Gazda et al., 2012 |
| *GATA1* | <1 | SNV (missense, splice site mutations) | Sankaran et al., 2012 |
| *RPL15* | <1 | Large deletion | Landowski et al., 2013 |
| *RPS29* | <1 | SNV (missense mutations) | Mirabello et al., 2014 |
| *RPS28* | <1 | SNV (missense mutations) | Gripp et al., 2014 |
| *TSR2* | <1 | SNV (missense mutation) | Gripp et al., 2014 |
| *RPL31* | <1 | Single copy deletion | Farrar et al., 2014 |
| *RPS27* | <1 | Microdeletion | Wang et al., 2015 |
| *RPL27* | <1 | SNV (splice site mutation) | Wang et al., 2015 |

Genes ordered according to the percentage of DBA-affected individuals who carry a certain mutation. RPL, RP from large ribosomal subunit; RPS, RP from small ribosomal subunit.

## 1.3. DBA Diagnosis

DBA patients usually present early in life (median age at diagnosis 2-month-old). Typical presentation includes pallor, failure to thrive, and feeding difficulties. After collecting the family history of the patient, the first test is typically a blood smear and a complete blood count together with a reticulocyte count. DBA may be suspected if haemoglobin (Hb) concentration is low for the age of the patient (as low as 2-3g/dl).

Blood smear shows macrocytosis with absent or low reticulocyte count (<20 × $10^9$/L), normal granulocyte counts with occasional mild-to-moderate neutropenia. Foetal Hb might also be increased. An elevated eADA level prior to transfusion is seen in 80-90% of DBA patients.

A high erythropoietin level may help with the diagnosis indicating an intrinsic defect of bone marrow. A positive family history for anaemia and/or syndromic features is present in at least 50% of DBA patients.

Bone marrow aspiration shows normocellular marrow with low erythroid precursors (below 5%) and normal appearing myeloid and megakaryocyte lineages.

Although DBA is characterised by genetic defect in ribosomal protein genes, to date hydrops fetalis is a very rare event in DBA and requires more work to understand the currently unknown mechanism by which embryonic and early foetal erythropoiesis is most often spared in DBA. (17,24)

In cases of late-onset DBA and non-classical DBA (e.g. in adolescents or adults), hypocellularity of the bone marrow with dysplasia and megaloblastic changes might be displayed which should be differentiated from low grade MDS or 5q- syndrome. Other presentation for non-classical DBA might include congenital anomalies without anaemia or mild haematological abnormalities without congenital malformations (19,22).

Differential diagnosis of DBA includes **parvovirus B19**-associated pure red cell aplasia which can be identified by cytology, IgM/IgG serology and is usually seen in immune deficient patients, by a PCR analysis of blood and/or bone marrow samples. Additionally, Parvovirus B19 is neither associated with positive family history nor with congenital anomalies.

Also, DBA needs to be differentiated from **transient erythroblastopenia** (TEC) which usually manifests beyond the first year of age and is associated with normal MCV, eADA and HbF values. Additionally, it can run in families and is not associated with congenital anomalies. Other **inherited bone marrow failure syndromes (IBMFS)** needs to be differentiated from DBA, such as **Fanconi anaemia** or **other ribosomopathies** (24)**,** these include **Shwachman Diamond syndrome** (SDS) which is caused by mutations in the Shwachman-Bodian-Diamond syndrome (*SBDS*) gene, which plays a role in ribosomal subunit joining. This disease is manifested by decreased production of white blood cells, exocrine pancreatic insufficiency, and skeletal problems.(41)

Another differential diagnosis includes **Treacher Collins syndrome** (TCS) which is caused by mutations in genes responsible for rRNA transcription such as *POLR1D*, *POLR1C* and *TCOF1* which encode subunits of RNA polymerase I and III. TCS is characterized by craniofacial deformities such as absent cheek bones(42).

**Dyskeratosis congenita** (DC) which is charachterised by mucocutaneous features: nail dystrophy, oral leukoplakia, and abnormal reticulate skin pigmentation(43).

**North American Indian childhood cirrhosis** (NAIC) is another form of ribosomopathy that is caused by mutation in *CIRH1A*, which encodes a protein that functions in rRNA processing. The phenotype is restricted to neonatal jaundice that progresses to biliary cirrhosis (44)

## 1.4. Molecular Diagnosis of DBA

There are many molecular tests used to confirm the diagnosis of DBA. Most screening analyses begin with Sanger sequencing of *RPS19*. This involves PCR amplification and sequencing of each exon and promoter region by specific primers; however, this approach can only detect 25% of the cases. Some labs investigate for the most common genetic defects by sequencing the most frequently reported genes responsible for DBA using Sanger sequencing or comparative genomic hybridization (CGH) array. In other labs, however, Next Generation Sequencing (NGS) has replaced Sanger sequencing. This includes high throughput sequencing technology (involving a bespoke target enrichment platform) to screen all 80 known RP genes. This allows rapid and cost-effective identification of DBA-associated mutations. Furthermore, other Labs use NGS for whole exome sequencing if no mutation was detected using the NGS ribosomal protein gene targeted approach (24).

## 1.5. Molecular Genetics in DBA

To date, nearly all the genetic lesions driving DBA have been found to be mainly located in ribosomal protein (RP) genes and result in a pre-ribosomal RNA (rRNA) maturation defect. DBA is considered a “ribosomopathy”. Ribosomes, the cellular machinery responsible for protein translation, are made of ribosomal proteins (RP) and rRNA. RP are not only required as structural components of ribosomes but also for the efficient processing and cleavage of rRNA in the nucleolus from a single RNA transcript to several smaller fragments. These fragments are subsequently incorporated into the small or large ribosomal subunits (8,28,45). Figure (1.1).

The single RNA transcript 45S rRNA will then undergo endonucleolytic and exonucleolytic processing steps that will result in the end products of maturation of the pre-rRNA which are the 18S, 5.8S and 28S rRNA. All of the RPs involved in DBA so far have proved essential in ribosome production, and pre-rRNA processing defects can be detected in patient cells (3) as shown in figure (1.1), (1.2) and (1.3).

Choesmel et al in 2008 studied the effect of *RPS24* mutation compared to *RPS19* mutation on rRNA processing in DBA and found that *RPS24* affects maturation of the 18S rRNA, but not at the same stage as mutations in *RPS19*. Northern blot analysis of RNA extracted from lymphoblastoid cell lines derived from DBA patients with deleterious mutations in *RPS24* and *RPS19* revealed that cells with mutations in *RPS19* showed increased levels of 21S and 41S pre-rRNAs when compared with control cells, in addition to high 21S/18S-E and 41S/30S ratios. On the other hand, cells with mutation in *RPS24* showed that the level of 41S pre-rRNA was lower than in control cells, whereas the 30S species accumulated and consequently a lower 41S/30S ratio. In parallel, there was less 21S and 18S-E pre-rRNAs compared to control cells(46). Furthermore, Choesmal et al found that *RPS24* is essential for the cleavage of 5'-ETS which is crucial for the production of 18S rRNA resulting in lower 18S rRNA and 18S/28S rRNA ratios compared to controls, but has no effect on the level of 28S, 5.8S rRNA (46). Mutations of *RPS24* in human cells alter maturation of the 18S rRNA at an earlier stage than do mutations in *RPS19* (46).

Wang et al in 2014 (29), discovered novel mutations in *RPL27* and *RPS27* in sporadic patients with DBA. In vitro knockdown of these genes disturbed pre-ribosomal RNA processing. Zebrafish models of *RPL27* and *RPS27* mutations showed impairments of erythrocyte production and

abnormal brain and/or tail development. This group studied the effect of knocking down some genes in human erythroid cell line (K562) by introducing siRNA then Northern blot analysis was applied as a read out of the consequences. It was found that a knockdown of *RPS19* was associated with a decrease in *18S-E rRNA* with accumulation of a 21S rRNA precursor. Furthermore, knock down of *RPS27* led to the accumulation of 30S rRNA and a decrease in the 21S rRNA and 18S-E rRNA. Their findings suggested that *RPL27* and *RPL5* are essential for 28S, 5.8S rRNA processing while *RPS19* and *RPS27* both play an important role in 18S rRNA processing (29).

By knocking down gene expression of *RPS10* and *RPS26* in Hela cells with siRNA, Doherty et al (9) found that there was an accumulation of 18S-E rRNA,26S, and 43S rRNA with a reduced expression of each of these genes resulting in a decrease in the mature 18S rRNA and an increase in the mature 28S rRNA (9). This phenotype suggested that *RPS10* and *RPS26* play a key role in rRNA processing at a downstream step compared to the effect of *RPS19*, *RPS24* and *RPS7* in which 18S-E rRNA was reduced (9).

To determine whether DBA mutations in *RPL5* and *RPL11* affects pre-rRNA maturation, Hanna et al (1) analyzed pre-rRNAs from lymphoblastoid cells generated from DBA patients by Northern blotting. Accumulation of 32S pre-rRNA was observed in cells mutated in either of these genes especially *RPL5*. In parallel, a higher amount of 12S pre-rRNA, as well as accumulation of smaller precursors of 5.8S rRNA was observed. These results suggest defective splicing of the internal transcribed spacer 2 (ITS2), both at the initial endonucleolytic cleavage in the 32S pre-rRNA and in the subsequent processing steps (1). Additionally, this study showed that these genes are also responsible for

the cleavage of the ITS1 (internal transcriber spacer1) as knock out of *RPL5* resulted in the accumulation of 30S rRNA and 18S-E rRNA (1). Moreover, Hanna et al demonstrated reduced levels of 28S and 5.8S mature rRNA (1). Furthermore, the consequences of *RPS7* on rRNA processing was studied in the same way and it appeared that *RPS7* is crucial in the 5' ETS splicing. Therefore, knockdown of *RPS7* resulted in accumulation of 45S and 30S rRNA together with a reduction in 41S, 21S, and 18S-E rRNA (1). Their findings were consistent upon transfecting Hela cells with siRNA and using Northern blot analysis as a read out(1). *RPL5* and *RPL11* knock down induces accumulation of precursors to 5.8S rRNA, whereas siRNA-mediated knockdown of *RPL35A* transcripts results in a decrease of these rRNA species, indicating that RPs play different roles in pre-rRNA processing (1).

Gazda et al in 2012 (3) determined the role of *RPL3, RPL19*, and *RPL26* in pre-rRNA maturation using siRNA knockdown in Hela cells. Upon *RPL3* knockdown, a general decrease of the amount of 28S, 5.8S rRNA and the presence of an unusual pre-rRNA migrating between the 32S and the 12S pre- rRNAs were noticed. *RPL19* depleted HeLa cells accumulated 32S rRNA which is a precursor to the 28S and 5.8S rRNAs. Depletion of *RPL26* with siRNAs in HeLa cells resulted in decreased amount of 30S, 21S, 12S rRNA and a higher 41S rRNA suggesting a defect in the cleavage at the internal transcriber spacers. Additionally, the effect of mutations in these genes was further studied by Northern blot analysis of RNA extracted from lymphoblastoid cell lines derived from DBA patients with *RPL3, RPL19* and *RPL26* mutations. rRNA processing defect was consistent except for *RPL3* where the patients' cells did not show a similar phenotype as the siRNA transfected HeLa cells(3).

Landowski et al reported a novel *RPL15* mutation in 2013 (27). They studied the effect of transfecting HeLa cells with siRNA to knock down *RPL15*. This mutation was associated with decreased levels of 32S and 12S pre-rRNA, the precursors to the 5.8S and 28S rRNAs. In addition, accumulation of the 41S, 30S and 18S-E pre-rRNAs relative to the 21S pre-rRNA was observed, suggesting a defect in ITS1 cleavage. Other mutations were also tested in patients' lymphoblastoid cells. RNA from the patients with *RPS17* and *RPS26* mutations showed higher levels of 21S and 18S-E pre-rRNA, which matches the pattern observed upon knockdown of these proteins. On the other hand, the deletion in *RPS24* resulted in higher levels of 30S pre-rRNA (27).

Farrar et al, investigated pre-rRNA processing defect in Con A stimulated peripheral blood mononuclear cells (PBMNC) by northern blot in DBA patients with *RPL5, RPL11* and *RPL31* mutations (28). It was found that these mutations are associated with an accumulation of 32S pre-rRNA. Then, in the same study this finding was confirmed on K562 erythroleukemia cell after transduction with lentivirus expressing a small hairpin RNA (shRNA) to *RPL13. .RPS26* mutation, in contrast, showed no evidence of an increased 32S pre-rRNA precursor but rather shows an accumulation of 18SE-pre-rRNA (28).

In the same study (28), the Bioanalyzer 2100 (RNA capillary electrophoresis system) was used to analyze RNA from Con A-stimulated PBMC of DBA patients and compare it to normal control. Remarkably, a minor peak corresponding to 32S rRNA was evident in Bioanalyzer trace data from all DBA specimens with a 60S subunit gene alteration. The ratio of 32/28S rRNA was significantly higher in the 60S gene DBA group than the 40S gene group which was slightly higher than the control group (28).

Using the same RNA capillary electrophoresis technology in 2016 Quarello et al revealed that 18S rRNA is reduced in patients with RPS mutations and the ratio of 28S/ 18S rRNA was significantly higher in patients with mutations in RPS genes than normal control (48). Conversely, the ratio of 28S/18S rRNA was significantly lower in patients with RPL mutations as compared to the control group. This is most likely explained by the lower levels of 28S rRNA in these patients. Additionally, the 32S/28S rRNA ratio was also assessed. An increase in 32S/28S rRNA ratio was evident in most patients with mutations in RPL genes relative to the control group. However, there were outliers in this analysis including mutations in *RPL35A* that did not show an increase in 32S/28S ratio (48). In summary, nearly all the genetic lesions responsible for DBA have been found to be mainly located in ribosomal protein (RP) genes and result in a pre-ribosomal RNA (rRNA) maturation defect. Ribosomal proteins are key elements in rRNA processing, hence, mutations in RP genes can result in a decreased amount of mature rRNAs and an increased rRNA intermediates. This is the conclusion of many studies; some of which performed Northern blot analysis on cell lines with specific RP genes being knocked down, others performed Northern blot analysis on RNA extracted from lymphoblastoid cell lines made from DBA patients, whereas a third group of studies performed RNA capillary electrophoresis on DBA samples.

rRNA processing defect can be utilized to confirm the pathogenicity of RP gene mutations in mutational validation studies and has the potential to be used as a diagnostic tool to aid in DBA diagnosis.

Our aim will be to find a more robust and applicable diagnostic tool to assist DBA diagnosis and confirm the pathogenicity of some novel RP gene mutations. In this project we are going to study rRNA processing

defect in DBA cells using capillary electrophoresis as well as q-RT-PCR. This will be performed on resting and stimulated T cells as we hypothesized that proliferative stress will make the aberrant rRNA processing more prominent.

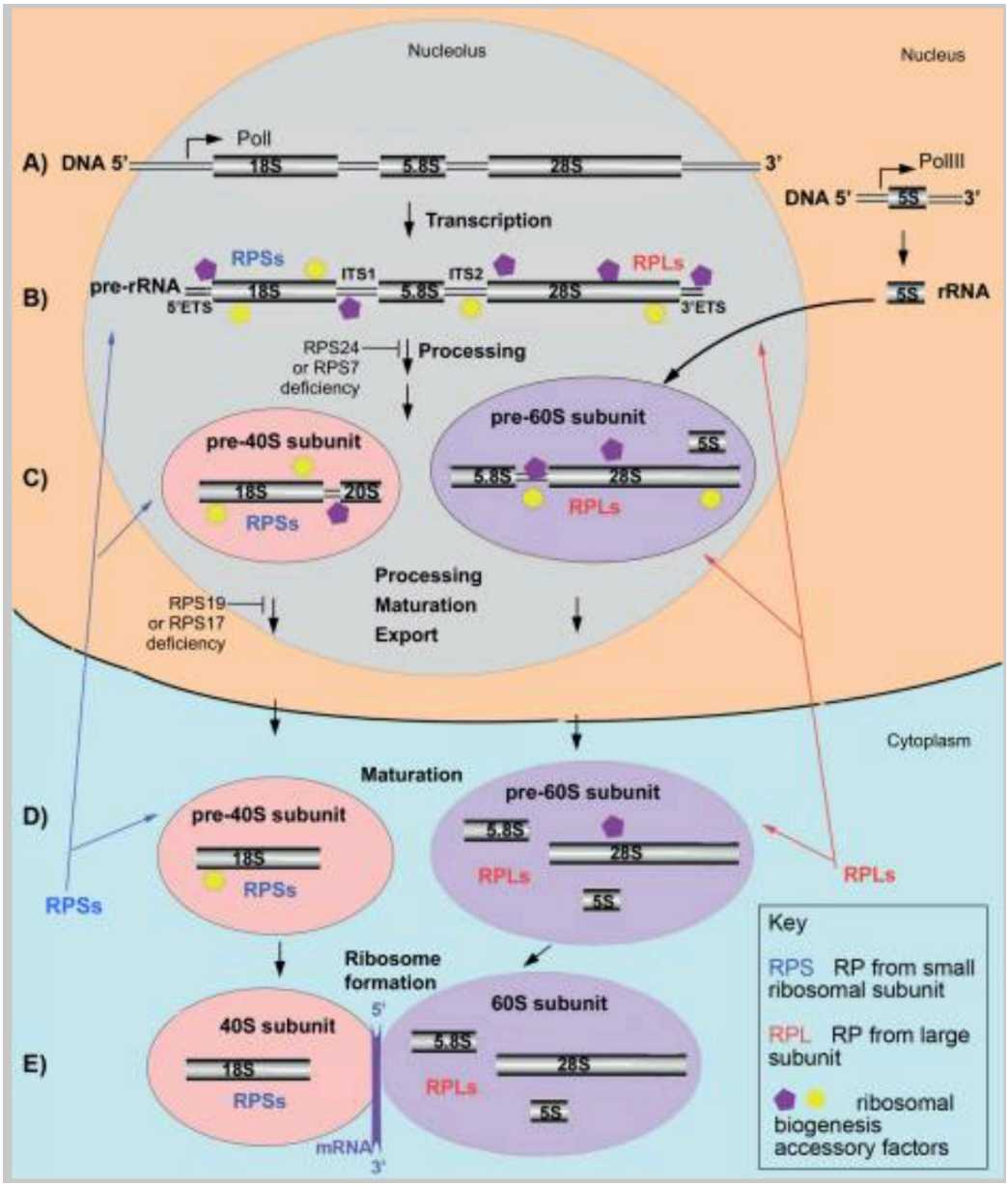


**Figure (1.1): The composition of ribosomes.** (A) 18S, 5.8S and 28S rRNAs are transcribed by Pol1 in the nucleolus as segments of a long precursor pre-rRNA, which also includes two externally transcribed spacers 5′ETS and 3′ETS and two internally transcribed spacers, ITS1 and ITS2. 5S rRNA is transcribed independently by Pol III in the nucleus. (B) The pre-rRNA assembles with accessory factors and a subset of ribosomal proteins (RPSs and RPLs). This facilitates the formation of a secondary structure necessary for the correct folding, modification and

cleavage of pre-rRNA. (C) After removal of the 5′ETS and cleavage in the ITS1 site, pre-40S and pre-60S subunits are formed and continue to mature. 5S rRNA incorporates into pre-60S subunit. Subunits are then exported to the cytoplasm.
(D) Once in the cytoplasm, small and large subunits undergo final maturation, which involves the removal of remaining accessory factors and incorporation of missing RPs.
(E) A functional ribosome forms after transcribed mRNA binds to the 40S subunit, which triggers association of the 60S subunit with this complex. More than 200 accessory factors, which include helicases, nucleases, small nucleolar RNAs, chaperones and transporters, temporally associate with the maturing ribosomal subunits at various steps. In human cells, pre-rRNA processing is differentially affected by deficiency of various RPs. This figure is adapted from Gazda et al(39).

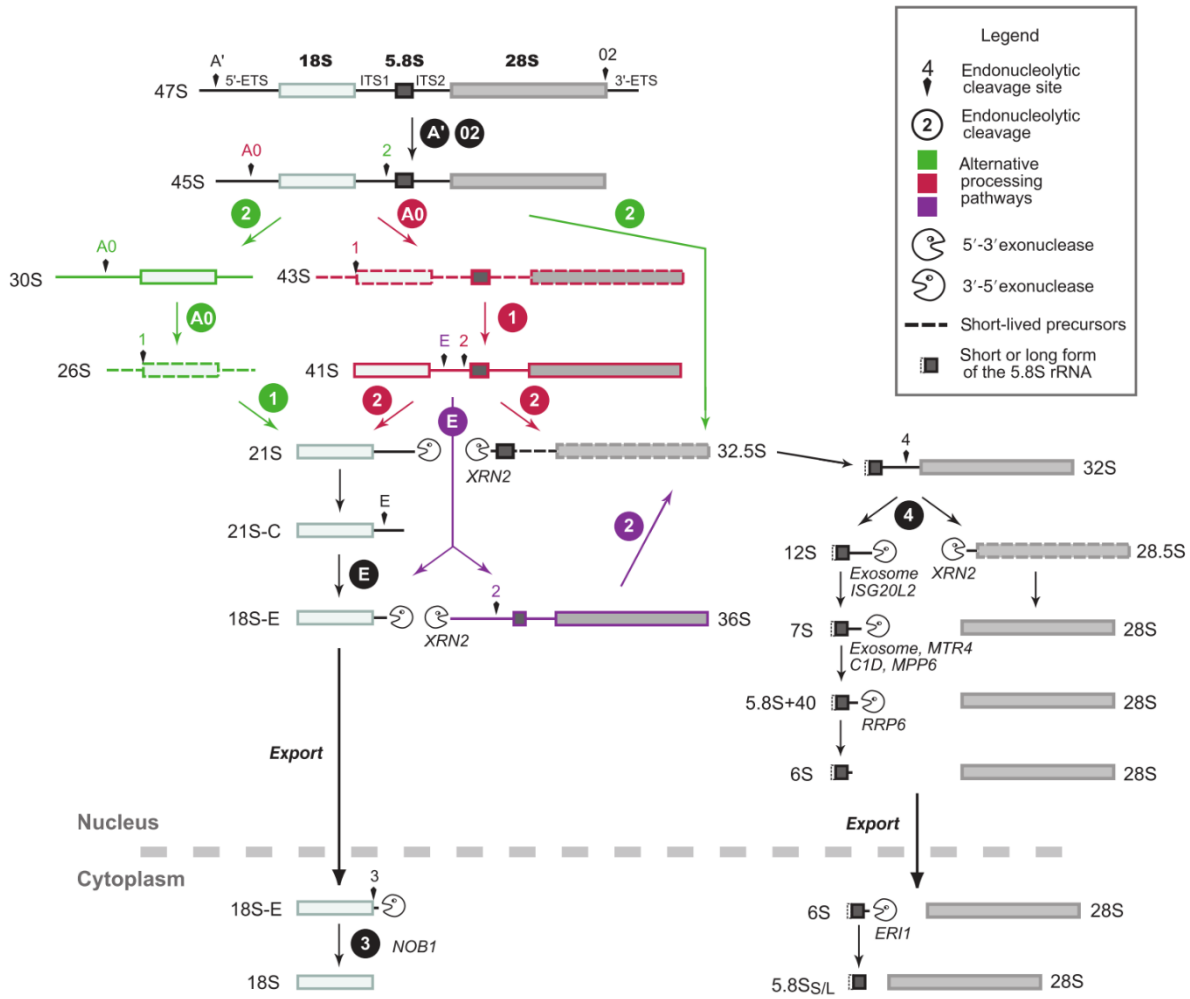


**Figure (1.2): Pre-ribosomal rRNA processing in mammalian cells from a single transcript.** Alternative cleavage sequences are shown in different colors. Short lived intermediates are represented with dotted lines. This figure is adapted from Gleizes et al(47).

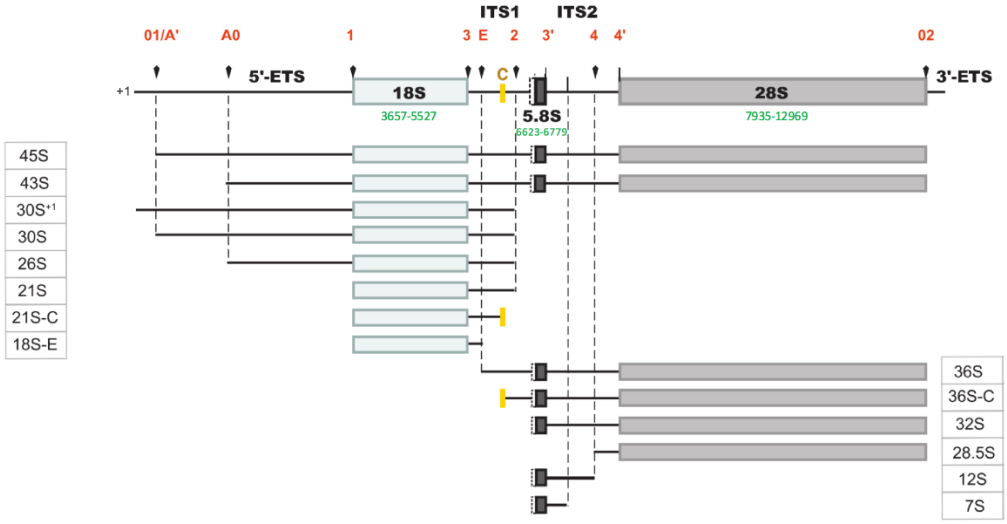


**Figure (1.3): The different pre-rRNAs detected in human and their nomenclature.** Arrowheads indicate the endonucleolytic cleavage sites. The yellow box corresponds to a highly conserved domain in ITS1 among mammals. The 36S-C and the 30S+1 species are examples of precursors that are only detected upon perturbation of ribosome biogenesis. The nucleotides correspond to the residue located 5′ to the cleavage site. Human site E was determined by primer extension and 3′-RACE.The numbering of the nucleotides (in green) refers to GenBank sequences U13369.1 (human rDNA). This figure is adapted from Gleizes et al (47).

## 1.6. Pathogenesis

Ribosomopathies are a heterogeneous group of human disorders characterized by ribosomal dysfunction. This group broadly comprises two groups: the first one includes disorders caused by ribosomal proteins gene mutations, an example of which is DBA; the second group encompasses disorders associated with defects in the factors that are responsible for ribosome biogenesis.

Despite the ubiquitous presence of ribosomes in all cells these disorders affect hematopoietic and skeletal tissues sparing other tissues. Several studies were trying to explain the variable expression of the protein dysfunctional machinery at the phenotypic level (49).

DBA is associated with ribosomal protein gene mutations in 55% of cases resulting in haploinsufficiency of ribosomal proteins and impairment of ribosomal function. This reduces general protein translation in addition to abnormalities of erythroid transcription factors functions, more predominantly (50,51).

RP haploinsufficiency activates P53 and thereby inhibits cell proliferation (52) This was also exemplified by the finding of Dutt et al (2011), who revealed that RPL5 and RPL11 are essential in P53 stabilization through binding HDM2 and inhibiting it from suppressing P53.Hence, RPL5 and RPL11 are essential in P53 stabilization(53)(54)(55).

The wild type p53 protein is involved in cell homeostasis, including the cell cycle, DNA maintenance, and apoptosis. On the other hand, mutant p53 protein is seen in most cancers (56). In the RPS e19 deficient mouse model mutations in MDM2 that blocks its binding to the ribosomal protein

uL18 and uL5 resulted in stabilization of P53 and prevented cell death. This finding supports the P53 dependent cell cycle arrest hypothesis (57) This hypothesis is further supported by many studies that revealed deletion of P53 rescues many of the phenotypes caused not only by RP mutations but also mutations in other factors involved in ribosomal biogenesis.(58)(59)(57) However, the mechanism behind the tissue specificity that is seen in DBA is still not understood. One possibility is that different cell types have different thresholds for P53 activation in response to RP gene mutation. In other words, the tissues that can compensate for the RP gene defect will not show the phenotype of DBA and in certain tissues failure of this compensatory mechanism triggers p53 activation and cell death. Another reasonable explanation is that cells with high protein synthesis (and therefore high ribosomal activity) are more sensitive to mild changes in RP levels(53) as illustrated in (Figure 1.4)

An argument against the above is that many tissues like skin, liver, gastrointestinal tissues and muscles have high protein synthesis; yet they are not affected in DBA(60). This can point to the presence of other features of translational dysfunction must be involved in determining the tissue specificity in DBA phenotype(49). This is supported by the zebrafish study that showed P53 deletion in zebrafish DBA models does not rescue the erythropoietic defect(61).

## Translational dysfunction

### The specialized ribosomal hypothesis

Ribosome heterogeneity refers to the idea that within individual cells or across tissues, ribosomal composition varies over time or under different conditions. The translation of certain mRNA requires certain structural variation of the ribosome. The loss of such specialized ribosomes has been proposed to mediate the tissue specificity that results from deficiencies in some RPs. (62)(63)

### The Concentration of Ribosomes

This hypothesis implies that defective RPs can lead to reduced ribosomal concentration and thus reduced translation of specific complex mRNAs like GATA 1 that requires high ribosomal concentration. This is supported by the fact that engineered increase in GATA 1 protein level can rescue the DBA phenotype and overcome the effect of RP defect. (50)In addition it is found that translation of mRNA with high initiation rates ex mRNA encoding RPs or haemoglobin can be efficient even at lower ribosomal concentration. Conversely, this same analysis revealed that mRNAs with low initiation rates (hormones, transcription factors and GATA1 will be poorly translated when ribosomal availability is limited.(64) Several analyses of RP expression levels from global mass spectrometry data have suggested that total tissue ribosome abundance may vary by a factor of 3 to 10 among different tissues. (65)(66)

Most RPs are strictly required for the pre-rRNA processing. It is thought that their role is to assist the proper folding of the pre-rRNA(47). RPs act in a hierarchical order. Approximately half of RPs from the 40S ribosomal subunit, including RPS24 and RPS7, associate early with the 5′ end of the

nascent pre-rRNA and are required for the initiation of cleavages at the 5′ external transcribed spacer 5′ETS; and ITS. The second set of RPs, which includes RPS19 and RPS17 along with other RPs, is not required for 5′ETS removal but is necessary for ITS1 cleavage. Similarly, depletion of RPS24 or RPS7 in yeast and human cells results in the failure of maturation of the 5′ end of 18S rRNA, whereas, in the absence of RPS19 or RPS17, the 3′ end of the 18S rRNA cannot mature (67)(68)(45)(69)(8). In all cases, pre-40S particles that contain non-cleaved pre-rRNA cannot be exported and accumulate in the nucleus.

Similarly, deficiency of most other RPs that are mutated in DBA, both from the small and the large ribosomal subunits, affects pre-rRNA processing in a unique way, leading to the accumulation of different rRNA precursors and disruption of ribosome biogenesis at different steps(8).

It is found that RPs play an important role in selectively enhancing the translation of specific mRNAs(49). Ribosomal protein gene mutation in DBA resulted in an abnormal translation of specific transcripts that are essential in erythroid differentiation (51). Although DBA is characterized by red cell aplasia, recent studies have shown that haploinsufficiency of ribosomal proteins can affect other cell types in humans leading to different diseases including congenital asplenia and T cell lymphocytic leukemia (70,71).

A study done by Horos et al (2018) revealed that reduced expression of *RPS19* and RPL11 in mouse erythroblasts lead to deregulated translation of Bag1 and Csde1 transcripts resulting in impaired proliferation and differentiation of erythroid blasts (51). However, the mechanism underlying the selective mRNA translation is an area of active investigation.

Ludwig et al 2014 found that ribosomal protein deficiency can result in decreased *GATA1* translation despite normal *GATA1* mRNA level (50). Additionally, the impaired haematopoiesis in DBA patients could be partially reversed by increasing GATA1 protein levels (50).

Allam et al 2018, found that ribonuclease inhibitor (RNH1) binds to ribosomal 40S subunit and plays an important role in GATA1 translation (14). Ribonuclease inhibitor 1 (RNH1, also known as RI) is a protein localized in the cytosol mainly in addition to the nucleus and mitochondria. It is 50 kDa leucine-rich repeat (LRR) protein. In the same study, this group has revealed that RNH1 deficient mice show normal GATA1 mRNA but decreased GATA1 proteins resulting in an impaired production of mature erythroid cells from progenitors (72).

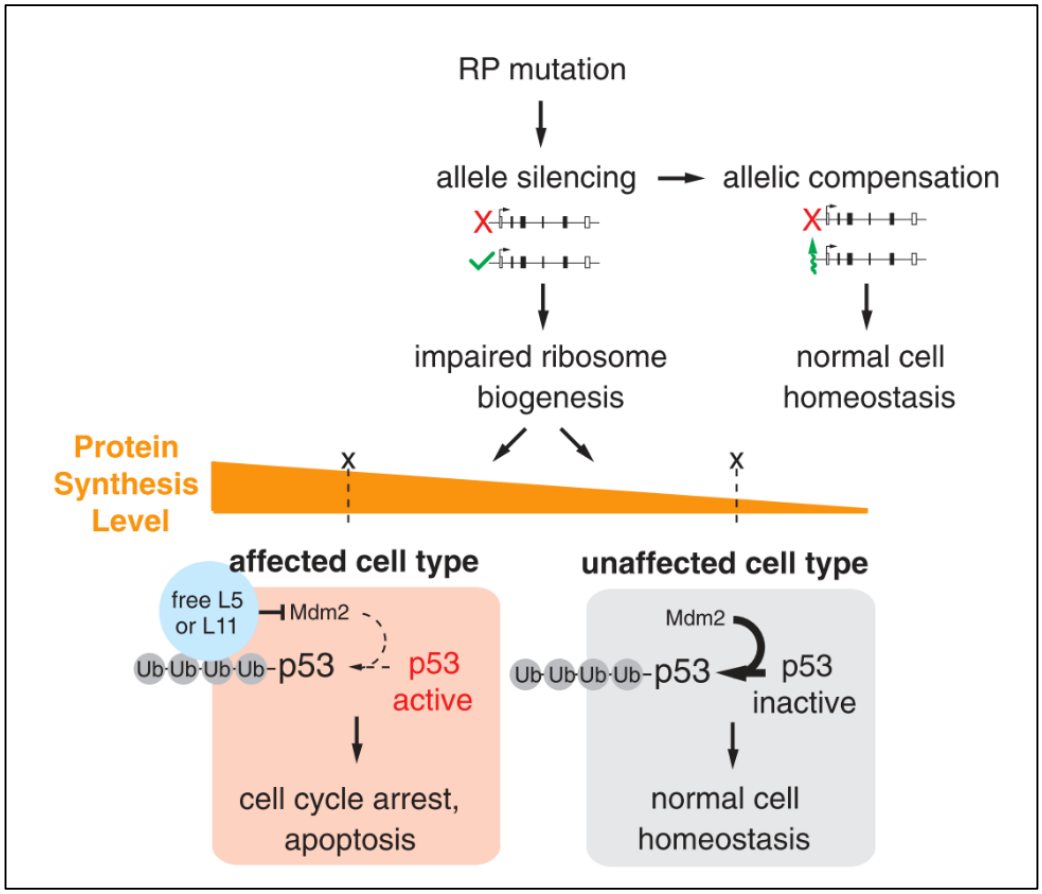


**Figure (1.4) RP mutation is associated with tissue specific phenotype.** Adapted from Green et al (49).

## 1.7.Treatment of DBA

The current mainstays of treatment of DBA are: red cell transfusion (10-15ml/kg of leukocyte depleted), corticosteroid therapy, and haematopoietic stem cell transplantation (HSCT). Approximately 80% of patients respond to initial corticosteroid therapy, whereas 20% do not respond and will require chronic red cell transfusion. However, at a later stage only 37% of patients will remain on steroid therapy, 31% on transfusion, about 13% will end up in remission and 9% will require HSCT(23).

Steroids are usually not started only until one year of age as it has a deleterious effect on the physical and neurocognitive development. The target of steroid and transfusion therapy is to maintain a stable Hb of 8-11 g/dl which is adequate to ensure growth and cognitive development. Family members are not encouraged to be blood donors in order not to jeopardize using them as donors for future HSCT(23).

### Steroid therapy

A starting dose of prednisolone 2mg/kg per day and to keep monitoring the Hb and reticulocyte count on weekly bases for 4 weeks. The target of Hb 9g/dl without the need for transfusion is regarded as a response to steroids. Response to steroids is usually within the first few weeks from initiation of therapy. Otherwise, if the Hb continues to drop despite using steroids then the patient is considered as a non-responder. Another scenario to consider steroid therapy as a failure is if the patient shows a partial response then symptoms relapse upon tapering the dose. Our aim is to keep the patient on a maximum of 1mg/kg every other day or 0.5mg/kg daily. As DBA patients will be on lifelong steroids starting from infancy, side effects need to be monitored. Periodic bone densitometry,

ophthalmologic examination to diagnose and treat cataracts and glaucoma before the onset of significant visual damage. Hypertension and diabetes mellitus need to be monitored together with a regular growth follow up on a growth chart.

Some children can subsequently respond to steroids upon recommencing steroid therapy in 12-18 months. Therefore, patients who failed the steroid trail will need to be re started on steroids after 12-18 months(23).

