

**Palacky University Olomouc**

**Diploma thesis**

**Olomouc 2014**

**Zuzana Macečková**

**Palacky University Olomouc**  
**Faculty of Science**  
**Departement of Cell Biology and Genetics**



**Molecular pathology of Diamond-Blackfan anemia**

**Diploma thesis**

**Zuzana Macečková**

Study program: Biology

Field of study: Molecular and cell biology

**Olomouc 2014**

**Supervisor: Mgr. Petr Vojta**

**Declaration**

I declare that I elaborate this diploma thesis independently under supervision of Mgr. Petr Vojta using information sources mentioned in the references.

In Olomouc:

.....

## Souhrn

Diamond-Blackfanova anémie je vzácná vrozená aplázie červené krevní řady, která je často doprovázena vrozenými malformacemi, zejména horních končetin a kraniofaciální oblasti. Příčinou tohoto onemocnění jsou mutace v genech pro ribozomální proteiny, které byly detekovány u přibližně 50 % případů. Mutace v ribozomálních genech vedou k defektům během ribozomální biogeneze a tím k ribozomálnímu stresu, který může vyústit v zastavení buněčného cyklu, případně v apoptózu erythroidních progenitorů. Tato práce měla dva hlavní cíle. Prvním byla identifikace molekulárních kauzalit vedoucích k fenotypu Diamond-Blackfanovy anémie u dvanácti pacientů suspektních na DBA. Pro tyto účely byla zvolena metoda celoexomového sekvenování. Po anotaci dat byly identifikovány mutace potenciálně vedoucích k tomuto onemocnění u šesti z nich. Druhým cílem bylo vyvinutí detekčního systému pro testování mutací identifikovaných jako možná příčina Diamond-Blackfanovy anémie na základě evaluace translační aktivity buněk. Tento systém by úspěšně otestován, ačkoli pro absolutní ověření spolehlivosti, je nutné provést ještě několik kontrolních měření.

## **Summary**

Diamond-Blackfan anemia is a rare congenital red cell aplasia, often associated with congenital malformations, particularly of upper limbs and craniofacial region. As a cause, mutations in genes for ribosomal proteins have been reported in approximately 50% of the cases. Mutations in ribosomal proteins lead to defects during ribosome biogenesis and thus ribosome stress, which in turn leads to cell cycle arrest or apoptosis in erythroid progenitors. This study was focused on two main objectives. First, identification of mutations leading to Diamond-Blackfan anemia phenotype in twelve DBA suspected patients. For this purpose whole exome sequencing was used. After data analysis, mutations potentially leading to this disease in six of them were identified. Development of detection system for testing mutations identified as a possible cause of Diamond-Blackfan anemia on the basis of the evaluation of translational activity of cells was the other objective. This system has been successfully tested, although for absolute effectiveness verification it needs to be done a more control measurements.

## **Acknowledgment**

I would like to thank Mgr. Petr Vojta for guidance, laboratory technicians for sample preparation, Mgr. Gabriela Rylová for introduction to cell cultures handling, Mgr. Petr Konečný for flow cytometry output analysis, IT crowd for IT support and Khushboo Agrawal for advice and grammar correction.

## Contents

1. Background .....	8
2. Review .....	9
2.1 Ribosome structure, biogenesis and function .....	9
2.2 Ribosomal stress .....	14
2.3 Diamond-Blackfan anemia .....	20
3. Aim of work .....	26
4. Material and methods.....	27
4.1 Material.....	27
4.2 Exome sequencing .....	29
4.3 Functional study.....	31
5. Results .....	36
5.1 Exome sequencing .....	36
5.2 Functional study.....	44
6. Discussion.....	46
7. Conclusion .....	50
8. References.....	51
9. Abbreviations .....	64
Attachment .....	66

## 1. Background

Diamond-Blackfan anemia (DBA; #105650, OMIM) is a rare congenital erythroid aplasia first described in 1938 [1]. DBA is usually diagnosed in early infancy as a serious anemia characterised by reticulocytopenia and normocellular bone marrow with lack of erythroid precursors [2]. DBA patients can also present physical malformarities, particularly of the upper limb and craniofacial region, and an increased risk of cancer development, especially acute myeloid leukemia and osteogenic sarcoma [3]. Due to hard prediction of the disease, treatment is challenging. The disease can be suppressed by steroid therapy and blood transfusions, however, the only currently available permanent treatment is the bone marrow transplantation [4].

DBA is the first described disease caused by defects in the ribosomal genes. In humans mutations in genes encoding ribosomal proteins of both small (RPS7, RPS17, RPS19, RPS26) and large (RPL5, RPL10, RPL11, RPL15, RPL26, RPL31) subunits of ribosome have been reported in approximately 50 % of cases, in other cases the cause is still elusive [5-13]. Mutations in ribosomal genes lead to incorrect ribosome assembly [14]. Free ribosomal proteins lead to p53 dependent and independent ribosomal stress, apoptosis in erythroid progenitor cells [15] and amended translation compared to healthy individuals [16, 17]. Recently, a new mutation in non-ribosomal protein was described as the cause of DBA [18].

## 2. Review

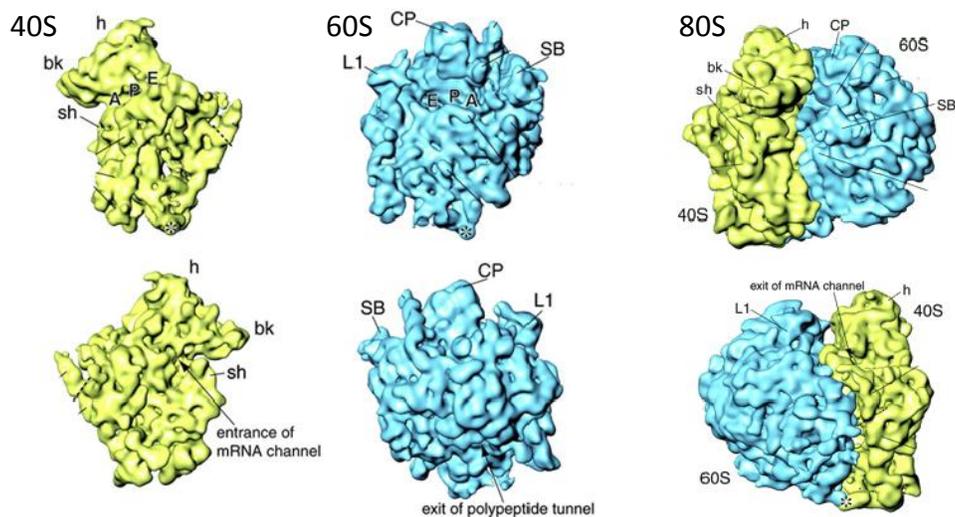
### 2.1 Ribosome structure, biogenesis and function

#### 2.1.1 Ribosome structure

Ribosome found in all living cells is a large multi-molecular manufactory responsible for protein synthesis and its correct assembly and function is essential for the cell cycle progression [19, 20].

Eukaryotic ribosome (80S) consists of two subunits, large subunit (60S) and small subunit (40S). Each composed of rRNAs and ribosomal proteins [21]. Large subunit consists of three rRNAs (5S, 28S and 5.8S) and 46 ribosomal proteins and small subunit is composed of one rRNA (18S) and 33 proteins. In both cases rRNA, on which appropriate ribosomal proteins anneal, is backbone of ribosomal structure. Characteristic structural features of 60S subunit are central protuberance and two stalks. 40S subunit possesses shoulder, neck and beak [22].

Main functional domains of ribosome are three binding sites (A, P and E). Site A is the position where aminocyl-tRNA interacts with mRNA, P holds extending polypeptide during elongation and site E is an exit place for used tRNA. Structure of ribosome is presented on Fig. 1 [23].



**Fig.1:** The 12-Å resolution cryo-EM density map of the ribosome: bk-beak, h-head, sh-shoulder, CP-central protuberance, L1-L1 stalk, SB-stalk base [23]

### 2.1.2 Ribosome assembly

Ribosome assembly is complex process, requiring many trans-activating factors like non-ribosomal proteins and snoRNA [24, 25] and it is initialised by rRNA genes expression.

Transcription of 47S rRNA, containing mature 18S, 5.8S and 28S rRNA separated by internal transcribed spacers (ITS), is performed by polymerase I and occurs in nucleolus in so-called NORs (nucleolus organizer regions). NORs contain several tandem copies of genes encoding ribosomal 47 rRNA. In humans, these regions are localized on the short arms of chromosome 13, 14, 15, 21 and 22 [27].

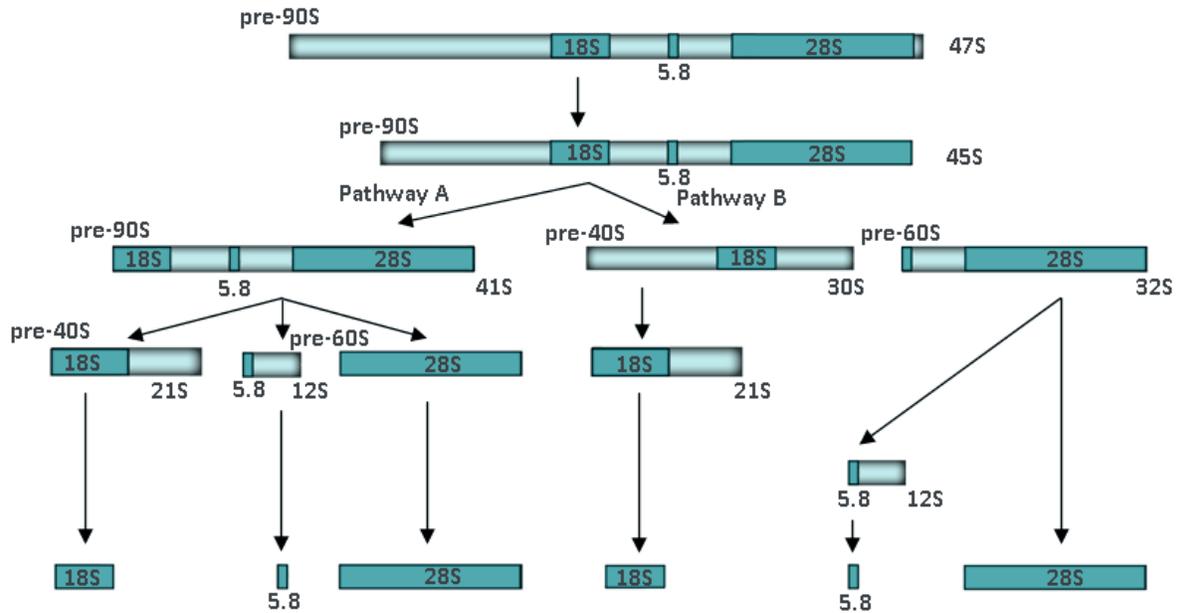
In contrary, transcription of 5S rRNA by is performed by polymerase III and takes place in nucleus [26]. After transcription, 5S rRNA is associated with ribosomal proteins RPL5 and RPL11 [28], which carry nucleolus location signal allowing transfer of 5S rRNA to nucleolus. In nucleolus 5S RNA assembles with 47S rRNA and its coupled proteins and together creates the first precursor of ribosome, 90S ribosome [29].

90S ribosome mainly consists of proteins and factors necessary for 40S formation and almost completely lacks components of 60S [30]. Cleavage of ITS between sequences of pre-18S and 5.8S rRNA leads to decay of 90S ribosome into two parts, pre-40S and pre-60S ribosome. Advanced maturation of these particles is independent and final assembly occurs in cytoplasm [31].

In this stage, pre-40S subunit attained the typical features of mature 40S ribosome (only beak structure is missing) and is quickly transferred to cytoplasm where beak is formed by stable association of pre-40S ribosome with RPS3 [32]. The final step of 40S-ribosome assembly is transformation of 20S rRNA to 18S rRNA by cleavage [33]. All together, the process is simple and requires only low number of factors [34].

However, maturation of 60S subunit is far more complicated [35]. In contrast to small subunit, processing of large subunit rRNAs, except for 5.8S rRNA, takes place in nucleolus. Into cytoplasm pre-60S subunit is transferred with already mature rRNAs [36]. In cytoplasm pre-60S ribosome undergoes wide range of modifications before it is prepared to join 40S ribosome. Last step in 60S subunit maturation is trimming of 3'end of 5.8S rRNA [37]. Processing of 47S rRNA is summarised on Fig.2.

For the formation of fully functional ribosome, after separate maturation of both subunits in cytoplasm, both subunits are joined during final steps of translation initiation [38]. Besides, the process of ribosome assembly requires high energy and therefore has to be under strict control [39], for instance by Pol I activity regulation [40].



**Fig.2:** Summarizes scheme of 47S pre-rRNA processing in human cells

### 2.1.3 Translation mechanism

Translation is generally composed of three parts: initiation, elongation and termination.