**Red cell transfusion and iron chelation**

In general DBA patients are usually transfused with 10-15ml/kg every 3 to 5 weeks to maintain a Hb level of more than 8g/dl. Some growing children require a target of 9g/dl. Iron overload is an important threatening complication in patients on regular transfusions. Cardiac hemosiderosis, insulin dependent diabetes, hepatic cirrhosis, hypothyroidism, hypoparathyroidism and delayed puberty needs to be monitored. Therefore, iron chelation therapy is started after transfusing the patient with 10 to 20 units or after age 2. Monitoring with serum ferritin, liver iron concentration, hepatic, pancreatic and cardiac MRI is essential(23).

**Remission in DBA**

Patients are considered in remission if they achieve an adequate Hb level lasting 6 months without any treatment. There is a 72% chance of remission within the first decade of life and 20% chance of remission by the age of 25 years. Some patients experience more than one remission during their life. It is therefore advised to encourage patients who are on twice weekly steroids to try discontinuation(23).

**Haematopoietic stem cell transplantation (HSCT):**

This is the only definitive treatment for the haematological complications in DBA. The best time to perform HSCT is below 9 years old and using HLA matched sibling(17).

**Emerging Therapeutic Approaches for DBA**

To date, the only definitive cure for the hematological manifestations of DBA is HSCT, however, this is associated with life-threatening side effects like infections and graft versus host disease GvHD. Clinical trials employing several DNA and RNA-based procedures and new pharmacological options are now in progress(73).

DNA based strategies:

This involves the introduction of the wild type version of the gene of interest into the target cells (like HCTs) to overcome the intrinsic genetic defect. This is usually done by the use of viral vectors. Clinical trials in Wiskott-aldrich syndrome and metachromatic leukodystrophy showed remarkable success after ex vivo gene therapy of HSCs(74,75). Additionally, gene therapy is progressing in the treatment of B thalassemia and sickle cell anaemia(76,77).

The use of viral vectors ensures high efficiency of gene delivery but it is associated with various drawbacks depending on the type of viral vector. Adenoviral Vectors (ADVs) are highly immunogenic and the transgene will not be integrated into the host genome. Hence its expression will be diluted over time upon cell proliferation. Adeno-Associated Viral Vectors (AAVs) are a safer alternative as they show low immunogenicity and are able to be integrated into the genome with a limited transgene capacity of about 4.5kb(73). Retroviral and Lentiviral Vectors(RVs,LV,) lead to the integration of the viral genome into the host genome which allows stable transgene expression but is associated with a risk of insertional

mutagenesis like the activation of proto-oncogenes, hence, inducing carcinogenesis(73). LVs are considered safer than RVs as they are not reported to produce clonal expansion or leukemic transformation(74)(78).

Gene therapy in DBA cells and animal models:

Gene therapy in DBA is promising as it would be able to cure DBA without the need of an HLA-matched donor and prolonged immunosuppressive therapy. Furthermore, the preconditioning regimen could be reduced or even absent because of the proliferative advantage of the gene-corrected HSCs.

In the past, several studies tried to assess the feasibility of gene therapy in DBA. The enforced expression of RPS19 in RPS19-deficient cells rescued the pathological phenotype of RPS19-mutated lymphoblastoid cell lines derived from DBA patients(79). However, when the same strategy was used for RPL5-haploinsufficient cells a partial rescue was achieved, suggesting that specific investigation will be needed for each DBA gene(79). Another group reported that transfer of RPS19 cDNA using oncoretroviral or LVs into RPS19-mutated CD34 cells isolated from patients with DBA promoted the formation of erythroid colonies both in solid and liquid cultures (80,81). Moreover, Flygare et al. used such corrected CD34 cells to transplant irradiated mice and demonstrated that a high level of RPS19 expression is essential for cell survival and favored engraftment(82).

The use of animal models for DBA is vital to investigate the therapeutic efficacy and the safety of gene therapy. The hematological phenotype of DBA was recapitulated in mice using transgenic RNA interference that allowed a doxycycline-inducible downregulation of RPS19 (83). This

model developed macrocytic anemia and BM failure that were recovered in vitro and in vivo by RPS19 gene transfer using LVs(83,84).

Another revolution towards gene therapy in DBA is reprogramming mature cells to pluripotency this is not associated with the risk of immune rejection and will not require the ethical concerns of using embryonic cells. Induced Pluripotent Stem Cells (iPSCs), are an unlimited source of autologous cells that can be genetically manipulated, differentiated into specialized cells and entirely characterized before transplant. IPSCs from skin fibroblasts were obtained from two patients with DBA who carried mutations in RPL5 or RPS19, thus providing for the first time a renewable reservoir of cells that display ribosomal and hematopoietic defects(85). DBA fibroblasts generated iPSC colonies at a frequency of 0.0045%, whereas the efficiency for control fibroblasts was 0.03%. The ribosomal and hematopoietic abnormalities were recovered via DNA transfer of a wild-type copy of the haploinsufficient gene(85).

### Ex Vivo versus in Vivo Gene Therapy

Gene transfer can be either ex vivo or in vivo. Ex vivo is directed to the cells of interest (e.g. HSC) before their reinjection into the patient and therefore acts selectively on target cells preventing the transduction of cells that would not benefit from the genetic modification. In vivo gene transfer is an alternative method for gene delivery that avoids some of the drawbacks of ex vivo transfer, in particular, the need to collect a sufficient number of HSCs from BM or peripheral blood, and to manipulate them in ex vivo cultures. With the in vivo technique the viral vectors carrying the therapeutic gene can target HSCs directly in their environment, thus ensuring the maintenance of physiological conditions. Specific promoters or microRNA (miRNA) target sequences can be added to increase the specificity of gene therapy(86,87).

Future Strategies to Correct DBA by Gene Editing

The three most commonly used genome editing technologies are Zinc Finger Nucleases (ZFNs), Transcription Activator Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated Cas9 (CRISPR/Cas9). Several research groups have successfully applied CRISPR-Cas9 technology to correct β-thalassemia mutations in patient-derived iPSCs(88,89).

**RNA based strategies:**

One of the RNA-based technologies potentially useful for therapeutic purposes in DBA is Spliceosome-Mediated RNA Trans-splicing (SMaRT) can modify a target mRNA sequence at the post-transcriptional level. SMaRT exploits the ability of the spliceosome to carry out trans-splicing between two different RNA molecules: the mutated endogenous transcript and a synthetic RNA template delivered into the cell by gene transfer. This will result in a chimeric mRNA encoding a sequence without mutations(90). The most important added value of this technology is the conversion of mutant transcripts into wild-type mRNAs for the correction of certain disorders (90). Although this technology is not well established in DBA, this possible mechanism can be taken into account.

Another RNA-based technology with a potential for therapeutic purposes is represented by SINEUPs which can enhance translation of target mRNAs. A functional class of long non-coding antisense RNAs that can increase the translation of a specific transcript by partially overlapping the 5' UTR of the target mRNA(91). The antisense sequence in synthetic SINEUPs can be designed to enhance expression of any gene of interest,

the upregulation induced by SINEUPs is within a physiological range (approximately 2 fold), this is regarded as an advantage to avoid possible side effects due to overexpression. Overexpression of some RPs such as RPL5 and RPL11 is expected to be detrimental for the cell because it can activate p53(92), and SINEUP technology would overcome this issue. Having said that, the mechanism of action of SINEUPs has not been fully understood and needs to be further investigated.

Another technology with a therapeutic potential is to activate the expression of a target gene by the use of Small Activating RNA (saRNAs). This is a class of RNA molecules able to activate the expression of a target gene by binding its promoter region. SaRNAs are double-stranded, 19-21 nucleotides long molecules(93,94). It is known that saRNAs can associate to the protein Argonaute (Ago) 2, forming a nucleoprotein complex called RNA- induced transcriptional activation (RITA) complex. The RITA complex recognizes complementary sequences on the promoter of the target gene and induces histone modification and transcription initiation (95). This technology has not been implemented in DBA yet.

## 1.8. CRISPR/Cas9 in Genome Editing

The Cas9 protein (CRISPR-associated protein 9) derived from type II CRISPR (clustered regularly interspaced short palindromic repeats) bacterial immune systems, is a great tool that is used in gene editing in diverse organisms. Furthermore, it is revolutionizing many areas of medical research and one of the key areas is its gene therapy potentials. As an RNA-guided DNA endonuclease, Cas9 can be easily programmed to target new sites by altering its guide RNA sequence, this has made sequence-specific gene editing more applicable than before. Furthermore, the nuclease-deactivated form of Cas9, which is a mutant Cas9 that retains DNA-binding activity but lacks the gene cutting ability, can be engineered as a programmable transcription repressor by preventing the binding of the RNA polymerase (RNAP) to promoter sequences or as a transcription terminator by blocking the running RNAP. This represents an additional use of CRISPR/Cas9 technology in rewriting the epigenetic status, all in a sequence-specific manner (96).

### 1.8.1. Tools for Programmable Genome Editing

Precise editing of genes is essential to understanding the function of a given gene. During the past decade, the revolution in genetic techniques has made genome editing and regulation significantly easier.

Recently CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated protein) bacterial immune system has been adapted as a simple, RNA-guided technique for highly efficient and specific genome editing, hence creating profound tools for biomedical research. In general, the precise editing or regulation of genomic

information at the DNA level involves the action of a molecular machine composed of two major parts: a DNA-binding domain that mediates sequence-specific DNA recognition and binding, and an effector domain that enables DNA cleavage or regulates transcription near the binding site. Creating a double-stranded break (DSB) by using a sequence-specific endonuclease will be followed by genetic repair either by the error prone non homologous end joining repair (NHEJ) which increase**s** the rate of gene modification at the desired sequence in the form of inactivating mutations in most of the cases or by homologous recombination which is of a very low efficiency(96).

Meganucleases, also called home nucleases, are among the first classes of nucleases that were designed. They are designed to target specific genomic sites for gene editing purposes (15, 16, 21). They recognize long nucleotide sequences and induce a double stranded break (DSB) at their targeted site. Having a long recognition sequence which may occur only once within a genome increases the specificity and thus facilitates its use for site-specific genome editing (97–99).

Other examples of programmable genome editing machines include zinc-finger nucleases (ZFNs) (100–102) and transcription activator-like effector nucleases (TALENs) (103,104), in which the nuclease domain of the restriction enzyme FokI recognizes and fuses with the DNA binding domains of the transcription factors, the FokI domains of these programmable, site-specific nucleases form a dimer that activates the nuclease activity, thus creating a DSB near their binding sites. Researchers can utilize the cell's endogenous DNA repair pathways to create mutations at the desired DSB sites. However, because these tools are dependent in their function upon protein–DNA interactions, targeting

to a new site will require engineering and cloning a new protein which takes a long time, hence impedes ZFNs and TALENs from being used widely (96).

### 1.8.2. CRISPR: An Adaptive Immune Mechanism

The CRISPR system is an adaptive immune mechanism present in many bacteria and most characterized Archaea. CRISPR-containing organisms acquire DNA fragments, to be incorporated in the genome, from invading bacteriophages and plasmids to guide cleavage of invading RNA or DNA (104). CRISPR is an immune system which allows bacteria to both prevent foreign DNA from being inserted into the genome and target the invasive DNA for destruction. It is a repeat and non-repeat sequence. These repeat arrays are referred to as CRISPR, or Clustered Regular Interspaced Short Palindromic Repeats(105).These sequences, or “spacers”, actually contained DNA from bacteriophages. Furthermore, it was found that there are Cas genes, which encode for a DNA endonuclease, in close proximity to CRISPR structures, strongly suggesting that foreign DNA degradation may be a primary function of CRISPR/Cas (106). There are certain motifs within the genome, just upstream of the “protospacers” or target genomic sequences on the foreign DNA, not on the bacterial DNA. These conserved motifs are called protospacer adjacent motifs (PAM). These motifs are preferential targets for the Cas endonucleases and allow the system to discern between self- and non-self-DNA(107). When a virus infects a bacteria, its’ genome will be cut into small fragments, then it will be implanted in the bacterial genome between these clusters so that it will be recognized in the future and hence the bacteria will be immunized against that specific virus (107).

Type II CRISPR systems from *S. Thermophilus* and *S. Pyogenes* (SpCas) could be engineered to edit mammalian genomes. Two major components include (1) a Cas9 endonuclease and (2) the crRNA-tracrRNA complex. When both are co-expressed, they form a complex that is recruited to the target DNA sequence. The crRNA and tracrRNA can be combined to form a chimeric guide RNA (gRNA) with the same function, to guide Cas9 to target gene sequences(107).

The Cas9–gRNA complex can bind DNA that is complementary with the gRNA and is adjacent to a Protospacer Adjacent motif (PAM) sequence, this will result in cleavage ~3-4 nucleotides upstream (5') of the Motif. Thus, by designing 20-nucleotide region of the gRNA to pair with the DNA sequence of interest, Cas9 can be utilized to target any genomic locus containing a PAM sequence. This makes CRISPR an easily programmable platform for specific genomic editing (96).

These components can then be delivered to mammalian cells via transfection or lentiviral transduction. DSB resolution occurs by homology directed repair (HDR) or error-prone non homologous end joining (NHEJ). If there is no donor DNA, resolution will occur by NHEJ, resulting in (indels) that can ultimately knock-out gene function. If donor DNA sequences are available, the DSBs will be repaired by HDR which will result in gene knock-in (108). Additionally, Plouegh and his group successfully employed SCR7 to inhibit NHEJ by inhibiting the ligase enzyme and thus increasing the efficiency of genome editing by HDR (109).

The guide RNA enables Cas9 to cut a specific genomic locus adjacent to a Protospacer Adjacent motif (PAM) sequence. To target different genomic sequences, a large variety of Cas9 proteins exist each with a different PAM requirement. In type II CRISPR systems, Cas9 nucleases has three subclasses(A,B and C) the protein range from about 900 to 1,600 amino acids (110,111). Figure (1.5)

The most commonly used Cas9 for genome targeting has been adapted from *Streptococcus pyogenes* (Sp). This is type II-A CRISPR system. The Sp Cas9 has a simple PAM (NGG), or a weaker NAG, where N is any nucleotide (113). Kleinstiver et al (114) has identified three novel spCas 9 variants that can recognize other PAM sequences, these are VQR variant, EQR variant and VRER variant that can identify (NGAN or NGNG, NGAG, NGCG) respectively. Furthermore, these variants are as selective and specific as the wild type SpCas9, since the number of off-target breaks for VQR and VRER is similar to the wild-type SpCas9 (114,115).

Sa Cas9 is another Cas9 derived from *Staphylococcus aureus* (Sa), which is 1,053 aminoacid (AA) in length, with NGRRT or NGRRN PAM (where R is an A or G)(113). The relatively small size of Sa Cas9 compared to SpCas9 allows it to evade some of the delivery issues caused by the larger Sp Cas9 (1,368 AA).The gene editing efficiency with Sa Cas9 is comparable to that of Sp Cas9 (113).

Cas9 from *Neisseria meningitidis* (Nm; PAM = NNNNGATT) and *S. thermophilus* 1 (St1; PAM = NNAGAAW, where W is an A or T), were both used in mammalian cells (116,117). Another form of Cas9 is *Treponema denticola* (Td) Cas9 which recognizes NAAAAC as a PAM sequence (107).

CRISPR Cas9 is emerging as a powerful tool for engineering the genome in diverse organisms. CRISPR Cas9 can be easily programmed to target new sites by altering its guide RNA sequence, and its development as a tool has made sequence-specific gene editing easier compared to other gene editing methods.

This work will explore the possible applications of CRISPR Cas9 in knocking out certain RP genes in HEK 293T cells and in k562 cells, in addition this technique will be utilized to knock in certain novel mutations that were discovered in our lab.

The effect of RP gene editing on rRNA processing in cell lines will then be studied by analyzing rRNA intermediates using RT-qPCR and RNA capillary electrophoresis which were validated in the first aim of this project

CRISPR Cas9 knock in experiments will be further optimized in the future to knock in the wild type sequence into CD34 positive primitive cells to correct the mutation and hence cure anaemia which is the long term goal of this project.

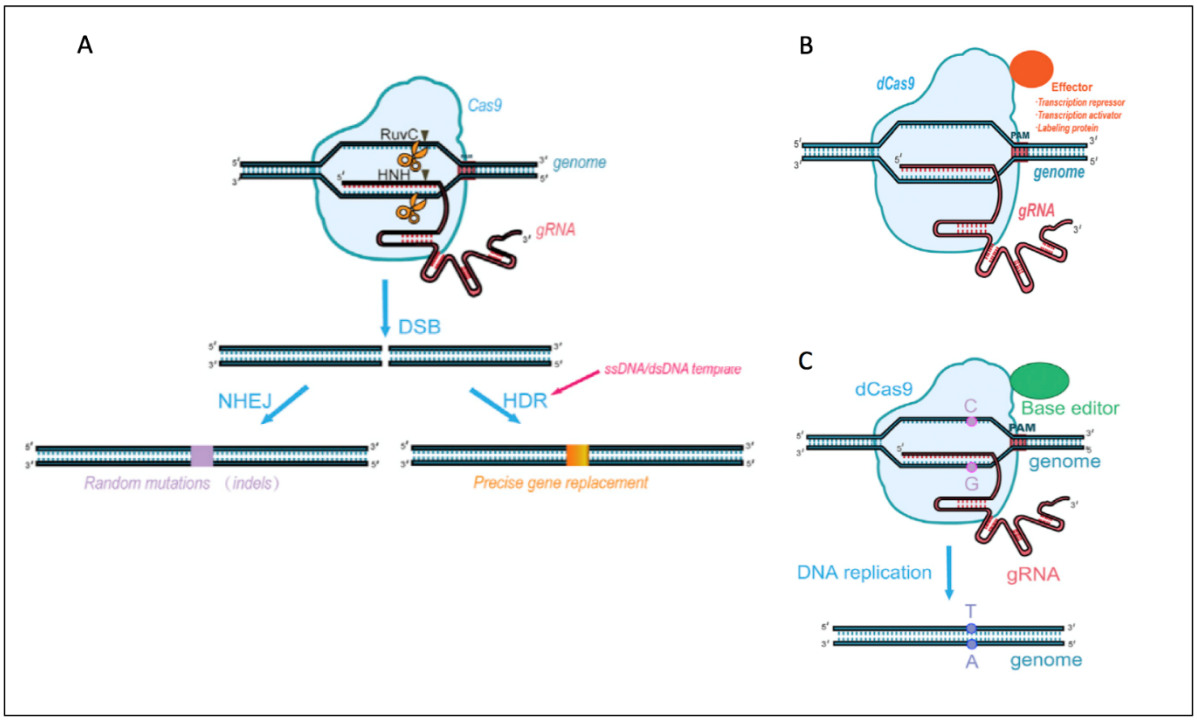


**Figure (1.5): (A) Cas9 nuclease cleaves double-stranded DNA via RuvC and HNH domains to introduce double-strand breaks (DSBs) that are then repaired by NHEJ or HDR.** Error-prone NHEJ repair pathways always introduce random insertions or deletions (indels) but with the use an exogenous DNA donor, the HDR pathway can introduce precise insertions. (B) Dead Cas9 (dCas9) contains inactivating domains in HNH or RuvC. It can be tethered with transcriptional factors to mediate downregulation or activation of target genes. (C) dCas9 can be fused to deaminase for catalytic conversion of C to U, thus achieving single-base editing during DNA replication. Similarly, dCas9 can be tethered with epigenetic modification enzymes to obtain desired edits. (E) dgRNA can guide Cas9 to regulate gene expression. Adapted from You et al (112)

# Materials and Methods

## 2.1. Isolation and stimulation of primary T cells

Forty-eight Blood samples in EDTA were collected at St Mary's Hospital in addition to 17 age matched normal donors. Mononuclear cells were isolated by gradient density centrifugation after layering anti-coagulated blood onto Histopaque-1077 (Sigma-Aldrich Company Ltd., Dorset, UK). T cells were isolated by magnetic columns and Human CD3 Positive Selection Kit (Stemcell Technologies) according to the manufacturer's protocol. This was confirmed by Flow cytometry through staining the cells with anti-CD3 monoclonal antibody(MoAb). (See appendix for patients' details).

Next, $1x10^5$ T- cells (CD3 positive) were plated in each well of 96-well plate with 120 µl T cell medium (TCM) at 37°C, 5% $CO_2$.T cells were then activated and cultured with CD3/CD28-coated paramagnetic beads (beads covalently coupled to anti-CD3 and anti-CD28 antibodies) (Life Technologies, Paisley, UK). These paramagnetic beads were added to obtain a bead-to-cell ratio of 1:1. Interleukin 2 (IL-2) 10ng/mL was added on the following day and then twice weekly. TCM consisted of RPMI 1640 (Invitrogen) supplemented with 10% Fetal Bovine Serum (FBS) (Sigma-Aldrich Company Ltd., Dorset, UK), 10ml/L Penicillin-Streptomycin (Stem Cell Technologies, Vancouver, Canada), L-glutamine 4mM, 1:100 nonessential amino acids 1%, sodium pyruvate 1% and 10mM Hepes buffer.

Growth of cells was monitored daily and T cell purity assessment was performed by flow-cytometry after staining the cells with anti-CD3 fluorescent antibody CD3-PerCP Cy5.5 (eBiosciences). RNA was extracted and studied at two-time points: before stimulating the cells (day 0) and six days after stimulation (day 6).

## 2.2. RNA Isolation

RNA isolation from selected T cells was performed using the Nucleospin RNA isolation Kit (Macherey-Nagel) according to the manufacturer's protocol.

This kit is recommended for isolation of RNA from cultured cells, tissue, cell-free biological fluids. This kit allows purification of up to 70 µg of highly pure, DNA-free RNA.

RNA prepared with this kit is suitable for applications like reverse transcriptase-PCR (qRT-PCR), Northern blotting, and primer extension.

With the NucleoSpin RNA kit, cells were lysed by incubation in a solution containing large amounts of chaotropic ions. This lysis buffer immediately inactivates RNases (which are present in virtually all biological materials). After cell lysis, homogenization and reduction of viscosity were achieved by filtration with specific filter units provided with the kit. Next, the lysate was applied on silica membrane (columns), as the lysis buffer has created appropriate binding conditions which favor adsorption of RNA to the silica membrane. Contaminating DNA, which is also bound to the silica membrane, was removed by DNase solution which was directly applied onto the silica membrane during the preparation. Optimal conditions for the rDNase were achieved by washing the silica membrane with a specific desalting buffer before treatment. Salts, metabolites and macromolecular cellular components were removed by simple washing steps with two different buffers. RNA was finally eluted with RNase-free water supplied with the kit. After elution the RNA was stored in −80°C or in liquid nitrogen to provide protection from degradative reactions.

## 2.3. RNA Capillary electrophoresis (Agilent Bioanalyzer/2100)

The Agilent RNA 6000 Nano Kit reorder number (reorder number 5067-1511) was used in this experiment. At the beginning, RNA Nano reagents need to equilibrate to room temperature for 30 minutes and the electrodes must be decontaminated using RNase free water for 10 seconds.

Filtered RNA gel matrix should be prepared according to the kit instructions and then RNA dye concentrate was added to it in order to have the “Gel-Dye” mix ready to be used. RNA samples were denatured by incubation at 70°C for 2 minutes, then they were placed on ice during the experiment. After having the RNA samples denatured and the “Gel-Dye” mix ready then priming of the RNA 6000 Nano chip with 9µl of the Gel-Dye mix was done. Nine µl of gel-dye mix was pipetted into the two remaining wells on the chip marked **G**, and 5µl of RNA 6000 Pico Marker was added into the 13 remaining empty wells including the ladder’s position. After priming the chip and adding the Nano marker, RNA samples were added to the chip. 1µl of each sample was pipetted into its assigned well and finally the chip was loaded into the Bioanalyzer.

## 2.4. Quantitative real time polymerase chain reaction (Q-RT-PCR)

Complementary DNA was synthesized from 90 ng of RNA using the Revert Aid First Strand cDNA Synthesis Kit (Thermo Scientific) according to the manufacturer’s instruction. Amplification was performed using TaqMan Multiplex Master Mix (Applied Biosystems).

rRNA taqman assays that were used in this work were:
18S rRNA, 28S rRNA, 32S rRNA , 5S rRNA. The primer sequences were designed in our lab as illustrated in figure (2.1), and sent to Thermo Fisher company to manufacture the assay.Because of the overlap between 28S rRNA and 32S rRNA as illustrated in figure (2.1). The expression of 32S rRNA was deducted from that of 28S rRNA.
5s rRNA F:ctagctgcgagaattaatgtg
5s rRNA R::gaagtgtcgatgatcaatgtg
18s rRNA F:ttgcgttgattaagtccct
18s rRNA R:tcactaaaccatccaatcgg
28s rRNA F:cctcacgatccttctgac
28s rRNA R:ccacaagccagttatccc
32S rRNA F(ITS-28s) F: CGATTCCGTCCGTCCGTC
32S rRNA R(ITS-28s) R: TTAAATTCAGCGGGTCGCCA
Afterwards Thermo Fisher has made the assays accordingly.

| Name | Assay ID |
|---|---|
| 18S_ | AP322NG |
| 28S_ | AP47V9E |
| 5S__ | AP7DPUC |
| ITS-28S | AP9HJD9 |

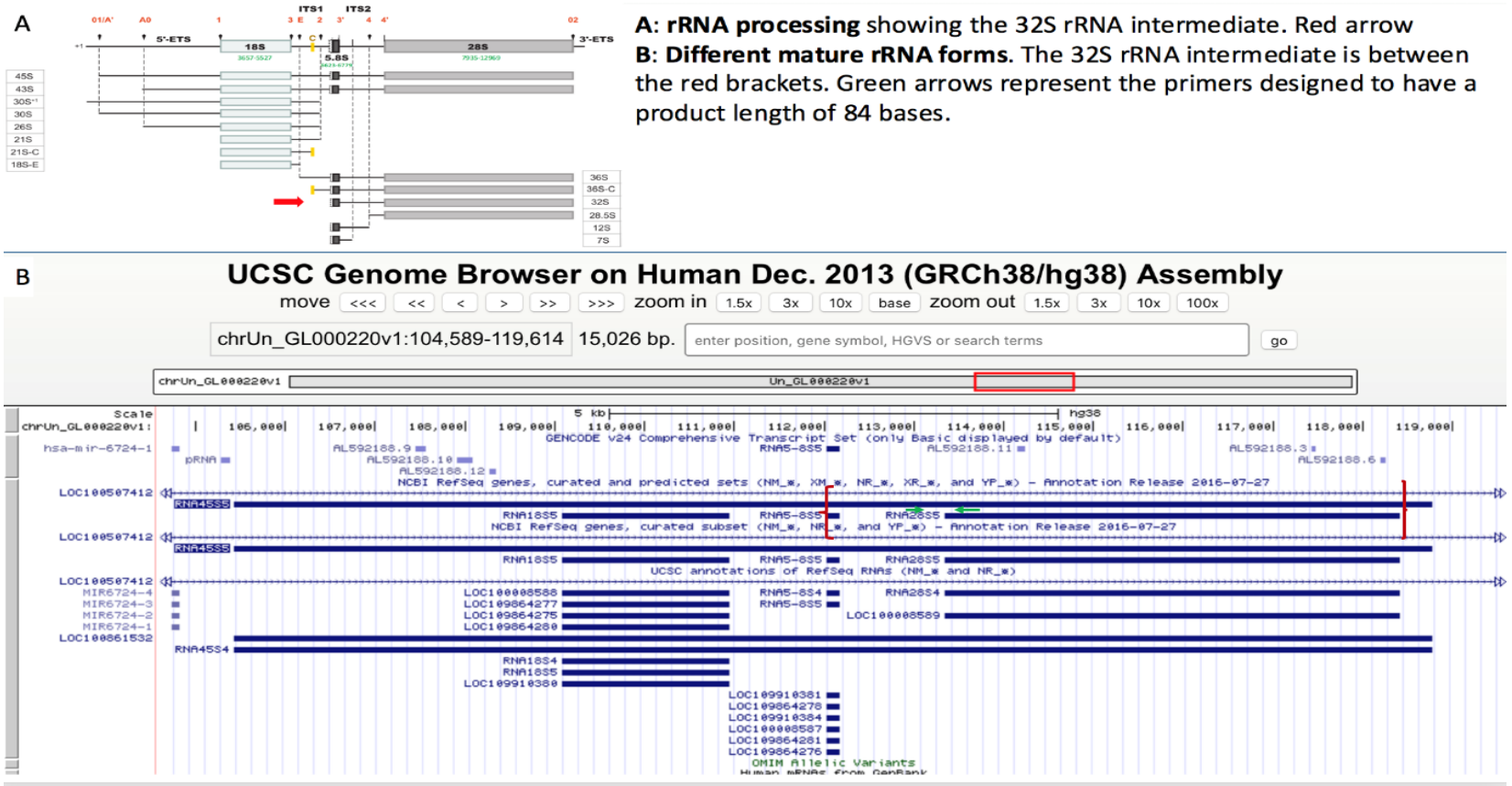


**Figure (2.1): 32s rRNA primer design.** The primers (green arrows) can amplify 32S rRNA without overlapping with 28S or 5.8s rRNA. The forward and reverse primers are indicated by the green arrows, the amplicon length is 84 bases.

Internal reference genes that were used in this study were: Importin 8 (IPO8) this was shown to be the most stable reference gene for internal validation. Other reference genes used in this study include: actin beta (ACTB), TATA box binding (TBP), transferrin receptor (TFRC), and ribosomal protein lateral stalk subunit P0 (RPLP0). These genes were tested on 10 samples on unstimulated T cells (day 0) and six days following stimulation. In all of our experiments we used the equation 2^ (Ct housekeeping gene-Ct gene of interest), Ct is the cycle threshold.

## 2.5. Software determination of reference gene stability

Reference gene expression stability was evaluated using the universally available mathematical software NormFinder version 0.953 (Andersen *et al.*, 2004) (available from Aarhus University, Denmark; http://moma.dk/normfinder-software).

## 2.6. CRISPR/CAS9 transfection

**2.6.1. LentiCRISPv2 puro (Addgene Plasmid #98290)** This plasmid was provided from Feng Zhang Lab. It contains two expression cassettes, hSpCas9 and the chimeric guide RNA as shown in figure (2.2). However, GFP was cloned after removal of puromycin resistance gene in order to be used in our experiment.The resultant vector was digested using BsmBI, and a pair of annealed oligos were cloned into the single guide RNA scaffold as illustrated in figure (2.3). The oligos are designed based on the target site sequence (20bp) and needs to be flanked on the 3' end by a 3bp NGG PAM sequence.

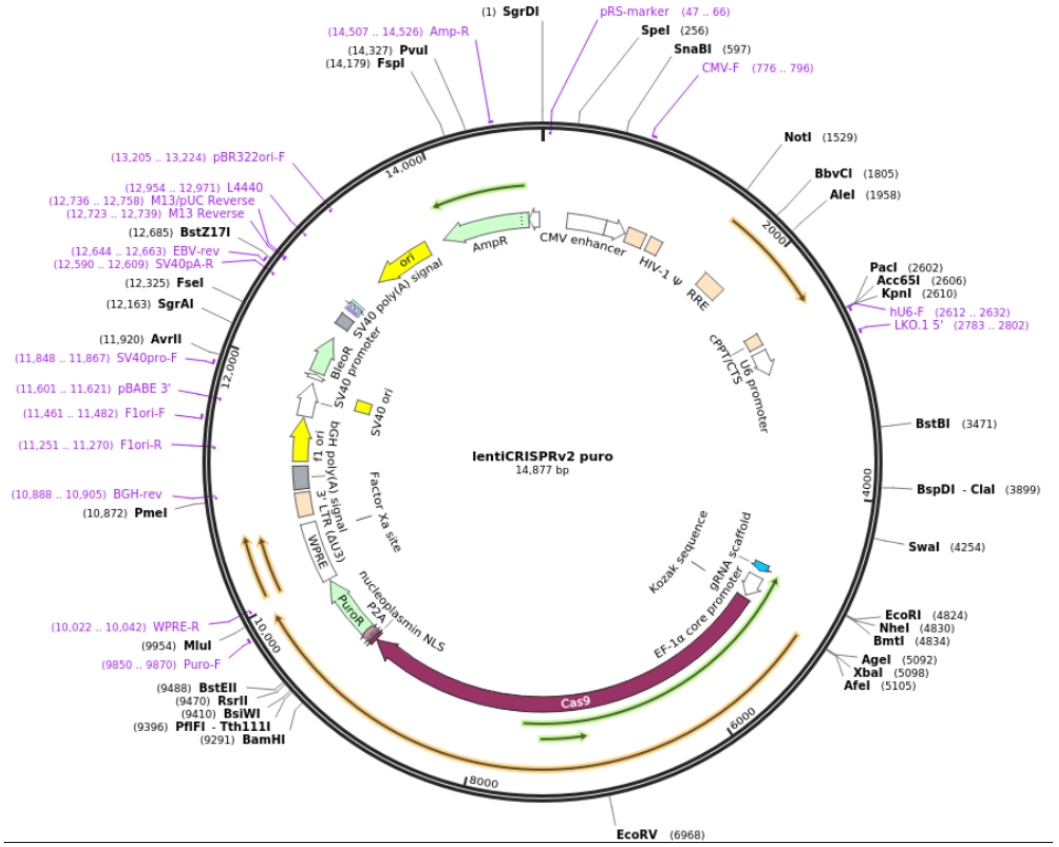


**Figure (2.2): LentiCRISPv2 puro (Addgene Plasmid #98290)** from Brett Stringer. GFP gene was cloned after removing puromycin resistance gene.

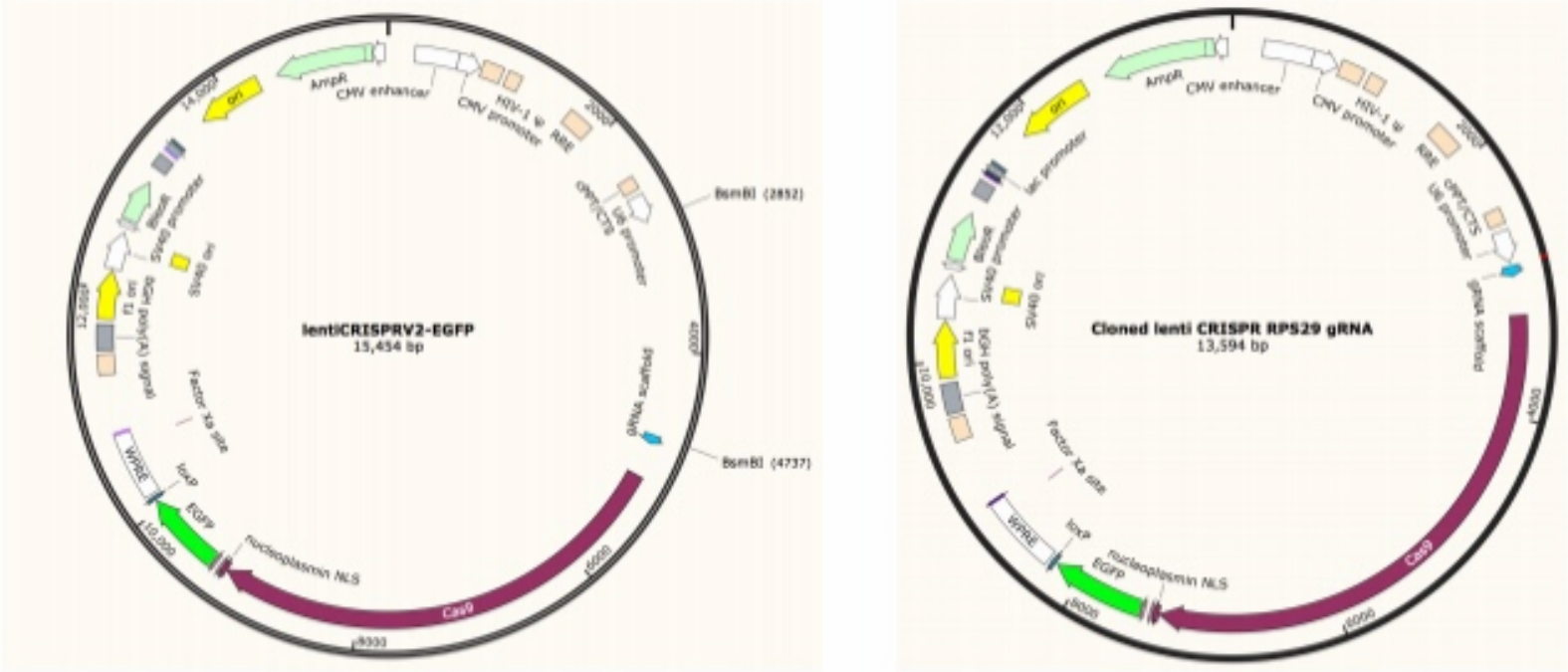


**Figure (2.3): LentiCRISPRv2-EGFP plasmid map**. (A) Showing the cloning site. BsmBI is the restriction enzyme used to clone the double stranded gRNA (B) gRNA cloned into LentiCRISPRv2-EGFP plasmid. **Abbreviations are as follows: EGFP**: Green fluorescent protein. **ORI**: Origin of replication is the place where DNA replication begins. **NLS:** Nuclear Localization signal. **U6 promoter** is used for gRNA production. **AmpR** is the ampicillin resistance gene**. bGH poly A** signal: promotes both polyadenylation ad termination.

### 2.6.2. Target Guide Sequence Cloning Protocol

Following digestion of LentiCRISPRv2 with BsmBI, this plasmid will have a certain overhang. Based on this overhang gRNA was designed. Firstly, the gRNA is designed using the Gecko software (http://crispr.mit.edu) to ensure the high specificity of the guide and avoid off targets. Two gRNA oligo**s** were ordered and cloned into the LentiCRISPRv2 vector. Standard de-salted oligos were diluted to 100 µM in sterile water.

| gRNA | | 20 bp specific to the sequence to be edited | |
|---|---|---|---|
| 1st Forward Oligo | 5' | CACCGnnnnnnnnnnnnnnnnnnn | 3' |
| 2nd Reverse Oligo | 3' | Cnnnnnnnnnnnnnnnnnnn**CAAA** | 5' |

**Lentiviral vector digestion and cloning into digested vector:**

**1.** From the lentiviral CRISPR plasmid, 5µg was digested with *BsmB*I for 30 min at 37°C and dephosphorylated according to the protocol shown in table (2.1).

**Table (2.1): LentiCRISPRv2 digestion and dephosphorylation protocol.** Reagents used were from ThermoFisher.

| | |
|---|---|
| 5 ug | LentiCRISPRv2 |
| 3 ul | Fast Digest BsmB1(ThermoFisher) |
| 3 ul | Fast AP |
| 6 ul | 10X Fast Digest Buffer |
| 0.6 ul | 100 mM DTT (freshly prepared) |
| X ul | H2O |
| 60 ul | Total |

On the gel: a 2kb filler piece should be present in addition to the larger band 13,454 bp which will be extracted and purified as illustrated in figure (2.4)

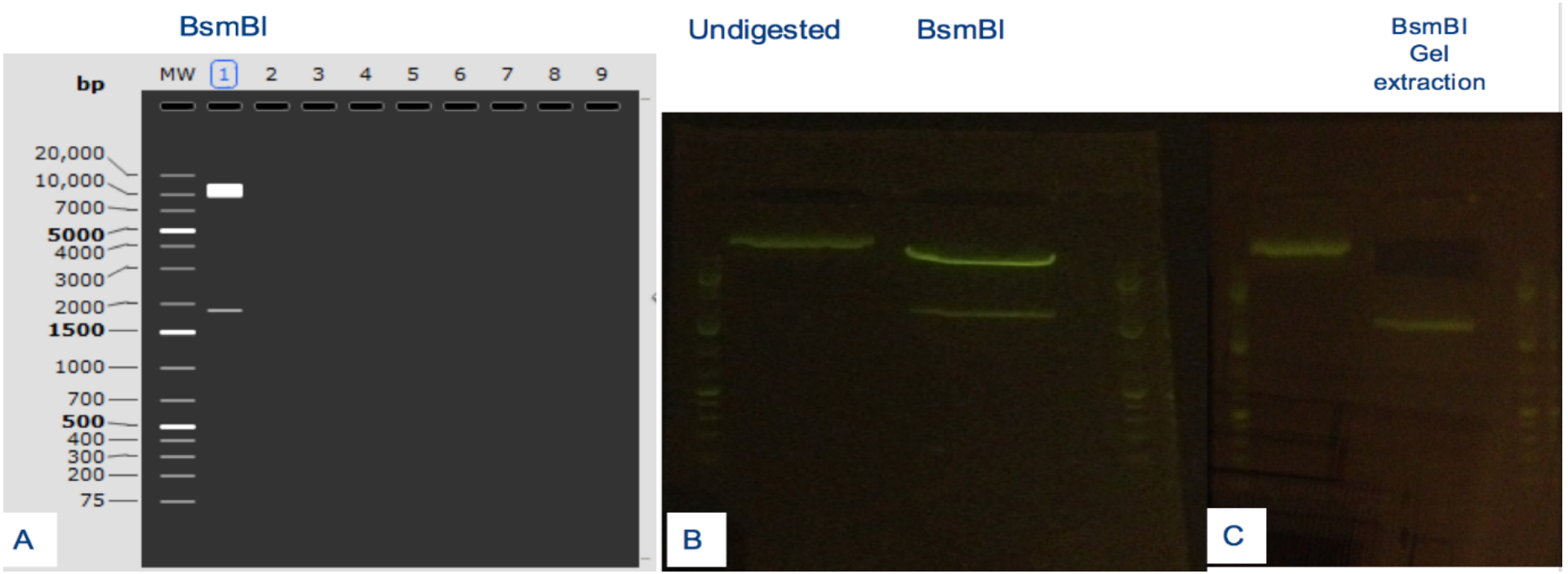


**Figure (2.4): LentiCRISPRv2 plasmid cloning.** (A) Snap gene software showing an in-silico image for the Lenti CRISPRv2–EGFP plasmid (B) Gel image of LentiCRISPRv2 before and after digestion with BsmB1 restriction enzyme. (C) Same Gel image as B following gel extraction of the digested part of the plasmid 13,454 bp.

**Agarose gel electrophoresis of DNA**

DNA agarose gels were cast by melting between 1% (w/v) and 1.2 % (w/v) Molecular grade agarose (Sigma) in 1x TAE (40 mM Tris (pH 7.6), 20 mM acetic acid, 1 mM EDTA). Syber Safe DNA stain 1x (Invitrogen) was added to the gel after allowing the gel to cool and cast in an Owl D3-14 Horizontal Electrophoresis System (ThermoScientific).

Six times loading buffer (ThermoFisher) which contains two different dyes bromophenol blue and xylene cyanol for visual tracking of DNA migration during electrophoresis was added to DNA samples to a final 1x concentration. Then loaded into the cast agarose gel submerged in 1x TAE. Generuler 1 kb Plus (ThermoFisher) DNA standard was used as a ladder. Electrophoresis was run at 70-90 V using the Owl D3-14 Horizontal Electrophoresis System (ThermoScientific). UV light on a transilluminator (BioDoc-It imaging system, UVP) was used to see DNA bands.

**2.** The digested plasmid was cut from the gel as shown in the figure (2.4) and purified using GeneJET Gel Extraction Kit (ThermoFisher, K0691) according to the manufacturer's protocol and eluted in elution buffer.

Thermo Scientific GeneJET Gel Extraction Kit is designed for rapid purification of DNA fragments from standard agarose gels run in either TAE buffer. The kit utilizes a silica-based membrane technology in the form of a spin columns. The kit can be used to purify DNA fragments from 25 bp to 20 kb in size.

**3.** Phosphorylation and annealing of each pair of oligos (guide RNA) was performed according to the following protocol:

**Table (2.2): annealing and phosphorylation of gRNA oligos.**

| | |
|---|---|
| 1 ul | Oligo 1(100uM) |
| 1 ul | Oligo 2(100uM) |
| 1 ul | 10X T4 Ligation Buffer (with ATP) (ThermoFisher) |
| 6.5 ul | H2O |
| 0.5 ul | T4 PNK (BioLabs) |
| 10 ul | Total |

The tube was placed in the thermocycler for 30 minutes at 37°C followed by 5 minutes at 95°C and then ramp down to 25 at a rate of 5/minute.

**4.** Annealed oligos were diluted at a 1:100 dilutions into sterile water.

**5.** The diluted annealed oligos were then ligated with the digested CRISPRv2 plasmid and incubated at room temperature for 10 minutes according to the following protocol.

**Table (2.3): Ligation of the annealed gRNA to the digested LentiCRISPRv2 plasmid.**

| | |
|---|---|
| X ul | BsmB1 digested plasmid (50ng) |
| 1 ul | Diluted annealed oligos |
| 5 ul | 2X Quick Ligase Buffer(BioLabs) |
| X ul | H2O |
| 1 ul | Quick Ligase(BioLabs) |
| 11 ul | Total |

**6.** Transformation into NEB 5-alpha competent *E. coli* which is a derivative of the popular DH5α

Firstly, the 50μl competent cells were thawed on ice for 30 minutes. Then, 5 μl of the ligated CRISPR/CAS9 vector were added to the competent cells and mixed by flicking the tube 4 times, then the mixture was kept on ice for 30 minutes. This step was followed by heat shock at exactly 42°C for 45 seconds. The tube was then placed on ice for 2 minutes. Stable Outgrowth Medium (SOC) (BioLabs) medium was added to the mixture (350μl) and the tube was placed in the shacking incubator at 37°C for one hour. Selection plates were pre warmed to 37°C. Finally, 150 μl of the cells were spread onto a selection plate and incubated overnight at 37°C. In addition, a negative control ligation (digested CRIPR vector-only) was also used for bacterial transformation.

**7.** Growth of Bacterial Cultures: A single colony was picked from a freshly streaked selective plate and inoculated in 5mL of LB medium supplemented with ampicillin and incubated for 12-16 hours at 37°C while shaking at 200-250 rpm. The culture is then diluted in 300mL LB medium supplemented with 100 μg/ml ampicillin (Sigma) and incubated for 12-16 hours at 37°C while shaking at 200-250 rpm.

**8.** Isolation of high-quality plasmid DNA from recombinant E.coli culture. This was done using the GeneJet Plasmid maxiprep Kit (Thermo Scientific GeneJET Plasmid Maxiprep Kit) according to the kit protocol.

In summary, the bacterial cells were pelleted to get rid of the LB medium then resuspended and subjected to lysis solution which contains buffering and ionic salts to regulate pH and osmolarity of the lysate in addition to the detergent sodium dodecyl sulfate (SDS) to help breaking break up the membrane structure and liberate the plasmid DNA. Next, the

chromosomal DNA and cell debris were pelleted by centrifugation to result in a supernatant that contains the plasmid DNA which was loaded onto the purification column. Plasmid DNA will bind to the silica membrane by the aid of the appropriate conditions provided by the high salt concentration of the lysate. The adsorbed DNA specifically binds to silica membrane in the presence of certain salts and at a particular pH. The cellular contaminants are removed by wash steps. DNA is eluted in a low salt buffer or elution buffer. Chaotropic salts are included to aid in protein denaturation and extraction of DNA.

**9.** DNA quantification was done using the Nano drop and a concentration of 700-1200 ng/ ul was obtained.

**10.** The insertion of the guide RNA into the CRISPR vector was confirmed using Sanger sequencing.

## 2.7. Media preparation

LB Agar (Sigma) Solid Media was used to grow NEB 5-alpha competent *E. coli*. Media was always autoclaved and left to cool, 100 µg/ml ampicillin (Sigma) was then added. LB agar was melted and cooled to around 50 °C before antibiotics were added. 9 cm petri dishes (Corning) were then poured in laminar flow hood to dry for 20 min. Plates are then stored at 4°C.

Lysogeny Broth (LB) (Sigma) Liquid Media is prepared, autoclaved then 100 µg/ml ampicillin (Sigma) was added before the use for bacterial expansion.

## 2.8. Preparation of Competent Cells

Competent bacterial strain DH5α (NEB) were expanded and made chemically competent via the following method. Bacteria were spread onto LB Agar plate containing no selectable antibiotics and incubated for 16 hours. From this plate a single colony was picked and added to 5 ml of antibiotic free LB Broth for a further 16 hours. The 5 ml culture was then moved into an Erlenmeyer flask containing 400 ml LB, containing no antibiotic, and incubated at 37 °C until an OD600 of 0.6 was reached. The flask was placed for 15 min on ice and then centrifuged for 20 min at 3,500 x g at 4 °C. The resulting pellet was resuspended in 100 ml of 0.1 M $MgCl_2$. Cells were then centrifuged for 20 min at 3,500 x g at 4°C. The pellets were resuspended in 200ml of 0.1 M Ca $Cl_2$, and placed on ice for a further 30 min. The cells were again centrifuged for 20 min at 3,500 x g and pellet resuspended in 8 ml ice cold 0.1 M $CaCl_2$ 15% (v/v) glycerol and aliquoted into 100µl aliquots and snap frozen for 15 min on dry ice. The resulting competent cells were stored at -80°C for future work.

## 2.9. Nucleofection Protocol

For K562 cell nucleofection, Amaxa Cell Line Nucleofector Kit V(Lonza) was used as per manufacturer's protocol. K562 cells were grown in RPMI-1640 media supplemented with 10 % (v/v) heat inactivated Foetal Bovine Serum (FBS) (Lot: 08Q9057K Gibco) and 2 % (v/v) Penicillin/Streptomycin. After 4 tissue culture passages K562 cells were reseeded at an appropriate density of $5x10^5$. Following two days of culture in a humidified 37°C/5% $CO_2$ incubator, cells were washed, pelleted and resuspended in complete medium and then counted. Approximately 1-$2x10^6$ cells were re-pelleted and resuspended in 100µl Nucleofector

solution in addition to 1ug of the cloned GFP positive CRIPSR-Cas9 vector (with or without the addition of 5 ul of template. The resulting nucleofection buffer-cells-template mixture was then placed into specific cuvettes that fit in the Amaxa Nucleofector I device. T016 program was then selected as per manufacturer's recommendations for K562 cells. Immediately 500µl of pre-warmed culture medium was added to the cuvette. The cells were gently transferred into prepared 12-well plate which had 1.5 ml **medium** per well (so that the final volume will be 2 ml medium per well). Cells were incubated in humidified 37°C/5% $CO_2$ incubator for 24 hours and then cells were pelleted, washed and resuspended in medium. GFP expression was checked using flowcytometry. Three days later, single cell sorting into 96 well plate based on GFP expression was performed. Cells were then kept in humidified 37°C, 5% $CO_2$ incubator. After three weeks, the colonies were obvious and were transferred to 24 well plates with 2ml **medium** in each well. After a further one week, PCR was performed directly from each colony followed by Sanger sequencing to select the edited clone.

For CRISPR/Cas9 Knock in, the same nucleofection protocol is used with the addition of double stranded template that is 120 base pairs long. Template annealing is done according to the following protocol

**Table (2.4): Annealing of homology directed repair template.**

| | |
|---|---|
| 1 µl | Oligo 1(100µM) |
| 1 µl | Oligo 2(100µM) |
| 1 µl | 10X T4 Ligation Buffer (with ATP) (ThermoFisher) |
| 6.5 µl | $H_2O$ |
| 0.5 µl | T4 PNK (BioLab) |
| 10 µl | Total |

This should be placed in a thermocycler: 30 minutes at 37°C followed by 5 minutes at 95°C and then ramp down to 25 at a rate of 5/minute.
After annealing the template, 5ul of the annealed product is used for nucleofection to be mixed with the nucleofection solution prior to transfection.

## 2.10. Flow Cytometric Analysis

LSR Fortessa (Beckton-Dickinson, Oxford, UK) with FACSDiva v6.2 acquisition were used to analyze samples. Non-overlapping fluorophores where used. Additionally, in each experimental technique and cell line the data was compensated to remove cellular auto-florescence. Fluorophores and fluorescent proteins were excited and detected using the laser and filter combination detailed in Table (2.5)

### 2.10.1 Flow-cytometry for T cells

Flow-cytometry was performed on day 0 to confirm CD3 selection. Staining of peripheral blood mononuclear cells was performed at 4°C for 20 minutes in the presence of directly fluorescently conjugated mononuclear antibodies (MoAbs). Cells were resuspended in PBS. The MoAbs used were CD3-PerCP Cy5.5 (eBiosciences). DAPI was also added to the samples to exclude dead cells. Samples were analyzed on a LSR Fortessa flow-cytometer. Doublets were excluded based on FSC-W/FSC-A gating. Data analysis was performed with FlowJo software v7.6.5 (Tree Star). Gating strategy is shown below in Figure (2.5).

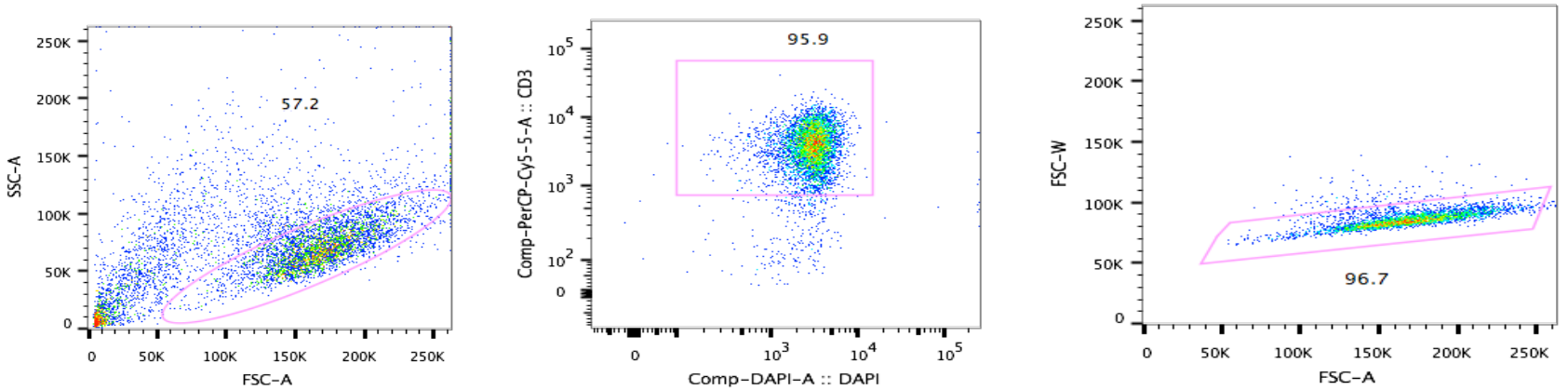


**Figure 2.5: Gating strategy used for analysis of T cells by flow-cytometry.** Representative plots illustrating the method used for the analysis and identification of CD3+ T-cells. Gating on live CD3+ cells identified with anti-CD3 MoAb (PerCP-cy5 is the fluorochrome). **Forward versus side scatter** (**FSC vs SSC**) gating is commonly used to identify live cells based on size and granularity (complexity). It is often suggested that forward scatter indicates cell size whereas side scatter relates to the complexity or granularity of the cell**.** **Forward scatter Area versus Width (FSC-A vs FSC-W)** is used to exclude doublets.

### 2.10.2. Flow Cytometric Analysis and Gating Strategy for CRISPR experiments

Seventy-two hours following electroporation, confirmation of CRISPR transfection is carried out by Flowcytometry to ascertain GFP expression in the transfected cells. Transfection efficiency following electroporation was ranging between 20-30% on the basis of GFP expression. Additionally, the viability was ranging from 45-60%. Figure (2.6) and (2.7) illustrates the gating strategy following electroporation.

### 2.10.3. Fluorescence-Activated Cell Sorting (FACS)

FACSAria (Beckton Dickinson, Oxford, UK) machines utilising a 100 µm nozzle, was used for single cell sorting, the eGFP was excited by 488 nm laser and detected using a 525/50 filter. Control of the FACSAria was done with use of BD FACSDiva v8.03.

### 2.10.4. Analysis of flow cytometry data

Data was analysed on FloJo vX software (Tree Star, Oregon, USA). Gating strategies are as described in the results and Appendix D. In all cases, an initial gate viable cells was followed by a second gate to exclude doublets. Finally, simple sequential gating has been used to identify specific cell populations based on their specific fluorophore.

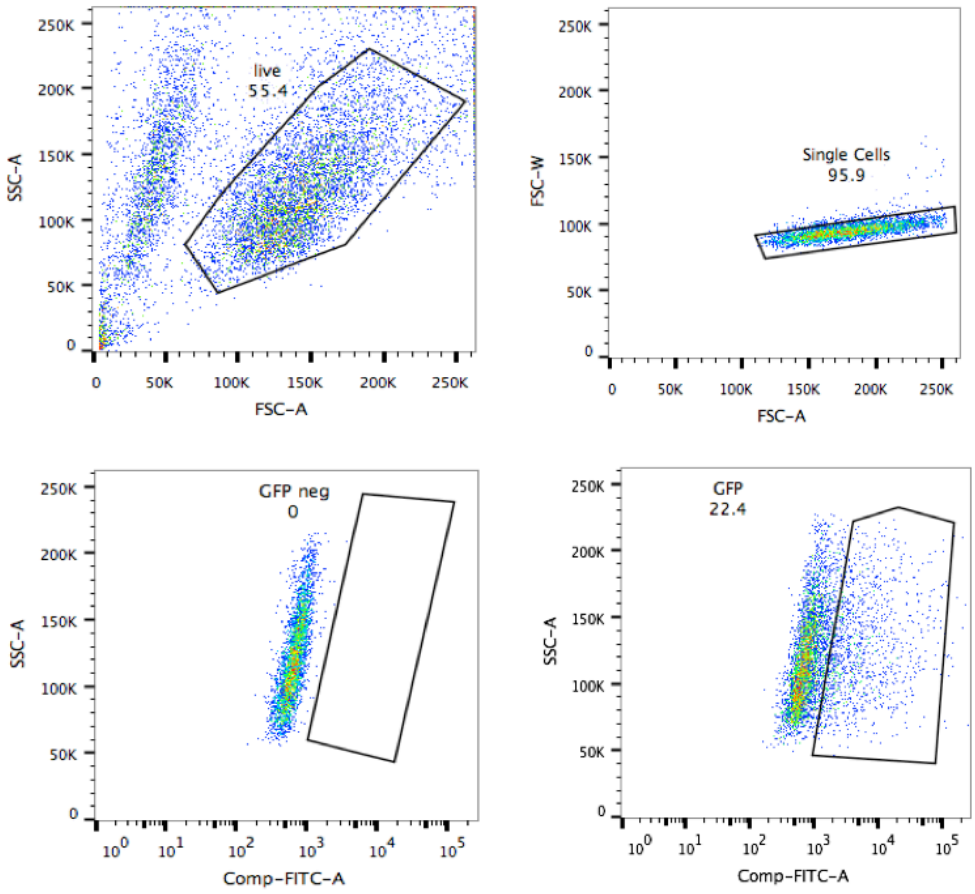


**Figure 2.6: GFP expression following electroporation**. Untransfected cells were used as a negative control to set up the GFP gate. K562 cells showing 22.4% GFP expression 72 hr. following electroporation. Single cell sorting was carried out using FACSAria (Beckton Dickinson, Oxford, UK) machines utilizing a 100um nozzle. The eGFP was excited by 488 nm laser and detected using a 525/50 filter. Forward versus side scatter (FSC vs SSC) gating is commonly used to identify live cells based on size and granularity (complexity). It is often suggested that forward scatter indicates cell size whereas side scatter relates to the complexity or granularity of the cell. Forward scatter Area versus Width (FSC-A vs FSC-W) is used to exclude doublets.

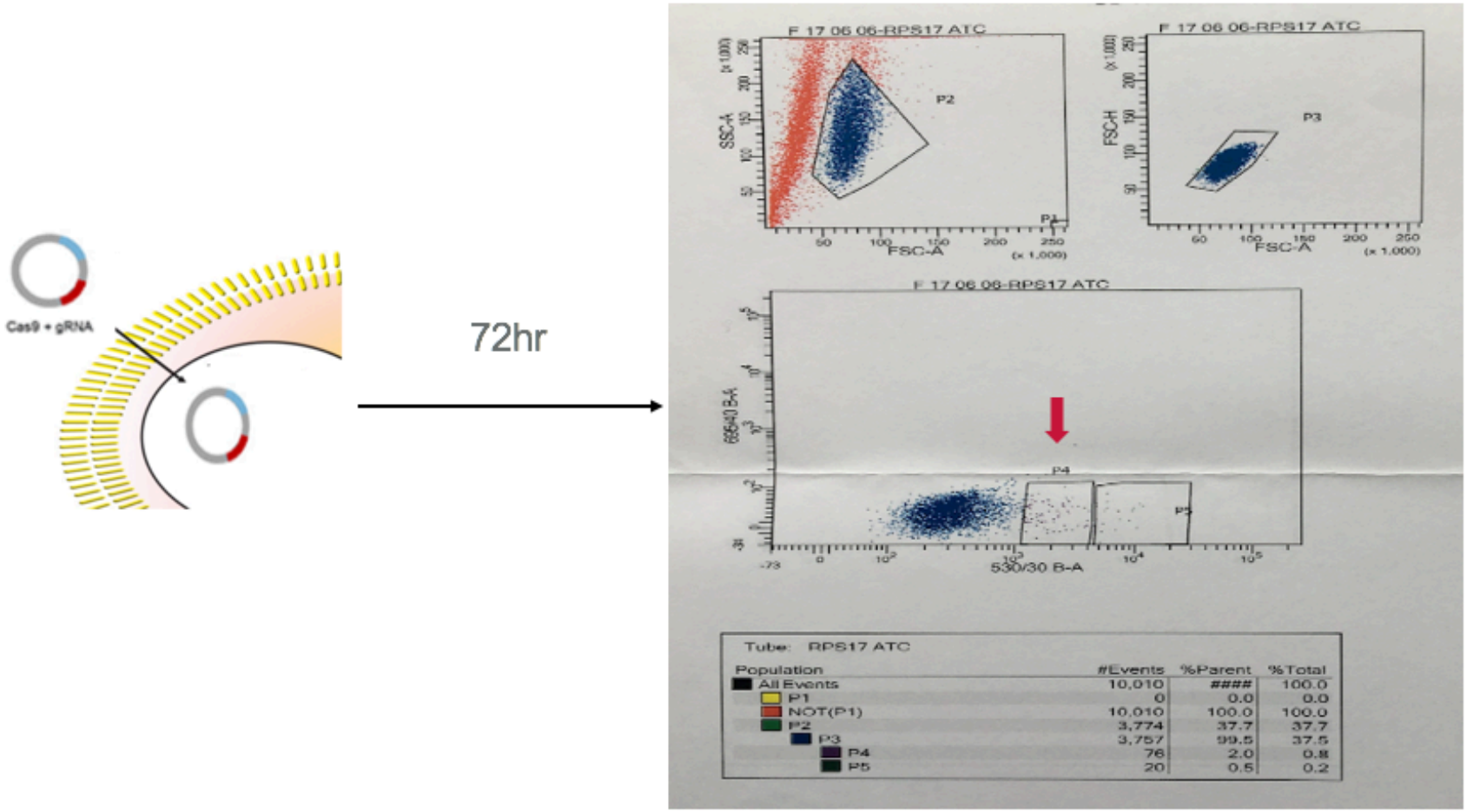


**Figure 2.7: Gating strategy to sort single cells 72 hours following electroporation.** Gating on low to intermediate GFP expression (red arrow) to ascertain transient transfection

## 2.11. Viral production, concentration and transduction

In order to validate certain novel mutations, CRISPR CAS9 editing tool was used to knock out gene expression. To achieve this a retroviral transduction technique was utilized.

For each viral construct to be designed, around ten 10 cm cell culture dishes were each seeded with $3x10^6$ HEK-293T cells (with a passage of no more than 12) that were grown in DMEM medium (Sigma) with 10 % (v/v) FBS, and 1 % (v/v) Penicillin/Streptomycin**.** The next day, the media was replaced with 9 ml fresh media and returned to the incubator. After a further hour, a transfection mixture with the following composition per plate: 12 µg psPAX2 (addgene plasmid #12260) , 4 µg p.MD2.G (plasmid #12259) and 15 µg of Lenti v2 CRISPR CAS 9 plasmid DNA was diluted to a final volume of 450 µl in Low TE (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) and 50 µl of 2.5 M $CaCl_2$ added. Finally, 500 µl of 2x HEPES Buffered Saline (274 mM NaCl, 10 mM KCl, 1.4 mM Na2HPO4, 15 mM D-Glucose and 42 mM HEPES) (HBS) was added dropwise to the existing mixture whilst vortexing. 2XHBS was stored at -20°C and thawed on ice and used immediately.

The resulting 1 ml mixture was added dropwise across the 10 cm plate and the plate gently swirled to mix then each plate was placed back to the incubator.

Sixteen hours later, the media which contains calcium phosphate precipitates is removed and replaced by 8 ml of new fresh medium and the transfected cells were then returned to the incubator. 24 hours later the media which contains the designed virus was collected and a fresh 7ml of media was added to the HEK-293T cells to be processed next day. The collected media that has the designed virus was spun at 500 RPM in

a refrigerated centrifuge and filtered with a 0.45 µm filter to remove cell debris, if present.

The virus containing media was then balanced in volumes of 36 ml and spun at 23,000 *x g* in a Hitachi ultra-centrifuge for 1 hour 40 min. The supernatant was then removed and 200 µl of DMEM media, without any supplementation, was added to resuspend the viral pellet for 4 hours at 4 °C with light aggregation.

The resulting resuspended virus was either directly added to cells or flash frozen in Liquid Nitrogen and stored at -80°C for a maximum of a month. 24 h after the first collection and concentration a second round of collection and concentration was performed. Transduction of HEK-293T cells was done by resuspending 850,000 cells with a viability of at least 70% in 2 ml of fresh media before being plated in a 6 well plate and incubated in a cell culture incubator. One hour later 200 µl of concentrated virus was added. 16 hours later the media was replaced by a fresh media and 24hours later the cells were washed, spun and reseeded in a 6 well plate and left to expand in cell culture incubator.

Transduced cells which express GFP were sorted after 48 hours. The cells of interest are isolated into each well of 96-well assay plate, with one cell in each well. Three weeks later, the grown colonies were studied by DNA extraction and sequencing of the gene of interest.

## 2.12. Polymerase Chain Reaction and Sequencing

Thermo Scientific Phire Tissue Direct PCR Master Mix (#F-170) was used to perform PCR directly from cultured cells with no prior DNA purification. The dilution and storage protocol was used. According to this protocol, cells were placed in a dilution buffer then a DNA release enzyme was added to improve the release of DNA from the tissues. Heating in 98 C

block for two minutes was performed to help releasing the DNA. This was performed before starting the PCR reaction as per kit protocol.

## 2.13. Sanger Sequencing Protocol

PCR products were diluted by 10 times (45µl H2O and 5µl DNA). A master mix with the forward primer was prepared according to the calculations mentioned below in Table (2.6), and another master mix with the reverse primer was also prepared to confirm sequencing.

**Table (2.6): The concentration of the reagents used for the big dye terminator sequencing Reaction**

| | |
|---|---|
| Primer Forward/Reverse 10mM (Sigma/Aldrich) | 1 µl |
| 5x Big Dye Terminator v1.1, v3.1 5x Sequencing Buffer(Applied Biosystems)(4336697) | 2 µl |
| Big Dye Terminator V3.1 Cycle Sequencing RR-100 (Applied Biosystems) (4337455) | 0.5 µl |
| Water (DNAse/RNAse-Free) | 5.5 µl |

In a 96 well plate, an aliquot of 9 µl of mastermix together with with1 µl of the diluted PCR product was added to each well. The plate was then sealed tightly with Starlab plate sealing plate and pulse spun in a plate centrifuge. A heating pad was then placed on the top of the plate before running the thermocycler: Denaturation: 96ºC for one minute, followed by 25 Cycles of: denaturation 96ºC for 10 seconds, annealing at 50ºC for 5 seconds, extension at 60ºC for 4 minutes and finally the thermocycler was set on 4ºC to store the samples.

PCR clean-up is essential to remove the excess PCR primers and dNTPs. This was done by adding 45µl of SAM solution (Applied Biosystems) and 10µl of x-terminator (Applied Biosystems) then the whole plate was placed in a shaker for 15 minutes. This was followed by spinning in a plate centrifuge at 1,000 RPM for 2 minutes resulting in a clear supernatant ready to get into the Applied Biosystems 3500 Dx / 3500xL Dx Genetic Analyzer. DNA sequencing was then analyzed using SnapGene 2.3.2 software.

## 2.14. Cloning of PCR Products

Following CRISPR/Cas9 gene editing, the targeted gene was PCR amplified using Phire Tissue Direct PCR Master mix from ThermoFisher according to the kit protocol. This was followed by cloning of the PCR product into a selection cloning vectorpJET1.2/blunt (ThermoFisher/ CloneJET PCR Cloning Kit) ). Firstly, a blunting reaction was performed on the PCR product before ligating it to the pJET1.2blunt cloning vector. This was then followed by bacterial transformation.

# Chapter 3

# Results

# Objective 1

## rRNA study in DBA cells

Ribosomes, the cellular machinery responsible for protein translation, are made of ribosomal proteins (RPs) and rRNA. RP are required as structural components of ribosomes. In addition, RPs are essential for the efficient processing and cleavage of rRNA in the nucleolus which involves converting a single RNA transcript to several smaller fragments. These fragments are then incorporated into the small or large ribosomal units. Any defect in the 40S and (or) 60S ribosomal proteins will interfere with the maturation of RNA component of either subunit and thus, disrupting the ribosomal biogenesis.(8) .

In this project the abundance and type of rRNA in DBA patients was analyzed and compared to normal controls to assist more efficiently in the diagnosis of DBA. Additionally, this assay can be used as a read out of our next objective, which is mutational validation using CRISPR/CAS9.

In this study, rRNA analysis was performed in two ways. In the first approach, RNA capillary electrophoresis was implemented to compare the ratio 32S/28S rRNA peak between DBA patients and controls. The second approach is to study rRNA by the aid of Quantitative real-time PCR with rRNA gene-based specific primers.

## 3.1.RNA Capillary Electrophoresis

In this study, peripheral blood mononuclear cells (PB-MNC) were isolated from patients' samples. T cells were selected using columns and magnetic beads. The purity of the T cell population was assessed by flowcytometry which confirmed CD3 positive cells were above 96%. Part of the T cells were used to extract and purify total RNA (200,000 cells). From part of the T cells collected, 250,000 T cells were grown by using an inert superparamagnetic bead which are similar in size to antigen-presenting cells and are covalently coupled to anti-CD3 and anti-CD28 antibodies. These two antibodies provide primary and co-stimulatory signals, optimized for efficient T cell activation and expansion. RNA was extracted at three different time points (day 0, 6 and 12). As we tried to study the effect of stimulating the cells on the rRNA profile, we hypothesized that stimulation will put the ribosomes under stress and will cause changes in the rRNA profile.

Figure (3.1) shows the three different rRNA species that were studied in normal control first (figure 3.1A). The profile in two different types of patients with DBA are also illustrated and compared to controls. Two different types of DBA patients were used as a comparison. One group of patients with the large 60S ribosomal gene subunit mutation (RPL), (Figure 3.1B), another group of patients with the small ribosomal gene 40S subunit mutation (RPS) (Figure 3.1C) and a third group of patients with non-identifiable ribosomal gene mutation. In general, this group of non-identifiable mutation showed a similar RNA capillary electrophoresis pattern to the RPL and RPS groups. The 28S/18S peaks are higher in DBA patients with RPS gene mutation than in controls who also showed a higher ratio than patients with RPL mutations.

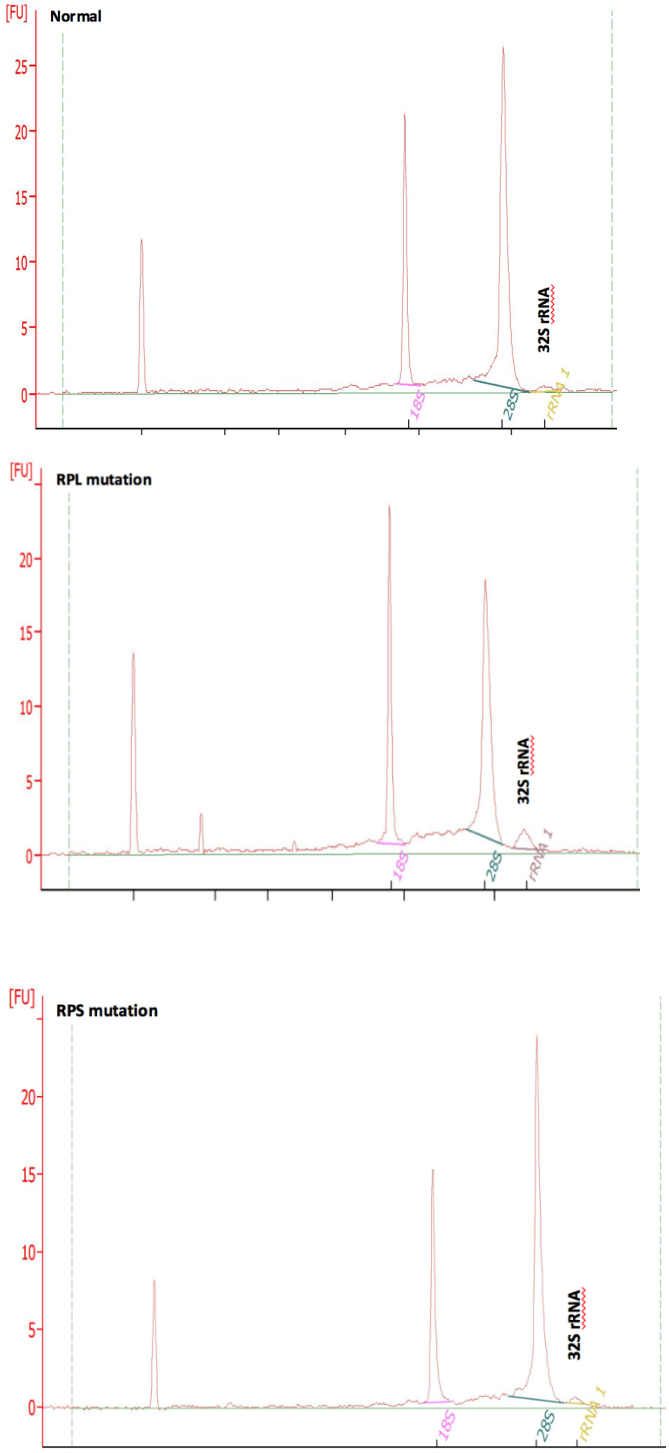


**Figure 3.1: Agilent Bioanalyzer 2100 electropherogram** obtained from total RNA of selected T cells from peripheral blood samples of a healthy control and from DBA patients. The bioanalyzer data showed 18S, 28S and a less prominent 32S peaks. The 32S peak in DBA patients with 60S abnormality is more prominent than in patients with 40S subunit alteration which is more prominent than in controls. The 28S/18S in DBA patients with RPS gene mutation was higher than in controls who also showed a higher ratio than patients with RPL mutations.s

Two biological replicates were run, there was a difference of +/- 0.1 in the 28S/18S rRNA. However, the overall outcome when comparing different groups was the same.