Cap-dependent initiation where 5' end of mRNA is modified by 7-methylguanosine cap is a common mechanism in eukaryotic cells. Ribosomal subunits which are not translationally active are separated in cytoplasm and their binding is blocked by initiation factor eIF3, which is bound to 40S ribosome, and eIF6, which lays on the surface of 60S ribosome [41].

Before initiation two independent steps are necessary. First step is association of Met-tRNA, eIF2-GTP, eIF3 and 40S ribosome resulting in pre-initiation complex (43S ribosome). Second step includes annealing of translation factors eIF4E, eIF4A and eIF4G on 5' cap of mRNA resulting in formation of the eIF4F complex. Interaction between IF4G of eIF4F complex and

eIF3 protein on surface of 43S ribosome results in the complete formation of the initiation complex [42], see Fig.3.



**Fig.3:** Scheme of fully formed cap-dependent initiation complex (eIF4F)

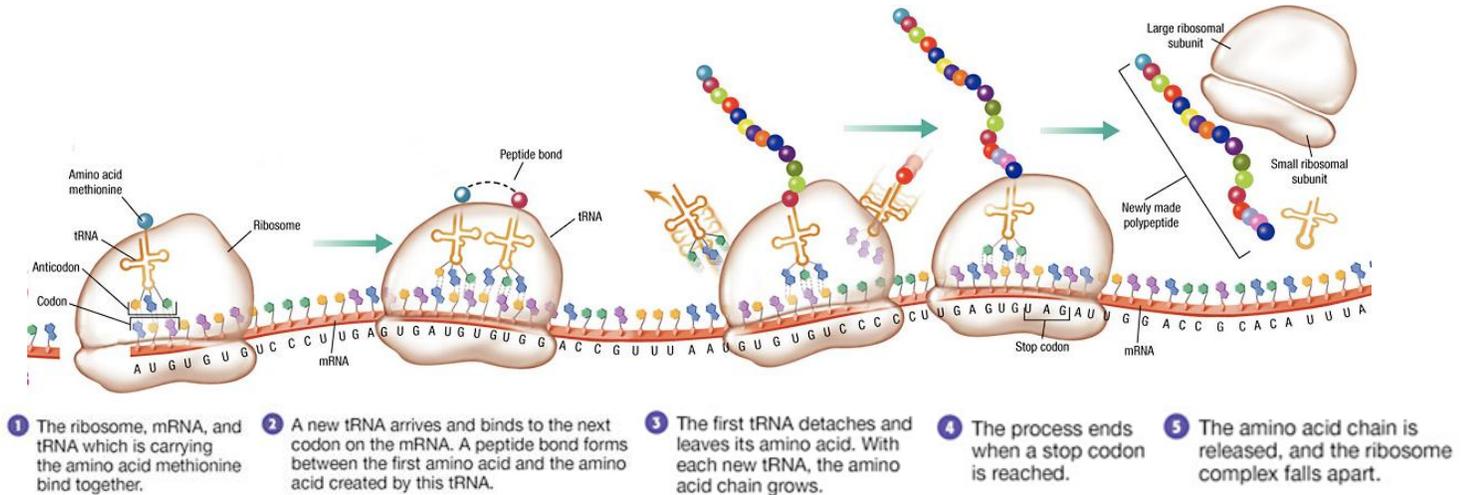
Subsequently, initiation complex moves along mRNA towards 3' end until start codon AUG is located. Hydrolysis of GTP from eIF2-GTP, followed by disengagement of several initiation factors from initiation complex stops the movement of the complex. In this stage, 60S ribosome anneals. This completes the whole process of translation initiation and marks the beginning of translation elongation [43].

Firstly, tRNA carrying corresponding amino acid anneals to the second codon and the peptide bond is formed. Then, ribosome is shifted by one codon towards 3' end of mRNA and on third codon annealing of matching tRNA takes place again and whole process continues [44]. Formation of peptide bond is facilitated by two elongation factors, eEF-1 and eEF-2[45].

Termination occurs when translation machinery encounters stop codon. In this stage, termination factors anneal, newly formed polypeptide and mRNA are released from the ribosome and ribosome itself falls apart into subunits [46]. Translation process is represented on Fig. 4.

Apart from cap-dependent process exists alternative way of translation initiation mediated by internal ribosome entry site (IRES). This strategy does not require multiple translation factors and scanning for start codon and is widely used by viruses [47]. The interaction between mRNA and 40S ribosome is driven by secondary structure of 5' end of mRNA, until now, four types of

IRES have been identified [48]. Transcripts which translation is driven by IRES elements, usually belongs to genes responsible for cell stress response, mitosis and cell death [49].



**Fig.4:** Translation in eukaryotic cell ([missevrardbio1.blogspot.cz/2012/04/intro-to-translation.html](http://missevrardbio1.blogspot.cz/2012/04/intro-to-translation.html))

#### 2.1.4 Extra-ribosomal functions of ribosomal proteins

Ribosomal proteins are mainly perceived as the building blocks of the ribosomes. Their additional functions, except for p53 association which will be discussed later, are left aside.

Ribosomal proteins, for example, modulate miRNA mediated repression of translation initiation [50]. Furthermore, presence of 54 different ribosomal proteins was proven in stress granules and P bodies, where mRNA is decomposed for purposes of translation regulation. This indicates that ribosomal proteins play an important role in mRNA turn over [51]. Ribosomal proteins also regulate splicing and processing of their own mRNA, for example, RPS13 binds to first intron of its mRNA and inhibits splicing [52].

GAIT complex mediates translation inhibition of range of transcripts and its activity is induced by gamma-interferon  $\text{IFN}\gamma$  [52]. Ribosomal protein RPL13a is component of this complex. RPL13a is phosphorylated under external signal, leaves 60S ribosome, joins GAIT complex and inhibits translation of specific transcripts [53]. Even that under these conditions,

ribosomes almost completely lack RPL13a, they still retain translation capacity and cell growth is not affected [54].

RPS3 protein is another example of extra ribosomal function of ribosomal proteins. RPS3 is involved in DNA base excision repair mechanism [55], it was revealed that increased activity of uracil-DNA glycosylase, the enzyme responsible for this kind of DNA repair, is initiated by RPS3 [56]. Additionally, RPS3 is partner of NF- $\kappa$ B complex and allows its binding to some parts of the genome and therefore the expression of wide range of the genes involved in immune reaction, development, cell growth and death [57].

Connection between ribosomal proteins and cancer development is also well established [58]. Haploinsufficiency of various ribosomal proteins leads to tumorigenesis in zebrafish through selective translation. Specifically, in spite of the presence of the p53 mRNA, p53 is not translated and therefore p53 cannot protect the cell [58].

## **2.2 Ribosomal stress**

Efficient and correct assembly of ribosome is essential for every living cell. This is reason why during whole process of ribosome biogenesis precursor particles are tested [59]. If ribosome biogenesis is impaired, various cascades of cell responses may start. Cell responses can have different executors and can lead from defective ribosome decomposition [59] to ribosomal stress leading to cell cycle arrest and apoptosis, if defect persists [60].

### **2.2.1 P53 dependent ribosomal stress**

P53 is the core protein of ribosome stress response. Balanced level of this protein is held by its systematic degradation in proteasome. MDM2 protein holds important position in this degradation, its E3 ubiquitin ligase activity allows binding of ubiquitin to p53 [61]. Ubiquitin is signalling molecule predetermining proteins for destruction in proteasome [62].

Furthermore, MDM2 binds to p53 DNA binding site and masks it. Under certain circumstances (DNA damage, oncogene activation, hypoxia, cell-cell contact and ribosomal stress) different molecules can bind to MDM2 and terminate its interaction with p53, this results in increased level of p53 within the cell [63].

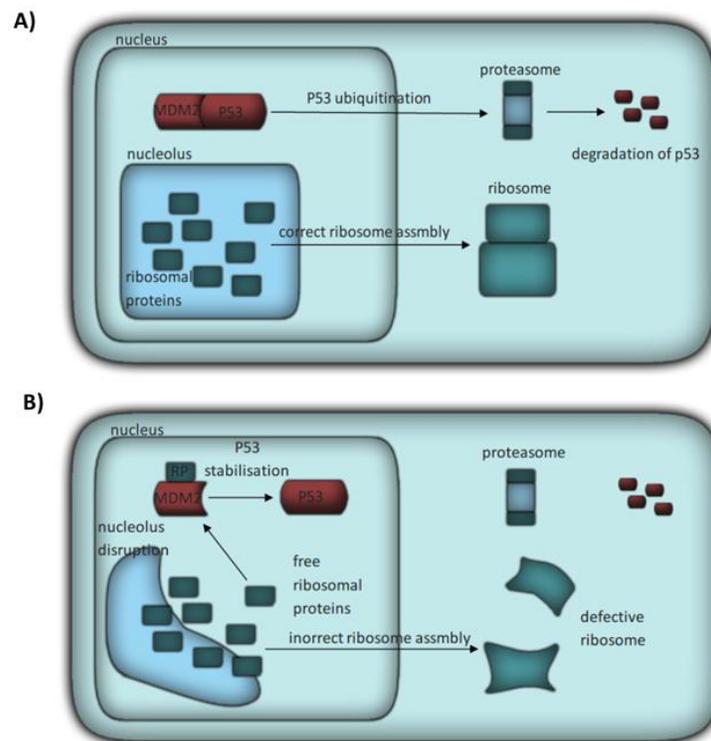
Human MDM2 is mainly nuclear protein composed of 491 amino acids with four major regions. The most important domain of MDM2 is localized on N-terminus at position 18-101, this domain interacts with p53 protein, masks p53 trigger point and prevent transcriptional co-activators annealing [64]. Acidic side and Zn finger side are localized at central region, these two domains are responsible for interaction between MDM2 and various proteins which modulate its activity [65]. At C-terminus of MDM2 fourth specific region, so called RING finger, is localized, this region is core of ubiquitin ligase activity of MDM2 [66].

RPL5 is the first described ribosomal protein interacting with MDM2 [67], followed by RPL11, RPS7, RPL26, RPS3, RPS14, RPS26 and RPL23 [68-73]. These proteins interact with MDM2 through its central region, although they bound to different sites and with various affinities.

It was revealed, that binding site of both RPL5 and RPL11 is Zn finger domain of MDM2. Natural variant C305F exists on this domain [74]. It prevents interaction between MDM2 and RPL5 and RPL11, on the other hand, it does not alter binding capacity of another ribosomal protein RPL23. Another variant in this site C305S was also tested [75]. This variant also removes annealing possibility for RPL11, however, it retains the binding capacity of RPL5. Complex associations between interacting ribosomal proteins and MDM2 reflect the importance of relationship between correct ribosome assembly and cell cycle progression.

MDM2 is nucleo-cytosolar protein, while biggest reservoir of the ribosomal proteins is in nucleolus. Under normal conditions amount of free ribosomal proteins does not have potential to trigger ribosomal protein-MDM2-p53 pathway, but under ribosomal stress, nucleolus is disrupted and free pool of ribosomal proteins is released to nucleus where they can associate with MDM2 [76].

Cause of nucleolus disruption can be divided into 3 categories: disruption of rRNA synthesis, disruption of rRNA processing and ribosomal protein imbalance [77]. In each case unbalanced level between rRNA and ribosomal proteins leads to accumulation of free ribosomal proteins and triggers cell response (Fig. 5). Moreover, disruption of rRNA gene expression together with down regulation of ribosomal protein synthesis does not activate p53 response [78]. It is the evidence that the cell is more sensitive to imbalance between rRNA and ribosomal proteins following by impair ribosome assembly than slower ribosomal synthesis.



**Fig.5:** p53-dependent nucleolar stress response: A) cell under normal circumstances, B) cell under nucleolar stress

For simulation of rRNA synthesis disruption models lacking necessary compartments of polymerase I machinery have been used. For instance, TIF-1A is essential polymerase I transcriptional cofactor. It was shown that inactivation of TIF-1A in mouse fibroblasts results in disintegration of nucleolus [79], p53 mediated cell cycle arrest and apoptosis. The importance of TIF-1A for preservation nucleolus integrity during mouse development has been also proven [80].