RIN of <9 was rejected.

The 32/28S rRNA ratio (Mean±SEM) detected and measured by 2100 Agilent Bioanalyzer in DBA patients is also compared to controls and illustrated in Figure 3.2

On day 0, the 32S/28S rRNA in resting T cells was not significantly different between DBA patients and the controls (*P:* NS). On day 6, however, there was a significant difference in the 32S/28S rRNA between the three different groups of DBA patients and the controls. This ratio was significantly higher (*P*=0.0001) in patients with RPL gene mutation (n=10) than the control (n=17). Similar but to a lesser extent, there was a significantly higher ratio in RPS mutation patients (n=10) compared to controls (*P*=0.02). Likewise, this ratio was significantly higher in DBA patients with no RP gene mutation (n=16) compared to controls (*P*=0.01). On day 12, the ratio of 32S/28S rRNA was significantly higher (*P*=0.004) in patients with RPL mutation (n=4) compared to controls (n=10). However, this ratio was not significantly different between DBA patients with RPS gene mutations (n=24) compared to controls (*P*: NS).

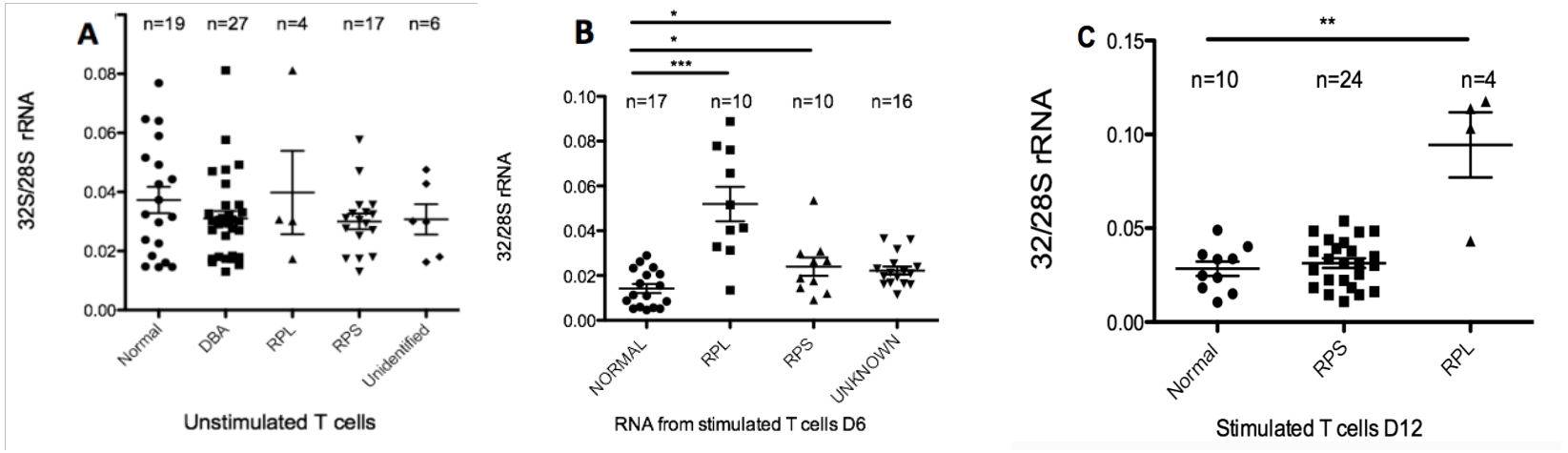


**Figure 3.2: 32/28S ratio (Mean±SEM) detected and measured by 2100 Agilent Bioanalyzer at three different time points following isolation and stimulation of T cells** (A) Day 0: No statistical difference between DBA patients and the control (*P*>0.05). (B) Day 6: After stimulating T cells with anti-CD3 Anti-CD28 activator, the 32S/28S ratio in DBA patients with RPL mutation was significantly higher than the control (*P* =0.0001)***. Additionally, the 32S/28S ratio was significantly higher in DBA patients with RPS mutation and in patients with undetected mutation (n=16) compared to controls (*P*=0.02 and 0.01, respectively). (C)Twelve days after T cell stimulation: there was a significant difference between DBA patients with RPL mutation and the control group (*P*=0.004)**. Group comparison was made with the Mann Whitney U-test. A p-value ≤0.05 was considered statistically significant. Statistical analyses were performed using Graph Pad prism version 5.0a

The 28S/18S rRNA ratio was also studied both in DBA patients and healthy controls on resting T cells and six days after stimulation. Results are illustrated in figure (3.3)

It was found that in DBA patients with RPL mutation (n=9), on day 6, the 28S/18S rRNA ratio was significantly lower than the controls (*P*=0.009). Additionally, the 28S/18S ratio was significantly higher in DBA patients who showed non-identifiable RP gene mutations (n=18) compared to control (*P*=0.04). Furthermore, in DBA patients with RPS mutation (n=10), the 28S/18S rRNA ratio was higher than in the control group (n=20); however, the difference was not significant (*P*=NS)

In resting T cells, this ratio was not significantly different between DBA patients and the control group. Figure 3.3 demonstrates a comparison of 28S/18S rRNA ratio among DBA patients and healthy controls.

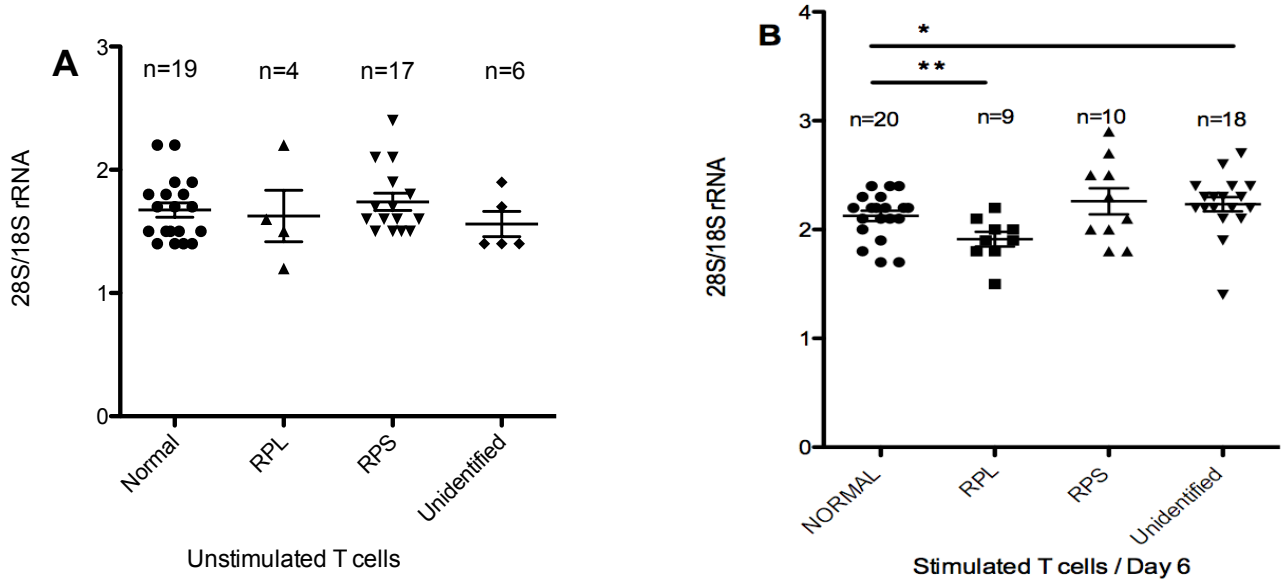


**Figure 3.3: 28S/18S rRNA ratio in pure selected T cells from DBA ptients and healthy controls, before stimulation and six days after stimulation.** RNA was analyzed using Agilent bioanalyzer 2100. (A)Resting T cells showed non-significant difference in the ratio of 28S/18S rRNA in DBA patients compared to control group. (B)On day 6, the 28/18S rRNA ratio was significantly lower in patients with RPL mutation compared to healthy controls (*P*=0.009). On the other hand, this ratio was significantly higher in DBA patients of non-identifiable RP gene mutation compared to control (*P*=0.04). Group comparison was made with Mann-Whitney *U*-test. A *P*-value ≤0.05 was considered statistically significant.

## 3.2.Quantitative real-time PCR

To increase the sensitivity of detection of rRNA and improve on the previous method, rRNA gene-based real time quantitative PCR using specific primers were designed to detect and quantify 18S, 28S, 32S, 5.8S and 5S rRNA and applied to both DBA patients and healthy controls.
Like the previous method, namely the capillary electrophoresis, peripheral blood mononuclear cells (PBMNC) were isolated from patients' samples. T cells were selected, stimulated and expanded. RNA was extracted at two different time points, day 0 and day 6 after stimulation.

**Selection of reliable reference genes for quantitative real-time PCR**

There has been a paucity of methods to accurately analyze rRNA in DBA patients especially in resting and stimulated T cells. Hence, we chose to investigate five housekeeping genes commonly used as internal controls in expression studies, reference genes include actin beta (ACTB), tata box binding (TBP), transferrin receptor (TFRC), importin 8 (IPO8) and ribosomal protein lateral stalk subunit P0 (RPLP0). The cycle threshold (Ct) value for each gene was determined and data obtained were analyzed using the software programs NormFinder. The most stable reference gene was selected using NormFinder software. This software ranks the set of candidate normalization genes according to their expression stability in a given sample set and in given experimental conditions. In this project, ten DBA samples were studied to assess the best reference gene before and after T cell stimulation. Firstly, PB-MNC were isolated from peripheral blood samples. Next, T cells were isolated followed by RNA extraction from resting T cells. In the next step, T cells were grown, and RNA was extracted from stimulated T cells on day 6.

This was followed by cDNA synthesis and quantitative real time PCR using the five different reference genes mentioned earlier.

By the aid of Normfinder software which demonstrates and compares the expression stability of the studied genes, where the lower stability value corresponds to the more reliable gene It was found that IPO8 is the most stable reference gene both in resting and stimulated T cells. Likewise, RPLP0 is also stable. Figure 3.4 illustrates the expression stability of these five different reference genes.

| Gene name | Stability value | | Best gene | rplp0 |
|---|---|---|---|---|
| rplp0 | 0.027 | | | IPO8 |
| tfrc | 0.196 | | | |
| ipo8 | 0.027 | | | |
| tbp | 0.352 | | | |
| actb | 0.769 | | | |

**Figure 3.4: Comparative expression stability of five different housekeeping genes.** Normfinder software demonstrates the expression stability where the lower stability value corresponds to the more reliable gene. This program compares the consistency (stability) of the CT values of a different house keeping genes in different samples. Because the concentration of cDNA is constant then the CT value for the reference gene should be the same in different samples. This software determines the most stable CT value of a reference gene when RT-qPCR is done in different samples. Normfinder showed the CT value of IPO8 and RPLP0 were the most stable when measured in different DBA patients. The concentration of cDNA is constant in all experiments.

**Ribosomal RNA Quantitative Real-time PCR**

We designed, ordered and tested rRNA specific probes to study the relative expression of 18S, 28S, 32S, and 5S rRNA. This test was performed on RNA extracted from resting and stimulated T cells isolated from DBA patients and healthy controls, 2^(delta Ct) was calculated according to this equation: 2^deltaCt= 2^(Ct(housekeeping gene)-Ct(gene of interest)). The different rRNA species were tested together with the 32S/28S rRNA which was measured as follows:

32S rRNA (2^deltaCt) / 28S rRNA (2^deltaCt).

### 3.2.1.QPCR in Resting, Unstimulated T cells

As shown in figure 3.5, the 18S rRNA detected by real time quantitative PCR was significantly lower in DBA patients (n=48) than the control group(n=13) (*P*=0.0004). Similarly, 28S rRNA was significantly lower in DBA patients as compared to the control group (*P*=0.03). Conversely, 32S/28S ratio was significantly higher in DBA patients compared to controls (*P*=0.019). The differences in 32S and 5S rRNA expression between DBA and control was non-significant (*P*>0.05).

Likewise, these findings were confirmed by repeating the qPCR experiment and quantifying the rRNA relative to another reference gene.

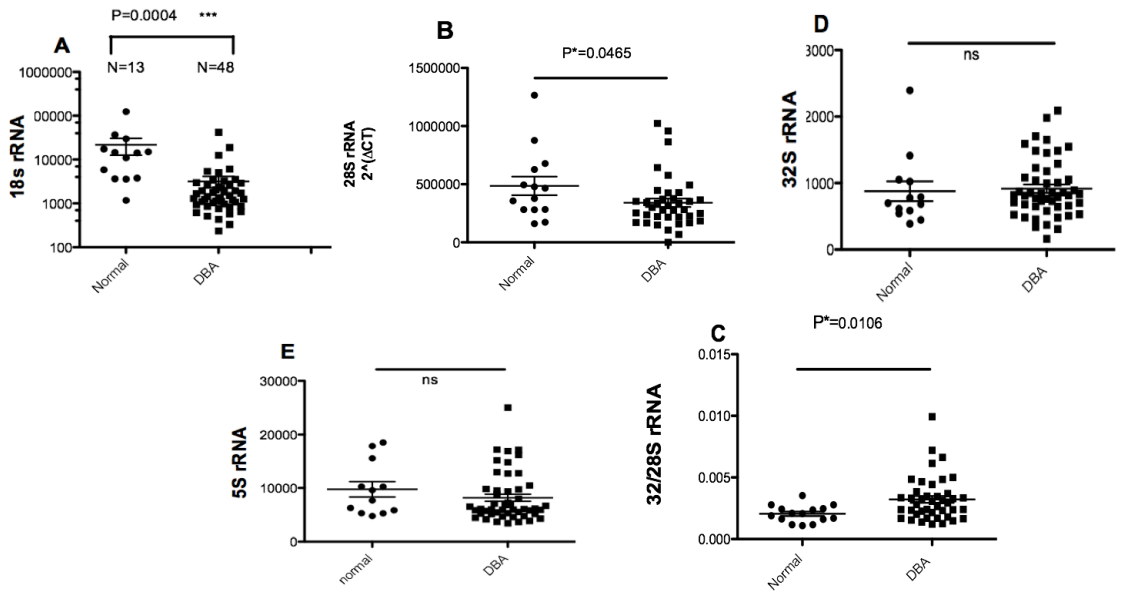


**Figure 3.5: Comparison of 18S, 28S, 32S, 32/28S and 5S rRNA gene expression among DBA patients and healthy controls.** cDNA was reverse transcribed from RNA extracted from non-stimulated, resting T cells. IPO8 was used as an internal control; and 2^(ΔCT) was calculated. The 18S, 28S rRNA was significantly lower in DBA patients if compared to controls (*P*=0.0004, *P*=0.03), respectively. 32S/28S rRNA ratio was significantly higher in DBA patients as compared to healthy controls (*P*=0.019). There was non-significant difference in 5S and 32S expression in DBA compared to the control group (*P*<0.05). Group comparison was made with Mann-Whitney *U*-test. A *P*-value of ≤0.05 was considered statistically significant. Data represent the average value of two technical replicates for each sample.

Additionally, rRNA gene expression in resting T cells was exploited in DBA patients with specific common mutations and compared to healthy controls. As shown in figure 3.6, it was found that there was a significant difference in the level of 18S rRNA between normal and DBA patients with *RPS19* (*P*=0.0004). However, there was no significant difference in the level of 28S, 32S, 32/28S and 5S rRNA between the patients and the controls (NS).

Furthermore, rRNA gene expression in resting T cells was compared between patients with *RPL11* mutation and healthy controls. As illustrated in figure 3.7, it was found that there was a significant difference in the level of 18S rRNA between normal and DBA patients with *RPL11* (*P*=0.004). However, there was no significant difference in the level of 28S, 32S, 32/28S and 5S rRNA between those patients and the controls. (*P*=NS).

In the same way, rRNA gene expression was compared among patients with *RPS26* mutation and healthy controls. As demonstrated in figure 3.8, it was found that there was a significant difference in the level of 18s rRNA between normal and DBA patients with *RPS26* (p=0.004). However, there was no significant difference in the level of 28S, 32S, 32/28S and 5S rRNA between patients with *RPS26* mutation and controls (*P*>0.05).

Moreover, rRNA gene expression was compared among patients with *RPS17* mutation and healthy controls. Figure 3.9 shows that there was a non-significant difference in the level of 18S, 28S, 32S, 32/28S and 5.8S between patients with *RPS17* and controls. (*P*>0.05).

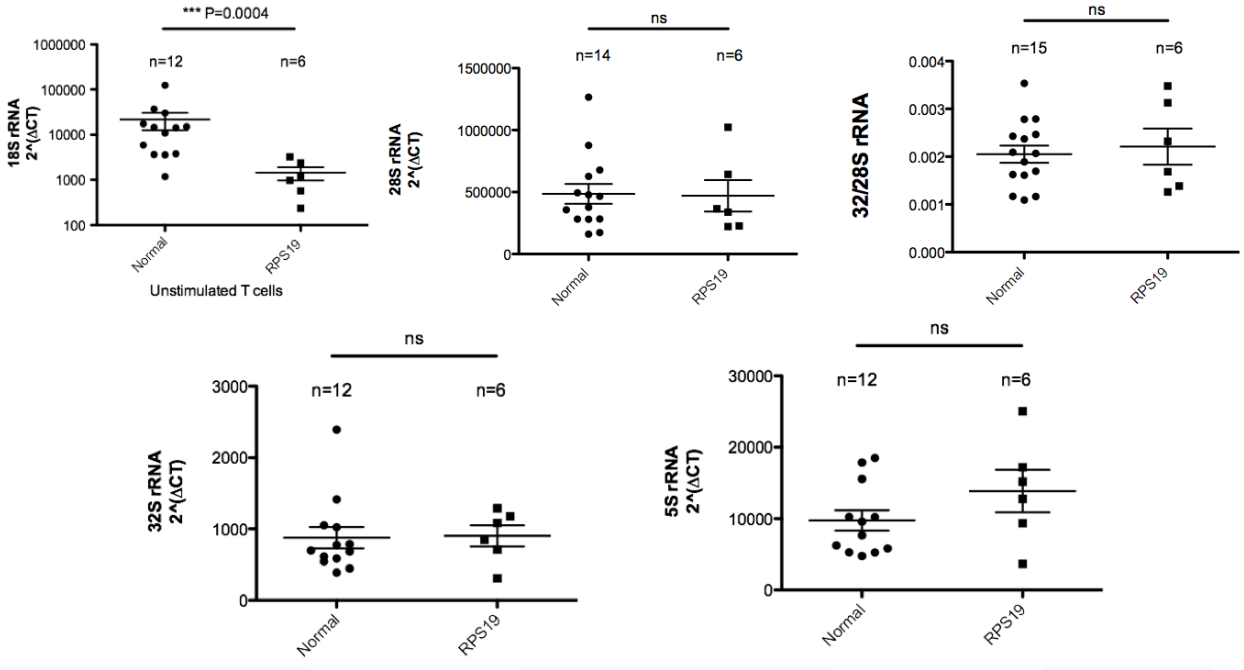


**Figure 3.6: Quantitative real time PCR comparing rRNA gene expression among DBA patients with *RPS19* mutation and control.**

**18S rRNA was significantly lower in patients with *RPS19* mutation when compared to healthy controls. Conversely, there was no significant difference in 28S, 32S, 32S/28S, 5S rRNA gene expression between *RPS19* patients and the control group. *P*<0.05. Group comparison was made with Mann-Whitney *U*-test. A *P*-value ≤0.05 was considered statistically significant.**

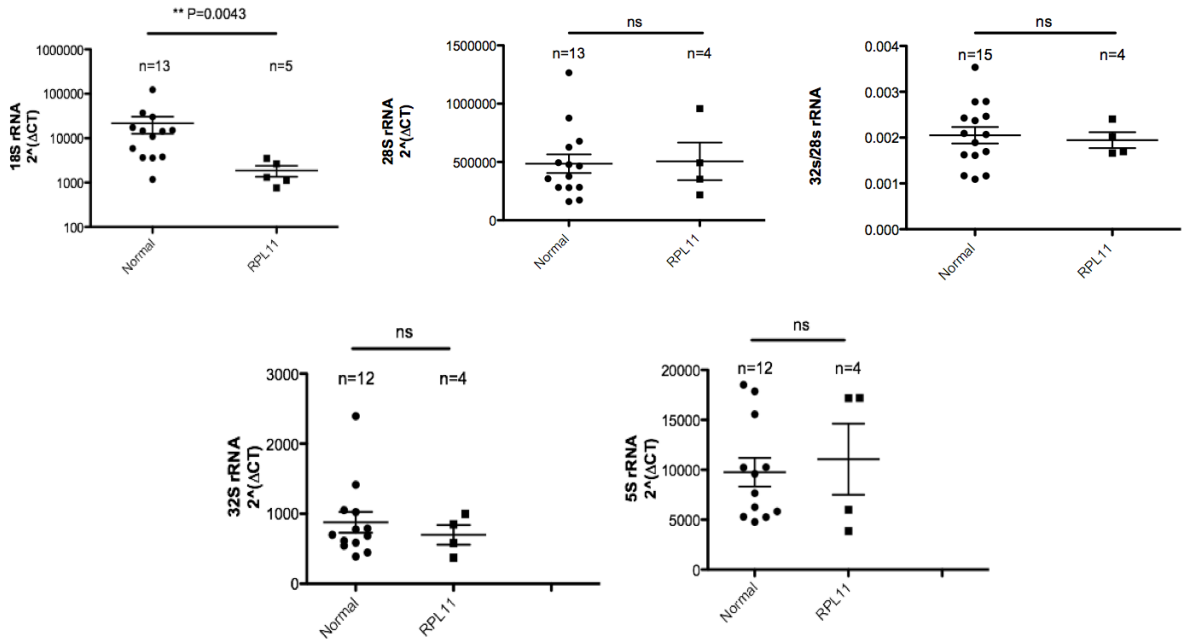


**Figure 3.7: Quantitative real time PCR comparing rRNA gene expression among DBA patients with RPL11 mutation and the control.** 18S rRNA was significantly lower in patients with *RPL11* mutation when compared to healthy controls. Conversely, there was no significant difference in 28S, 32S, 32S/28S, 5S rRNA gene expression between *RPL11* patients and the control group. $P<0.05$. Group comparison was made with Mann-Whitney *U*-test. A *P*-value ≤0.05 was considered statistically significant.

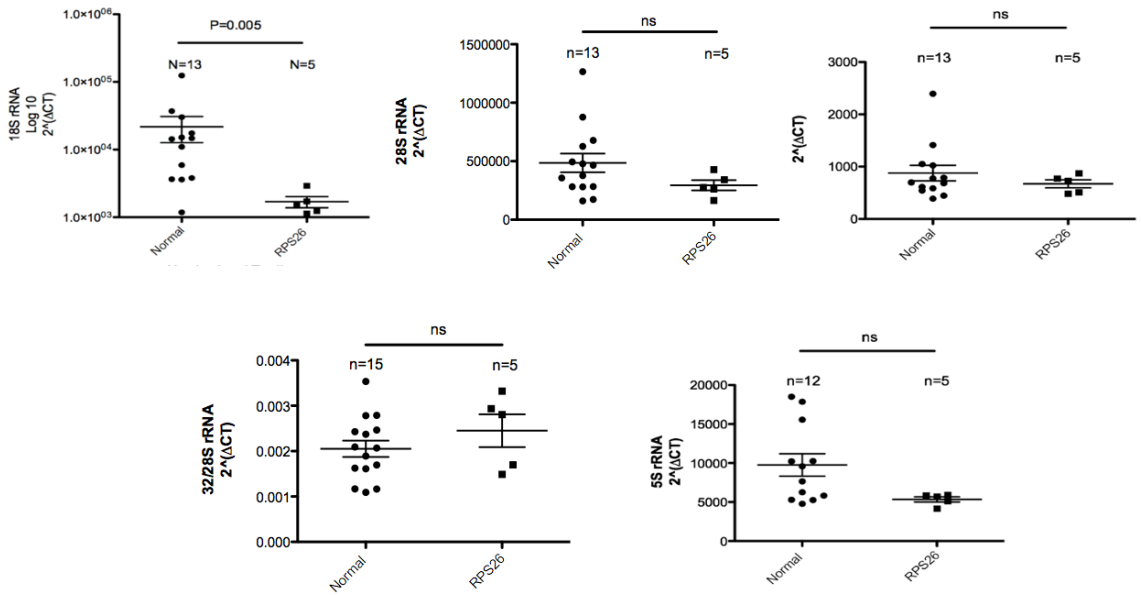


**Figure 3.8: Quantitative real time PCR comparing rRNA gene expression among DBA patients with RPS26 mutation and the control.** 18S rRNA was significantly lower in patients with *RPS26* mutation when compared to healthy controls. Conversely, there was no significant difference in 28S, 32S, 32S/28S, 5S rRNA gene expression between *RPS26* patients and the control group. (*P*>0.05). Group comparison was made with Mann-Whitney *U*-test. A *P*-value ≤0.05 was considered statistically significant

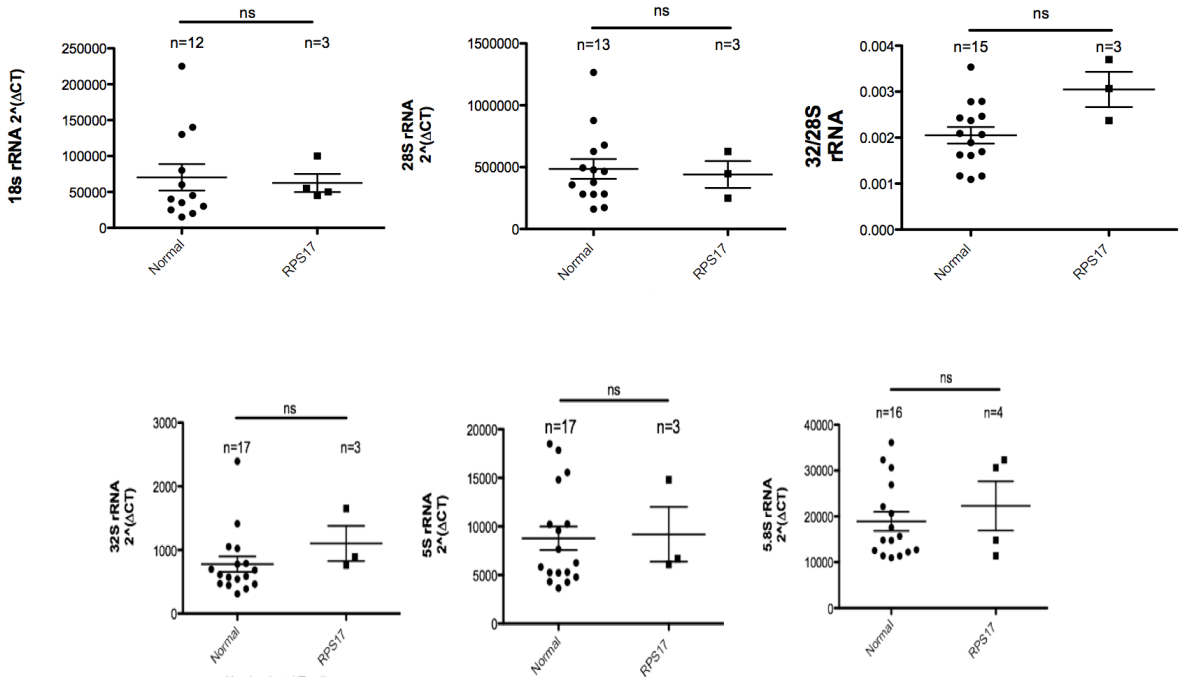


**Figure 3.9: Quantitative real time PCR comparing rRNA gene expression among DBA patients with RPS17 and control.** There was no significant difference in 18S, 28S, 32S, 32S/28S, 5S, 5.8S rRNA gene expression between normal and *RPS17* patients. $P<0.05$. Group comparison was made with Mann-Whitney *U*-test. A *P*-value ≤0.05 was considered statistically significant.

### 3.2.2.QPCR on Stimulated T cells, Day 6

Following stimulation of T cells using inert superparamagnetic bead which were covalently coupled to anti-CD3 and anti-CD28 antibodies namely Dyna beads, RNA was extracted on Day 6 both from DBA patients (n= 26) and the control group (n=8).

As illustrated in figure 3.10, Quantitative Real-time PCR showed that 18s sRNA was significantly higher in stimulated T cells from DBA compared to control (*P*=0.0003).

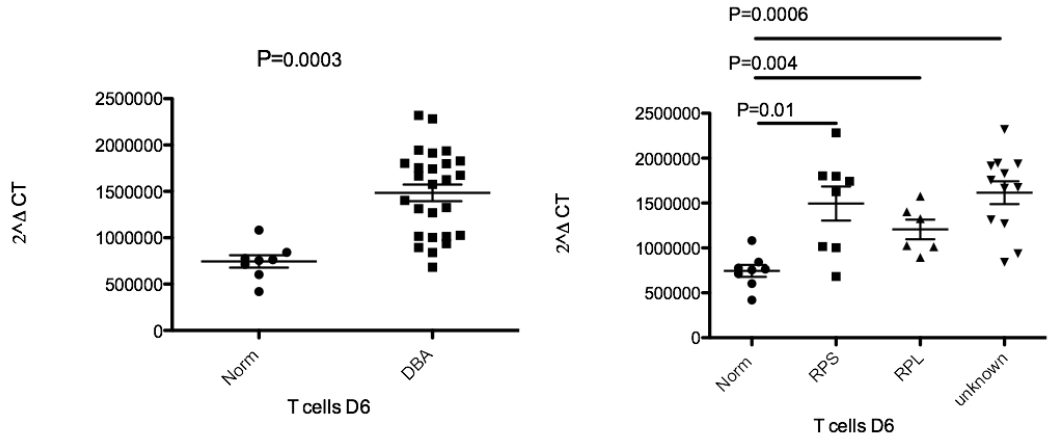


**Figure 3.10: Quantitative Real-time PCR** demonstrates that (A):18S rRNA was significantly higher in DBA patients compared to controls (p=0.0003). (B) The 18S rRNA in eight tested DBA samples with 40S abnormalities is significantly higher than in eight normal specimens (*P*=0.01). Similarly, the 18S rRNA was significantly higher in patients with 60S abnormality (n=6) when compared to controls (*P*=0.004). Patients with unidentified mutation in the RP genes also showed a significantly higher 18S rRNA relative to normal. (*P*=0.0006). Group comparison was made with Mann-Whitney *U*-test. A *P*-value ≤0.05 was considered statistically significant.

To summarise, after the detection of the most stable reference gene in resting and stimulated T cells from DBA patients, we studied rRNA defect in a cohort of fifty DBA patients by using our designed specific rRNA primers. We then compared the rRNA profile between DBA patients and the control and identified that the 18s rRNA, 28s rRNA and 32s/28s rRNA are significantly different in the rRNA profile in DBA patients compared to the control group as illustrated in the table below.

Additionally, stimulating T cells has made this difference more obvious as the 18S rRNA was significantly lower than normal then became significantly higher following stimulation.

This method together with CRISPR CAS 9 technology, electroporation, flow- cytometry, and timed cell-sorting was used to validate two novel heterozygous mutations discovered in our Lab, these are: *RPS17* c.3G>C and *RPL11* c.475_476 del AA. As discussed in the next chapter. We optimized a method to successfully introduce a mono allelic knock in mutation using CRISPR/Cas9 technology in k562 erythroleukemic cell line.

**Table (3.1): Summary of rRNA capillary electrophoresis and Q-RT- PCR analysis in DBA compared to control.**

| Using The Agilent bioanalyzer on day 6 | |
|---|---|
| 32S/28S rRNA | Significantly higher (p=0.0001) in RPL mutation compared to control |
| | Significantly higher (p=0.02) in RPS mutation compared to control |
| | Significantly higher (p=0.01) in unknown mutation compared to control |
| 28S/18S rRNA | Significantly lower (p=0.009) in RPL mutation compared to control |
| | Significantly higher (p=0.04) in unknown mutation compared to control |
| Q-RT-PCR studies on resting T cells | |
| 18S rRNA | Significantly lower (p=0.0004) in DBA compared to controls |
| 28S rRNA | Significantly lower (p=0.046) in DBA compared to controls |
| 32S/28S rRNA | Significantly higher (p=0.01) in DBA compared to controls |
| Q-RT-PCR studies on stimulated T cells, Day 6 | |
| 18S rRNA | Significantly higher (p=0.0003) in DBA compared to controls. |

# CHAPTER 3

# Objective 2

# Mutational Validation Using CRISPR/CAS9

In this project, forty-eight DBA patients were studied, their diagnosis was confirmed by the aid of NGS. Sequencing of 83 ribosomal protein genes in addition to GATA1, which was known to be involved in DBA showed that twenty-eight patients (58%) had a mutation in the RP genes, of which four were novel. In silico studies have confirmed pathogenicity.

To confirm the deleterious effects of some of these novel mutations, CRISPR/CAS 9 technology was implemented. Table 3.1 illustrates some of the novel mutations discovered by NGS.

To validate these mutations using CRISPR/Cas9, we successfully optimized a protocol to perform a monoallelic knock out to resemble the autosomal dominant mutations reported in DBA. This protocol was further optimized to carry out a monoallelic knock in. Two novel mutations discovered in our lab were successfully knocked in using the erythroleukemic cell line (K562 cell line).

First, we transduced HEK-293T cells with the CRISPR plasmid ligated to a guide RNA which is specific to the gene we are studying. This viral transduction technique resulted in biallelic knock out and further details are described below. To achieve a monoallelic knock out, we secondly tried HEK-293T and K562 cells transfection, but this had resulted in

biallelic knock out. This technique was further optimized until a monoallelic knock out and knock in were successfully introduced.

**Table 3.1: Some of the novel mutations discovered in this study**

| | | | | | |
|---|---|---|---|---|---|
| RPS17 initiation codon | c.3G>C; p.Met1Ile | Gene ID: 6218 NM_001021 | Non Familial | Spinal problems | Steroid responsive |
| RPL11 frame shift deletion | c.475_476 del AA (p.Lys159Argfs*11) | Gene ID: 6135 NC_000001 | Non Familial, | Patent Ductus Arteriosus hypoplastic thumbs; recurrent chest infections; Vitamin D deficiency | Steroid responsive |
| RPS29 Missense | c139G>C (P.Ala47Pro) | Gene ID: 6235 NG_050638 | Familial | No known congenital malformations | Transfusion dependent |
| RPS29 Splice site | c.63-3C>A | Gene ID: 6235 NG_050638 | Non Familial | No known congenital malformations | Transfusion dependent |

### *1-Viral transduction of HEK-293 T cells*

For each viral construct that was designed, HEK-293T cells were transfected with LentiCRISPRv2 which is a vector expressing GFP and CAS9 protein that can introduce a double stranded DNA break at the site of the gRNA that is cloned into this vector. Two different LentiCRISPRv2 plasmids were cloned each ligated to a specific gRNA. One gRNA was designed to target *RPS19* exon 5 while the second targeted *RPL11* exon 5. This resulted in two LentiCRISPRv2 plasmids into which two different gRNAs were cloned. From this, CRISPR plasmid lentivirus was produced. LentiCRISPRv2 plasmid is illustrated in the appendix.These viral constructs express CAS9 and GFP and can introduce double stranded break into *RPS19* and *RPL11*.

HEK-293T cells were transduced with three different viral vectors. The first one had a gRNA targeting *RPL11*, the second one had a gRNA that is specific to *RPS19* and the third one has a gRNA against *RPS19* but at a different location. Three days following transduction, cells were analyzed by flow cytometry for GFP expression to confirm viral transduction as illustrated in figure 3.11. This was followed by single cell sorting on GFP positive cells into 96 wells plates. It was observed that 60% of the single cells/clones produced colonies. On the recovered colonies, flow cytometry was performed and revealed variable degrees of GFP expression. Some clones lost their GFP expression while other clones showed a 100% GFP expression figure 3.12. The colonies with 100% stable GFP expression were further studied and sequenced to validate gene editing. Although Sanger sequencing revealed successful gene editing at the required site, this was the result of a biallelic modification/mutation rather than

monoallelic as shown in figure (3.12). Therefore, another method was used to introduce a monoallelic knock out.

CRISPR/Cas9 viral vectors were also transduced into Jurkat T cells by the same (RPL11, RPS19) viral constructs with the inclusion of polybrene as shown in figure 3.13.

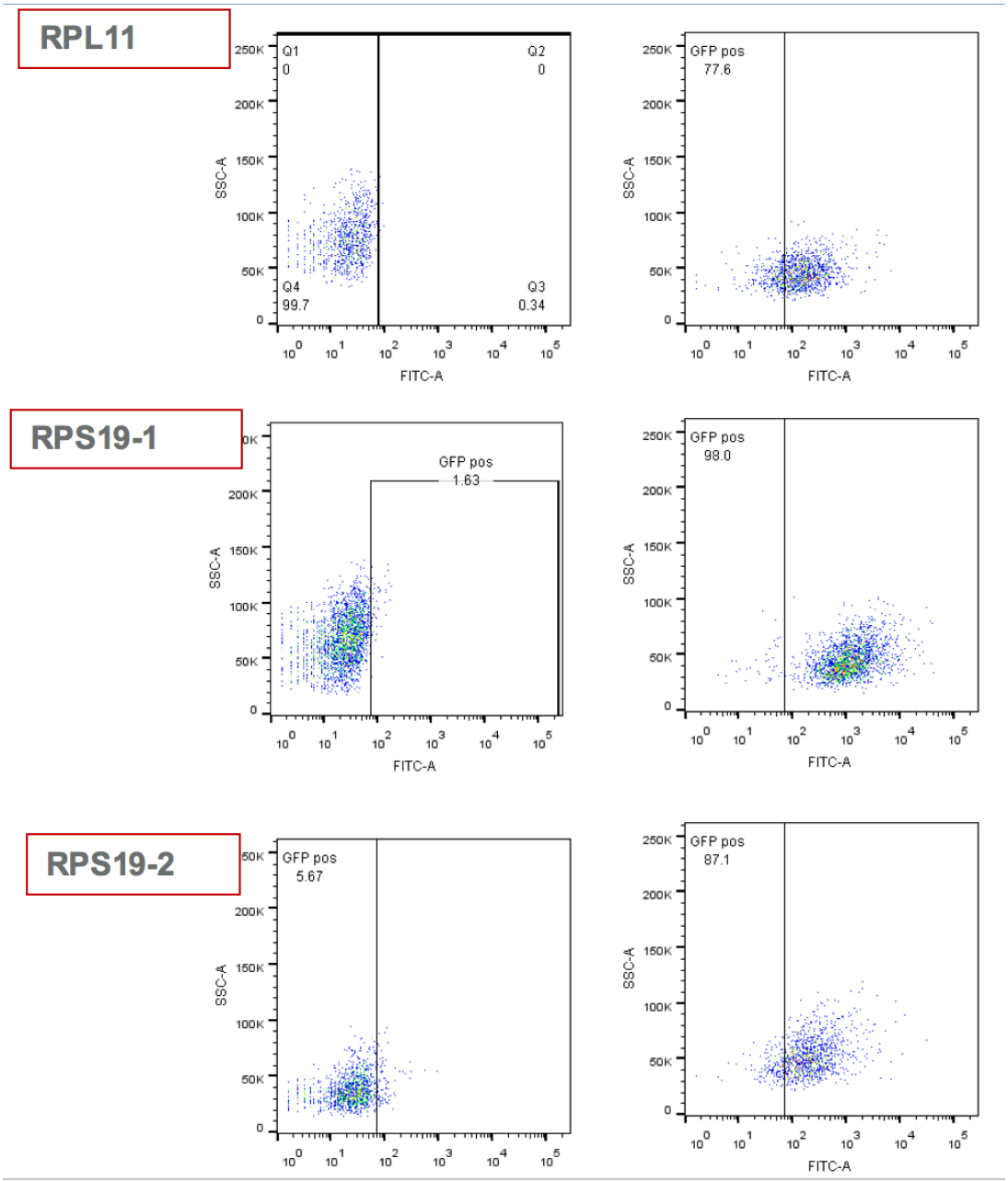


**Figure 3.11: Flow cytometry plots showing the transduction efficiency of the three different viral constructs in HEK 293T cells:** RPL11, the first RPS19 and the second RPS19 with a transduction efficiency of 77.6%, 98% and 87.1% respectively.

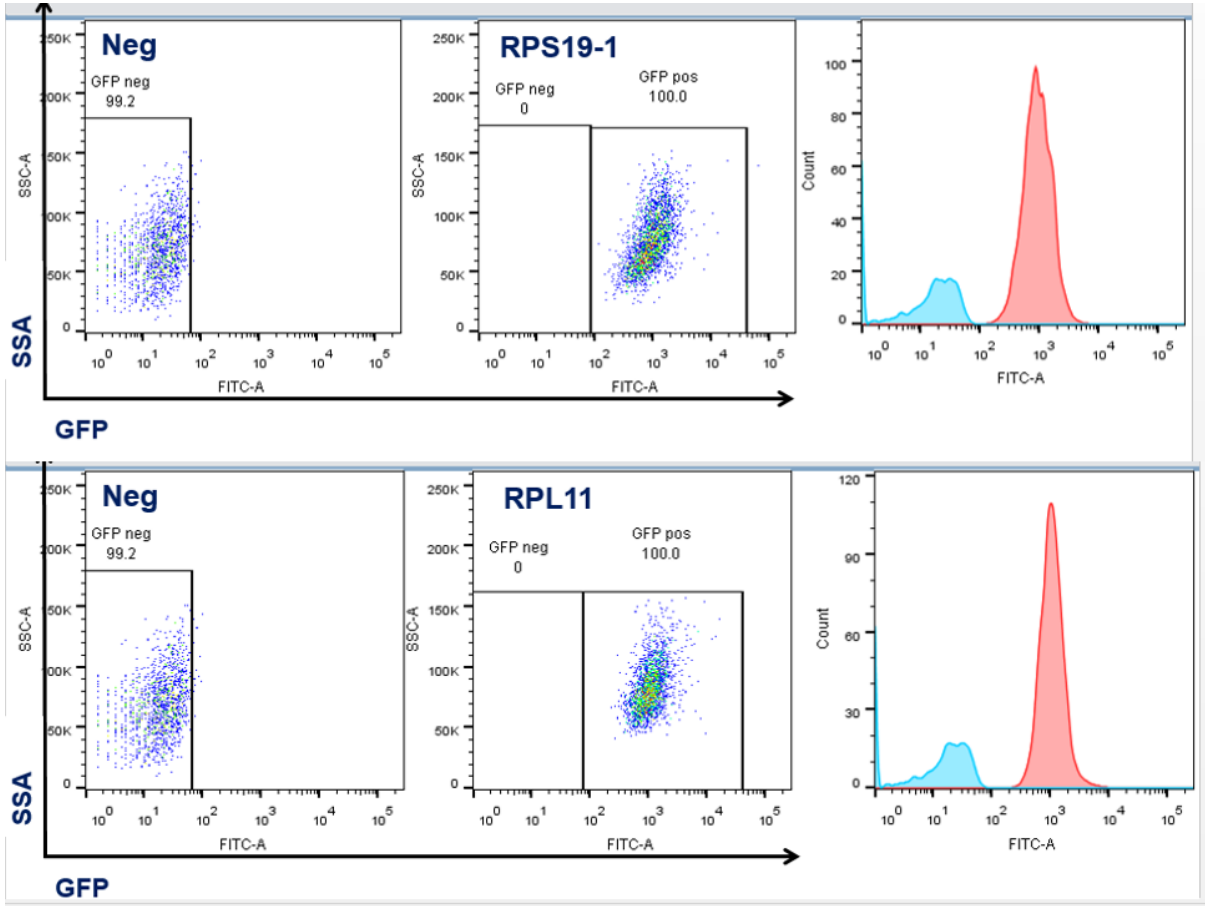


**Figure 3.12: (A) Two different HEK-293T clones following single cell sorting**. Showed 100% GFP expression were sequenced.


**Figure 3.12: (B) Sanger sequencing of RPL11 exon 5 in one clone of edited HEK-293T cells** demonstrating the sequence recognized by the gRNA in addition to biallelic gene editing compared to the wild sequence. Confirmation was done by aligning the edited sequence with the wild type sequence using snap gene software that can detect overlapping similar nucleotides. As demonstrated, there is a gap of three nucleotides CAG which are deleted from both alleles indicating biallelic gene editing. Additionally, this finding was further confirmed by designing complementary PCR primers with the last three bases at the 3' end of the primer complementary to the three deleted bases (CAG).This confirmatory PCR reaction did not result in PCR amplification

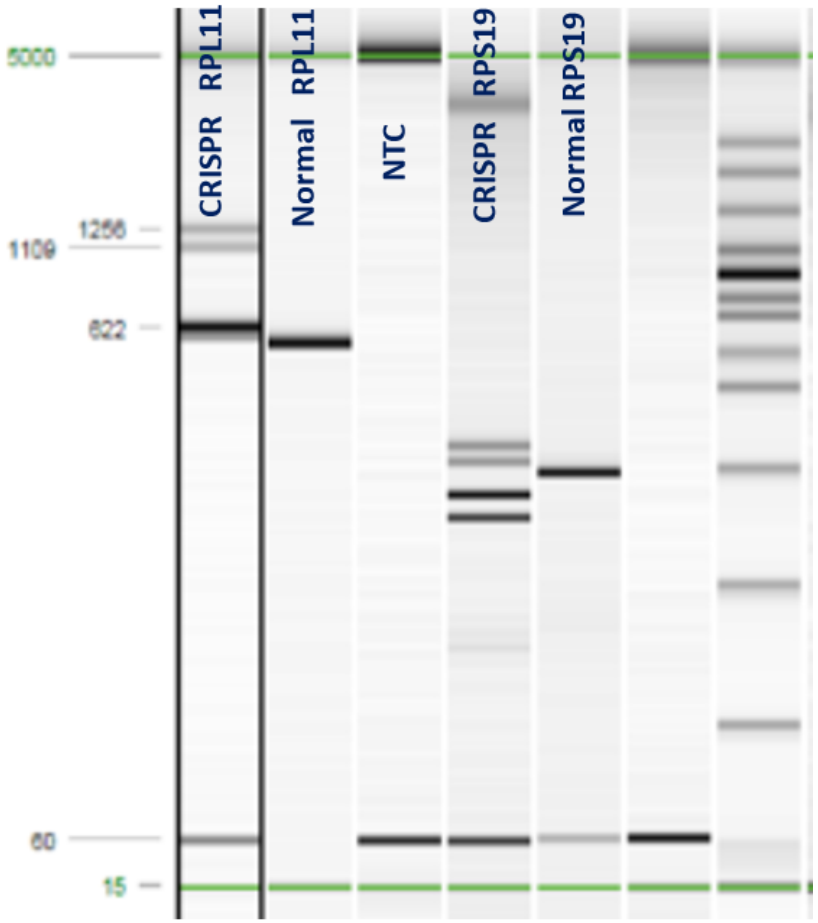


**Figure 3.12: (C) DNA capillary electrophoresis confirmed the existence of two bands following PCR amplification of the edited clones of HEK-293T RPL11 and RPS19 genes in the same clones shown in (A).**

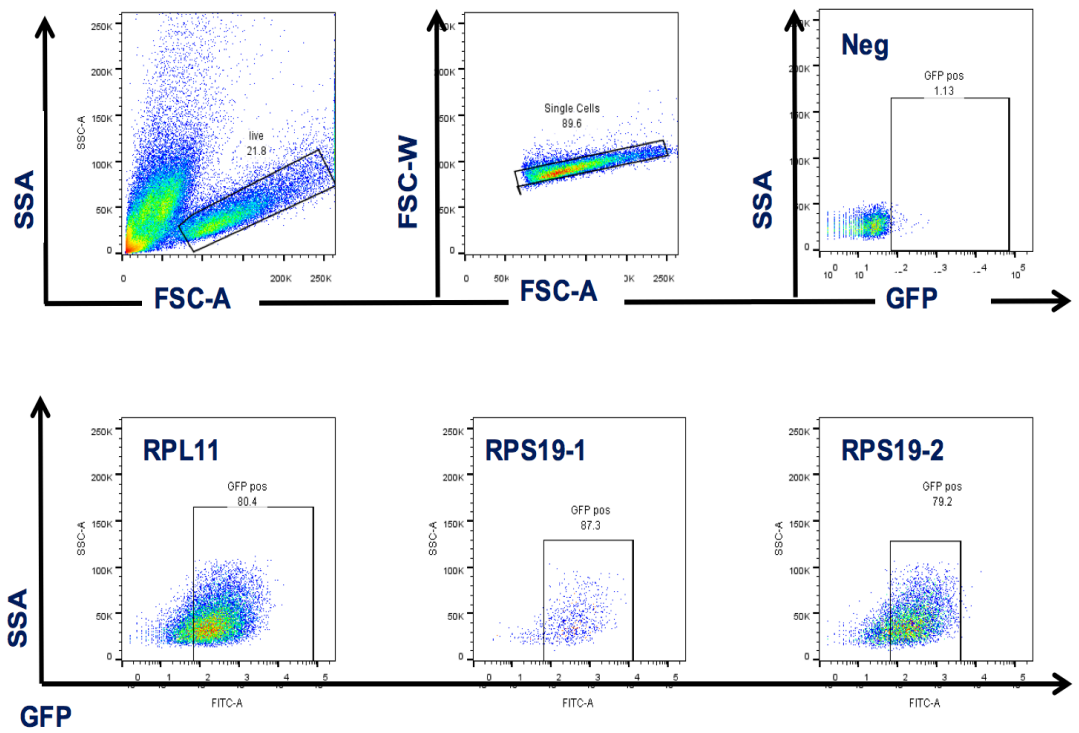


**Figure 3.13: The transduction efficiency of three different CRISPR/Cas9 viral vectors into Jurkat T cell line.** The transduction efficiency was 80.4%, 87.3% and 79.2% for *RPL11*, *RPS19-1* and *RPS19-2* respectively. RPS19-1 induces a DSB at exon 5 of *RPS19* while RPS19-2 induces DSB at exon 2

**2- CRISPR/CAS9 transfection with double sorting strategy**

To introduce mono allelic gene editing which is essential to validate DBA mutations (as DBA is inherited as autosomal dominant affecting RP genes), CRISPR/CAS9 transient transfection was tried and optimized.

Here, the specific gRNA was cloned into the LentiCRISPRv2 vector. Next, HEK-293T were transfected with the cloned LentiCRISPRv2. Twenty-four hours later the cells were sorted based on GFP expression detected by flow cytometry. Aiming to perform a transient transfection, GFP positive were double sorted after 48 hours gating on the cells with negative GFP expression. Finally, this population which was expressing GFP after 24 hours and had lost the GFP in the following 48 hours was sorted as single cells into 96 wells plate.

The resultant single clones were expanded in 37°C cell culture incubator supplemented with 5% $CO_2$, three weeks later 45% of the sorted clones were recovered and then each of the recovered clones was sequenced. This double sorting strategy resulted in none of the alive clones being edited. Figure 3.14 illustrates the technique performed and Sanger sequencing plots for some representative clones. We concluded that the technique of double sorting (i.e transfection of HEK 293T cells and sorting GFP expressing cells followed by sorting the GFP negative cells), was unsuccessful as none of the clones analyzed was edited.

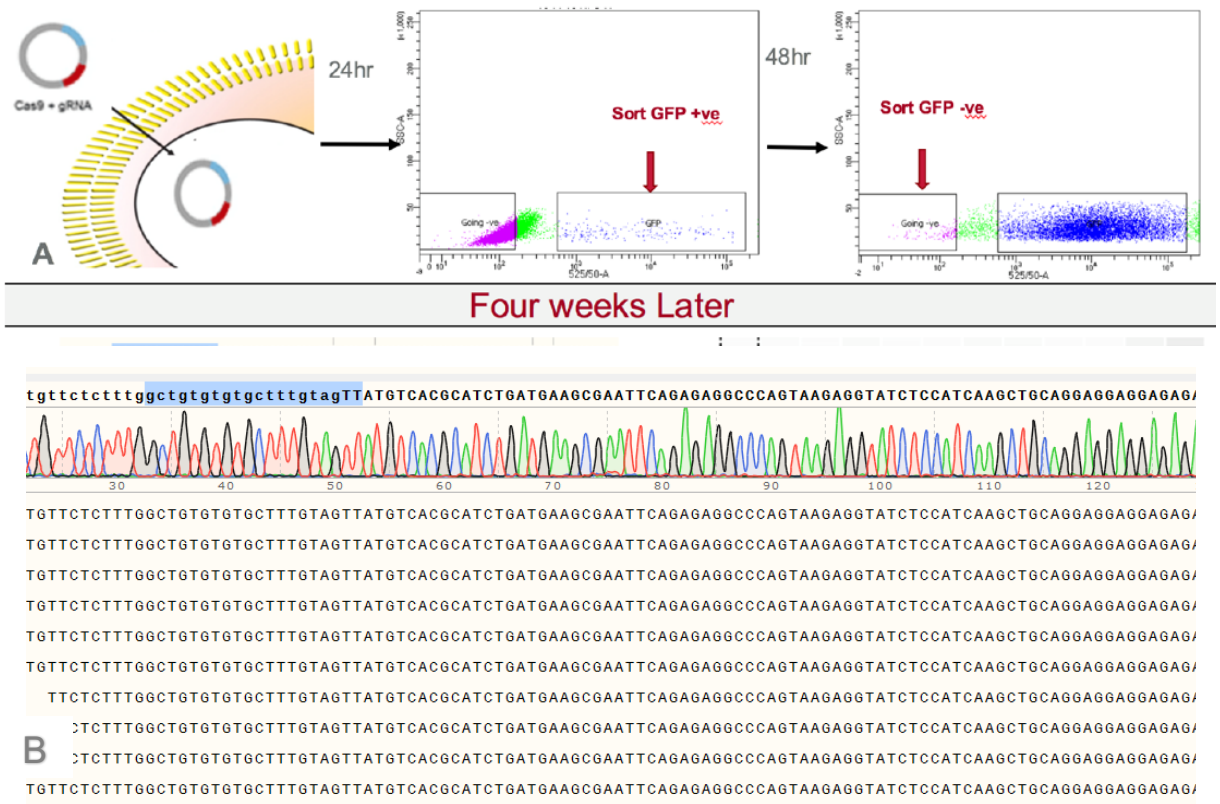


**Figure 3.14: (A) CRISPR transfection and double sorting technique. (B) Representative Sanger sequencing plots of eleven different clones four weeks after being sorted.** all showing wild type sequence, non-edited.

### 3- CRISPR/CAS9 transfection with timed single sorting strategy

To overcome the problem of either having both alleles cut or none, further optimizations were carried out.

HEK-293T cells were transfected with the cloned LentiCRISPRv2 plasmid. 72 hours later a single cell sorting, (on cells expressing low to intermediate level of GFP), was performed. Sanger sequencing was performed three weeks following the expansion of the sorted single clones in 37°C cell culture incubator supplemented with 5% $CO_2$. Among the 45% recovered clones, 40-60% showed monoallelic gene editing. Furthermore, biallelic editing was observed in 10% of the clones while the rest of the clones (40-50%) were non-edited.

In conclusion, the method of transfection and sorting after 72 hours for low to intermediate levels of GFP expression resulted in a successful introduction of a heterozygous knock out of HEK-293T cells as well as K562 cells. This technique was further optimized to introduce two successful knock in mutations. Figure 3.15 illustrates this technique.

This technique was further validated by performing colony PCR on the edited clones where the edited gene of interest was amplified by PCR specific primers. This was followed by ligation of this PCR product into a CloneJET PCR Cloning Kit (ThermoFisher) according to the manufacturer's protocol which was then transformed into recombination-impaired *Escherichia coli* strains (DH5α) bacteria.

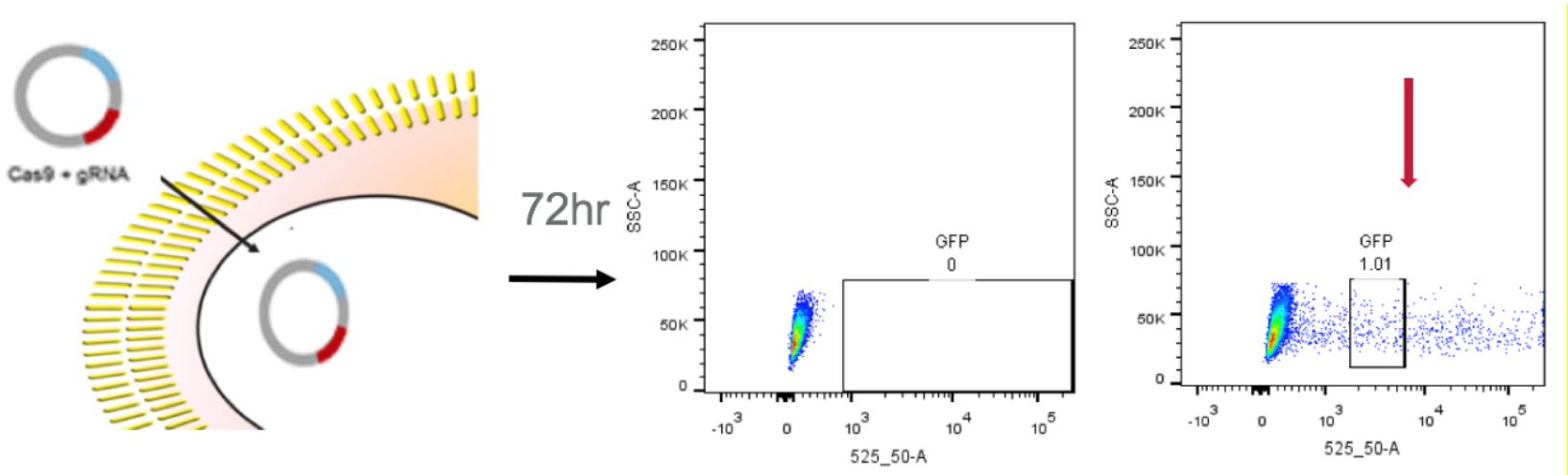


**Figure 3.15: Transfection and sorting strategy to introduce monoallelic gene editing by CRISPR/CAS9 technology.**

The bacteria was spread onto a selection plate and incubated overnight at 37°C. Next, twenty colonies were picked from a freshly streaked selective plate and inoculated in 50 ul of LB medium supplemented with ampicillin and incubated for 30 minutes at 37°C. Next PCR for the cloned plasmid was performed on each colony using the primers supplied with the CloneJET PCR Cloning Kit. Sanger sequencing was then carried out followed by Sanger sequencing using the big dye terminator BigDye™ Terminator v3.1 Sequencing Kit (ThermoFisher) according to the protocol mentioned in the materials and methods section. Cloning, colony PCR and sequencing confirmed the presence of two different alleles, the wild type and the edited allele. Figures 3.16 and 3.17 illustrate Sanger sequencing plots following cloning and colony PCR for some of the clones that were knocked out using the methodology mentioned above.


**Figure 3.16: Sanger sequencing of two bacterial clones following cloning and colony PCR of one RPS17 knocked out clone.** The upper plot is the knocked-out allele showing deletion of 33bp including the initiation codon ATG. The lower plot is the wild type allele.


**Figure 3.17: Sanger sequencing of two bacterial clones following cloning and colony PCR of one RPL11 knocked out clone.** The upper sequence is the knocked-out allele showing deletion of 13bp. The lower sequence is the wild type allele.

Wild: cagc aaa gag-gag-gcc-atg-cgc-tgg-ttc-cag-cag-aag-gta-aag-ctga.

Wild type Amino acids: KEEAMRWFQQKYDGIILPGK

Knocked out: cagc aaa-gag-gag-gcc-atg-cgc-tgg-ttc-cag-cag-aag-gta .The resulting amino acids: KEEAMRR* c.491-503 del gctggttccagca. p.Trp165Argfs*2

### *RPL11* and *RPS17* knock out in HEK-293T cells.

We have previously observed and reported above (aim 1) that the rRNA profile was significantly different between DBA patients and the control group. We decided to apply this method to validate the deleterious effect of some novel mutations tested in this study.

HEK-293 T cells were transfected by calcium phosphate method then 72 hours later the cells expressing GFP were sorted as single cells into 96-well plates.

Three weeks after growing these single clones, 60% were viable. Viable clones were analyzed by sequencing *RPS17*. Thirty percent of the clones showed monoallelic *RPS17* editing, 30% of the clones showed biallelic gene editing and the remaining 40% yielded only wild type sequence.

Using the same method *RPL11* was knocked out in HEK-293T cells as described below.

### *RPL11* knock out in HEK-293T cells

Using the method of transfection and timed cell sorting, *RPL11* was knocked out in HEK-293T cells. After single cell sorting 8 clones out of 192 survived. Sequencing RPL11 in these 8 clones showed 5 clones to be wild type (non-edited) and 3 clones showed monoallelic cut. Furthermore, two knocked out clones were assessed for the functional impact of this gene was using qPCR for 18S rRNA which was reduced together with the relative expression of RPL11 mRNA compared to control as demonstrated in figure 3.18.

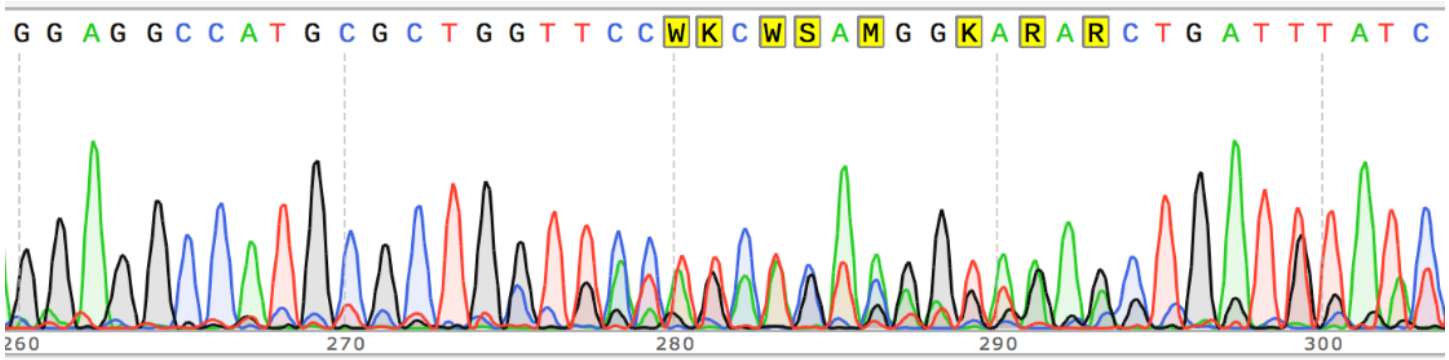


**Figure 3.18(A): Sanger sequencing plot confirming *RPL11* monoallelic knock out with deletion of 19 bases including the splice donor site at the exon-intron junction.**

Wild allele: atc-agc-aaa-gag-gag-gcc-atg-cgc-tgg-ttc-cag-cag-aag-gta-aag ctgatttatc. Wild type amino acids: KEEAMRWFQQKYDGIILPGK

Knocked out allele: aatcagc aaa gag-gag-gcc-atg-cgc-tg----Deletion--ctga.

c.del 495-507 gttccagcagaaggtaaag (p.Trp165CysfsTer1). Amino acid sequence is: KEEAMRW*

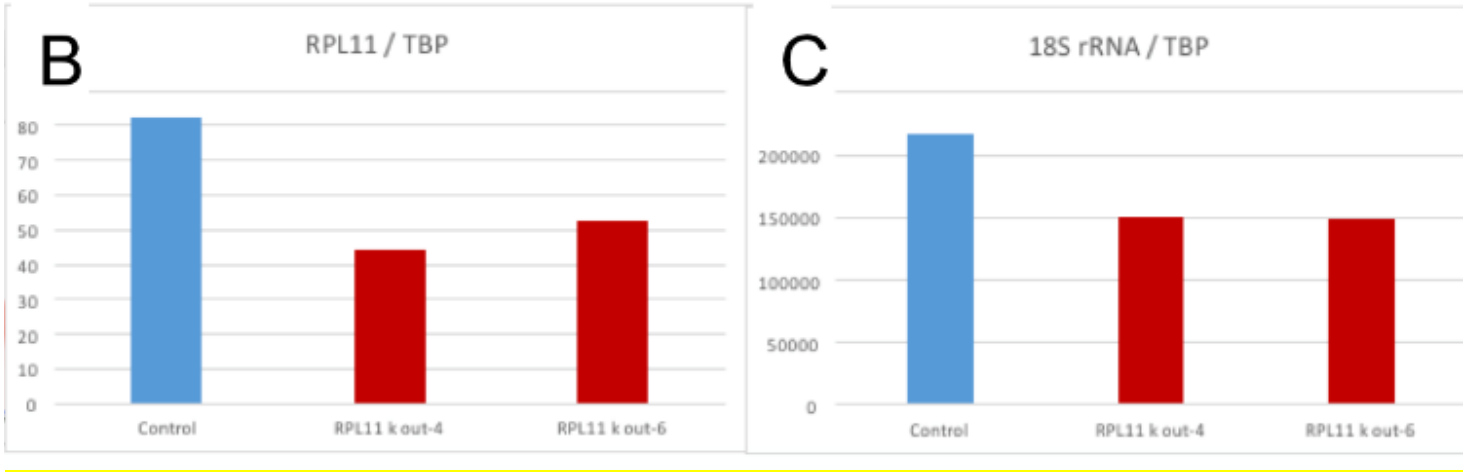


**Figure 3.18(B): qPCR studies applied on two RPL11 knocked out clones of HEK-293T.** It shows reduction in the level of *RPL11* mRNA as well as the level of 18S rRNA compared to control (n=2 experiments per clone). *TBP* was used as an internal control.

### *RPS17* knock out in HEK-293T cells

In the same way, in vitro knock out of the *RPS19* expression disturbed pre-ribosomal RNA processing. Using the same transfection and sorting technique *RPS17* was knocked out in HEK-293T cells. Single HEK-293T cells expressing GFP were sorted. After 3 weeks 10 clones out of 96 were proliferating. These ten clones were sequenced for *RPS17.* Sanger sequencing showed 5 clones to be wild type (non-edited), two clones showed homozygous mutation and 3 clones showed heterozygous mutation. Furthermore, pathogenicity of *RPS17* was studied using qPCR for 18S rRNA which was increased compared to control (as shown in figure 3.19).


**Figure 3.19 (A): Sanger sequencing plot applied on two *RPS17* knocked out clones of HEK-293T**.Sanger sequencing plots confirms *RPS17* monoallelic knock out in two different clones of HEK-293.(The upper and middle sequences). In both clones there is a monoalllelic loss of the translation initiation codon in addition to the presence of a normal wild type sequence in the background as compared to the wild type sequence which is the lower chromatogram

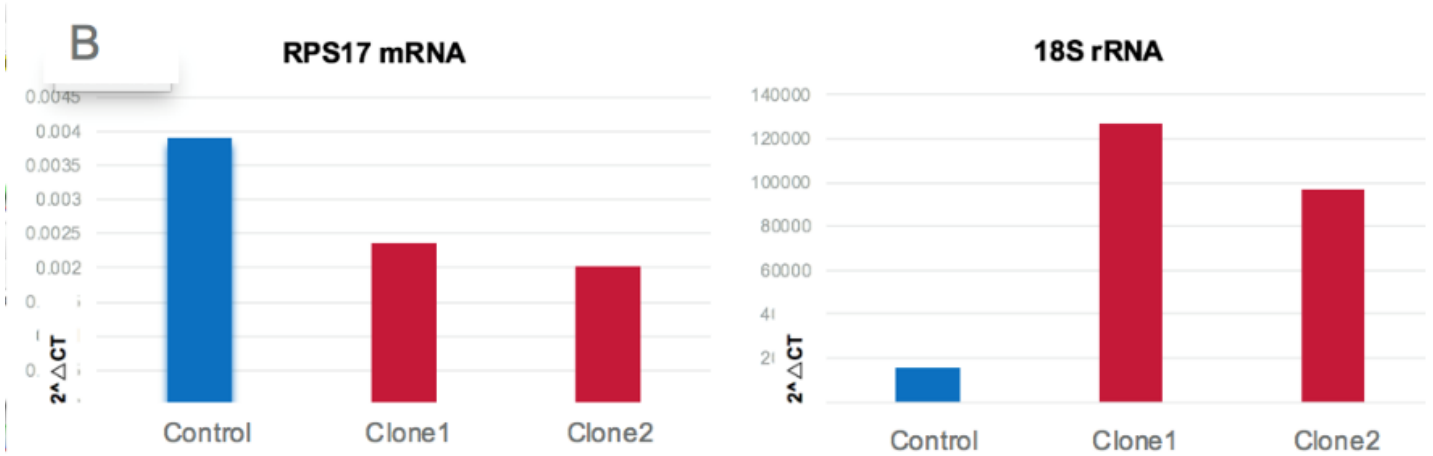


**Figure 3.19 (B): q-RT-PCR studies applied on two RPS17 knocked out clones of HEK-293T.** q-RT-PCR demonstrates reduction in the level of *RPS17* mRNA. There was an increase in 18S rRNA in two clones of HEK-293T cells compared to control (n=2 experiments per clone)

### *RPL11* heterozygous knock out of K562 cells

To validate certain novel mutations in DBA, RPL11 c.475_476 del AA**,** in vitro knock out of the *RPL11* gene followed by an in vitro knock-in was introduced. To confirm the role of *RPL11* in DBA development, gene expression was studied together with pre-ribosomal RNA processing which was disturbed in the edited cells.

Firstly, K5662 cells were transfected with LentiCRISPRv2 plasmid which carries an *RPL11* specific gRNA ligated to it. Next, Amaxa's nucleofection system of K562 cells was performed. Seventy-two hours later, single GFP expressing K562 cells were sorted into 96 wells plates. After three weeks of expanding these single clones, they were analyzed, and the viability was shown to be 30%. Sanger sequencing was carried out to confirm the presence of heterozygous gene editing. It showed three clones to be wild type (non-edited), five clones were showing heterozygous mutation and two clones were showing homozygous mutation.

Furthermore, the pathogenicity of *RPL11* was studied using qPCR for 18S, 28S, 32S, and 5S rRNA which were deranged compared to control (as shown in figure 3.20)


**Figure 3.20 (A): Sanger sequencing of one RPL11 knocked out clone of K562 cell line. sequence of RPL11, ACC. NG_011741. It** demonstrates the monoallelic *RPL11* knocked out allele and the wild type.

Wild: Gcc-atg-cgc-tgg-ttc-cag-cag-aag-gta-aag-ctg-att-tat-ctc-aa

Knocked out: Gcc-atg-cgc-tgg-ttc- cagcagaaggtaa agc-tga (**stop**)

The edited allel will result in frame shift deletion of 13 bases including the splice donor. C.167-179 del CAGCAGAAGGTAA (p.Trp165Serfs*2): amino acid sequence is: KEEAMRWFS*

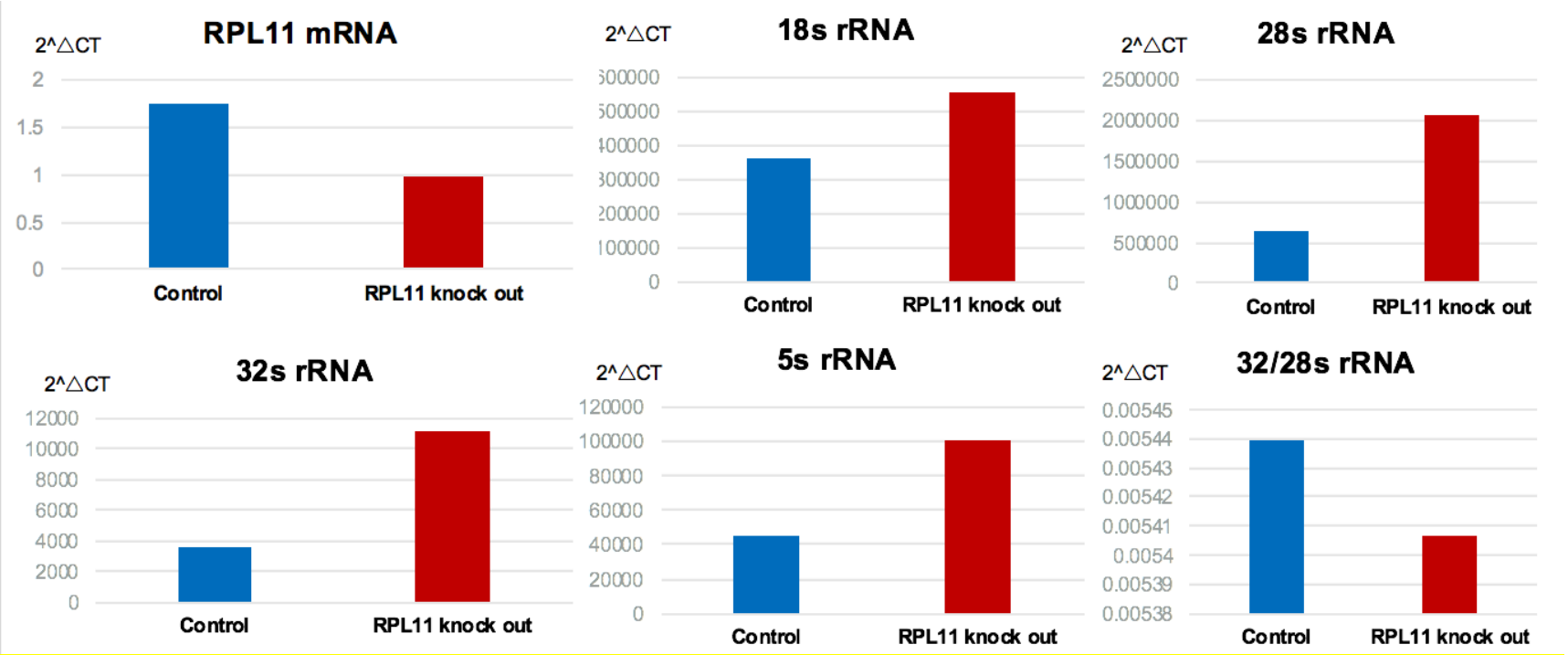


**Figure 3.20 (B): q-RT-PCR studies applied on one *RPL11* knocked out clone of K562 cell line.**

It demonstrates reduction in the level of RPL11 mRNA and increased levels of 18S, 28S, 32S, 5S rRNA compared to control with a decrease in the 32S/28S rRNA ratio. (n=2 experiments per clone).