As described in ribosome assembly section, precursors of 18S, 28S and 5.8S are transcribed at once as components of 47S rRNA and this long precursor has to be precisely cleaved to obtain mature rRNAs. If this process is crippled, ribosome subunit assembly is damaged and accumulation of unprocessed precursors leads to nucleolus stress and disruption. Evidence can be seen in depletion of ribosome biogenesis protein Bop1. This protein is part of complex Pes1-Bop1-WDR12 [81], which is responsible for cleavage of 47S rRNA. Negative form Bop1

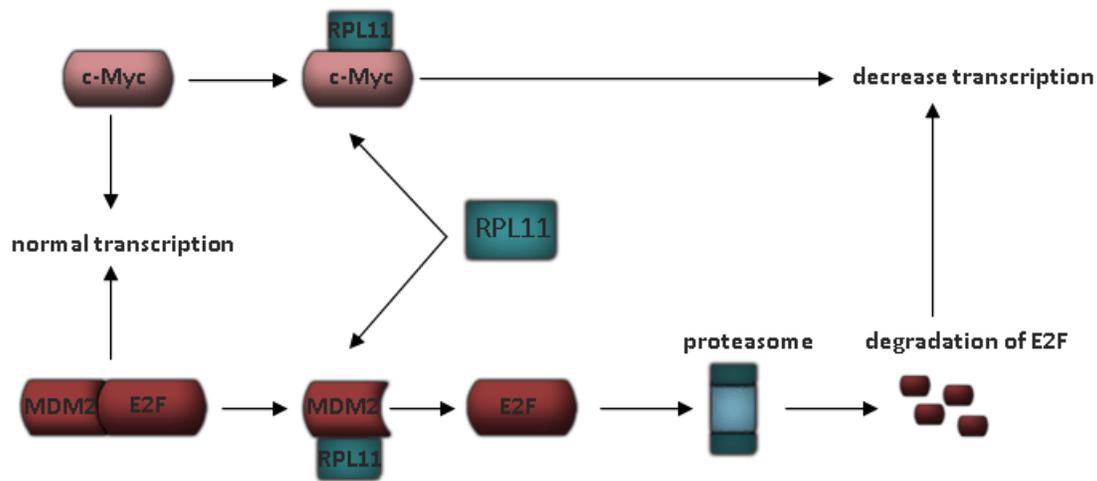
results in p53 dependent cell cycle arrest [82], supporting data showed that also depletion of partner of Bop1, WDR12, leads to same output [83].

Ribosome is composed of 4 rRNAs and 79 ribosomal proteins. Balanced level between these components is crucial for correct ribosome assembly. Insufficiency any of proteins leads to delay in maturation of ribosome, accumulation of ribosome precursors in nucleolus, nucleolar disruption and p53 activation. Since ribosomal proteins imbalance is the main cause of DBA, more attention is paid to this topic in section molecular pathology of DBA.

### **2.2.2 P53 independent ribosomal stress**

Besides p53 dependent ribosomal stress response cell also possess other mechanisms of ribosome stress response, which do not involve p53 signalling and result in cell cycle arrest [84]. Free RPL11 is mediator molecule in first two p53 independent mechanisms (Fig. 6). RPL11 binds to MDM2 and increases its protective function over transcription factor E2F. E2F promotes progression from G1 to S phase of the cell cycle, if unprotected, E2F is degraded by proteasome and cell cycle is stopped. In the other case RPL11 binds directly to transcriptional factor c-MYC [85]. c-MYC facilitates expression of the genes responsible for cell cycle progression and correct ribosome biogenesis. Interaction between c-MYC and RPL11 decreases c-MYC transcriptional activity and therefore it reduces cell proliferation.

Another p53 independent mechanism has PIM1 in its core. PIM1 is constitutively active serotonin/theronin-protein kinase regulating survival, cell cycle and proliferation, especially in hematopoietic cells [86, 87]. PIM1 is stabilized by interaction with RPS19 and binding to ribosomes, lack of available ribosomes results in decreased level of PIM1, stabilisation of CDK inhibitor p27<sup>Kip1</sup> and G1 cell cycle arrest [88]. Moreover, PIM1 also induces increase activity of c-MYC by its phosphorylation. Decreased level of PIM1 reduces activity of c-MYC and decline of cell proliferation [89].



**Fig.6:** Presentation how RPL11 trigger p53 independent ribosomal stress response

### 2.2.3 Ribosomopathies

Over years, several diseases and syndromes caused by impaired biogenesis and function of ribosomes have been described. Symptoms including anemia, skeleton malformations and higher risk of cancer development are mostly typical for these disorders. Main symptoms are summarized in Tab.1 [92].

TCOF1 gene encodes protein treacle which co-localizes with upstream binding factors of polymerase I in nucleolus. Treacle is an essential protein for rRNA genes transcription. Mutations in TCOF1 gene result in haploinsufficiency of treacle protein and decreased level of rRNA production which leads Teacher Collins syndrome [91].

Dyskeratosis congenital is heterogenous disease and has several causes and models of inheritance. Although this condition is caused by defects in several genes, all of them encode some protein of telomerase complex. Most extreme form of this disease is the result of mutation in dyskeretin, this protein acts as nucleolar protein and is associated with snoRNAs involved in rRNA processing [92].

Cartilage hair hypoplasia is typical for Amish families and Finnish population and it is associated with mutation in RMPR gene. RMPR encodes snoRNA which is part of mitochondrial RNA-processing complex. This complex can be also found in nucleolus where it processes pre-5S rRNA [93].

Schwanman-Diamond syndrome is disorder with wide range of symptoms. As cause of this syndrome, mutations in SBDS gene were identified. After translocation of 60S ribosome to cytoplasm, eIF6 from its surface has to be released to form complete 80S ribosome. SBDS is protein responsible for it. If SBDS is damaged, cleavage of eIF6 is not possible and neither formation of mature ribosome [94].

5q- syndrome is caused by deletion of short arm of chromosome 5, on which ribosomal protein RPS14 lies. Deletion of this arm leads to RPS14 haploinsufficiency [95]. Diamond-Blackfan anemia will be discussed in next chapter.

**Tab.1:** Summarised table of ribosomopaties, their causes and symptoms.

<b>Name</b>	<b>Gene</b>	<b>Clinic</b>
<b>DBA</b>	RPL5, RPL11, RPL15, RPL26, RPL31, RPL35a, RPS7, RPS10, RPS17, RPS19, RPS24, RPS26, GATA1	Macrocytic anemia, short stature, Craniofacial defects, thumb abnormalities
<b>5q- syndrome</b>	RPS14	Macrocytic anemia, hypolobudated micromegakaryocytes
<b>Schwanman-Diamond syndrome</b>	SBDS	Neutropenia, Pancreatic insufficiency, short stature
<b>Dyskeratosis congenita</b>	CKC1	Cytopenais, skin hyperpigmentation, nail dystrophy, oral leukoplakia
<b>Cartilage hair hypoplasia</b>	RMRP	Hypoplasticanemia, short limbed dwarfism, hypoplatic hair
<b>Teacher Collins syndrome</b>	TCOF1	Craniofacial abnormalities

## 2.3 Diamond-Blackfan anemia

### 2.3.1 Molecular pathology of DBA

Defects in 13 different genes responsible for causing DBA phenotype were described until today. Twelve out of these genes are encoding ribosomal proteins [5-13], six of them are genes for large ribosome subunit proteins (RPL5, RPL11, RPL15, RPL26, RPL31, RPL35a) and six for small subunit proteins (RPS7, RPS10, RPS17, RPS19, RPS24, RPS26). Thirteenth gene encodes GATA1 protein [18].

Generally, mutations in these genes are very diverse, ranging from single nucleotide changes leading to missense and nonsense mutations or splicing defects to whole gene deletion. Recently, even variant in non-coding region of RPS19 resulting in DBA, was described [96]. Types and counts of defects in ribosomal genes described in DBA database (dbagenes.unito.it) are enlisted in Tab.3.

**Tab.3:** Types and amounts of variant identified in DBA patients:

Gene	Substitution	Deletion	Insertion	Insertion/deletion	2 variants in 1 allele	complex	total
RPS19	145	47	19	1	4	2	218
RPL5	23	18	13	4	0	0	58
RPL11	12	21	1	0	0	0	34
RPS26	12	0	1	0	0	0	13
RPL35a	4	1	0	0	0	0	5
RPS10	4	0	1	0	0	0	5
RPS24	9	1	0	0	2	0	12
RPS17	2	1	0	0	0	0	3
RPS7	1	0	0	0	0	0	1

The most extensively studied gene connected with DBA is RPS19. Defects in RPS19 can be detected in nearly 25 % of DBA patient [97]. In all studied cases, mutation in RPS19 resulted in incapability of mutated protein incorporation into 40S ribosome [98]. Interestingly, different

mutations in this gene have different output at cellular level. Some mutations lead to decreased level of RPS19 mRNA [99] or protein stability [100], which leads to haploinsufficiency. These mutations also result in to impaired nucleolar localisation [101]. On the other hand, some mutations do not lead to haploinsufficiency, but exhibit impaired ribosomal localisation [101].

Ribosomal genes responsible for DBA phenotype play an important role in ribosome biogenesis and rRNA processing. Mutation or insufficiency of RPS19, RPS24, RPS26, RPS17, RPS7 and RPS10 leads to defects in pre-18S-rRNA processing and accumulation of 21S precursors. Defective 21S precursors cannot be used for formation of correct pre-40S ribosome, this result in increased level of free 60S ribosome and lack of 80S ribosome [6, 7, 102-103]. In case of mutation or insufficiency of RPL34, RPL11, RPL5 and RPL26, accumulation of 5.8S and 28S rRNA precursors and 40S ribosome together with decreased level of 60S ribosome was observed [6, 103-104]. In both cases, presence of halfmers and decreased level of polysomes was detected [103].

Only accumulation of ribosome precursors itself can start p53 mediated ribosome stress response [77], but some of the ribosomal proteins also have other regulatory roles within the cell. Several ribosomal proteins interact directly with MDM2 protein and can trigger p53 stabilisation by themselves, of which five are connected with DBA (RPL5, RPL11, RPL26, RPS26 and RPS7) [105-107].

Mutations in RPS19 and RPS24 also lead to the cell cycle arrest [108]. Specifically, mutations in RPS19 result as G1 arrest, probable cause of this is impaired association between RPS19 and PIM1 [88]. Mutation in RPS24 result in S/G2 arrest [108], molecular mechanism underlying this is still unclear. It was assumed that mutations in other DBA genes will also lead to cell cycle arrest. It was discovered that mutations in RPL5 and RPL11 do not result in cell cycle arrest but only suppress cell cycle progression [109].

Altered expression level of many gene clusters in both patients sample and DBA animal models was also revealed. Depletion of ribosomal genes leading to significant decrease of erythropoietic genes expression [110-112] is the most probable cause of anemia phenotype in

DBA. Furthermore, transcript levels of genes involved in cell death, cancer and tissue development are also impaired [113].

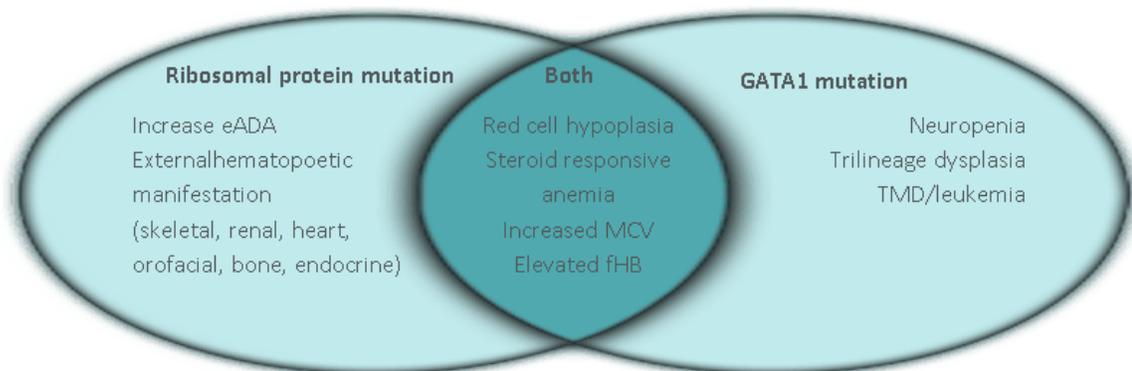
Reduced expression of all clusters responsible for efficient translation, particularly, in genes encoding ribosomal proteins, translation factors and tRNA transferases were described [114]. Moreover, haploinsufficiency of ribosomal proteins result in down regulation of transcripts of proteins belonging to same ribosomal subunit [103].

In eukaryotes two basic mechanisms of translation initiation exist, cap-dependent and IRES mediated. It has been discovered that IRES mediated translation is much more sensitive to knock down of ribosomal proteins than more competitive cap-dependent translation [115].

Another general feature of DBA is decreased level of translational capacity [116].

Precise mechanism, through which defects in ribosomal proteins causing DBA, is still unknown. Over years many theories explaining how mutated ribosomal proteins trigger anemia have been proposed, but none of them has been proclaimed as the only one. Due to the fact, that erythropoiesis is dynamic process which needs to be tightly regaled and for which site- and time-specific expression of certain gene is necessary, the most probable scenario is that all previously mentioned defects in cell contribute to DBA phenotype [117].

GATA1 is a zinc finger transcriptional factor expressed mainly in blood cells precursors. Several blood disorders connected with mutation in this protein were described [118]. Until now it remains unclear, if mutation in this gene is really responsible for causing DBA or similar phenotype [119] (Fig.7).



**Fig.7:** Differences of mutations in ribosomal genes and GATA1 in clinical appearance

### 2.3.2 Clinical appearance and diagnosis

DBA in most of cases (92%) is successfully diagnosed in first year of patient life [120]. This disease is mainly characterised by macrocytic or normocytic anemia with reticulocytopenia, normal platelet, neutrophil counts and deficiency of red blood cells precursors in bone marrow [2]. In approximately 35-50 % of the cases these symptoms are accompanied by congenital and developmental abnormalities [121]. Malformations typical for DBA are enlisted in Tab.2.

Majority of DBA patients have specifically elevated fetal haemoglobin HbF [121] and in 80-85% of the cases increased activity of cADA was reported [122]. When DBA is suspected, further clinical tests should be done to investigate other IBMFS (inherited bone marrow failure syndromes), especially Fancony anemia and Swachan-Diamond syndrome [2]. Next step in diagnosis includes screening for mutation in known DBA genes [2].

**Tab.2:** Malformations in DBA patients [2]:

<b>Location</b>	<b>Anomalies</b>
<b>Craniofacial</b>	Hypertelorism, Broad, flat nasal bridge, Cleft palate, High arched palate, Microcephaly, Micrognathia, Microtia, Low set ears, Low hair line, Epicanthus, Ptosis
<b>Ophthalmological</b>	Congenital glaucoma, Strabismus, Congenital cataract
<b>Neck</b>	Short neck, Webbed neck, Sprengel deformity, Klippel-Feil deformity
<b>Thumbs</b>	Triphalyngeal, Duplex or bifid, Hypoplastic, Flat thenar eminence, Absent radial artery
<b>Urogenital</b>	Absent kidney, Horseshoe kidney, Hypospadias
<b>Cardiac</b>	Ventricular septal defect, Atrial septal defect, Coarctation of the aorta, Complex cardiac anomalies
<b>Other musculoskeletal</b>	Growth retardation, Syndactyly
<b>Neuromotor</b>	Learning difficulties

### 2.3.3 Genotype-phenotype correlation

Unfortunately, only approximately 50 % of variants causing DBA enlisted in DBA database is connected with phenotype. Due to low number of mutation in many genes, proper genotype-phenotype correlation analysis cannot be performed. Tab.4 represents counts of reported patients with proper malformation report.

Mutations in RPL5 and RPL11 are mainly connected with malformations. Risk of malformations development was computed to 6.5-7.6 fold higher in RPL5 mutated and 2.7-4.5 fold higher in RPL11 mutated patients than in patient with mutation in any other gene [123]. Similar conclusion also has been reached in study performed in Italy [124].

Elevated risk of malformation development may be due to the fact that RPL5 and RPL11 have transporter function. RPL5 and RPL11 are associated with extra nucleolus 5S rRNA and carry nucleolar location signal [125]. When this signal is altered 5S rRNA is not transferred to nucleolus and cannot be incorporated into ribosome, this causes much bigger defect to ribosome structure than loss of one protein [126].

**Tab.4:** Relation between DBA genes and malformations development

Gene	RP	Mal
RPS19	104	58
RPL5	53	46
RPL11	36	26
RPS26	10	3
RPS24	5	2
RPL35a	7	4
RPS10	4	1
RPS17	3	1
RPS7	1	0

RP: reported patients, Mal: count reported malformations

#### 2.3.4 DBA treatment options

Glucocorticoid therapy and blood transfusion together with iron chelation are standard treatment strategies in DBA treatment. Glucocorticoids increase proliferation potential of erythroid progenitor cell, but according to DBA database this therapy works only in approximately 63% of patients. With this treatment severe side effects like physical and neurocognitive developmental defects are also associated [2]. Due to this, blood transfusion is first treatment choice, especially in infants. But serious side effects are associated with this therapy as well.

With transfusion and chelation, it is nearly impossible to restore healthy balance of iron in organism, iron through its high oxidative potential leads to mitochondrial and cell damage. This results in so called iron overload effect, iron overload is one of the major causes of death of DBA patients [2].

Neither of these treatment approaches is optimal and both of them are potentially harmful to patients, this is the reason why, new strategy in DBA treatment is needed. There are two ongoing trials for alternative DBA therapies. First agent is anti-cancerous drug lenalidomide [127], which is successfully used as a cure for 5q- syndrome. The other one is L-leucine, which initiates translation via mTOR pathway [128], its beneficial effect on translation rate has been proven *in vitro*, *in vivo* [129] and also in patients it has demonstrated great potential [130].

Until today the only permanent DBA treatment is bone marrow transplantation [2].

### **3. Aim of work**

Goal of theoretical part was to summarise the basis of Diamond-Blackfan anemia. First section was focused on the ribosome, its importance in cellular processes, with great impact on causes and consequences of the ribosomal stress and roles ribosome plays in human disorders. In second section, the attention was paid to Diamond-Blackfan anemia, mainly to the mechanism how mutations in ribosomal proteins result in their incorrect structure and function and how they can trigger Diamond-Blackfan phenotype. Moreover, clinical appearance of DBA and possibilities of its treatment were also discussed.

Practical part has in two goals. First, to determine mutations causing DBA phenotype in subset of DBA suspected patients. For this purpose whole exome sequencing approach was used. Secondary, to develop screening method for detection of translation efficiency in Diamond-Blackfan anemia which can be subsequently used for prediction whether newly identified mutation in DBA patients can lead to DBA phenotype or not.

## 4. Material and methods

### 4.1 Material

#### 4.1.1 Biological material

Patient DNA samples (P1-P12), control DNA samples of healthy parents (C1-C5), sample P6 was internal control with known mutation RPS19 R56Q.

Plasmid pCMV6-Entry with incorporated RPS19 cDNA (OriGene)

Plasmid pM-KAT-2-N (Addgene)

Cell line MRC-5 (ATCC)

Competent cells DH5 $\alpha$  (Sigma-Aldrich)

#### 4.1.2 Kits

Nextera Rapid Capture Exome Kit (Illumina)

Expanded Exome Enrichment Kit (Illumina)

TruSeq Rapid SBS Kits-HS (Illumina)

High Sensitivity DNA Analysis Kit (Agilent Technologies)

QIAquick PCR Purification Kit (Qiagene)

SigmaSpin™ Sequencing Reaction Clean-Up (Sigma-Aldrich)

QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies)

GenomeLab DTCS Quick Start Kit (Beckman Coulter)

NeonTransfection System 100  $\mu$ L Kit (Life Technologies)

Click-IT AHA (L-Azidohomoalanine) (Life Technologies)

#### 4.1.3 Enzymes and solutions

Taq polymerase (Thermo Scientific), 10X taq polymerase buffer (Thermo Scientific), 25 mM MgCl<sub>2</sub> (Thermo Scientific), dNTP set (Bioline)

RNAse (Top Bio)

LB medium: 20 g LB powder (Amresco), 4.17 g NaCl (Sigma-Aldrich), fill to 1 l by distilled water

Buffer 1: 50 mM glucose (Sigma-Aldrich), 25 mM TrisHCl (pH8) (Sigma-Aldrich), 10 mM EDTA (pH8) (Lach-Ner)

Buffer 2: 0.2 M NaOH (Sigma-Aldrich), 1 % SDS (Sigma-Aldrich)

Buffer 3: 3 M CH<sub>3</sub>CO<sub>2</sub>K (Sigma-Aldrich), 11.5 % v/v AcOH (Lach-Ner)

TBE buffer: 17.2 g TBE buffer (Duchefa Biochemie) dissolved in 1 l by distilled water

Cell line Medium: 10 % FBS (Sigma-Aldrich) in Dulbecco's Modified Eagle Medium (Sigma Aldrich)

Ethanol (Lach-Ner)

DEPC H<sub>2</sub>O (Ambion)

Primers (Generi Biotech):

rps19F	CTGTAAAAGACGTGAACCAGC
rps19R	GCCATCTTGGTCCTTTTCCA
rps19seqF	TGTAAAAGACGTGAACCAGC
rps19seqR	TCTTGGTCCTTTTCCACCAT
rps19mutF	CTGGTTCTACACGCAAGCTGCTCCACAG
rps19muR	CTGTGGAAGCAGCTTGCCTGTAGAACCAG

#### 4.1.4 Instruments:

Illumina HiSeq2500 (Illumina)

NeonTransfection System (Live Technologies)

2100 Bioanalyzer Instrument (AgilentTechnologies)

Beckman CEQ 8800 Genetic Analysis System (Beckman Coulter)

BD FACSAria II (bdbiosciences)

Thermal cycler C1000 (Biorad)

Incubator shaker C24 (New Brunswick scientific)

Transilluminator Odyssey (Li-Cor)

Incubator Trigon-plus (Thermo Scientific)

Centrifuge: Minispin (Eppendorf), Z383K (Hermle)

Vortex mixer VX-200 (Labnet)

## 4.2 Exome sequencing

### 4.2.1 Library preparation and sequencing run

Pair-end library was prepared using DNA from 22 individuals. P1-P12 were DBA suspected patients, C1-C5 were control samples, the others were patients with nonspecific anemia. Library was prepared by Nextera Rapid Capture Exome and Expanded Exome Enrichment Kit according to Nextera Rapid Capture Enrichment Guide (version November 2013). Sequencing data was obtained after run from BaseSpace analysis workspace (Illumina).

### 4.2.2 Raw data conversion

Data gained by sequencing were in bcl format. Demultiplexing and index removal was performed by Casava (Illumina). Resulting fastq files were align to indexed human reference genome, hg19, by BWA algorithm [131], by this alignment files in sam format were obtained. Sam files were further processed by samtools [132], data were transformed to bam files, subsequently sorted and indexed. Whole pipeline is represented on Fig.8.

```
$bwa mem -p hg19.fa sample1.fastq sample2.fastq > sample.sam
$ samtools view -bS sample.sam > sample.bam
$ samtools sort sample.bam sample.bam.sorted
$ samtools index sample.bam.sorted
```

**Fig.8:** Raw data conversion orders: -p:paired-end mode; -bS: output in the BAM format, input is in SAM format

### 4.2.3 Variant annotation

Mpileup algorithm of samtools package was used for variant calling, followed by annotation by annovar [133]. Pipeline is represented on Fig.9.

```

$ samtools mpileup -uf hg19.fa sample.bam | bcftools view -bvg - > sample.bcf
$ bcftools view sample.bcf > sample.vcf
$ perl annovar/convert2annovar.pl sample.vcf --format vcf4 > sample.vcf4
$ perl annovar/summarize_annovar.pl -ver1000g 1000g2012apr --verdbnp 135 --veresp
6500si --alltranscript --remove -buildver hg19 -outfile path sample.annovar
/annovar/humandb/

```

**Fig.9:** Variant annotation pipeline: -uf: output is uncompressed BCF, reference file in the FASTA format; -bvg: output in the BCF format, variant sites only, call per-sample genotypes at variant sites; for more information about annovar settings see: [openbioinformatics.org/annovar/annovar\\_accessary.html](http://openbioinformatics.org/annovar/annovar_accessary.html)

First step in variant analysis was extraction of variants found in ribosomal genes responsible for DBA and diseases with similar phenotype (Fanconi anemia, Swachmann-Diamond syndrom) into separate files. Genes involved in these conditions were stored in Gene\_panel (see attachment). Extraction of variants from genes of Gene\_panel was performed by our own script (see attachment). From identified variants in Gene\_panel were considered as relevant only those, which were not synonymous variants and which QUAL (see Tab.5) was higher than 50.

In case, possible variant responsible for anemia phenotype could not be determine in previous step, all variants which fulfilled certain conditions (Tab.5) were filtered. After this filtration relevant variants, which were subsequently used for additional analysis were obtained. These variants were divided into rs and other fraction.

**Tab. 5:** Exonic variants filtration table

Item	Value	Meaning
<b>QUAL</b>	>=50	Quality of variants (1000genomes.org/node/101 )
<b>Exonic function</b>	≠synonymous	Removal of synonymous variants
<b>SegDup</b>	empty	Removal of segmental duplications (e.g. pseudogenes) variants (humanparalogy.gs.washington.edu)
<b>dbSNP, 1000g</b>	<0.01	Frequency of variant in population lower than 1 % (ncbi.nlm.nih.gov/SNP/, 1000genomes.org, )

Clinvar database ([ncbi.nlm.nih.gov/clinvar/](http://ncbi.nlm.nih.gov/clinvar/)) consists of rs variants connected with clinical impact. Essential data were extracted from clinvar database and used for developing our own database. Fraction rs from previous step was compared with this database by our own script (see attachment). Output files were then searched for anemia phenotype.