### *RPS 17 heterozygous knock* out of K562 cells

To validate certain novel mutations in DBA, RPS17 c.3 G>C**,** in vitro knock out of the *RPS17* gene in K562 cell line (in addition to knock in of this mutation in particular) was successfully introduced. To confirm the role of *RPS17* in the pathogenicity of the disease, *RPS17* gene expression was studied together with pre-ribosomal RNA processing in the edited cells.

First, K562 cells were transfected with LentiCRISPRv2 plasmid that has an *RPS17* specific gRNA ligated to it.

Next, this plasmid was transfected into K562 cells by Amaxa's nucleofection system. Seventy-two hours later, single GFP expressing K562 cells were sorted into 96 wells plates. After three weeks of growing these single clones, they were analyzed, and the viability was shown to be 55%. The recovered clones were sequenced to confirm the presence of heterozygous gene editing. It showed 30% of the clones to be wild type (non-edited), 35% of the clones showed heterozygous mutation and the remaining 35% showed homozygous mutation. Furthermore, the pathogenicity of *RPS17* was studied using qPCR for 18S, 28S, 32S, and 5S rRNA which were deranged compared to control as shown in figure 3.21.

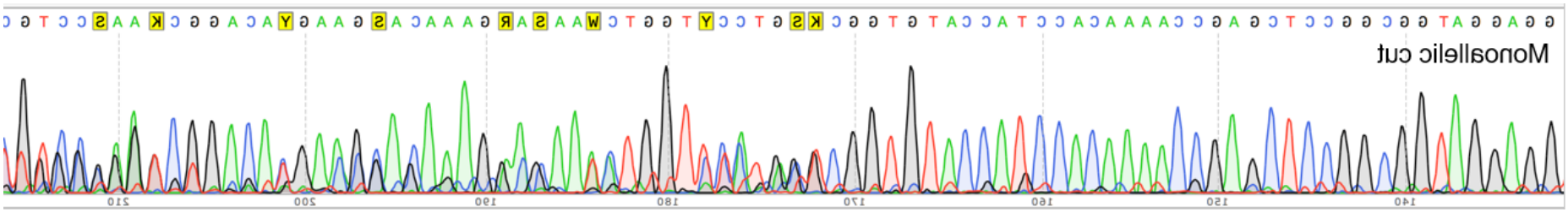


**Figure 3.21(A): Sanger sequencing plot applied on one RPS17 knocked out clone of K562 cell line**. Sanger sequencing plot confirming *RPS 17* monoallelic knock out in one clone of K562 cells which resulted in deletion of the translation initiation codon.

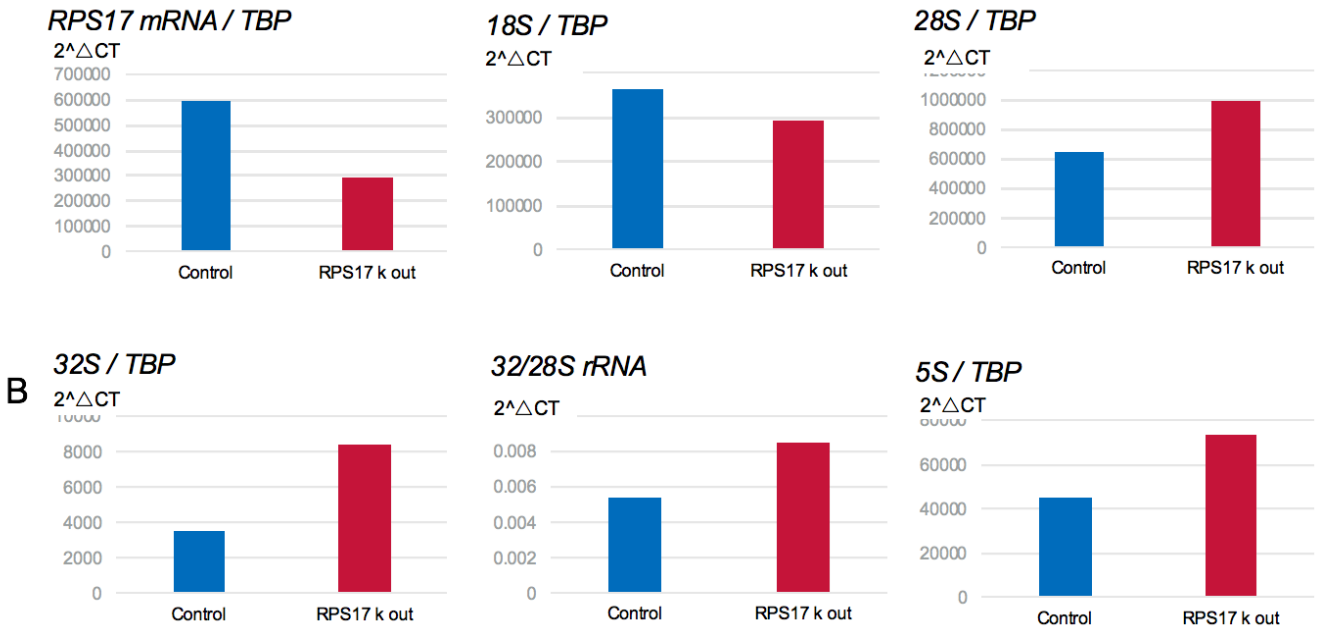


**Figure (3.21)(B):Q-RT-PCR studies applied on one RPS17 knocked out clone of K562 cell line.** It demonstrates reduction in the level of RPS17 mRNA and an increased level of 28S, 32S, 5S rRNA compared to control with an increase in the 32S/28S rRNA ratio. (n=2)

### *RPS29 heterozygous* knock out of K562 cells

To validate certain novel mutations in DBA, a missense *RPS29 c.139G>A*, P. Ala 47 Pro**,** in vitro knock out and knock in of the *RPS29* gene in K562 cell line were carried out. To confirm the role of *RPS29* in the pathogenicity of DBA, *RPS29* gene expression was studied together with pre-ribosomal RNA processing in the edited cells.

First, K562 cells were transfected with LentiCRISPRv2 plasmid that has an *RPS29* specific gRNA ligated to it. The structure of the plasmid is illustrated in the appendix. Next, this plasmid was transfected into K562 cells together with the homology directed repair template by Amaxa's nucleofection system and 72 hours later, single GFP expressing K562 cells were sorted into 96 wells plates. After three weeks of growing these single clones, 60% of these recovered clones were viable. These viable clones were analyzed. On the recovered viable clones, Sanger sequencing was performed to confirm the presence of heterozygous gene editing. Sanger sequencing revealed 30% of the clones to be wild type (non-edited), while 35% of the clones showed heterozygous mutation and the remaining 35% showed homozygous mutation. None of the clones showed a successful knock in of the desired *RPS29* c.139G>A. Furthermore, the pathogenicity of *RPS29* was studied using qPCR for 18S, 28S, 32S, and 5S rRNA which were deranged compared to control as shown in figures 3.22and 3.23

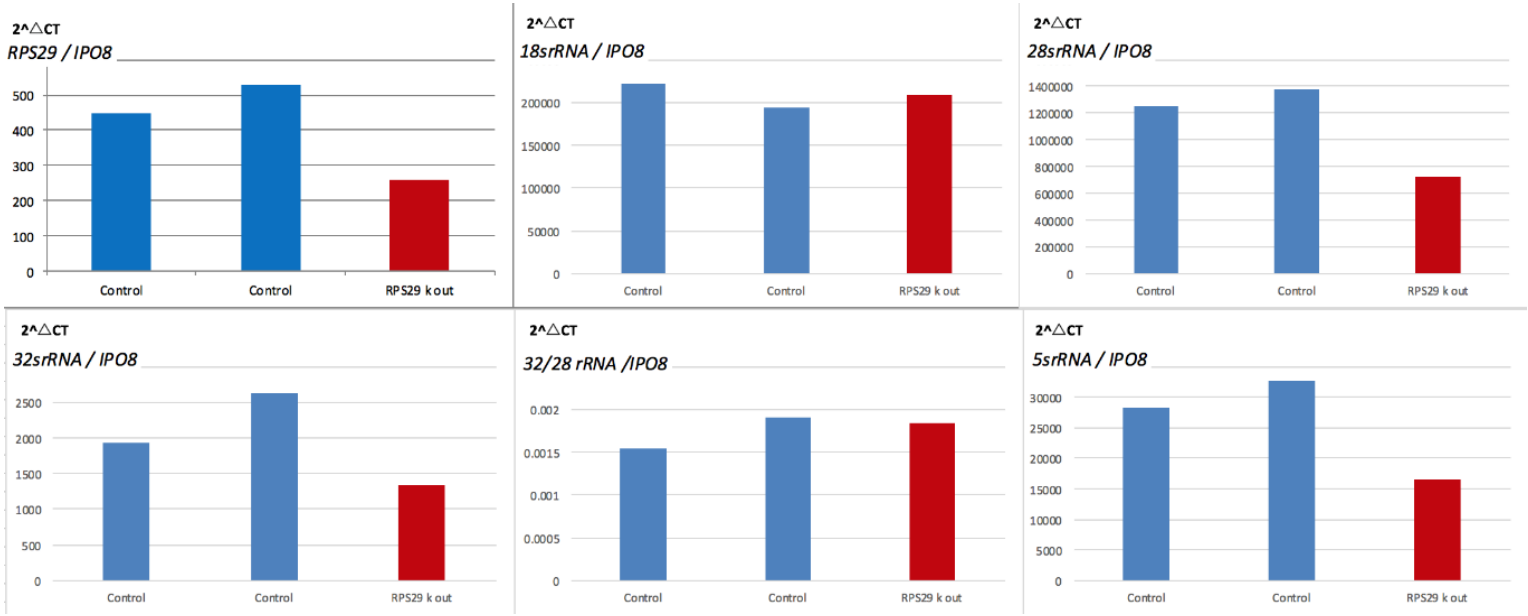


**Figure 3.22: Quantitative PCR performed on K562 RPS29 knocked out clone.** qPCR demonstrates reduction in the level of RPS29 mRNA clone. Additionally, there is a detectable decrease in the levels of 28S, 32S and 5S rRNA in the edited clone compared to the control. The control is a k562 cell clone that was transfected with an empty CRISPR vector without a ligated gRNA. IPO8 was used as a reference gene. (n=2)

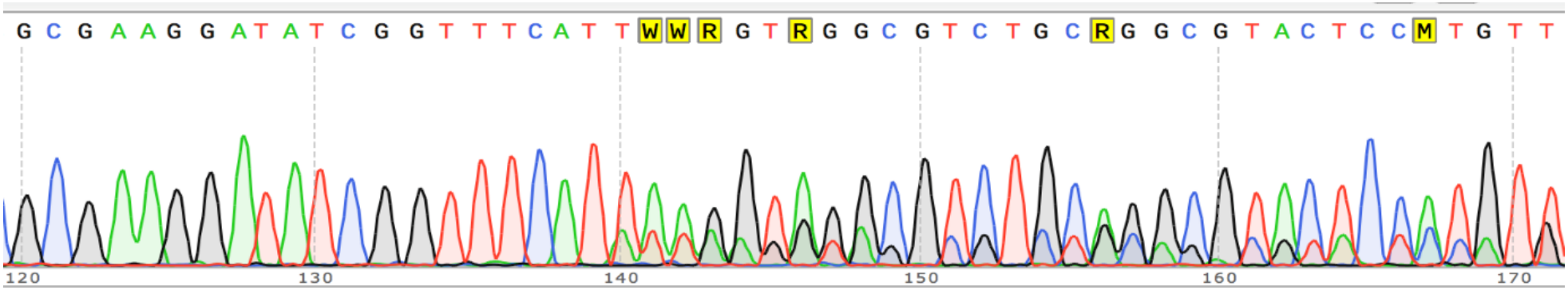


**Figure 3.23: Sanger sequencing plot of one of the clones of RPS29 knock out experiment.** It showed *RPS29 monoallelic* knock out of K562 cells.

Wild: TTC-ATT-AAG-GTA-GGC-GTC-TGC-AGG. Wild protein: GFIKLD*

Mutant: TTC-AT(AT)TA-AGG-TAG-GCG-TCT-GCA-GG

C.158-159insAT resulting in p.Lys54Leufs*

### *RPS19 heterozygous* knock out of K562 cells

To validate another novel mutation in DBA which was characterized in our Lab, a splice site *RPS19* c.71+5G>T, in vitro knock out and knock-in of the *RPS19* gene in K562 cell line was carried out. To confirm the role of this specific mutation in the pathogenicity of DBA, *RPS19* gene expression was studied together with pre-ribosomal RNA processing in the edited cells.

First, K562 cells were transfected with LentiCRISPRv2 plasmid that has an *RPS19* specific gRNA ligated to it. Next, this plasmid was transfected into K562 cells together with the homology directed repair template by the Amaxa's nucleofection system and 72 hours later, single GFP expressing k562 cells were sorted into 96 wells plates. After three weeks of expanding these single clones, the viable recovered clones were 55%. These viable clones were analyzed. On the recovered viable clones, Sanger sequencing was performed to confirm the presence of heterozygous gene editing. Sanger sequencing revealed 35% of the clones to be wild type (non-edited), while 35% of the clones showed heterozygous mutation and the remaining 30% showed homozygous mutation. None of the clones showed a successful knock in of the desired *RPS19* c.71+5G>T. Furthermore, the pathogenicity of *RPS19* gene mutation was studied using qPCR for 18S, 28S, 32S, and 5S rRNA which were deranged compared to control as shown in figures 3.24and 3.25.

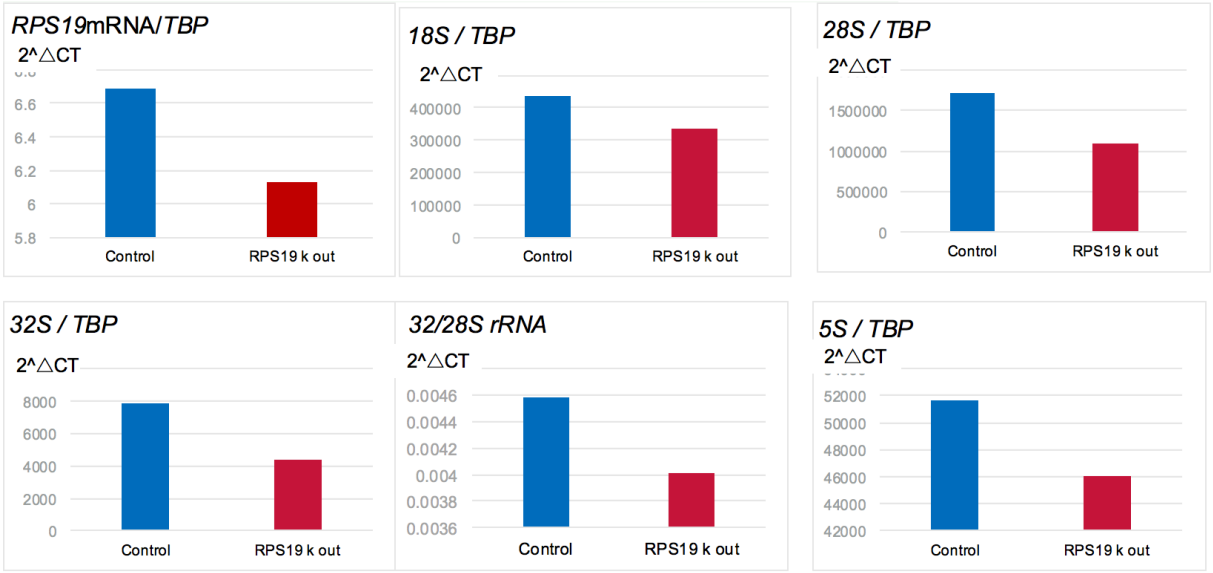


**Figure 3.24: Quantitative PCR performed on K562 RPS19 knocked out clone.** qPCR demonstrates reduction in the level of RPS19 mRNA clone. There is a detectable decrease in the levels of 18S, 28S, 32S, 5S rRNA and 32S/29S rRNA in the edited clone compared to the control. The control is a k562 cell clone that was transfected with an empty CRISPR vector without a ligated gRNA. TBP was used as a reference gene. (n=2)

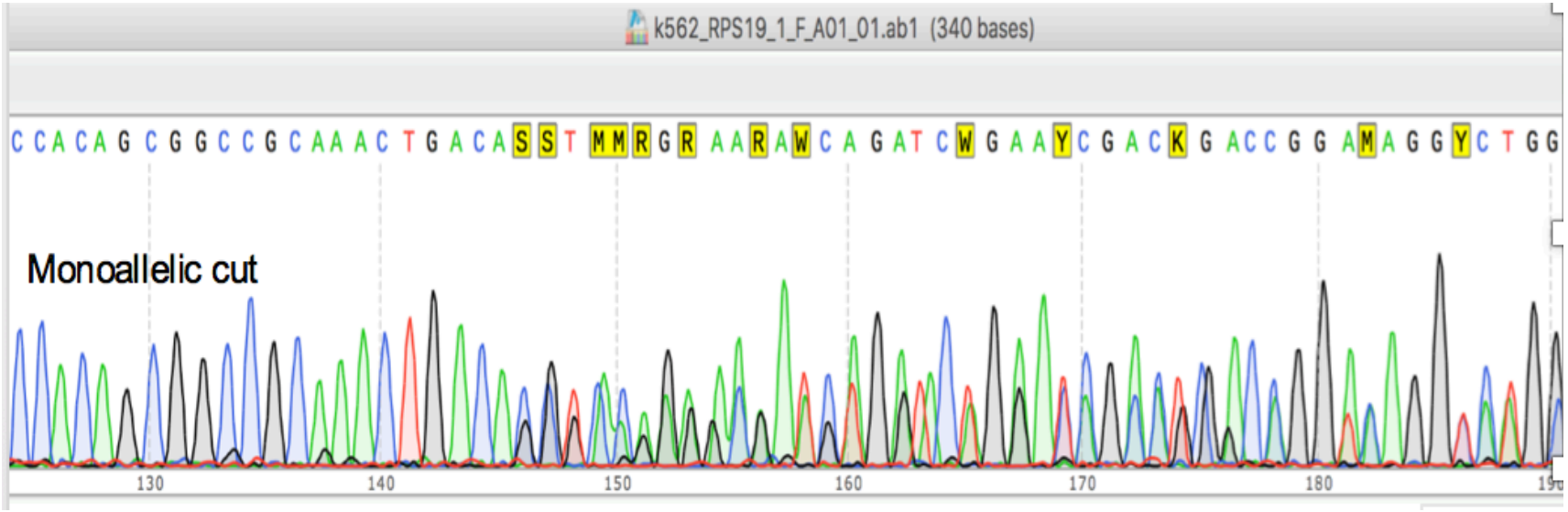


**Figure 3.25: Sanger sequencing plot of one of the clones of RPS19 knock out experiment**. It showed *RPS19 monoallelic* knock out of K562 cells.

Wild: GGC-CGC-AAA-CTG-ACA-CCT-CAG-GGA-CAA-AGA-GAT-CTG-GAC-AGA

Mutant :GGC-CGC-AAA-CTG-ACA-GGG-ACA-AAG-AGA-TC

c.373-377 del CCTCA p.Pro126Glyfs*29

**K562 cell line Knock in studies**

***RPS17 c.3G>C***

The *RPS17* novel mutation, *RPS17 c.3G>C* was knocked-in into K562 cells using the same Amaxa's nucleofection system of the CRISPR plasmid that is ligated to a gRNA specific to *RPS17* with the addition of the homology directed repair template followed by sorting single cells which expressed GFP into 96 wells plates 72 hours after electroporation. Three weeks later single colonies viability was 60%. On the recovered clones Sanger sequencing was performed on each clone and revealed 35% clones to be wild type and 35% with monoallelic editing in addition to 30% showing biallelic gene editing. One clone was successfully knocked in. This specific clone was utilized to study the impact of the knocked in mutation on rRNA processing. It was found that 18S, 28S, 32S 32/28S ,5S rRNA are deranged in this specific clone compared to control. Figure 3.26 and 3.27.

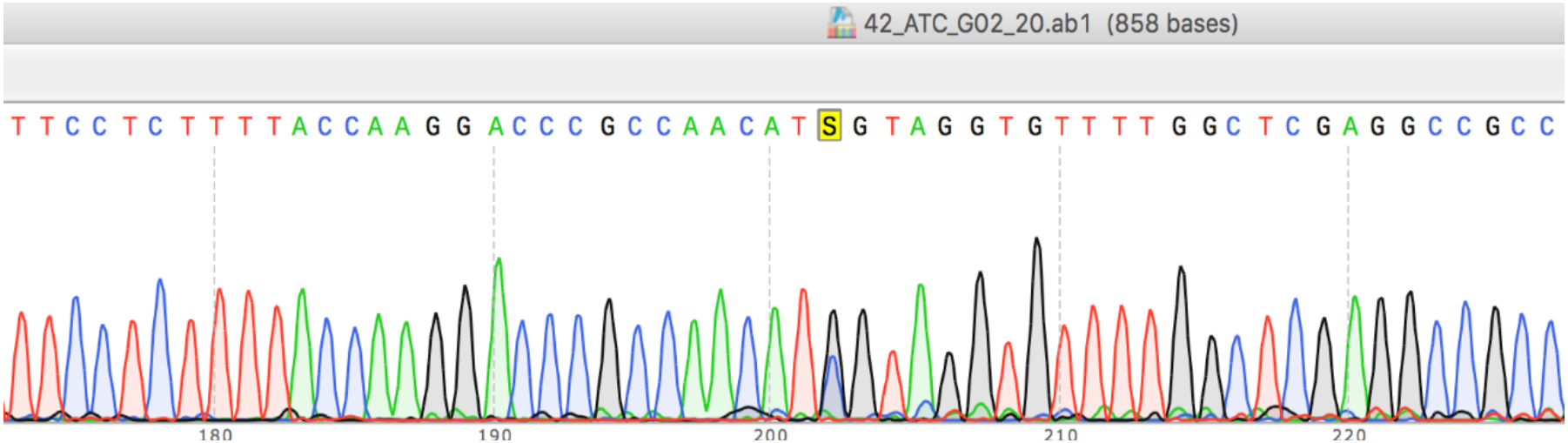


**Figure 3.26: Sanger sequencing of one clone of K562 confirming RPS 17 monoallelic knock in.** A novel *RPS17 c.3G>C.* p.Met1Ile

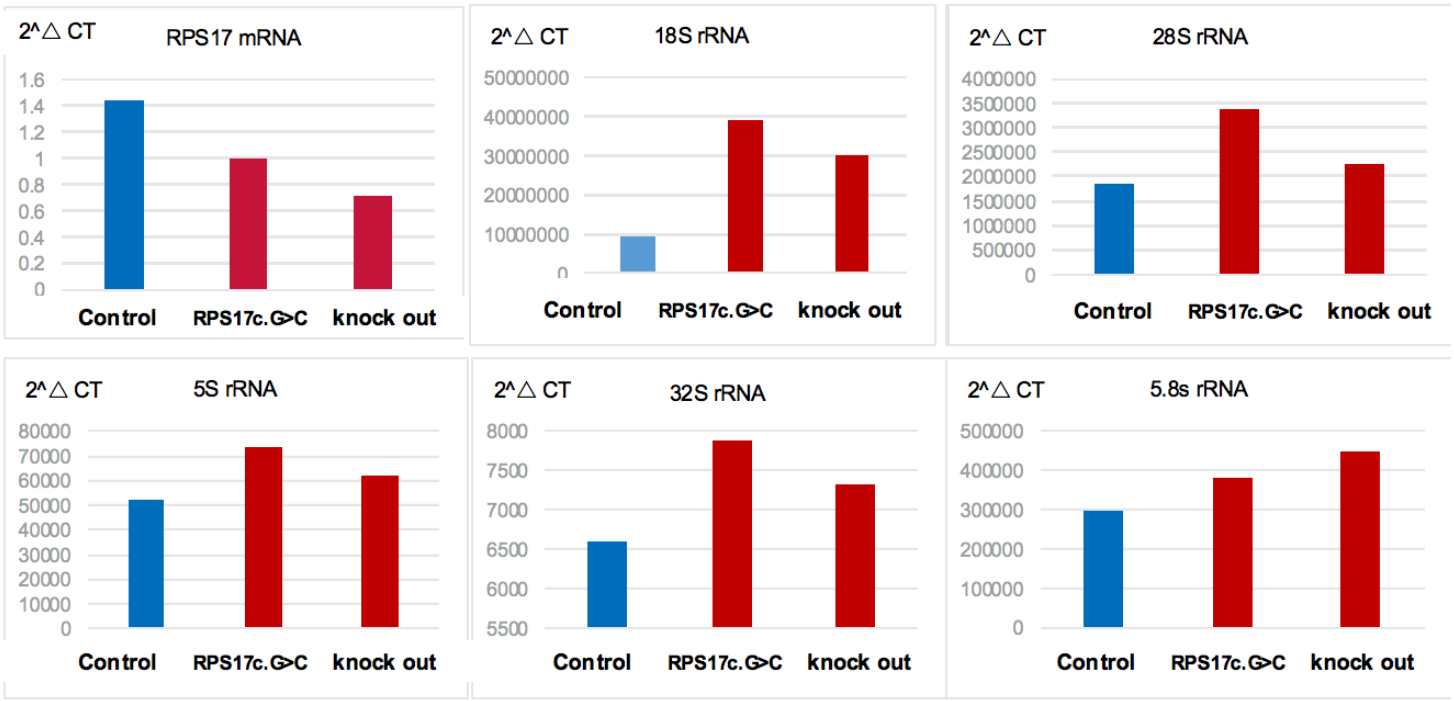


**Figure 3.27: Quantitative PCR performed on K562 RPS17 knocked out clone and RPS17 c.3G>C knocked in clone.** qPCR demonstrates reduction in the level of *RPS17* mRNA in both clones. Additionally, there is a detectable increase in the levels of 18S, 28S, 32S, 5S and 5.8S rRNA in both edited clones compared to the control. The control is a K562 cell clone that was transfected with an empty CRISPR vector without a ligated gRNA. TBP was used as a reference gene. (n=2).

### *RPL11* c.475_476 del AA

Secondly, the *RPL11* novel mutation, *RPL11* c.475_476 del AA, was knocked in into K562 cells using the Amaxa's nucleofection system followed by sorting single cells which were expressing GFP into 96 wells plate 72 hours after electroporation.

Three weeks following expansion in 37°C cell culture incubator supplemented with 5% $CO_2$, these single colonies were analyzed. The recovered viable colonies accounted for 10% of the original colonies. Sanger sequencing of each viable clone revealed four clones with wild type sequence and six clones with monoallelic editing in addition to two expressing biallelic gene editing. One clone was successfully knocked in with *RPL11* c.475_476 del AA mutation.

This specific clone was utilized to study the impact of the knocked in mutation on rRNA processing. It was found that 18S, 28S, 32S, and 5S rRNA were higher in the knocked in clone compared to control. However, the 32S/28S rRNA ratio was lower in the *RPL11* c.475_476 del AA knocked in clone compared to control as shown in figure 3.28 and 3.29.

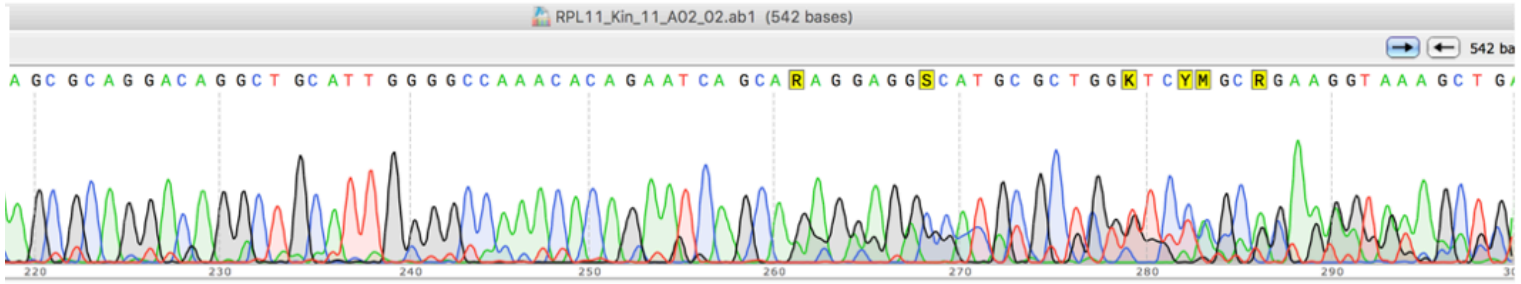


**Figure 3.28(A): Sanger sequencing and rRNA study in the K562 cell clone with RPL11 c.475_476 del AA knock in mutation.** Sanger sequencing plot confirming monoallelic knock in of one clone of K562 cells.

**Wild**: AGA-ATC-AGC-AAA-GAG-GAG-GCC-ATG…

**Knocked in**: AGA-ATC-AGC-AGA-GGA-GGC-CAT. p.Lys159Argfs*11.

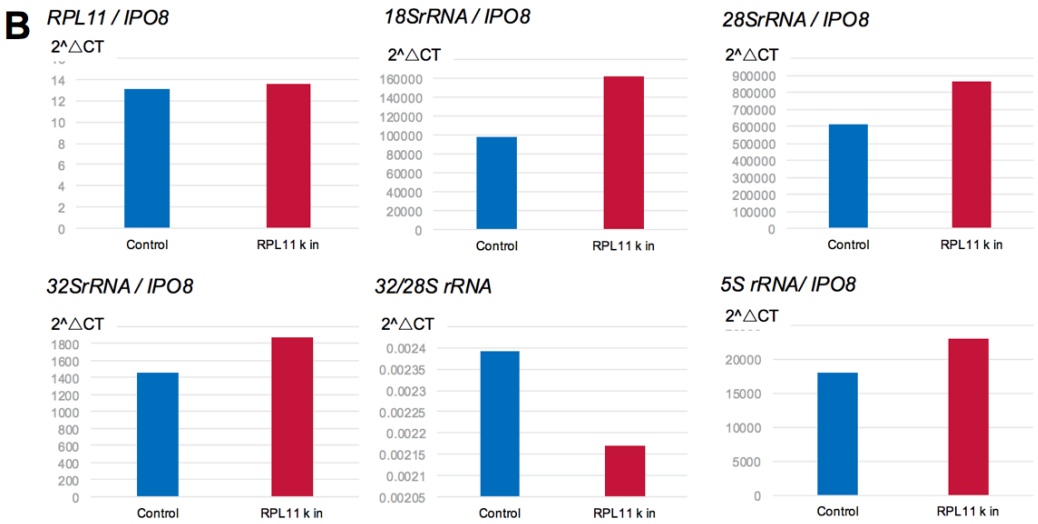


**Figure 3.28 (B) Quantitative RT-PCR on RPL11 knocked in clone of K562 cells.** demonstrates an increase in the levels of 18S, 28S, 32S, 5S rRNA compared to control with a reduction in 32/28S rRNA ratio in the knocked in clone compared to control. The control is a k562 cell clone transfected with an empty CRISPR vector without a ligated gRNA. IPO8 was used as a reference gene. (n=2).


**Figure 3.29: Sanger sequencing of 11 bacterial clones following cloning and colony PCR of one clone with *RPL11* knocked in mutation.** Sanger sequencing revealed one clone with a silent mutation and the second clone showed the knocked in allele RPL11 c.475_476 del AA.

**Hemin Treatment of K562 Cells**

Hemin can induce differentiation in K562 cells which stimulates exclusive production of fetal and embryonic haemoglobins since hemin can significantly stimulate RNA and protein synthesis. In this study, to validate the effect of mutations in RP genes, Knocked-out K562 cells were treated with 25 microM hemin. This was followed by monitoring the growth and differentiation of the edited K562 clones.

**Hemin stimulation of RPS17, RPS19 and RPL11 K562 edited clones**

It was found that upon treatment with 20 uM hemin, *RPS17* Knocked out cells and the *RPS17* c.3G>C Knocked in clone showed higher expression of α, ß, ζ, and ε globin genes compared to the control. Furthermore, ɣ globin expression was also increased in *RPS17* c.3G>C knocked in clone compared to the control as shown in figure 3.30.

Similarly, it was found that treatment of *RPS19* knocked out clone with 20 uM hemin resulted in higher expression of α, ß, ζ, and ɣ globin genes compared to the control clone as illustrated in figure 3.31. Having said that, *RPL11* Knocked out K562 cells, likewise *RPL11* c.475_476 del AA, showed a different globin expression pattern compared to *RPS17* and *RPS19*. It showed lower α, ɣ, ζ and ε globin gene expression compared to the control clone as shown in figure 3.32. These findings were further confirmed by repeating the RT-qPCR studies relatives to a second different reference gene.

Furthermore the patient with, *RPL11* c.475_476 del AA mutation had a normal HbF level and a low HbA. These findings are consistent with what was observed from studying the *RPL11* c.475_476 del AA Knocked in mutation on K562 cells upon treatment with hemin. This can indicate

hemin stimulation of edited K562 cells is a useful way for mutational validation in DBA.

As for the RPS19 knocked out clone, this result can explain the elevation in HbF that is seen in many patients with *RPS19* mutations who are involved in this study.

Although there has been a paucity of studies on the effect of RP gene mutations on Hb differentiation following stimulation of the cells with hemin, this finding further confirms the pathogenicity of this novel mutation in addition to agreeing and explaining the elevated level of HbF ($\alpha_2\gamma_2$) which is reported in DBA. This results is also concordant with Hb electrophoresis which was performed on one of the patients with *RPS17* mutation. Hb electrophoresisl showed elevated HbF in a patient with *RPS17* mutation.

Interestingly, the increased expression of ζ, and ε globin genes might well explain the absence of life threatening anemia in utero, as this might be associated with an increased level of embryonic Hb like Gower 1 ($\zeta_2\varepsilon_2$) and Gower 2 ($\alpha_2\varepsilon_2$).

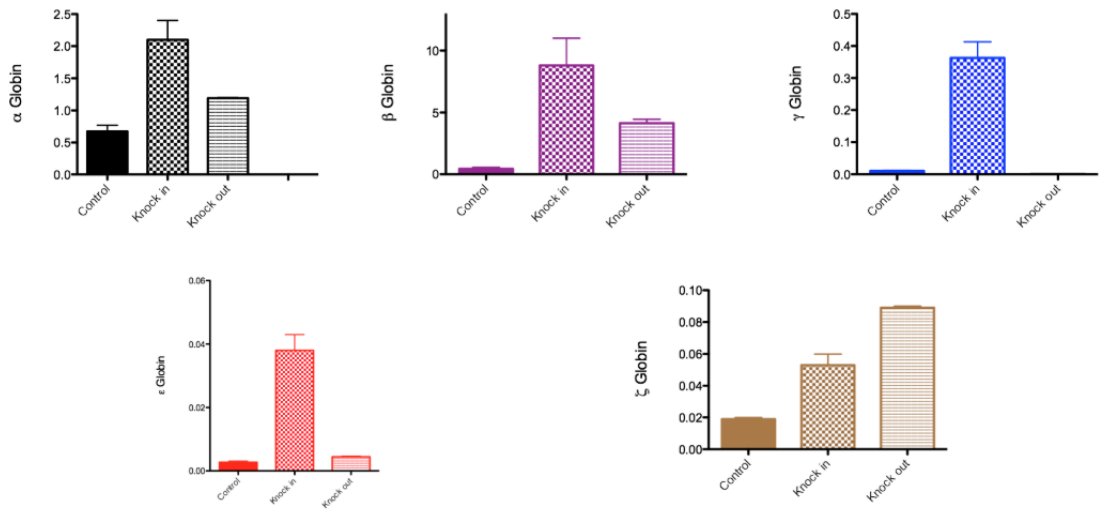


**Figure 3.30: Globin gene expression in the RPS17 Knocked out and RPS17 c.3G>C Knocked in K562 clone.** 72 hours following treatment with 20 uM hemin. (n=2). Gene expression was measured relative to IPO8 and GAPDH.