GeneDestiller database [134] was used for the final analysis of other variants fraction, all genes enlisted in other variants fraction were uploaded to this database and output was screened for genes involve in ribosome biogenesis, erythropoiesis and anemia phenotype. All relevant nonsynonymous variants were then analysed by polyphen2 [135] and all findings were visualised in IGV software [136].

#### 4.2.4 CNV detection

Contra algorithm [137] was used for detection of copy number variation polymorphisms. Every patient without relevant variant in Gene\_panel was compared with 5 control samples (C1-C5). All findings in ribosomal genes were visualizes in IGV software and variations in coverage of selected ribosomal genes were searched.

```
$python contra.py --target TruSeq_exome_targeted_regions.hg19.bed.chr --test  
sample.bam --control C1-C5.bam --fasta hg19.fa --o file --numBin 20
```

**Fig.10:** CNV detection by Contra in terminal: --target: targeted regions; --test: patient fil; --cotrol: control file; --fasta: reference genome; --o: output file; --numBin: bin number

### 4.3 Functional study

#### 4.3.1 Plasmid pCMV6-ENTRY preparation

Competent cells DH5 $\alpha$  Plasmid were transformed by pCMV6-ENTRY using heat shock. To 50  $\mu$ l of competent cell was added 30 ng of plasmid, followed by incubation (10 min on ice, 30s at 42°C, 2min on ice), followed by addition of 200  $\mu$ l of LB medium and incubation for 45 min in

37°C. Cell suspension was then spread on plate containing 1.5% agarose gel in LB medium with 50ng/ml of kanamycin.

#### **4.3.2 Plasmid pCMV6-ENTRY isolation**

Picked colony was transferred to 100 ml of LB medium containing 50 ng/ml of kanamycin and incubated overnight in 37°C in shaker. After incubation suspension was centrifuged for 15 min under 5000 g. Supernatant was removed, pellet was resuspended by vortexing in 300 µl of Buffer 1 and incubated on ice for 5 min. Cell lysis was achieved by addition of 600 µl of Buffer 2, solution was mixed and incubated for 5 min on ice. Proteins were precipitated by 450 µl of Buffer 3, followed by mixing and incubation for 5 min on ice. For protein fraction removal was solution centrifuged for 15 min at 15000 g after last incubation. Supernatant containing target plasmids was transferred to new tube. RNA was removed by addition of 1 U of RNase, followed by incubation for 10 min at room temperature and ethanol precipitation. Final clean-up was performed on QIAquick PCR Purification column according to its manual. Isolated plasmid is represented on Fig.11A, line 1.

#### **4.3.3 Preparation of mutant vector**

QuikChange II Site-Directed Mutagenesis Kit was used for purpose and PCR Site-direct mutagenesis. Whole procedure was performed according to manufacturer's protocol, only items which were not component of kit, were mutagenesis primers rps19mutF and rps19mutR. Site directed-mutagenesis PCR program can be seen in Tab.6. Sample was analyzed on 0.5% agarose gel in TBE (Fig.11A). Transformation of cells was done as described in manufacturer's protocol.

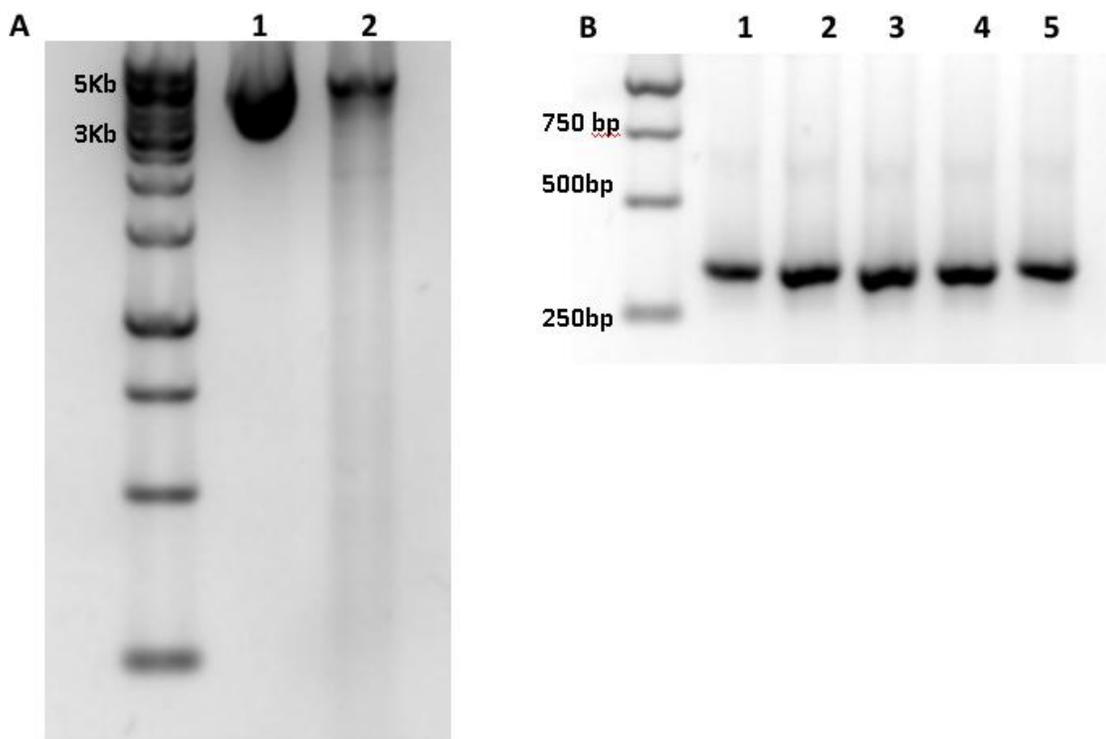
Colonies were picked and used for PCR reaction using primers rps19f and rps19r. PCR mix consisted of 1x PCR buffer, 2mM MgCl<sub>2</sub>, 0.2mM dNTPs, 0.5 µM each primer, 1 U Taq polymerase, 20-50 ng DNA. PCR program can be seen in Tab.7. Amplicons obtained by this PCR were confirmed on 0.5% agarose gel in TBE (Fig.11B) and Sanger sequencing was used for selection of successfully mutated plasmid carrying variant R56Q in RPS19.

**Tab.6:** Site-direct mutagenesis PCR program

Temperature	time	Number of cycles
95C	30 s	1x
95C	30 s	
62C	30 s	16x
68C	5 min	
12C	∞	1x

**Tab.7:** PCR program

temperature	time	Number of cycles
95C	30 s	1x
95C	30 s	
58C	30 s	35x
72C	1 min	
12C	∞	1x



**Fig.11:** Electrophoretogram: a) mutagenesis PCR reaction result – 1- original circularized plasmid, 2 – linear plasmid after mutagenesis PCR, B) PCR of 5 picked colonies result (1-5)

#### 4.3.4 Mutagenesis verification

Sanger sequencing was used for site direct mutagenesis confirmation. Target region was firstly amplified by standard PCR (see above) then internal primes, rps19seqF and rps19seqR, were used for Sanger sequencing reaction, which was prepared using GenomeLab DTCS Quick Start Kit (Tab.8).

Amplicons were cleaned on SigmaSpin™ Sequencing Reaction Clean-Up columns, dried in evaporator and dissolved in 20 µl of SLS sequencing buffer (Beckman Coulter), loaded to capillary sequencer and run was carried out using seq program (Tab.9). Resulting sequence was evaluated in Chromas programme.

Tab.8: sequencing reaction contents

component	Volume (µl)
Mastermix (Backmacultur)	8
Primer (10 uM)	4
Amplicon	~50ng
Water	to 20

Tab.9: seq program

Process	Temperature	Voltage	Time
Denaturation	90		120s
Injection		2kV	15s
Separation		4,2kV	60min

#### 4.3.5 Cell line transformation

Cell line was transformed by electroporation using NeonTransfection System according to manufacturer's protocol. Two millions cells MRC-5 and 1 mg of plasmids isolated in previous step were used for one transformation. Device setting was optimized to 1500 V for 30 ms. Cells were incubated overnight in medium without antibiotics after electroporation.

Efficiency of electroporation was verified by control plasmid pM-KAT-2-N carrying fluorescently labelled protein. Amount of fluorescently labelled cells was detected by flow cytometer.

#### **4.3.6 Translation efficiency analysis**

Translation efficiency of cells prepared in previous step was analyzed using Click-IT<sup>®</sup> (L-Azidohomoalanine) kit. Whole procedure was performed according to manufacturer's protocol. Final measurement was done on flow cytometer.

## 5. Results

### 5.1 Exome sequencing

#### 5.1.1 Library preparation

Quality and quantity of final library was verified using agilent chip (Fig.12). Prepared library consisted of DNA fragments in range from 180 bp to 1000 bp, median length was 288 bp and final concentration was 135.3 nmol/l.

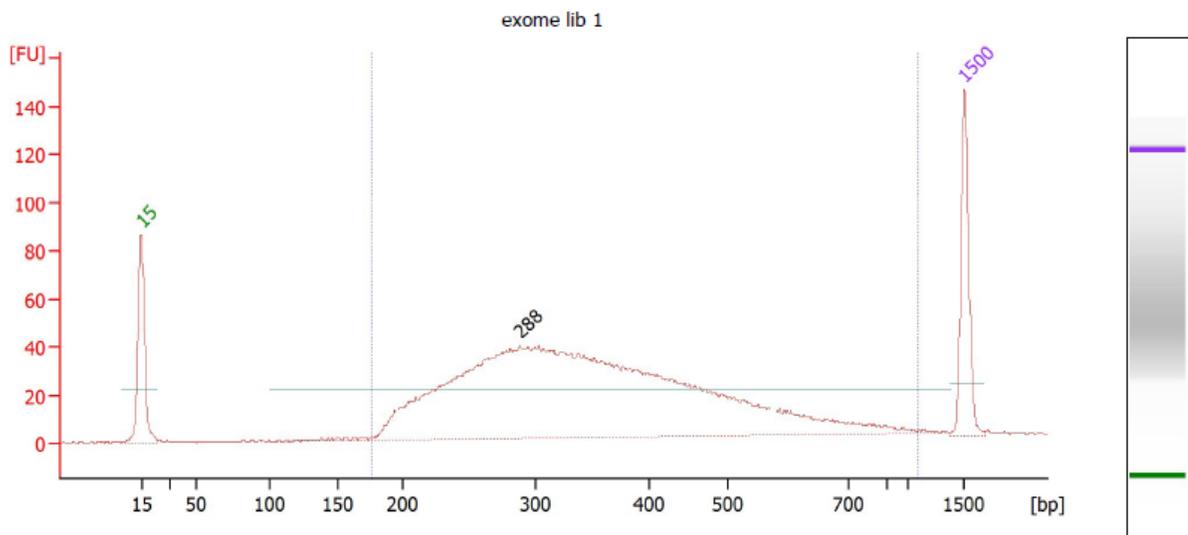
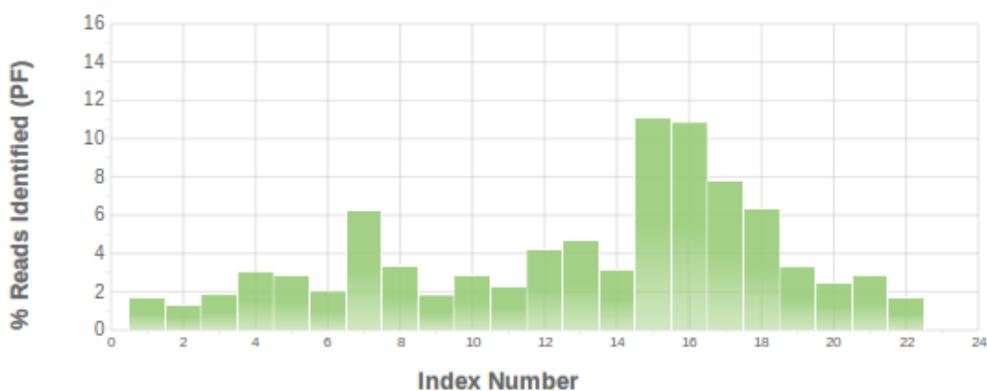


Fig.12: Visualization of exome library using agilent chip

#### 5.1.2 Sequencing run information

In total 27.7 Gb of bases in 137392 704 reads were identified. Since Nextera Rapid Capture Exome Kit capture region has approximately 37 Mb, total coverage was 748 and average coverage per sample was 34. Percentage distribution of reads per sample can be seen on Fig.13. Additional information for further analysed samples are enlisted in Tab.10. Although, the average coverage was 34, the real coverage per sample was very unevenly distributed and average coverage per sample ranged from 10 (C5) to 48 (P12).



**Fig.13:** Percentage distribution of reads per sample

**Tab.10:** Reads distribution per sample

Individual	Individual index (Fig.v)	Count of reads	Read bases (Kb)	Approximate coverage
P1	4	4149259	838	23
P2	1	2349415	474	13
P3	10	3970649	802	22
P4	12	5825450	1176	32
P5	13	6416239	1296	35
P6	14	4327870	874	24
P7	19	4533959	915	25
P8	17	4575177	924	25
P9	20	3407339	688	19
P11	22	2363154	477	13
P12	18	8738175	1765	48
C1	6	2789071	563	15
C2	5	3956909	799	22
C3	11	3091335	624	17
C4	3	2528025	510	14
C5	2	1799844	363	10

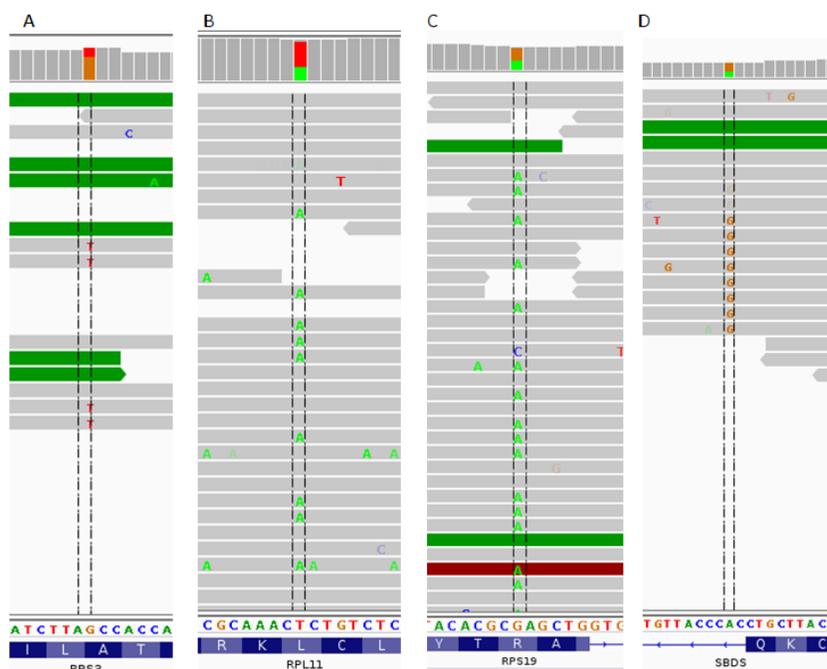
### 5.1.3 Variants in Gene\_panel

Number of detected variants with desired parameters (variant not synonymous, QUAL >50) in individual patients in the Gene\_panel was from 3 to 16, precise number per patient is enlisted in Tab.11.

Expected variant RPS19 exon3:c.G167A:p.R56Q was found in internal control P6. Variant in RPL11exon2:c.T59A:p.L20H, which is known cause of DBA, was identified in patient P4. Variant in splicing site of SBDS gene which leads to Swachman-Diamond syndrome was detected in patient P9. We also identified previously not described variant in ribosomal gene, RPS3 exon2:c.G154T:p.A52S, in patient P1. This variant was found in affected patient, but not in parents without DBA phenotype (controls C4, C5). Polyphen2 score for this variant was calculated to 0.702 which means that this variant is possibly damaging for protein structure. IVG view of all these variants is represented on Fig.14.

**Tab.11:** Summary of variants in Gene\_panel

Patient	Number of variants	Gene	Change	Previously described
P1	3	RPS3	exon2:c.G154T:p.A52S	no
P2	3			
P3	9			
P4	7	RPL11	exon2:c.T59A:p.L20H	yes
P5	16			
P6	8	RPS19	exon3:c.G167A:p.R56Q	yes
P8	5			
P9	5	SBDS	c.258+2T>C	yes
P10	7			
P11	3			
P12	7			



**Fig.14:** IVG view of all relevant variants in the Gene\_panel A) patient P1, B) patient P4, C) patient P6, D) patient P9

#### 5.1.4 Further variant analysis

Average number of variants in exome region in patients without any relevant variant detected in Gene\_panel at the beginning of analysis was 18 911. After filtration was reached average number 331 rs variants and 242 other variants. Reduction of number of relevant variants in exome region is represented in Tab.12.

**Tab.12:** Reduction of relevant variants during filtration

	EV	Q50	-Syn	-segDup	F 0.01	rs	other
<b>P2</b>	7571	2672	1293	1079	174	94	80
<b>P3</b>	15025	8043	4071	3413	467	270	198
<b>P5</b>	23260	15031	7472	6488	771	449	323
<b>P7</b>	18245	8202	3863	3339	550	294	257
<b>P8</b>	23384	17803	8805	7627	927	538	389
<b>P10</b>	22270	11678	5848	5010	565	341	225

<b>P11</b>	18826	5850	2855	2470	364	205	159
<b>P12</b>	22702	16891	8449	7239	767	463	304
<b>Average</b>	18910	10771	5332	4583	573	331	242

EV- number of exonic variants, Q50 – QAUL 50 and higher, Syn synonymous variants removal, -segDupsegmental duplications removal, F 0.01 – frequency in population (dbSNP, 1000g) lower than 0.01, rs - number of rs variants, other – number of other variants

Comparison of rs fraction with our clinvar database, provided from 2 to 5 rs variants per patient. But none of these variants was related to DBA phenotype.

Analysis of other fraction provided no variant that can be potentially responsible for DBA phenotype in patient P2, P3, P10, P11 and P12. In patient P5, nonsynonymous variant in translation initiation factor EIF4G3 exon18:c.G2916C:p.M972I in heterozygous form was detected. This variant had polyphen2 score 0.855, which means that this variant is possibly damaging for human health. Stop gain heterozygous variant in exon7:c.C1027T:p.R343X in BYSL gene, which encodes rRNA processing protein, was identified in patient P8. In this case Polyphen2 score could not be calculated, because polyphen2 algorithm is applicable only for missense mutations. Both relevant variants are summarized in Tab.13 and visualisation in IGV is represented on Fig.15.

**Tab.13:** Relevant variants in the others fraction per patient

<b>Patient</b>	<b>Gene</b>	<b>Type</b>	<b>Change</b>	<b>Genotype</b>	<b>Polyphen2 score</b>
<b>P2</b>					
<b>P3</b>					
<b>P5</b>	EIF4G3	NS	exon18:c.G2916C:p.M972I	Het	0.855
<b>P7</b>					
<b>P8</b>	BYSL	SG	exon7:c.C1027T:p.R343X	Het	
<b>P10</b>					
<b>P11</b>					
<b>P12</b>					

Ns = nonsynonymous, SG = stop gain



**Fig.15:** IVG view of relevant variants in the others fraction A) patient P5, B) patient P8

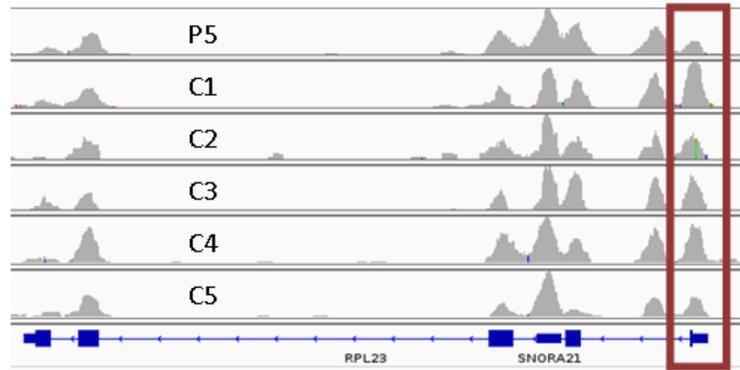
### 5.1.5 CNV analysis by Contra

Resulting genes, labelled as lost, in patients are enlisted in Tab.14. According to Contra algorithm patients, P2, P3, P10, P11 do not carry any deletion in the ribosomal genes.

**Tab.14:** Resulting genes which have been labelled as lost in patient in contrast to controls

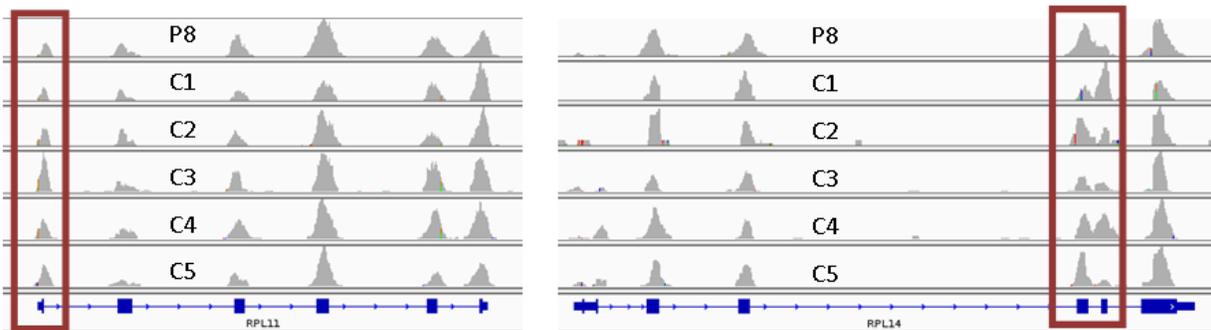
	C1	C2	C3	C4	C5
<b>P2</b>					
<b>P3</b>					
<b>P5</b>	RPL23				
<b>P7</b>	RPS24, RPS7	RPL11	RPS24, RPL22	RPS7, RPL11	RPS7, RPL11, RPL22
<b>P8</b>	RPL14			RPL11	RPL11
<b>P10</b>					
<b>P11</b>					
<b>P12</b>					

In patient P5 was detected deletion of the first exon of RPL23 gene in region chr17: 37009951-37010053 compares with healthy control C1. Compared with rest of controls this region had lower coverage profile as well (Fig.16).



**Fig.16:** Contra CNV variant in patient P5 in RPL23 gene (affected site is framed in red)

Two possible deletion sites were detected in patient P8, first one compared with C4 and C5 in RPL11 on coordinates chr1: 24018294-24018319, second one compared with C1 in RPL14 on position chr3: 40503099-40503152. Detailed observation in IGV (Fig.17) has revealed that in case of potential deletion in RPL11, peak in labelled area is smaller compared to C3-C5 but not compared to C1 and C2. In RPL14, profile of detected site was also significantly lower compared to two of controls, C1 and C4, but not compared to other controls.

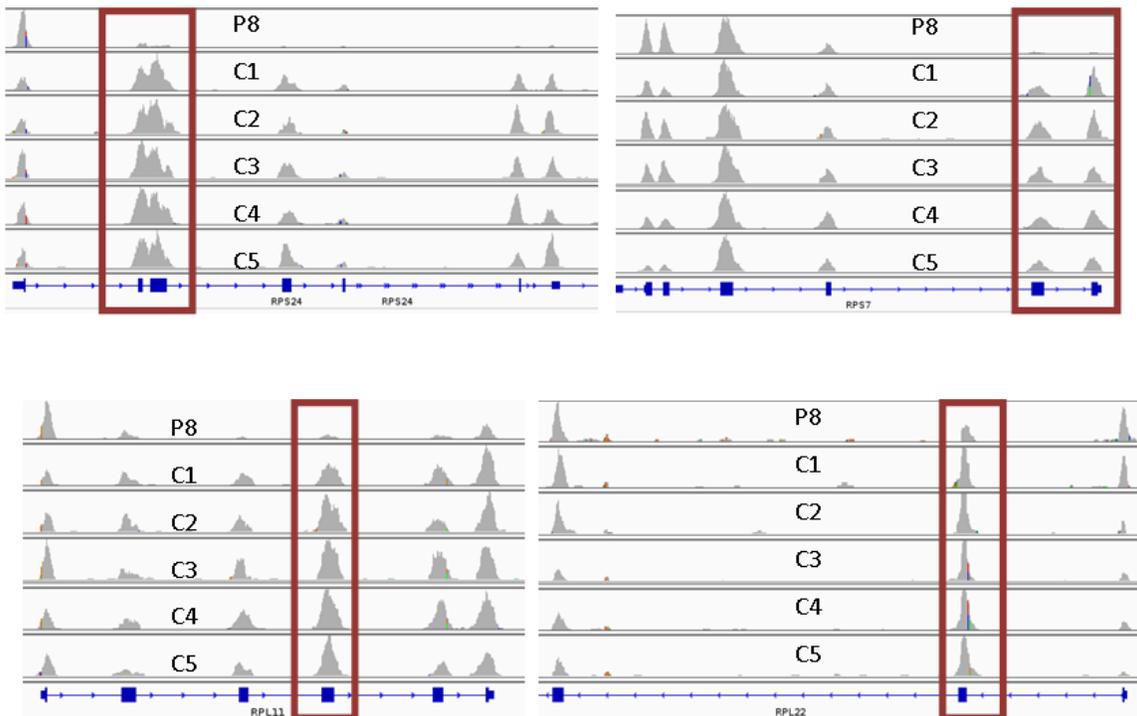


**Fig.17:** Contra CNV variants in patient P8 in RPL11 and RPL14 gene (affected sites are framed in red)

The largest number of possible deletion sites was detected in patient P7. All deletion sites detected by Contra are enlisted in Tab.15. Visualisation in IGV is represented on Fig.13. From visualisation in IGV it is evident, that coverage profile of all sites detected by Contra is significantly lower compared to all the controls.