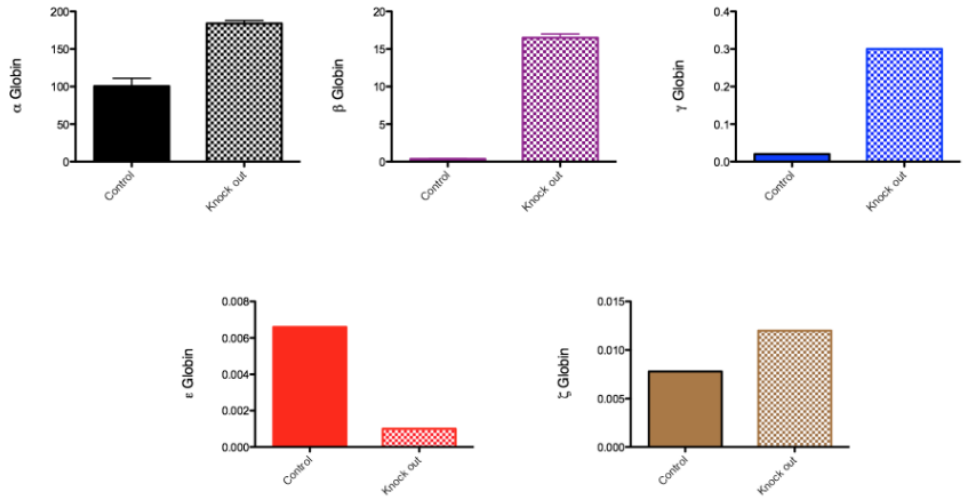


**Figure 3.31: Globin gene expression in the RPS19 Knocked out K562 clone.** 72 hours following treatment with 20 uM hemin. (n=2). Gene expression was measured relative to IPO8 and GAPDH.

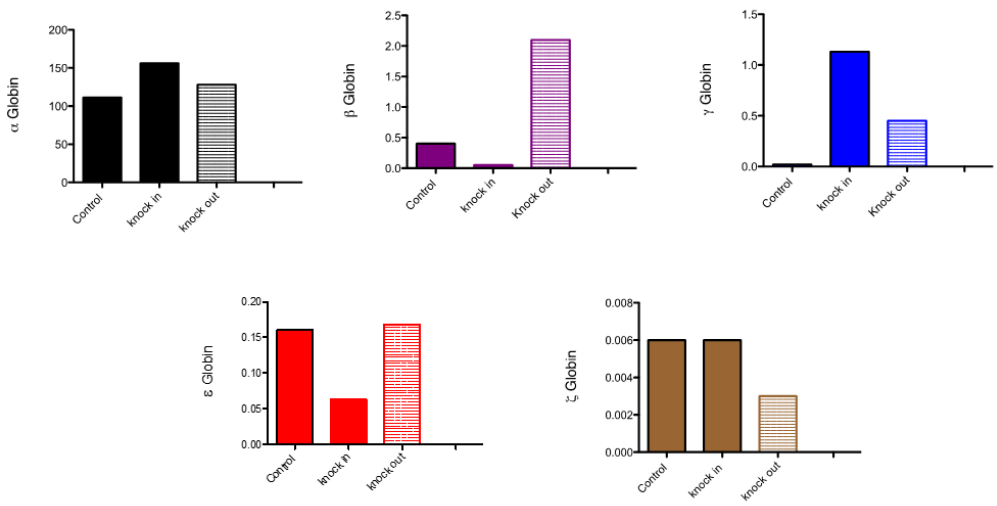


**Figure 3.32: RT-QPCR Globin gene expression studies on two clones of K562 cells.** RPL11 Knocked out clone and another RPL11 knocked in clone. RPL11 c.475_476 del AA. These clones were treated with 20 uM hemin. (n=2). Gene expression was measured relative to IPO8.

Preliminary data was showing that certain RP gene mutations like *RPL11* c.475_476 del AA, was associated with impaired growth potential of K562 cells compared to control while other RP gene mutations *RPS17* c.3G>C was not showing a clear impact on the proliferation of K562 cells. Furthermore, proteins were quantified form equal number of cells in the Knocked in clone of *RPL11* c.475_476 del AA and RPS 17 c.3G>C compared to the control non edited K562 clone. Our preliminary data showed that both of these mutations were associated with a reduction in the overall protein content of the cells. This result agrees with many studies carried out on DBA showing impairment in translation.

# CHAPTER 4

## Discussion

In this study all DBA patients underwent screening for a panel of 83 ribosomal protein genes in addition to *GATA1* gene analysis. Furthermore, Sanger sequencing for *RPS17* gene was also performed, separately because of the specific nature of this gene which made NGS analysis complex and prone to errors. A total of 48 patients with DBA were included in this study. Twelve patients (~25%) had mutations in genes encoding RPs of the large ribosomal subunit and sixteen patients (~33%) had mutations in genes encoding RPs of the small ribosomal subunit. Furthermore, there were twenty patients (~41%) involved in this study with no identifiable RP gene mutations.

To validate novel mutations identified in this study, we tested two protocols which utilize the disturbance in rRNA profile that was reported in DBA. In addition, we aimed to develop a more rapid method for the analysis of the rRNA as a diagnostic tool for DBA. These included: 1.RNA capillary electrophoresis using 2100 Agilent bioanalyzer and 2. quantitative real time PCR for the relative quantification of rRNAs. Finally, these two protocols could potentially be exploited as an alternative to Northern blot which has the disadvantages of possible mRNA degradation, high dose of radioactivity in addition to its lower sensitivity compared to RT-PCR (118)

**Capillary electrophoresis analysis of rRNA maturation**

The 40S ribosomal subunit is comprised of a single 18S rRNA and ribosomal proteins whereas the large subunit is comprised of three rRNAs, 28S, 58S and 5S rRNA in addition to the ribosomal proteins. The 28S and 18S rRNAs are the major peaks visualized by capillary electrophoresis and there is an additional minor peak which represent a rRNA intermediate, 32S rRNA. This 32S peak was frequently observed in the analysis of RNAs from most patients with RPL mutations.

Haploinsufficiency of RP can alter the balance between ribosomal proteins and the different rRNA species. Since RPs play a crucial role in the processing of rRNA from the 45S transcript, we hypothesized that putting patients' cells under proliferative stress would trigger a more detectable derangement in the rRNA profile which could be implemented as a diagnostic tool in DBA.

As shown in figure 3.2, six days following stimulation of T cells with anti-CD3 anti-CD28 activator the ratio of 32S/28S was significantly higher in patients with mutations in RPL genes (mean 0.05 ± 0.007) (*P*=0.0001) compared to the control group (mean 0.014± 0.002). Similarly, the 32S/28S rRNA was significantly higher in DBA patients with RPS gene mutations (mean 0.02± 0.004) *(P*=0.02) compared to the control group. Interestingly, the 32S/28S rRNA ratio was significantly higher in DBA patients without an identifiable RP gene mutation compared to the control group. This finding suggests that rRNA derangement is a characteristic finding in DBA which is not only related to RP gene defect. This finding is consistent with Ellis et al (28) who analyzed 32S/28S rRNA from stimulated PBMNC in a small group of DBA patients.

As shown in figure 3.3, six days after stimulating T cells with anti-CD3 anti-CD28 the 28/18S rRNA ratio was significantly lower in patients with RPL nutation (mean 1.9 ± 0.067) compared to healthy controls (mean 2.1 ± 0.05) (*P*=0.02). Conversly, this ratio was significantly higher in DBA patients of non-identifiable RP gene mutation (mean 2.22 ± 0.07) compared to the control group (*P*=0.04). This finding is concordant with Quarello et al (48) who studied 28S/18S rRNA in stimulated PBMNC from DBA patients compared to a healthy control group.

Although there was a statistical difference in the 32S/28S rRNA and 28S/18S rRNA ratio but there was an overlap in the range of both ratios between DBA patients and the control group.

In the healthy control group the range of 28S/18S rRNA was (1.7-2.4) while the range in DBA patients with RPS, RPL gene mutation and the group with nonidentifiable mutation was 1.8-2.9 and 1.5-2.2 and 1.4-2.7, respectively.

Regarding 32S/28S rRNA the normal range in the control group was 0.004-0.028 compared to 0.013-0.088, in patients with RPL mutation and 0.009-0.053 in patients with RPS mutation, while the ratio was 0.011-0.036 in patients with nonidentifiable RP gene mutation.

As shown in figures 3.2 and 3.3, resting and unstimulated, T cells showed no significant difference in the 32S/28S rRNA and also in the 28S/18S rRNA when compared to unstimulated T cells from the control group (*P*=NS). This might be due to the proliferative stress that the cells underwent following stimulation which make the aberrant rRNA processing more detectable. In addition, this might also explain the

paucity of studies on rRNA in unstimulated cells, as both Ellis et al (28) and Quarello et al (48) analyzed RNA electrophoresis on stimulated PB-MNC.

The assay reported here would be amenable for development as a diagnostic tool. Potential applications for this diagnostic tool emerges in patients who currently showed non identifiable RP gene mutation. the genes affected in these patients might be non-ribosomal protein genes, such as *TSR2*, that play a role in 40S subunit biogenesis and have recently been identified to play a role in DBA pathogenesis (31).

Moreover, our data reveal that the32S/28S rRNA and 28S/18S rRNA can differentiate patients with RPL mutation, RPS mutation and normal. Especially when studied before and after stimulation. Our findings showed that the 32S/28S rRNA will increase upon stimulation in DBA patients compared to control. Not only this ratio but also the 28S/18S rRNA will decrease in RPL patients compared to normal upon stimulation, but it will decrease in RPS patients. Thus, the relative ease and low expense of this test in comparison to next generation sequencing suggest that this approach may be amenable to further development to help in identifying affected parents or siblings in a faster and more cost effective way compared to other techniques. Additionally, further strategies can be studied to augment and increase the sensitivity of this test.

On this basis, we were prompted to ascertain whether the addition of another test like RT-PCR to assess the disruption of pre-ribosomal RNA processing can augment the sensitivity of RNA capillary electrophoresis and aid in the diagnosis of DBA.

**Quantitative real-time PCR**

To increase the sensitivity of detection of rRNA and improve on the previous method, rRNA gene-based real time quantitative PCR using specific primers were designed to detect and quantify 18S, 28S, 32S, 5.8S and 5S rRNA and applied to both DBA patients and healthy controls.

Like the previous method, namely the capillary electrophoresis, peripheral blood mononuclear cells (PBMNC) were isolated from patients' samples. T cells were selected, stimulated and expanded. RNA was extracted at two different time points, day 0 and day 6 after stimulation.

**Selection of reference gene**

Relatively few studies have quantified gene expression in DBA using real-time quantitative PCR (RT-qPCR), which is a reliable and sensitive technique for measuring gene expression. RT-qPCR requires the selection of reference genes to normalize gene expression data and control for internal differences between samples. In this study, five common housekeeping genes were selected for their suitability in normalizing gene expression using Normfinder. Following RNA extraction form pure DBA patients' T cells, RT-qPCR was performed. The expression levels of five candidate human reference genes: actin beta (*ACTB*); tata box binding (*TBP*) and transferrin receptor (*TFRC*), which are commonly used for gene expression studies in bone marrow mesenchymal stromal cells; importin 8 (*IPO8*), which is one of the best genes used for expression studies in adipose tissue; and ribosomal protein lateral stalk subunit P0 (RPLP0). The expression stability of these

selected reference genes was evaluated using the Normfinder software which is able to rank different HK genes on the basis of the stability of their Ct values. Results indicated that *IPO8* and *RPLP0* showed the most stable expression in DBA unstimulated T cells. As for stimulated T cells, the same analysis was performed on ten DBA samples on day 6 following stimulation with anti-CD3 and anti-CD28 antibodies and IL2. Results indicated that *IPO8* and *TBP* are the most stable genes to be used for internal control. Selection and validation of the most suitable reference gene is a critical step to improve the validity and accuracy of gene expression studies. Results of this study provide vital information on reference genes and are valuable in developing a standardized RT-qPCR protocol for functional genomics research in *DBA.*

**Ribosomal RNA Quantitative Real-time PCR**

We designed, ordered and tested rRNA specific probes to study the relative expression of 18S, 28S, 32S, 5.8S and 5S rRNA. This test was performed on RNA extracted from resting and stimulated T cells isolated from DBA patients and healthy controls.

**RT-qPCR in Resting, Unstimulated T cells**

As shown in figure 3.5, the 18S rRNA detected by real time quantitative PCR was significantly lower in DBA patients (n=48) than the control group (n=13) ($P$=0.0004). Similarly, 28S rRNA was significantly lower in DBA patients as compared to the control group ($P$=0.03).

Conversely, 32S/28S ratio was significantly higher in DBA patients compared to controls ($P$=0.019). The differences in 32S and 5S rRNA expression between DBA and control was non-significant ($P$>0.05).

Likewise, these findings were confirmed by repeating the qPCR experiment and quantifying the rRNA relative to another reference gene (ACTB).

**RT-qPCR in Resting, Unstimulated T cells with *RPS19***

From the cohort of patients included in this study, six patients (12.5%) had mutation in *RPS19.* RT-qPCR revealed a significant decrease in the level of 18S rRNA in DBA patients compared to the healthy control group (*P*=0.0004). This finding is in agreement with Choesmal et al (8) who transfected Hela cells with siRNA against *RPS19* and analyzed rrNA by Northern blot. Choesmal et al (5) showed that knock out of RPS19 is associated with a reduction in 18S rRNA and accumulation of its precursor 21S rRNA. Additionally, the same study did not notice any changes in the 28S rRNA nor its precursors like 32S rRNA (5). Similarly, in our project we did not find a significant difference in the level of 28S, 32S and the ratio of 32S/28S rRNA between *RPS19* patients and the control group (*P*=ns) as shown in figure 3.6.

**QPCR in Resting, Unstimulated T cells with *RPL11***

As shown in figure 3.7, in this study *RPL11* mutations were observed in five patients (10.5%). rRNA analysis data from those patients was compared with the control. It was found that in resting T cells the 18S rRNA was significantly lower in patients with *RPL11* mutation compared to the control group (*P*=0.004). However, there was no significant difference in the level of 28S, 32S, 32/28S and 5S rRNA between those patients and the controls (*P*=NS).

This findings is concordant with a study carried out by Gazda et al (1) who stated that lymphoblastoid cells established from *RPL11* DBA patients

were associated with a reduction in the level of 18S rRNA. However, they noticed an increase in the level of 32S/28S rRNA in addition to a reduction in the level of 5.8S rRNA with an increase in its precursor in the lymphoblastoid cells established from DBA patient with RPL11 mutation (6). Although our study did not show an increase in the mean value of 32S/28S rRNA in patients with RPL11 mutation compared to the healthy control; however, some of the *RPL11* patients showed a higher 32S/28S rRNA compared to some of the normal controls. There was an overlap in the range of 32S/28S in DBA patients with *RPL11* mutation (0.001-0.002) and the control (0.001-0.004).

**QPCR in Resting, Unstimulated T cells with *RPS26***

Five patients (~10%) with *RPS26* mutation were also included in this study. As demonstrated in figure 3.8, it was found that there was a significant difference in the level of 18s rRNA between normal and DBA patients with *RPS26* ($P$=0.005). However, there was no significant difference in the level of 28S, 32S, 32/28S and 5S rRNA between patients with *RPS26* mutation and the control group ($P$>0.05). This finding is concordant with Ferrar et al (28) who studied stimulated mononuclear cells from DBA patients and revealed that RPS26 mutation is associated with a decrease in the mature 18S rRNA and no increase in the level of the 32S rRNA by northern blot. Our finding is also in agreement with Landowski et al (27) who found that lymphoblastoid cells established from DBA patient with RPS26 mutation showed a decrease in the level of 18S rRNA without an increase in the level of 32S rRNA.

**QPCR in Resting, Unstimulated T cells with *RPS17***

Three patients with *RPS17* (~6%) were included in this study. rRNA gene expression was compared among patients with *RPS17* mutation and

healthy controlsas shown in figure 3.9. Although there was a non-significant difference in the level of 18S, 28S, 32S,32/28S and 5.8S between patients with *RPS17* and controls (*P*>0.05); however, the level of the mature 18S rRNA was lower than that of the control and the level of the 32S rRNA precursor was higher in patients with *RPS17* mutation compared to the control group which agrees with Landowski et al (1) who applied Northern blot on RNA extracted from lymphoblastoid cells established from DBA patients with *RPS17* mutation. Similar to our findings, Landowski noticed that *RPS17* mutation was associated with impaired maturation of 18S rRNA and increased 32S rRNA intermediate.

Following T cell stimulation, Q-RT PCR showed that 18s rRNA was significantly higher in stimulated T cells from DBA patients compared to control (p=0.0003). A striking finding was that the 18s rRNA in DBA patients was lower than normal in resting T cells. Hence, monitoring 18s rRNA in resting T cells and six days upon stimulation will help to identify DBA patients by showing a lower then a higher 18s rRNA level. This finding suggests that proliferative stress will exert a burden on ribosomes and make the aberrant rRNA processing more prominent. This can also add a justification to the selective translational defect that is seen in the tissues with more active proliferation like hematopoietic compared to others.

**Mutational Validation Using CRISPR/CAS9**

In this project, forty-eight DBA patients were studied. DBA diagnosis was made on the basis of the clinical symptoms with the aid eADA (erythrocyte adenosine deaminase) and HbF, in addition to the exclusion of the other IBMF syndromes. Furthermore, the diagnosis was supported by the aid of next generation sequencing to look for mutations in 83 RP genes and *GATA1* gene. The only RP gene not included in the DBA diagnostic panel is *RPS17*. This was because of the homology between *RPS17* and its pseudogene that made designing of primers for this gene too complex.

Therefore, as part of this project *RPS17* screening for DBA patients was carried out by Sanger sequencing following designing specific primers to amplify the all five exons together with the intron-exon junction. All primers sequencing included in this study are provided in the appendix. By the aid of Sanger sequencing we identified one novel *RPS17* mutation among our cohort of patients. This is *RPS17* c.3G>C; p.Met1Ile

Furthermore, in this study Sanger sequencing was also performed to validate some other novel mutations detected by NGS. These include *RPL11* frameshift deletion *RPL11* c.475 del *AA (*p. Lys159Argfs*11), *RPS2*9 missense c139G>C (P.Ala47Pro), *RPS29* Splice site *RPS29* c.63-3C>A. These are summarized in table (3.1).

To confirm the deleterious effects of some of these novel mutations, CRISPR/CAS 9 technology was implemented. The application of CRISP/Cas9 has many challenges.

The first challenge was the successful introduction of the CRISPR/Cas 9 plasmid into the cells. Hence, 293T was firstly selected as a model to transfect CRISPR/Cas9 and to analyze its editing potential. This is because HEK 293 T cells are characterized by a quick reproduction

potential and are amenable to transfection using a wide variety of techniques and most importantly because they exhibit a high efficiency of transfection and protein production (119).

A viral transduction of HEK 293 T cells was first endeavored following the production of a viral vector from CRISPR/Cas9 and other packaging plasmids. Three different viral vectors were produced each with a different gRNA specific to a specific gene to be knocked out. Knock out of *RPL11* and *RPS19* was attempted as they are the most common genes involved in DBA. One viral vector specific to *RPL11* exon 5 was developed and two other viral vectors specific to *RPS19* at two different sites (exon 2 and exon 5) were also produced. Following viral transduction to HEK 293T cells, the transduction efficiency was determined based on GFP expression by flow cytometry as seen in figure 3.11, which showed the efficiency to be 77.6%, 98% and 87.1% for *RPL11*, *RPS19* exon 5 and *RPS19* exon 2, respectively. This was followed by sorting single cells for GFP expressing cells. Three weeks later, the recovered clones were sequenced to explore the editing potential of CRISPR/Cas9 system. It was found that following viral transduction on HEK 293 T cells both alleles were edited as seen in figure 3.12. However, the cells were still viable and proliferating efficiently. Later, in this study homozygous gene editing was accidentally introduced in *RPS17*, *RPS19* and *RPL11* genes while trying to introduce a heterozygous knock out. Nevertheless, the cells with homozygous gene editing were viable. This is a novel finding that needs to be further elucidated. Future work will include PCR amplification of the edited region followed by cloning of the PCR amplification product into bacterial colonies. Colony PCR and sequencing will show the exact edited sequence. This will be followed by gene expression studies and western blot to study the effect of this homozygous RP gene editing on protein

production and hence cell survival. The result will then be compared with what has been stated that biallelic RP mutations in mice and zebrafish are always lethal (9,10). Additionally, to overcome the first challenge CRISPR/Cas 9 was also successfully introduced into Jurkat T cells as an example of suspension cells as shown in figure 3.13. From these results it was clear that the first challenge in the CRISPR work was resolved. This was the successful introduction of the CRISPR plasmid into the cells and testing gene editing by Sanger sequencing. However, the second challenge was to induce a heterozygous knock out of the gene of interest for mutational validation as DBA is inherited as an autosomal dominant disease.

From the previous viral transduction method, it was concluded that stable expression of Cas 9 protein will result in biallelic knock out. Therefore, transient transfection of CRISPR/Cas 9 was attempted. To ensure transient transfection we tested two protocols, the double sorting strategy and the timed single sorting strategy.

The double sorting strategy involves transfection of HEK 293T cells by calcium phosphate method with the GFP expressing CRISPR/Cas 9 plasmid. Twenty-four hours later, the cells expressing the plasmid were sorted based on the GFP expression. From this population, forty-eight hour later the cells which had lost the GFP plasmid were sorted into 96 wells plate as single cells and left to expand in 37°C cell culture incubator supplemented with 5% $CO_2$. This protocol should ensure that the plasmid was expressed transiently inside the cell. Four weeks later following expansion, Sanger sequencing was carried out on the recovered clones. Surprisingly, it was found that none of the clones was edited. Figure (3.14)

illustrates these findings. Hence, another protocol was tried which was the transfection and timed single cell sorting protocol.

The timed single cell sorting protocol involves transfection of HEK 293T cells by calcium phosphate method with the GFP expressing CRISPR/Cas 9 plasmid. Seventy two hours later, the cells expressing the plasmid were sorted based on the GFP expression. Sorting into 96 wells plate as single cells expressing GFP, next the cells were incubated at 37°C with 5% $CO_2$

Three weeks later, the viable recovered colonies were sequenced using Sanger sequencing. Interestingly, monoallelic gene editing was detected in 30% of the recovered clones. To further confirm monoallelic editing we cloned the PCR product, amplicon of interest, into CloneJET PCR Cloning Kit (ThermoFisher) and transformed the ligated plasmid into NEB 5-alpha competent *E. coli* bacteria that was plated and grown overnight. Sequencing bacterial colonies the next day further confirmed a successful monoallelic knock out. After establishing this protocol, the second challenge in the CRISPR work was achieved. This was the successful induction of monoalllelic knock out. The third challenge was enhancing homology directed repair to introduce a specific mutation which will be discussed later in this section.

After establishing the transfection protocol in HEK-293T cells, *RPL11, RPS19, RPS29* and *RPS17* were knocked out and the effect of gene knock out on rRNA profile was studied.

### *RPS17* knock out in HEK 293T cells and *RPS17* Knock out and *RPS17* c.3G>C knock in of K562 cells

Using calcium phosphate co-precipitation protocol for transfecting HEK-293T cells and the timed single cell sorting protocol which we optimized in the methodology mentioned above, *RPS17* was knocked out in HEK-293T cells. Single HEK-293T cells expressing GFP were sorted. After 3 weeks 10 clones out of 96 were proliferating. These ten clones were sequenced for *RPS17.* Sanger sequencing showed 5 clones to be wild type (non-edited), two clones showed homozygous mutation and 3 clones showed heterozygous mutation. Two clones with heterozygous *RPS17* knock out were studied using qPCR for 18S rRNA which was increased compared to control (as shown in figure 3.19). Similarly, when *RPS17* was knocked out in K562 cells a higher level of 18S rRNA compared to control was observed. Similarly, the *RPS17* c.3 G>C knock in of K562 cells showed an increase in the level of 18S rRNA as seen in figure 3.27).

These findings are novel and there are relatively few studies on the effect of *RPS17* on rRNA processing. However, it was found that 21S rRNa which is a precursor of 18S rRNA is increased upon knock down of *RPS17* (27,120,121), however, the level of the 18S rRNA following RPS17 knock out was not reported before. Having said that, higher level of 21S rRNA, a precursor of 18S rRNA may be associated with an increase in the mature 18S rRNA. The level of 21S rRNA intermediate was not included in our study due to the difficulty in designing primers specific to 21S rRNA to be used for RT-qPCR studies.

### *RPL11* knock out in HEK-293T and K562 cells

Using the method of transfection and timed cell sorting, *RPL11* was knocked out in HEK-293T cells. After single cell sorting 8 clones out of 192 survived. Sequencing *RPL11* in these 8 clones showed 5 clones to be wild type (non-edited) and 3 clones showed monoallelic cut. Furthermore, two knocked out clones were assessed for the pathogenicity of this gene using qPCR for 18S rRNA which was reduced together with the relative expression of *RPL11* mRNA compared to control as demonstrated in figure 3.18. This result is consistent with what we have observed in this study when comparing the 18S rRNA between T cells of RPL11 mutated DBA patients T and the control group. Additionally this finding is in agreement with Gazda et al who studied lymphoblastoid cells established from DBA patients(1). Having said that, when comparing the result of HEK-293T with K562, the 18S rRNA finding of HEK293T is neither consistent with our findings following knock out of *RPL11* in k562 cells nor with the findings of *RPL11c*.475_476 del AA Knock in of K562 cells. This might be attributed to the difference in the type of cells studied.

Using Amaxa's nucleofection system and time cell sorting, K562 cells were knocked out. Three weeks following sorting 10 clones out of 192 clones have survived. Sanger sequencing showed three clones to be wild type (non-edited), five clones were showing heterozygous mutation and two clones were showing homozygous mutation. Interestingly,as shown in figure 3.20 *RPL11* knock out was associated with an increased 32S rRNA which is consistent with the findings of Gazda et al (1) who studied lymphoblastoid cells derived from patients with *RPL11* mutation. Additionally, our findings concerning 32S rRNA elevation in *RPL11* knock out clone is also consistent with Farrar et al (28) who studied Con A stimulated MNC from DBA patients with *RPL11* mutation. Furthermore,

*RPL11* knock out clone in K562 cells was associated with an elevation in the levels of 5S rRNA and 28S rRNA in the knocked out clone compared to the control. Moreover, the rRNA profile of *RPL11* knock out in K562 cells was consistent with what was explored in this study from a successful *RPL11c*.475_476 del AA Knock in of K562 cells as illustrated in figure 3.28. In other words K562 cells showed a consistent rRNA profile both in the knocked out clone as well as the *RPL11*c.475_476 del AA Knock in clone.

### *RPS19* knock out of K562 cells

K562 cells were transfected with LentiCRISPRv2 plasmid that has an *RPS19* specific gRNA by Amaxa's nucleofection system and 72 hours later, single GFP expressing K562 cells were sorted into 96 wells plates. After three weeks of expanding these single clones, the viable recovered clones were 55%. These viable clones were analyzed. On the recovered viable clones, Sanger sequencing was performed. Sequencing revealed 35% of the clones to be wild type (non-edited), while 35% of the clones showed heterozygous mutation and the remaining 30% showed homozygous mutation. Furthermore, the pathogenicity of *RPS19* gene mutation was studied using RT-qPCR for 18S, 28S, 32S, and 5S rRNA which were deranged compared to control as shown in figures 3.24 and 3.25. It was found that 18S,28S, 32S, 5S rRNA and 32S/28S rRNA were all decreased compared to the control clone. This result agrees with what was demonstrated earlier in this study in patients' T cells compared to the control. As was shown in figure 3.6 regarding the rRNA profile in patients' T cells. It was found that 18S rRNA was significantly reduced in patients with *RPS19* mutation compared to the control group ($P$=0.0004). This

finding is consistent with other studies that demonstrated the role of RPS19 gene in 18S rRNA maturation (9,29,46).

***RPS29* knock out of K562 cells**

To validate certain novel mutations in DBA, a missense *RPS29 c.139G>A*, P. Ala 47 Pro**,** in vitro knock out and knock in of the *RPS29* gene in K562 cell line were carried out. To confirm the role of *RPS29* in the pathogenicity of DBA, *RPS29* gene expression was studied together with pre-ribosomal RNA processing in the edited cells.

K562 cells were transfected with LentiCRISPRv2 plasmid that has an *RPS29* specific gRNA to target RPS29. Amaxa nucleofection system and timed sorting was carried out according to the same methodology mentioned before. Three weeks following single cell sorting 60% of the clones were viable. These viable clones were analyzed. On the recovered viable clones, Sanger sequencing was performed to confirm the presence of heterozygous gene editing. Sanger sequencing revealed 30% of the clones to be wild type (non-edited), while 35% of the clones showed heterozygous mutation and the remaining 35% showed homozygous mutation. None of the clones showed a successful knock in of the desired *RPS29* c.139G>A. We reasoned that its unique position which is 20 bases away from the double stranded break made it a difficult mutation to be knocked in. RT-qPCR for 18S, 28S, 32S, and 5S rRNA was performed on the mooallelic knock out clone rRNA profile was deranged compared to control as shown in figures (3.23).

It was found that there was a detectable decrease in the level of 28S, and 5S rRNA compared to control. Additionally, the mean of 5S rRNA was lower in DBA T cells compared to the control group. Furthermore, this finding is concordant with Mirabello et al (30) who stated that *RPS29* mutation is associated with pre-RNA processing defect.

### *RPS17* c.G>C Knock in of K562

### *RPS17 c.3G>C*

DBA patients who showed no identifiable ribosomal protein gene mutation with Next generation sequencing panel where screened as part of this project for *RPS17* mutation by Sanger sequencing. A novel initiation codon mutation in *RPS17* was reported. The mutation affects the translation initiation start codon, changing G to C, therefore eliminating the natural start of RPS17 protein synthesis The next downstream start codon located at position +158 should give rise to a short peptide of only four amino acids (Met-Ser-Arg-Ile). This makes the mutation deleterious. The mutation arose de novo, since all family members carry the wild- type alleles. To confirm the pathogenicity of this novel mutation, *RPS17 c.3G>C,* p. Met1Ile.

was knocked-in into K562 cells using the same Amaxa's nucleofection system to transfect the CRISPR plasmid that is ligated to a gRNA specific to *RPS17* with the addition of the homology directed repair double stranded template followed by sorting single cells which expressed GFP into 96 wells plates 72 hours after nucleofection.

Three weeks later 200 clones (~20%) were recovered from 960 sorted clones. On 50 recovered clones Sanger sequencing was performed on each clone and revealed 26 clones (52%) clones to be wild type and 14 clones (28%) with monoallelic editing in addition to 9 clones (18%) showing biallelic gene editing. One clone out of 50 (2%) was successfully knocked in. This specific clone was utilized to study the impact of the knocked in mutation on rRNA processing. It was found that 18S, 28S, 32S 32/28S ,5S rRNA are deranged in this specific clone compared to the control. Figure 3.26 and 3.27.

It was found that the relative RPS17 expression was reduced both in the knocked in clone and another knocked-out clone. Additionally, 18S,28S,32S and 5S rRNA was higher both in the knocked in and knocked-out clone compared to the control. This finding is consistant with what we have found from knocking out RPS17 in K562 cells in a previous experiment as shown in figure (3.21). Furthermore, these findings are consistent many studies that discovered the role of *RPS17* in pre-RNA processing (6,27,120,121)

***RPL11* c.475_476 del AA**

One novel frameshift deletion in *RPL11 was* reported in our lab by the aid of next generation sequencing (7) and required further validation. In this project this mutation *RPL11* c.475_476 del AA was firstly validated by Sanger sequencing. Secondly, *RPL11* c.475_476 del AA, was knocked in into K562 cells using the Amaxa's nucleofection system followed by timed sorting of single cells which were expressing GFP into 96 wells plate 72 hours after nucleofection with the simulatneous introduction of a double stranded homology template which contains the intended sequence change to be incorporated into the edited genome in addition to DNA ligase IV inhibitor SCR7 to block NHEJ pathway and increase the efficiency of HDR.

Three weeks following expansion in 37°C cell culture incubator supplemented with 5% $CO_2$, these single colonies were analyzed. The recovered viable colonies accounted for 1.5% of the original colonies (fifteen colonies were recovered out of 960). Sanger sequencing of each viable clone revealed four clones with wild type (26%) sequence and six clones with monoallelic editing (40%) in addition to four expressing biallelic gene editing. One clone (~6.5%) was successfully knocked in with

*RPL11* c.475_476 del AA mutation. In this experiment the HDR efficiency was improved from 2% in the previous *RPS17* c.3 G>C knock in experiment to 6.5% in this experiment. This result can be explained by the simultaneous addition of DNA ligase IV inhibitor SCR7 together with the HDR template.

This specific clone was utilized to study the impact of the knocked in mutation on rRNA processing. It was found that 18S, 28S, 32S, and 5S rRNA were higher in the knocked in clone compared to control, while the 32S/28S rRNA ratio was lower in the *RPL11* c.475_476 del AA knocked in clone compared to control as shown in figure 3.28. Although this is a novel mutation but it agrees with several studies that stated that *RPL11* plays an essential role in pre-rRNA processing. Interestingly, this finding is consistent with the rRNA profile that was explored in patients' T cells. In this study, it was found that 32S rRNA and 28S rRNA are both elevated in patients' T cells compared to the control group. Furthermore, the rRNA profile observed in the knocked-in clone is consistent with what was found in this study with *RPL11* knock out clone. Moreover, RPL*11* c.475_476 del AA was associated with an increased 32S rRNA which is consistent with the findings of Gazda et al (1) who studied lymphoblastoid cells derived from patients with *RPL11* mutation. Additionally, our findings concerning 32S rRNA elevation in *RPL11* c.475_476 del AA clone is also consistent with Farrar et al (28) who studied Con A stimulated MNC from DBA patients with *RPL11* mutation. In conclusion, K562 cells showed a consistent rRNA profile both in the knocked out clone as well as the *RPL11c*.475_476 del AA Knock in clone.

Our third objective was to perform whole exome sequencing for all DBA patients with unidentified mutation. However, this objective was omitted later on because of the emergence of the 100,000 genome project and

the involvement of our patients to this project, in additional to the financial expenses which are beyond the budget of this study

**Future work**

Continuing the process of DBA sample collection, T cell isolation and rRNA studies both via the RNA capillary electrophoresis technique using the Agilent bioanalyzer and also by RT-qPCR studies. This will further validate our hypothesis of using rRNA as a way to diagnose DBA patients and affected family members as our study confirmed the presence of significant rRNA processing defect even in patients with non-identified RP gene mutation.

Trying to knock out and knock in more novel mutations on k562 cells using CRISPR Cas 9 technology, growing the resultant clones and treating them with hemin to study the impact of these mutations on embryonic, foetal and adult globin gene expression, as our preliminary data suggested a strong correlation between RP gene mutations and globin gene expression which can also explain the rationale behind the absence of anaemia in utero in most of DBA patients that warrant further investigations.

Whole genome sequencing was part of our project but we are awaiting the results of the 100,000 genome project as some of our patients with non RP gene mutation are involved in this project. The resultant mutations will be validated on k562 cells by CRISPR/Cas 9 technology for monoallelic knock out and knock in.

Mass spectrometry for the knocked out and knocked in clones will be essential to study the impact of RP gene mutation on the post translational modifications, protein complex formation and also ribosomal protein interactions. This can also help in understanding DBA

pathogenesis and hence to identify patients with non RP gene mutation. Another technique that will be of great help in understanding DBA pathogenesis and the interaction and association between RP and rRNA is to perform immunoprecipitation of the knocked in or knocked out RP and then perform RT-qPCR to the rRNA attached to this protein and further studies to localise the affected RP inside the cell to help understand the impact of the mutation on the protein movement inside the cell by con focal microscopy in a similar way to the study done by vollumy et al.

On the mRNA level, more studies will be carried out on the edited clones like RNA sequencing to enhance the understanding of the effect of *RPL11, RPS19, RPS29* and *RPS17* knock out on other genes which will put the light on more genes involved in DBA and can help in future diagnosis of DBA through adding more genes to the NGS panel.

ATAC-sequencing (Assay for Transposase-Accessible Chromatin using sequencing) to study the influence of RP gene mutation on chromatin accessibility leading to a better understanding of DBA pathogenesis.

General protein quantification in the edited clones together with studying specific proteins like GATA1 by western blot technique.This will validate the hypothesis that RP gene mutations can affect GATA1 at the translational level rather than the transcriptional level.

Protein crystallography to study protein structure in the edited k562 clones especially the clones with the knock in mutations. This will help in validating these mutations and understand the effect on the protein structure. Furthermore, mass spectrometry applied on the edited k562 clones will be of great benefit in exploring the effect of these mutations on the post translational modifications, protein complex

formation and protein-protein interactions.

Immunoprecipitation of the ribosomal proteins in k562 edited clones followed by RT-qPCR for rRNA will provide a better view of the relationship between RP and rRNA. Furthermore, visualising the RP as they move inside and outside the nucleus by the aid of confocal microscope can give a clue about the functional abnormality of that protein following genetic mutation.

Studying the effect of steroids on DBA. This project has resulted in the generation of a protocol to validate genetic mutation in k562 cells. Thus the resultant clones that simulate DBA can be used to study the effect of steroids on the proliferation, rRNA processing, gene expression together with the translational and post translational levels.