**Fig.15:** CNV variants detected by Contra in patient P7 in ribosomal genes per controls

gene	C1	C2	C3	C4	C5
<b>RPS24</b>	chr10:79795269-79795478		chr10:79795110-79795175		
<b>RPS7</b>	chr2:3628395-3628508		chr2:3624077-3624220		chr2:3628395-3628508
<b>RPL11</b>	chr1:24021150-24021281		chr1:24021150-24021281		chr1:24021150-24021281
<b>RPL22</b>			chr1:6257712-6257816		chr1:6257712-6257816

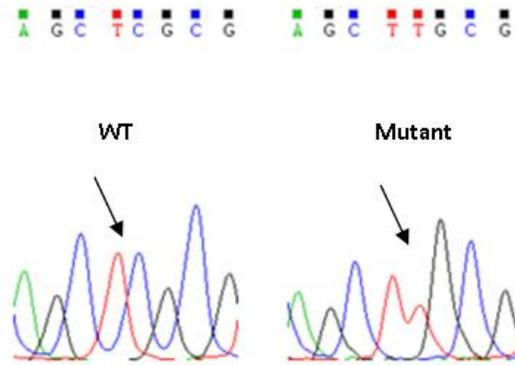


**Fig.18:** Potential deletion sites detected by Contra in patient P8 in RPS24, RPS7, RPL11 and RPL22 gene (affected sites are framed in red)

## 5.2 Functional study

### 5.2.1 Mutant plasmid preparation

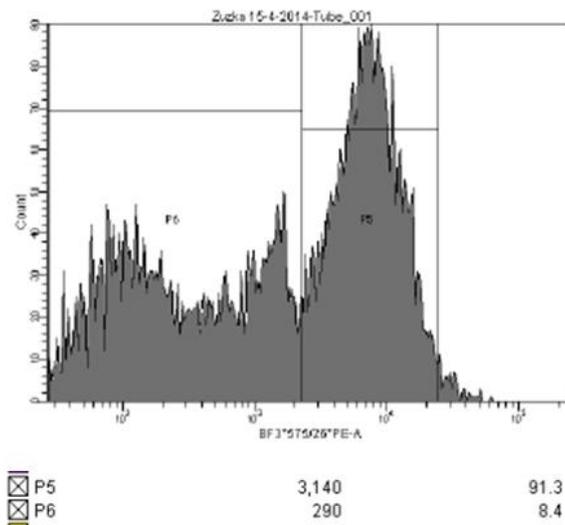
Mutant plasmid was successfully prepared and verified by Sanger sequencing (Fig.19).



**Fig. 19:** Cut out of electrophoretogram of WT and successfully prepared mutated plasmid

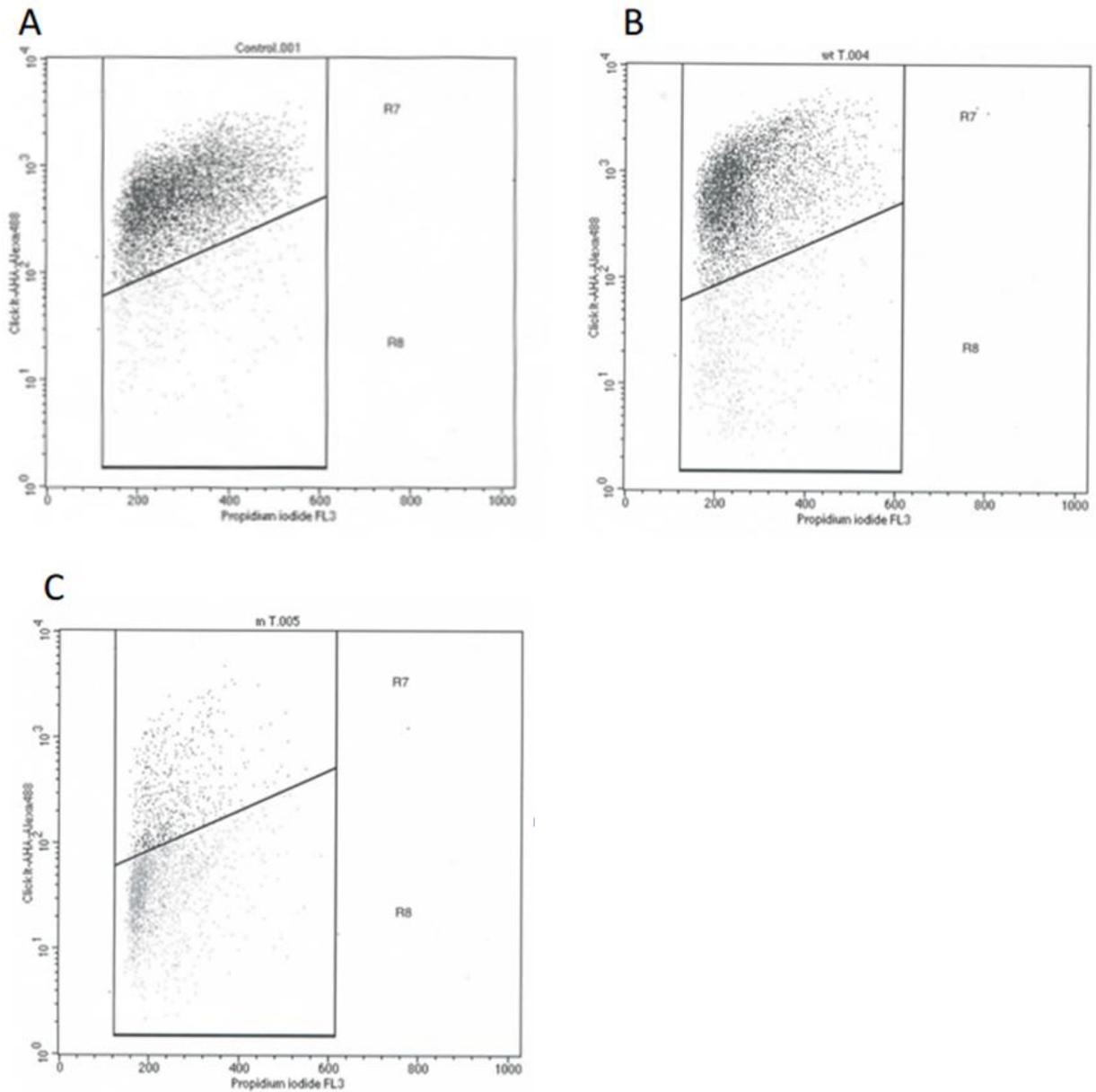
### 5.2.2 Translation analysis

Transformation efficiency was 91.3% (Fig.20).



**Fig.20:** Flow cytometer output: P6 – successfully transformed cells, P5 not transformed cells

Count of translation active cells was 93.5% in control cells, 89.5% in cells with incorporated WT plasmid and 25.1% in cells with incorporated plasmid with Q56R mutation. Output from flow cytometer is represented on Fig.21.



**Fig.21:** Flow cytometer output: R7 - translationally active cells, R8 - translationally inactive cells, A) control, B) WT, C) mutant Q56R

## 6. Discussion

Diamond-Blackfan anemia is a rare congenital disorder with symptoms including erythroid precursor deficiency, developmental malformations and increased risk of cancer development. The cause of this condition is heterogeneous and in approximately 50% of the cases, it is associated with mutation in one of the ribosomal genes, in other cases, the cause still remains unknown. Mutations in ribosomal genes usually results in impaired ribosomal assembly, ribosomal stress and cell cycle arrest in both p53 dependent and independent manner.

This work was focused on two main objectives. First was to establish a possible genetic condition causing DBA phenotype in twelve patients suspected of DBA. Second was to develop a detection system for possible mutations causing DBA screening based on translation efficiency detection.

From 3 to 16 variants per patient was identified in Gene\_panel. Most of these variants were identified in Fanconi anemia genes, but none of these variants was considered as cause of phenotype, due to the fact, that these variants were known variants with no connection with Fanconi anemia or was present in heterozygous form. Same applies for variants in ribosomal genes. Most of variants detected in ribosomal genes were known variants which are common in population, for instance variant in RPL10 exon6:c.A442G:p.I148V was identified in 10 out of 12 patients (all except P4 and P11). According to 1000g this variant is present in 71 % population and therefore cannot be responsible for such a rare condition as DBA.

Apart from the harmless variants, variants which are most likely causing anemia phenotype in four individuals were also identified. One of them has been internal control P6. In which known variant in RPS19 gene exon3:c.G167A:p.R56Q was detected. This validated correct analysis approach.

Out of three other variants identified using this method, two of them have already been described. In patient P4, variant in RPL11 exon2:c.T59A:p.L20H was indentified. This variant lies in conserved amino acid site and leads to loss of function in RPL11 protein [138].

In patient P9, mutation in SBDS gene c.258+2T>C connected with Swachman-Diamond syndrome in heterozygous composition was indentified. This mutation leads to disruption of

the splice site and 8bp frame shift deletion in transcript of this gene [139]. Although Swachman-Diamond syndrome is autosomal recessive disorder, this variant results in haploinsufficiency of SBDS protein and it was reported that this variant can cause milder symptom even in heterozygous form [140].

Last variant possibly responsible for DBA phenotype has been detected in patient P1. Variant in RPS3 gene exon2:c.G154T:p.A52S was identified in this patient. This variant was not detected in parents of P1, controls C4 and C5, which indicates that it is *de novo* mutation. Polyphen2 score has been calculated as 0.702 which means that this variant is possibly damaging for human health. Furthermore, RPS3 plays curtail roles within the cell. It is involved in DNA base extinction repair, modulates NF- $\kappa$ B activity and interacts with MDM2 [55-57]. All these facts emphasize seriousness of this variant to human health and mark RPS3 as a new DBA gene.

In patients with no detectable variant in Gene\_panel, analysis of all variants detected in whole exome was performed. Before filtration, average number of variants in eight individuals without mutation in Gene\_panel in exome region was 18 911, after filtration average number 331 rs and 242 other variants was reached.

Rs variants are SNP polymorphisms, detected and known across populations, most of them are common and have no medical impact. Little fraction of rs with known impact on health is stored in clinvar database. From 2 to 5 rs variants per patient which are enlisted in clinvar were detected in rs fraction. Unfortunately, these variants were not responsible for DBA or any other anemia related phenotype.

Online database GeneDestiller2 was used for analysis of other fraction. This database is very powerful toll for picking the genes of interest. In our search we focused on genes involved in ribosome assembly because it is presumed that unknown genes leading to DBA will be involved in ribosome biogenesis and will have power to trigger ribosomal stress, and genes involved in anemia and erythropoiesis phenotype.

In patient P5, variant in translation initiation factor EIF4G3 exon18:c.G2916C:p.M972I in heterozygous form was identified. Polyphen2 score was calculated as 0.855, which means that this variant is possibly damaging for human health. Translation initiation factors eIF4G3 refers

also to eIF4GII which plays important role in initiation complex formation [141]. According to Uniprot database (Uniprot.org) region 699 – 1019 of this protein is responsible for eIF3/EIF4A-interaction. This interaction allows binding of mRNA to ribosome and beginning of translation [141], if variant M972I in this region alters the binding, it may alter translation capacity and lead to anemia. Molecular mechanism would be more similar to Swachman-Diamond syndrome than DBA.

In patient P8, stop gain heterozygous variant exon7:c.C1027T:p.R343X in BYSL gene which encodes rRNA processing protein was identified. This stop variant reduces this protein by 94 amino acids, almost one quarter. It is most likely that this variant changes the conformation of whole protein and alters its function within the cell. It was shown that this protein is necessary for 40S ribosomal subunit formation [142]. Upon knock down of this protein accumulation of same 21rRNA precursor as in case of RPS19 knock down was detected [142,143]. Furthermore, connection between BYSL gene and ribosomal stress was revealed [144]. Since both crippled rRNA processing and ribosomal stress are typical features of DBA, this variant can be suspected of causing DBA.