## References


1. Gazda HT, Sheen MR, Vlachos A, Choesmel V, O'Donohue M-F, Schneider H, et al. Ribosomal Protein L5 and L11 Mutations Are Associated with Cleft Palate and Abnormal Thumbs in Diamond-Blackfan Anemia Patients. Am J Hum Genet. 2008 Dec;83(6):769–80.
2. Farrar JE, Nater M, Caywood E, Mcdevitt MA, Kowalski J, Takemoto CM, et al. Abnormalities of the large ribosomal subunit protein , Rpl35a , in Diamond-Blackfan anemia. Blood. 2008 Sep 1; 112(5): 1582–1592.
3. Gazda HT, Preti M, Sheen MR, O'Donohue M-F, Vlachos A, Davies SM, 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 Jul;33(7):1037–44.
4. Farrar JE, Vlachos A, Atsidaftos E, Carlson-donohoe H, Markello TC, Arceci RJ, et al. Ribosomal protein gene deletions in Diamond-Blackfan anemia. Blood. 2011 Dec 22;118(26):6943-51.
5. Doherty L, Sheen MR, Vlachos A, Choesmel V, O'Donohue M-F, Clinton C, et al. Ribosomal Protein Genes RPS10 and RPS26 Are Commonly Mutated in Diamond-Blackfan Anemia. Am J Hum Genet. 2010 Feb;86(2):222–8.
6. Cmejla R, Cmejlova ÃJ, Handrkova H, Petrak J, Pospisilova D. Ribosomal Protein S17 Gene ( RPS17 ) is Mutated in Diamond-Blackfan Anemia. Hum Mutat. 2007;28(July):1178–82.
7. 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 Aug;162(4):530–6.

8. Robledo S, Idol R a, Crimmins DL, Ladenson JH, Mason PJ, Bessler M. The role of human ribosomal proteins in the maturation of rRNA and ribosome production. RNA. 2008 Sep;14(9):1918–29.
9. Hanna T. Gazda1, MeeRie Sheen2*, Leana Doherty2*, Adrianna Vlachos3, Valerie Choesmel4*, Marie-Francoise O'Donohue4*, Hal E. Schneider2*, Catherine Clinton2*, Colin A. Sieff5, Peter E. Newburger6, Sarah E. Ball7*, Edyta Niewiadomska8*, Michal Matysiak8*, P-EG and AHB. Ribosomal Protein Genes S10 and S26 Are Commonly Mutated in Diamond-BlackfanAnemia. Oral Poster Abstr Oral Sess Bone Marrow Fail I. :293–296 (Ernest N. Morial Convention Center).
10. Draptchinskaia N, Gustavsson P, Andersson B, Pettersson M, Willig T, Dianzani I, et al. The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. Nat Genet. 1999;21(february):169–75.
11. Gazda HT, Grabowska A, Merida-long LB, Latawiec E, Schneider HE, Lipton JM, et al. Ribosomal Protein S24 Gene Is Mutated in Diamond-Blackfan Anemia.
12. Ellis SE. Shwachman Diamond Syndrome—Phenotypes and Genotypes: When Clinical Research Informs Biology. Pediatr Blood Cancer. 2008 Oct;51(4):449-50.
13. Parrella S, Aspesi A, Quarello P, Garelli E, Pavesi E, Carando A, et al. Loss of GATA-1 Full Length as a Cause of Diamond – Blackfan Anemia Phenotype. Pediatr Blood Cancer. 2014 Jul;61(7):1319-21.

14. Ramanjaneyulu Allam, Vijaykumar Chennupati, Diogo F.T. Veiga, Kendle M Maslowski, Aubry Tardivel, Manfredo Quadroni, Michel Duchosal, H Robson MacDonald, Nicolas Fasel AA-S. An Unexpected Role for Ribonuclease Inhibitor (RNH1) in Erythropoiesis. p. Oral Session:Hematopoiesis: Epigenetic, Transcript.
.December 8, 2014 in 2010.
15. Horos R, von Lindern M. Molecular mechanisms of pathology and treatment in Diamond Blackfan Anaemia. Br J Haematol. 2012 Dec;159(5):514–27.
16. Farrar JE, Dahl N. Untangling the phenotypic heterogeneity of Diamond Blackfan anemia. Semin Hematol. 2011 Apr;48(2):124–35.
17. Vlachos A, Blanc L, Lipton JM. Diamond Blackfan anemia: a model for the translational approach to understanding human disease. Expert Rev Hematol. 2014 Jun;7(3):359–72.
18. Iskander D, Psaila B, Gerrard G, Chaidos A, En Foong H, Harrington Y, et al. Elucidation of the erythroid progenitor defect in Diamond-Blackfan Anemia by characterization and prospective isolation of human erythroid progenitors. Blood. 2015 Apr 16;125(16):2553-7.
19. Vlachos A, Ball S, Dahl N, Alter BP, Sheth S, Ramenghi U, et al. Diagnosing and treating Diamond Blackfan anaemia: results of an international clinical consensus conference. Br J Haematol. 2008 Sep;142(6):859–76.
20. Backer K. Elevated red cell adenosine deaminase activity: a marker of disordered erythropoiesis. Br J Haematol. 1988 Feb;68(2):165-8.

21. Backer K, Alter BP, Glader B. Erythrocyte Adenosine Deaminase: Diagnostic Value for Diamond-Blackfan Anaemia. Br J Haematol. 2013 Feb;160(4):547-54.
22. Balaban EP, Graham M, Frenkel EP. Diamond-Blackfan Syndrome in Adult Patients. Am J Med.1985 Mar, 79: 533–8.
23. Vlachos A, Muir E. How I treat Diamond-Blackfan anemia. Blood. 2010;116(19):3715–23.
24. Da Costa L, O'Donohue M-F, van Dooijeweert B, Albrecht K, Unal S, Ramenghi U, et al. Molecular approaches to diagnose Diamond-Blackfan anemia: The EuroDBA experience. Eur J Med Genet [Internet]. 2017 Oct 26 [cited 2018 Mar 12];
25. Matsson H, Klar J, Draptchinskaia N, Gustavsson P, Carlsson B, Bowers D. Truncating ribosomal protein S19 mutations and variable clinical expression in Diamond-Blackfan anemia. Hum Genet. 1999 Nov;105(5):496-500.
26. Boria I, Garelli E, Gazda HT, Aspesi A, Quarello P, Pavesi E, et al. The ribosomal basis of Diamond-Blackfan Anemia: mutation and database update. Hum Mutat. 2010 Dec;31(12):1269–79.
27. Landowski M, O'Donohue MF, Buros C, Ghazvinian R, Montel-Lehry N, Vlachos A, et al. Novel deletion of RPL15 identified by array-comparative genomic hybridization in Diamond-Blackfan anemia. Hum Genet. 2013;132(11):1265–74.
28. Farrar JE, Quarello P, Fisher R, O'Brien K a, Aspesi A, Parrella S, et al. Exploiting pre-rRNA processing in Diamond Blackfan anemia gene discovery and diagnosis. Am J Hematol. 2014 Oct;89(10):985–91.

29. Wang R, Yoshida K, Toki T, Sawada T, Uechi T, Okuno Y, et al. Loss of function mutations in RPL27 and RPS27 identified by whole-exome sequencing in Diamond-Blackfan anaemia. Br J Haematol. 2015 Mar;168(6):854–64.
30. Mirabello L, Macari ER, Jessop L, Ellis SR, Myers T, Giri N, et al. Whole-exome sequencing and functional studies identify RPS29 as a novel gene mutated in multicase Diamond-Blackfan anemia families. Blood. 2014 Jul 3;124(1):24–32.
31. Gripp KW, Curry C, Olney AH, Sandoval C, Fisher J, Chong JX-L, et al. Diamond-Blackfan anemia with mandibulofacial dystostosis is heterogeneous, including the novel DBA genes TSR2 and RPS28. Am J Med Genet A. 2014 Sep;164A(9):2240–9.
32. 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.
33. Amsterdam A, Sadler KC, Lai K, Farrington S, Bronson RT, Lees JA, et al. Many ribosomal protein genes are cancer genes in zebrafish. PLoS Biol. 2004 May;2(5):E139.
34. Matsson H, Davey EJ, Draptchinskaia N, Hamaguchi I, Ooka A, Levéen P, et al. Targeted disruption of the ribosomal protein S19 gene is lethal prior to implantation. Mol Cell Biol. 2004;24(9):4032–7.
35. Kuramitsu M, Sato-Otsubo A, Morio T, Takagi M, Toki T, Terui K, et al. Extensive gene deletions in Japanese patients with Diamond-Blackfan anemia. Blood. 2012 Mar 8;119(10):2376–84.
36. Quarello P, Garelli E, Brusco A, Carando A, Mancini C, Pappi P, 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 Dec;97(12):1813–7.

37. Wegman-Ostrosky T, Savage SA. The genomics of inherited bone marrow failure: from mechanism to the clinic. Br J Haematol. 2017 May;177(4):526-542.
38. Sankaran VG, Ghazvinian R, Do R, Thiru P, Vergilio J, Beggs AH, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. J Clin Invest. 2012 Jul;122(7):2439-43.
39. Danilova N, Gazda HT. Ribosomopathies: How a common root can cause a tree of pathologies. Dis Model Mech. 2015 Sep;8(9):1013-26.
40. Miller C, Alikian M, Harrington Y, Al-oqaily Q, Roberts I, Karadimitris A. CLINICAL AND GENETIC DIVERSITY IN DIAMOND-BLACKFAN ANAEMIA: AN UPDATE FROM THE UNITED KINGDOM. EHA Library. Iskander D. Jun 23, 2017; 181520; P233
41. Warren AJ. Molecular basis of the human ribosomopathy Shwachman-Diamond syndrome. Adv Biol Regul. 2018 Jan;67:109-127.
42. Dauwerse JG, Dixon J, Seland S, Ruivenkamp CAL, Van Haeringen A, Hoefsloot LH, et al. Mutations in genes encoding subunits of RNA polymerases i and III cause Treacher Collins syndrome. Nat Genet. 2011 Jan;43(1):20-2.
43. Vulliamy TJ, Marrone A, Knight SW, Walne A, Mason PJ, Dokal I. Mutations in dyskeratosis congenita: Their impact on telomere length and the diversity of clinical presentation. Blood. 2006 Apr 1;107(7):2680-5.

44. Conferences JB. North American Indian childhood cirrhosis (NAIC). 2007;
45. Flygare J, Aspesi A, Bailey JC, Miyake K, Caffrey JM, Karlsson S, et al. Human RPS19, the gene mutated in Diamond-Blackfan anemia, encodes a ribosomal protein required for the maturation of 40S ribosomal subunits. Blood. 2007 Feb 1;109(3):980–6.
46. Legrand P, Gazda HT, Gleizes P. Mutation of ribosomal protein RPS24 in Diamond-Blackfan anemia results in a ribosome biogenesis disorder. Hum Mol Genet. 2008 May 1;17(9):1253-63.
47. Henras AK, Plisson-chastang C. An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip Rev RNA. 2015 Mar-Apr;6(2):225-42.
48. Quarello P, Garelli E, Carando A, Mancini C, Foglia L, Botto C, et al. Ribosomal RNA analysis in the diagnosis of Diamond-Blackfan Anaemia. Br J Haematol. 2016 Mar;172(5):782-5.
49. Mills EW, Green R. Ribosomopathies: There's strength in numbers. Science. 2017 Nov 3;358(6363).
50. Ludwig LS, Gazda HT, Eng JC, Eichhorn SW, Thiru P, Ghazvinian R, et al. Altered translation of GATA1 in Diamond-Blackfan anemia. Nat Med. 2014 Jul;20(7):748–53.
51. Horos R, Ijspeert H, Pospisilova D, Sendtner R, Andrieu-soler C, Taskesen E, et al. Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. Blood. 2012 Jan 5;119(1):262-72.
52. Deisenroth C, Zhang Y. Ribosome biogenesis surveillance : probing the ribosomal protein-Mdm2-p53 pathway. Oncogene. 2010 Jul 29;29(30):4253-60.

53. Dutt S, Narla A, Lin K, Mullally A, Abayasekara N, Megerdichian C, et al. Haploinsufficiency for ribosomal protein genes causes selective activation of p53 in human erythroid progenitor cells. Blood. 2011 Mar 3;117(9):2567-76
54. Marechal V, Elenbaas B, Piete J, Nicolas J, Levine AJ. The Ribosomal L5 Protein Is Associated with mdm-2 and mdm-2-p53 Complexes. Mol Cell Biol. 1994 Nov;14(11):7414-20.
55. Bursa S, Cokari M, Pfannkuchen M, Orsoli I, Golomb L, Zhu Y, et al. Mutual protection of ribosomal proteins L5 and L11 from degradation is essential for p53 activation upon ribosomal biogenesis stress. Proc Natl Acad Sci U S A. 2012 Dec 11;109(50):20467-72.
56. Tanaka T, Watanabe M, Yamashita K. Potential therapeutic targets of TP53 gene in the context of its classically canonical functions and its latest non-canonical functions in human cancer. Oncotarget. 2018 Mar 23;9(22):16234-16247.
57. Stadanlick JE, Zhang Z, Lee S, Hemann M, Biery M, Carleton MO, et al. Developmental arrest of T cells in RpL22-deficient mice is dependent upon multiple p53 effectors. J Immunol. 2011 Jul 15;187(2):664-75.
58. Barlow JL, Drynan LF, Hewett DR, Holmes LR, Lane AL, Jolin HE, et al. A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q − syndrome. Nat Med. 2010 Jan;16(1):59-66.
59. Fumagalli S, Thomas G. The role of p53 in ribosomopathies. Semin Hematol. 2011 Apr;48(2):97-105.
60. Kessel R, Vlachos A, Lipton JM. Ribosomopathy association with colorectal cancer. . Gastroenterology. 2015 Jan;148(1):258.

61. Torihara H, Uechi T, Chakraborty A, Shinya M, Sakai N, Kenmochi N. Erythropoiesis failure due to RPS19 deficiency is independent of an activated Tp53 response in a zebrafish model of Diamond-Blackfan anaemia. Vol. 152, British Journal of Haematology. 2011. p. 648–54.
62. Bolze A, Mahlaoui N, Byun M, Turner B, Trede N, Ellis SR, et al. Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. Science (80- ). 2013;340(6135):976–8.
63. Kondrashov N, Pusic A, Stumpf CR, Shimizu K, C. AHS, Shijima J, et al. Ribosome mediated specificity in Hox mRNA translation and vertebrate tissue patterning Nadya. 2014;497(7451):628–32.
64. Touw IP, Cmejla R, Andrieu-Soler C, IJspeert H, Sendtner M, Sendtner R, et al. Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. Blood. 2011;119(1):262–72.
65. Gupta V, Warner JR. Ribosome-omics of the human ribosome. Rna. 2014;20(7):1004–13.
66. Yang G, Sau C, Lai W, Cichon J, Li W, et al. A draft map of the human proteome. Nature.2014;509:575–581.
67. Choesmel V[1], Fribourg S, Aguissa-Touré AH, Pinaud N, Legrand P, Gazda HT, Gleizes PE.. Mutation of ribosomal protein RPS24 in Diamond-Blackfan anemia results in a ribosome biogenesis disorder. Hum Mol Genet. 2008 May 1;17(9):1253-63.
68. Choesmel V, Bacqueville D, Rouquette J, Noaillac-Depeyre J, Fribourg S, Crétien A, et al. Impaired ribosome biogenesis in Diamond-Blackfan anemia. Blood. 2007 Feb 1;109(3):1275-83.

69. Léger-Silvestre I, Caffrey JM, Dawaliby R, Alvarez-Arias DA, Gas N, Bertolone SJ, et al. Specific role for yeast homologs of the diamond blackfan anemia-associated Rps19 protein in ribosome synthesis. J Biol Chem. 2005 Nov 18;280(46):38177-85.
70. Bolze A, Mahlaoui N, Byun M, Turner B, Trede N, Ellis SR, et al. Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. Science. 2013 May 24;340(6135):976-8.
71. Keersmaecker K De, Atak ZK, Li N, Vicente C, Clappier E, Porcu M, et al. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. Nat Genet. 2013 Feb;45(2):186–90.
72. Chennupati V, Veiga DFT, Maslowski KM, Andina N, Tardivel A, et al. Ribonuclease inhibitor 1 regulates erythropoiesis by controlling GATA1 translation. J Clin Invest. 2018 Apr 2; 128(4): 1597–1614.
73. Aspesi A, Borsotti C, Follenzi A. Emerging Therapeutic Approaches for Diamond Blackfan Anemia. Curr Gene Ther. 2018;18(6):327–335.
74. Ferrua F, Cicalese MP, Galimberti S, Giannelli S, Dionisio F, Barzaghi F, et al. Lentiviral haemopoietic stem/progenitor cell gene therapy for treatment of Wiskott-Aldrich syndrome: interim results of a non-randomised, open-label, phase 1/2 clinical study. Lancet Haematol. 2019 May;6(5):e239– e253.
75. Biffi A, Montini E, Lorioli L, Cesani M, Fumagalli F, Plati T, et al. Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. Science. 2013 Aug 23;341(6148): 1233158.

76. Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, et al. Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. Nature. 2010;467:318–322.
77. Negre O, Eggimann AV, Beuzard Y, Ribeil JA, Bourget P, Borwornpinyo S, et al. Gene Therapy of the β-Hemoglobinopathies by Lentiviral Transfer of the β a(T87Q) -Globin Gene. Hum Gene Ther. 2016;27(2):148–65.
78. Ariga T. A possible turning point in the hematopoietic stem cell gene therapy for primary immunodeficiency diseases? Lentiviral vectors could take the place of retroviral vectors. Expert Rev Clin Immunol. 2013;9(11):1015–8.
79. Aspesi A, Monteleone V, Betti M, Actis C, Morleo G, Sculco M, et al. Lymphoblastoid cell lines from Diamond Blackfan anaemia patients exhibit a full ribosomal stress phenotype that is rescued by gene therapy. Sci Rep. 2017 Sep 20;7(1):12010.
80. Hamaguchi I, Ooka A, Brun A, Richter J, Dahl N, Karlsson S. Gene transfer improves erythroid development in ribosomal protein S19-deficient Diamond-Blackfan anemia. Blood. 2002 Oct 15;100(8):2724–31.
81. Hamaguchi I, Flygare J, Nishiura H, Brun ACM, Ooka A, Kiefer T, et al. Proliferation deficiency of multipotent hematopoietic progenitors in ribosomal protein S19 (RPS19) - Deficient Diamond-Blackfan anemia improves following RPS19 gene transfer. Mol Ther. 2003 May;7(5 Pt ):613–22.
82. Flygare J, Olsson K, Richter J, Karlsson S. Gene therapy of Diamond Blackfan anemia CD34(+) cells leads to improved erythroid development and engraftment following transplantation. Exp Hematol. 2008 Nov;36(11):1428–35.

83. Kamio T, Gu B, Olson TS, Zhang Y, Mason PJ. Mice with a Mutation in the Mdm2 Gene That Interferes with MDM2 / Ribosomal Protein Binding Develop a Defect in Erythropoiesis. PLoS One. 2016 Apr 4;11(4):e0152263.
84. Jaako P, Debnath S, Olsson K, Modlich U, Rothe M, Schambach A, et al. Gene therapy cures the anemia and lethal bone marrow failure in a mouse model of RPS19-deficient Diamond-Blackfan anemia. Haematologica. 2014 Dec;99(12):1792–1798.
85. Garçon L, Ge J, Manjunath SH, Mills JA, Apicella M, Parikh S, et al. Ribosomal and hematopoietic defects in induced pluripotent stem cells derived from Diamond Blackfan anemia patients. Blood. 2013 Aug 8;122(6):912–21.
86. Merlin S, Cannizzo ES, Borroni E, Bruscaggin V, Schinco P, Tulalamba W, et al. A Novel Platform for Immune Tolerance Induction in Hemophilia A Mice. Mol Ther. 2017 Aug 2;25(8):1815–1830.
87. Cantore A, Ranzani M, Bartholomae CC, Volpin M, Valle P Della, Sanvito F, et al. Liver-Directed Lentiviral Gene Therapy in a Dog Model of Hemophilia B. Sci Transl Med. 2015 Mar 4;7(277):277ra28.
88. Song B, Fan Y, He W, Zhu D, Niu X, Wang D. Improved Hematopoietic Differentiation Efficiency of Gene-Corrected Beta-Thalassemia Induced Pluripotent Stem Cells by CRISPR / Cas9 System. Stem Cells Dev. 2015 May 1;24(9):1053-65.
89. Xu P, Tong Y, Liu XZ, Wang TT, Cheng L, Wang BY, et al. Both TALENs and CRISPR/Cas9 directly target the HBB IVS2-654 (C > T) mutation in β-thalassemia-derived iPSCs. Sci Rep. 2015 Dec;5(1):12065.

90. D'Allard DL, Liu JM. Toward RNA Repair of Diamond Blackfan Anemia Hematopoietic Stem Cells. Hum Gene Ther. 2016 Oct;27(10):792–801.
91. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature. 2012 Nov 15;491(7424):454–7.
92. Donati G, Peddigari S, Mercer CA, Thomas G. 5S Ribosomal RNA Is an Essential Component of a Nascent Ribosomal Precursor Complex that Regulates the Hdm2-p53 Checkpoint. Cell Rep. 2013;4(1):87–98.
93. Li L, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, et al. Small dsRNAs induce transcriptional activation in human cells.Proc Natl Acad Sci U S A. 2006 Nov 14;103(46):17337-42
94. Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. Nature Chemical Biology.2007 January 28;3:166–173.
95. Portnoy V, Hua S, Lin S, Li KH, Burlingame A, Hu Z, et al. saRNA-guided Ago2 targets the RITA complex to promoters to stimulate transcription. Cell Research.2016;26:320–335.
96. Wang H, Russa M La, Qi LS. CRISPR / Cas9 in Genome Editing and Beyond. Annu Rev Biochem. 2016 Jun 2;85:227-64.
97. Rouet P, Smih F, Jasin M. Introduction of Double-Strand Breaks into the Genome of Mouse Cells by Expression of a Rare-Cutting Endonuclease. Mol Cell Biol. 1994 Dec; 14(12): 8096–8106.

98. Rouet P, Smih F, Jasin M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci U S A. 1994 Jun 21; 91(13): 6064–6068.
99. Silva G, Poirot L, Galetto R, Smith J, Montoya G, Duchateau P, et al. Meganucleases and Other Tools for Targeted Genome Engineering : Perspectives and Challenges for Gene Therapy. Curr Gene Ther. 2011 Feb;11(1):11-27. 100. Miller J, Mclachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 1985 Jun;4(6):1609-14.
101. Chandrasegaran S, Kim YG, Cha J .Hybrid restriction enzymes : Zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A. 1996 Feb 6; 93(3): 1156–1160.
102. Chandrasegaran S. Recent advances in the use of ZFN-mediated gene editing for human gene therapy. Cell Gene Ther Insights. 2017;3(1):33-41.
103. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. GENETICS. 2010 Oct 1; 186(2): 757-761.
104. Nicholas Burwick, Akiko Shimamura, and Johnson M. Liu. Non-DBA Disorders of Ribosome Function: Shwachman-Diamond Syndrome and 5q- Syndrome Semin Hematol. 2011 Apr; 48(2): 136–143.
105. Jansen R, Embden JDA Van, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. 2002;43(6):1565–1575.

106. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats ( CRISPRs ) have spacers of extrachromosomal origin. Microbiology. 2005 Aug;151(Pt 8):2551-61.
107. Church GM, Esvelt MK, Mali P, et al.Orthogonal Cas9 Proteins for RNA-Guided Gene Regulation and Editing. Nat Methods. 2013 Nov; 10(11): 1116–1121.
108. Wu Y, Liang D, Wang Y, Bai M, Tang W, Bao S, et al. Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9. Cell Stem Cell. 2013 Dec 5;13(6):659-62.
109. Hu Z, Shi Z, Guo X, Jiang B, Wang G, Luo D, et al. Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR – Cas9 and ssODN in human cancer cells. Cell Biosci. 2018 Feb 19;8:12
110. Hsu PD, Lander ES, Zhang F, Sciences C, Biology C. Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell. 2014 Jun 5;157(6):1262–78.
111. Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system : Biology , mechanisms and applications. Biochimie. 2015 Oct;117:119–28.
112. You L, Tong R, Li M, Liu Y, Xue J, Lu Y. Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. Mol Ther Methods Clin Dev. 2019 Mar 15;13:359–370.
113. Agarwala V, Li Y, Fine EJ, Wu X. DNA targeting specificity of RNA-guided Cas9 nucleases Patrick.Nature Biotechnology.2013 Jul;31:827–832.

114. Kleinstiver BP, Prew MS, Tsai SQ, Topkar V, Nguyen T, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities Benjamin. Nature.2015;523:481–485.
115. Topkar V V, Zheng Z, Joung JK. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol. 2015 Dec;33(12):1293–1298.
116. Horvath P, Romero DA, Cou A, Richards M, Moineau S, Boyaval P, et al. Diversity , Activity , and Evolution of CRISPR Loci in Streptococcus thermophilus. J Bacteriol. 2008 Feb;190(4):1401–12.
117. Schoen C, Vogel J, Sontheimer EJ. Processing-Independent CRISPR RNAs Limit Natural Transformation in Neisseria meningitidis. Mol Cell. 2013 May 23;50(4):488–503.
118. Ferrer M, Henriet S, Chamontin C, Lainé S, Mougel M. From Cells to Virus Particles : Quantitative Methods to Monitor RNA Packaging. Viruses. 2016 Aug; 8(8):239.
119. Thomas P, T TGS. HEK293 cell line : A vehicle for the expression of recombinant proteins. J Pharmacol Toxicol Methods. 2005 May-Jun;51(3):187–200.
120. Farrar JE, Vlachos A, Atsidaftos E, Carlson-Donohoe H, Markello TC, Arceci RJ, et al. Ribosomal protein gene deletions in Diamond-Blackfan anemia. Blood . 2011 Dec 22;118(26):6943–51.
121. Gleizes P. Functional dichotomy of ribosomal proteins during the synthesis of mammalian 40S ribosomal subunits. J Cell Biol. 2010 Sep 6;190(5):853–66.

# APPENDIX

| Number | Age(Years)/Gender | mutation |
|---|---|---|
| S1 | 10/M | RPS19 c.71+1 G>T |
| S2 | 20/F | RPS19 c.172+2T>G |
| S3 | 7/F | RPL5 c.367delG |
| S4 | 25/F | RPL11 BIG DELETION |
| S5 | 11/M | RPS15 c.244G>C |
| S6 | 22/F | RPL5 c.3+3G>C |
| S7 | 8/F | RPS19 c.361C>G |
| S8 | 19/M | RPS26 c.30+1G>A |
| S9 | 19/M | unknown |
| S10 | 6/F | rps26 BIG DELETION |
| S11 | 12/F | RPS10 c.337C>T |
| S12 | 11/F | None detected |
| S13 | 10/M | RPS24 c.46C>T |
| S14 | 13/F | None detected |
| S15 | 22/F | RPS17 c.159 T>G |
| S16 | 49/F | RPS17 c.159 T>G |
| S17 | 5/F | None detected |
| S18 | 33/M | RPS17 c.3G>C |
| S19 | 34/F | RPS26 deletion |
| S20 | 14/F | RPL5 c.535C>T |
| S21 | 15/M | RPS26 c.55C>T |
| S22 | 23/F | None detected |
| S23 | 5/M | None detected |
| S24 | 27/F | RPL11 c.44-45delTTinsCCCATC |
| S25 | 12/M | RPL11 c.475_476delAA |
| S26 | 12/F | not detected |
| S27 | 5/M | RPS19 |
| S28 | 2/F | RPS29 c.139G>C |
| S29 | 4/F | None detected |
| S30 | 35/M | None detected |
| S31 | 10/M | None detected |
| S32 | 32/F | RPL11 c.406-407insGGGACAGGT |
| S33 | 6/M | PS29 c.63-3C>A |
| S34 | 5/M | RPS19 c.71+5G>T, c.500G>A |
| S35 | | RPS19 big deletion |
| S36 | 8/M | None detected |
| S37 | 19/M | RPS17 c.159T>G |
| S38 | 17/F | RPS19 big deletion |
| S39 | 35/M | RPS29 c.139G>C |
| S40 | 4/M | RPS29 c.139G>C |
| S41 | 13/M | RPL11 c.475_476delAA |
| S42 | 12/F | RPL5 c.169_172delAACA |
| S43 | 40/F | RPS10 c.337C>T |
| S44 | 28/F | None detected |
| S5 | 44/F | RPL11 c.60_61delCT |
| S46 | 16/M | RPS26 c.3+1G>C |
| S47 | 8/M | none detected |
| S48 | 36/F | none detected |
| S49 | 10/M | rpl35A big deletion |

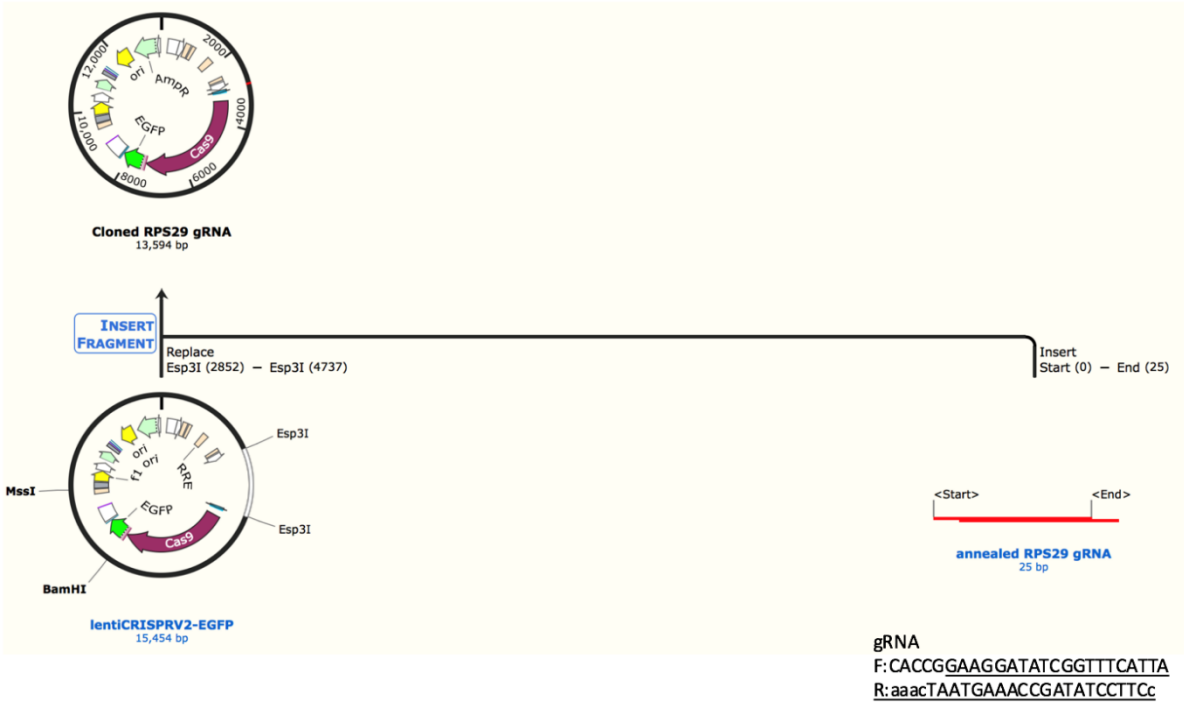


Figure: Illustration of cloning gRNA into the CRISPR Vector

## Table: Sequence of primers ordered from Sigma.

| | |
|---|---|
| 5s rRNA F | . CTAGCTGCGAGAATTAATGTG |
| 5s rRNA R | GAAGTGTCGATGATCAATGTG |
| 18s rRNA F | TTGCGTTGATTAAGTCCCT |
| 18s rRNA R | TCACTAAACCATCCAATCGG |
| 28s rRNA F | CCTCACGATCCTTCTGAC |
| 28s rRNA R | CCACAAGCCAGTTATCCC |
| 32S rRNA F | CGATTCCGTCCGTCCGTC |
| 32S rRNA R | TTAAATTCAGCGGGTCGCCA |
| TBP F | CAGCAACTTCCTCAATTCCTTG |
| TBP R | GCTGTTTAACTTCGCTTCCG |
| TFRC F | CCACCAAACAAGTTAGAGAATGC |
| TFRC R | TCAGAGCGTCGGGATATCG |
| ACTB F | TTGGCAATGAGCGGTTCC |
| ACTB R | GTTGAAGGTAGTTTCGTGGATG |
| IPO8 R | CGAAGTTGCGGATTGCAG |
| IPO8 F | GAATTCCACATGGTCAGAGACT |
| HBG1 F | AGACGCCATGGGTCATTCA |
| HBG1 R | ACCTTCTTGCCATGTGCCTT |
| HBZ F | ACTTCAAGCTCCTGTCCCAC |
| HBZ R | CTCAGCGGTACTTCTCGGTC |
| HBA 1F | AAGGTCGGCGCGCACGCT |
| HBA1 R | CTCAGGTCGAAGTGCGGG |
| HBB F | CTCATGGCAAGAAAGTGCTCG |
| HBB R | AATTCTTTGCCAAAGTGATGGG |
| RPS17 ex 1 F | AGGGCCCGGCTAAACAGT |
| RPS17 ex 1 R | GGAGCGGAAAGGGAGGAG |
| RPS17 ex 2F | GTTTAGCCGGGCCCTCTC |
| RPS17 ex 2 R | AGGCTAACGAAACCACCAAG |
| RPS17 ex 3 F | TGTGTAGGAGGTCCCAGGAT |
| RPS17 ex 3R | CCCACAGCTGGTTATGGAAT |
| RPS17 ex 4 F | GCTATCCTTGTGCAGCTCCT |
| RPS17 ex 4 R | GCACACAACTTCCGCTACAC |
| RPS17 ex 5 F | AGGCATTGAGAACCCTGTTG |
| RPS17 ex 5 R | GCACCTGAATCCCAGACAAG |
| RPL 11 F | TTCCAGCCTCCAAAGCAAAC |
| RPL 11 R | CTGAACTGCACTATCATCTGA |
| RPS 19 F | CCAGCTCGTTAGAATGCACC |
| RPS 19 R | TTTCTCCTGTCTCTGGGCAG |
| RPS29 F | TGCGGGTAGTAGCCGTCTGA |
| RPS29 R | CACTCCAGGGAAGATCTGTC |

**Table: Sequence of guide RNA used in CRISPR experiments ordered from Sigma-Aldrich.**

| gRNA | sequence | complementary sequence | PAM |
|---|---|---|---|
| RPS17 gRNA KNOCK IN | c.3G>C p.Met1Ile CACCG/TTACCAAGGACCCGCCAACA | AAAC/TGTTGGCGGGTCCTTGGTAAc | TGG |
| RPS17 gRNA KNOCK OUT | CACCG/TTACCAAGGACCCGCCAACA | AAAC/TGTTGGCGGGTCCTTGGTAAc | TGG |
| RPS19 gRNA KNOCK OUT | CACCG/GGCCGCAAACTGACACCTC | AAAC/GAGGTGTCAGTTTGCGGCc | AGG |
| RPS29 | CACCG/GAAGGATATCGGTTTCATTA | AAAC/TAATGAAACCGATATCCTTCc | AGG |

| | | | |
|---|---|---|---|
| gRNA KNOCK OUT | | | |
| RPL11 gRNA KNOCK IN | c.475-476Del AA CACCG/ACACAGAATCAGCAAAGAGG | AAAC/CCTCTTTGCTGATTCTGTGTc | AGG |
| RPL11 Knock out | CACCG/CATGCGCTGGTTCCAGCAGA | AAAC/TCTGCTGGAACCAGCGCATGc | AGG |

***RPS17 c.3G>C c.3G>C; p.Met1Ile Knock in template:***

CTGGCCTAAGCTTTAACAGGCTTCGCCTGTGCTTCCTGTTTCCTCTTTTACCAAGGACCCGCCAACATCGTAGG
TGTTTTGGCTCGAGGCCGCCATCCTCCACAATCGTCCCCCATCTG

***RPS17 c.3G>C c.3G>C; p.Met1Ile Knock in complementary template***

CAGATGGGGGACGATTGTGGAGGATGGCGGCCTCGAGCCAAAACACCTACGATGTTGGCGGGTCCTTGGTAA

AAGAGGAAACAGGAAGCACAGGCGAAGCCTGTTAAAGCTTAGGCCAG

***RPL11 c476del AA knock in Template***

TTCAGCATCGCAGACAAGAAGCGCAGGACAGGCTGCATTGGGGCCAAACACAGAATCAGCAGAGGAGGCCAT
GCGCTGGTTCCAGCAGAAGGTAAAGCTGATTTATCTCAAGTGAAGTGG

***RPL11 c476 del AA knock in complementary template :***

ACCACTTCACTTGAGATAAATCAGCTTTACCTTCTGCTGGAACCAGCGCATGGCCTCCTCTGCTGATTCTGTGTT
TGGCCCCAATGCAGCCTGTCCTGCGCTTCTTGTCTGCGATGCTGA