Large deletions in ribosomal genes are also responsible for DBA, this is reason, why we have apart from variant analysis also performed structural analysis by Contra software. By this approach we have identified possible deletion sites in three patients (P5, P7 and P8). As relevant were considered only variants in P4 (RPL23) and P7 (RPS7, RPS24, RPL11, RPL22), variants in P8 were excluded due to the fact, that coverage profiles in both detected sites were lower compared only with few controls. If proven, large deletions detected in P4 and P7 lead to haploinsufficiency of these ribosomal genes and in turn to ribosomal stress and possibly DBA phenotype [15].

Reliability of both variant and CNV analysis is dependent on coverage [145, 146]. Since, obtained coverage was low (in our experience, minimal acceptable coverage is 30) and unevenly distributed, all findings have to be verified (variants by Sanger sequencing and large deletions by PCR methods).

Apart from verification of the result obtained by whole exome sequencing, there is also need to perform functional study. For this purpose it was given effort to development of

system for translation efficiency detection. Decreased translation rate is one of main feature of DBA [116], therefore this system could ease us determination if newly identified variant can be responsible for DBA or similar phenotype.

The system was successfully tested. Translation efficiency was significantly decreased using mutated plasmid RPS19 Q56R compared to WT plasmid. For more convenient result it needs to be done more repetition. But even from this partial result is evident, that this system will be probably suitable for prediction whether newly identified mutation in DBA patients can lead to DBA phenotype or not.

## 7. Conclusion

This work was focused on two main goals. First, to determine the mutations causing DBA phenotype in twelve DBA suspected patients using whole exome sequencing approach. The second goal included the development of system for detection of translation efficiency in Diamond-Blackfan anemia which can be subsequently used for prediction whether newly identified mutation in DBA patients can lead to DBA phenotype or not.

The probable causes of DBA phenotype in six out of twelve patients were successfully identified, however, findings needs to be verified.

The development of detection system was also successful. The significant difference, detected in translation efficiency in variant Q56R in RPS19 gene leading to DBA, indicates suitability of the method for the detection of the variants, potentially responsible for DBA phenotype.

## 8. References

1. Diamond LK, Blackfan KD. Hypoplastic anemia. *Am J Dis Child*.1938; 56: 464–467.
2. Vlachos A, Muir E. How I treat Diamond-Blackfan anemia. *Blood*. 2010; 116: 3715–3723.
3. Ball S. Diamond Blackfan anemia. *Hematology Am Soc Hematol Educ Program*. 2011; 2011: 487-91.
4. Sjögren SE, Flygare J. Progress towards Mechanism-Based Treatment for Diamond-Blackfan Anemia. *Scientific World Journal*. 2012; 2012: 184362.
5. Willig TN, Draptchinskaja N, Dianzani I, et al. Mutations in ribosomal protein S19 gene and diamond blackfan anemia: wide variations in phenotypic expression. *Blood*. 1999; 94: 4294-306.
6. Gazda HT, Sheen MR, Vlachos A 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; 83: 769-80.
7. Doherty L, Sheen MR, Vlachos A, et al. Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. *Am J Hum Genet*. 2010; 86: 222-8.
8. Farrar JE, Nater M, Caywood E, et al. Abnormalities of the large ribosomal subunit protein, Rpl35a, in Diamond-Blackfan anemia. *Blood*. 2008; 112: 1582-92.
9. Gazda HT, Grabowska A, Merida-Long LB, et al. Ribosomal protein S24 gene is mutated in Diamond-Blackfan anemia. *Am J Hum Genet*. 2006; 79: 1110-8.
10. Cmejla R , Cmejlova J , Handrkova H, et al. Ribosomal protein S17 gene (RPS17) is mutated in Diamond-Blackfan anemia. *Hum Mutat*. 2007; 28: 1178-82.
11. Watkins-Chow DE, Cooke J, Pidsley R, et al. Mutation of the diamond-blackfan anemia gene Rps7 in mouse results in morphological and neuroanatomical phenotypes. *PLoS Genet*. 2013; 9: e1003094.

12. Landowski M, O'Donohue MF, Buros C, et al. Novel deletion of RPL15 identified by array-comparative genomic hybridization in Diamond-Blackfan anemia. *Hum Genet.* 2013; 132: 1265-74.
13. Gazda HT, Preti M, Sheen MR, et al. Frameshift mutation in p53 regulator RPL26 is associated with multiple physical abnormalities and a specific pre-ribosomal RNA processing defect in diamond-blackfan anemia. *Hum Mutat.* 2012; 33: 1037-44.
14. Choismel V, Bacqueville D, Rouquette J, et al. Impaired ribosome biogenesis in Diamond-Blackfan anemia. *Blood.* 2007; 109: 1275-83.
15. Dutt S, Narla A, Lin K, et al. Haploinsufficiency for ribosomal protein genes causes selective activation of p53 in human erythroid progenitor cells. *Blood.* 2011; 117: 2567-76.
16. Gazda HT, Kho AT, Sanoudou D, et al. Defective ribosomal protein gene expression alters transcription, translation, apoptosis, and oncogenic pathways in Diamond-Blackfan anemia. *Stem cells.* 2006; 24: 2034-44.
17. Avondo F, Roncaglia P, Crescenzo N, et al. Fibroblasts from patients with Diamond-Blackfan anaemia show abnormal expression of genes involved in protein synthesis, amino acid metabolism and cancer. *BMC genomics.* 2009; 10: 442.
18. Sankaran VG, Ghazvinian R, Do R, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J Clin Invest.* 2012; 122: 2439-43.
19. Thomas G. An encore for ribosome biogenesis in the control of cell proliferation. *Nat Cell Biol.* 2000; 2: E71-2.
20. Freed EF, Bleichert F, Dutca LM, et al. When ribosomes go bad: diseases of ribosome biogenesis. *Mol Biosyst.* 2010; 6: 481-93.
21. Spahn CM, Beckmann R, Eswar N, et al. Structure of the 80S ribosome from *Saccharomyces cerevisiae*--tRNA-ribosome and subunit-subunit interactions. *Cell.* 2001 107: 373-86.
22. Granneman S, Petfalski E, Swiatkowska A, et al. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA protein cross-linking. *EMBO J.* 2010; 29: 2026-2036.

23. Gao H, Ayub MJ, Levin MJ, et al. The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components. *Proc Natl Acad Sci USA*. 2005; 102: 10206-11.
24. Raska I, Shaw PJ, Cmarko D. New insights into nucleolar architecture and activity. *Int Rev Cytol*. 2006; 255:177-235.
25. Phipps KR, Charette J, Baserga SJ. The small subunit processome in ribosome biogenesis progress and prospects. *Wiley Interdiscip Rev RNA*. 2011; 2: 1-21.
26. Ciganda M, Williams N. Eukaryotic 5S rRNA biogenesis. *Wiley Interdiscip Rev RNA*. 2011; 2: 523-33.
27. Henras AK, Soudet J, Gerus M, et al. The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell. Mol. Life Sci*. 2008; 65: 2334-2359.
28. Zhang J, Harnpicharnchai P, Jakovljevic J, et al. Assembly factors Rpf2 and Rrs1 recruit 5S rRNA and ribosomal proteins rpL5 and rpL11 into nascent ribosomes. *Genes Dev*. 2007; 21: 2580-92.
29. Tschochner H, Hurt E. Pre-ribosomes on the road from the nucleolus to the cytoplasm. *Trends Cell Biol*. 2003; 13: 255-63.
30. Grandi P, Rybin V, Bassler J, et al. 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol Cell*. 2002; 10: 105-15.
31. Venema J, Tollervey D. Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet*. 1999; 33:261-311.
32. T. Schäfer, B. Maco, E. Petfalski, et al. Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature*. 2006; 441: 651–655.
33. Kressler D, Hurt E, Bassler J. Driving ribosome assembly. *Biochim Biophys Acta*. 2010; 1803: 673-83.
34. T. Schafer, D. Strauss, E. Petfalski, et al. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J*. 2003; 22: 1370-1380.
35. T.A. Nissan, J. Bassler, et al. 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. *EMBO J*. 2002; 21: 5539-5547.

36. Zemp I, Kutay U. Nuclear export and cytoplasmic maturation of ribosomal subunits. *FEBS Lett.* 2007; 581: 2783-93.
37. Ansel KM, Pastor WA, Rath N, et al. Mouse Eri1 interacts with the ribosome and catalyzes 5.8S rRNA processing. *Nat Struct Mol Biol.* 2008; 15: 523-30.
38. Pestova TV, Lomakin IB, Lee JH, et al. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature.* 2000; 403: 332-5.
39. Lempiainen H, Shore D. Growth control and ribosome biogenesis. *Curr Opin Cell Biol.* 2009; 21: 855-63.
40. Albert B, Perez-Fernandez J, Leger-Silvestre I, Gadal O. Regulation of Ribosomal RNA Production by RNA Polymerase I: Does Elongation Come First? *Genet Res Int.* 2012; 2012: 276948.
41. Valajsek LS. 'Ribozoomin'--translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). *Curr Protein Pept Sci.* 2012; 13: 305-30.
42. Voigts-Hoffmann F, Klinge S, Ban N. Structural insights into eukaryotic ribosomes and the initiation of translation. *Curr Opin Struct Biol.* 2012; 22: 768-77.
43. Jackson RJ, Hellen CUT, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol.* 2010; 10: 113.
44. Proud CG. Peptide-chain elongation in eukaryotes. *Mol Biol Rep.* 1994; 19: 161-70.
45. Sasikumar AN, Perez WB, Kinzy TG. The many roles of the eukaryotic elongation factor 1 complex. *Wiley Interdiscip Rev RNA* 2012; 3: 543-55.
46. Franckenberg S, Becker T, Beckmann R. Structural view on recycling of archaeal and eukaryotic ribosomes after canonical termination and ribosome rescue. *Curr Opin Struct Biol.* 2012; 22: 786-96.
47. Hahm B, Kim YK, Kim JH, et al. Heterogeneous nuclear ribonucleoprotein L interacts with the 3' border of the internal ribosomal entry site of hepatitis C virus. *J Virol.* 1998; 72: 8782-8.
48. Malys N, McCarthy JE. Translation initiation: variations in the mechanism can be anticipated. *Cell Mol Life Sci.* 2011; 68: 991-1003.

49. Hellen CU, Sarnow P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* 2001; 15: 1593-612.
50. Janas MM, Wang E, Love T, et al. Reduced expression of ribosomal proteins relieves microRNA-mediated repression. *Mol Cell.* 2012; 46: 171-86.
51. Han TW, Kato M, Xie S, et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell.* 2012; 149: 768-79.
52. Malygin AA, Parakhnevitch NM, Ivanov AV, et al. Human ribosomal protein S13 regulates expression of its own gene at the splicing step by a feedback mechanism. *Nucleic Acids Res.* 2007; 35: 6414-6423.
53. Mazumder B, Sampath P, Seshadri V, et al. Regulated release of L13a from the 60S ribosomal subunit as a mechanism of transcript-specific translational control. *Cell.* 2003; 115: 187-98.
54. Kapasi P, Chaudhuri S, Vyas K, et al. L13a blocks 48S assembly: role of a general initiation factor in mRNA-specific translational control. *Mol Cell.* 2007; 25: 113-126.
55. Jang CY, Lee JY, Kim J. RpS3, a DNA repair endonuclease and ribosomal protein, is involved in apoptosis. *FEBS Lett.* 2004; 560: 81-5.
56. Ko SI, Park JH, Park MJ, et al. Human ribosomal protein S3 (hRpS3) interacts with uracil-DNA glycosylase (hUNG) and stimulates its glycosylase activity. *Mutat Res.* 2008; 648: 54-64.
57. Wan F, Anderson DE, Barnitz RA, et al. Ribosomal protein S3: a KH domain subunit in NF-kappaB complexes that mediates selective gene regulation. *Cell.* 2007; 131: 927-39.
58. Kim TH, Leslie P, Zhang Y. Ribosomal proteins as unrevealed caretakers for cellular stress and genomic instability. *Oncotarget.* 2014; 5: 860-71.
59. Karbstein K. Quality control mechanisms during ribosome maturation. *Trends Cell Biol.* 2013; 23: 242-50.
60. Pestov DG, Strezoska Z, Lau LF. Evidence of p53-dependent cross-talk between ribosome biogenesis and the cell cycle: effects of nucleolar protein Bop1 on G(1)/S transition. *Mol Cell Biol.* 2001; 21: 4246-55.

61. Zhang Y, Lu H. Signaling to p53: ribosomal proteins find their way. *Cancer Cell*. 2009; 16: 369-77.
62. Micel LN, Tentler JJ, Smith PG, et al. Role of ubiquitin ligases and the proteasome in oncogenesis: novel targets for anticancer therapies. *J Clin Oncol*. 2013; 31: 1231-8.
63. Horn HF, Vousden KH. Coping with stress: multiple ways to activate p53. *Oncogene*. 2007; 26: 1306-16.
64. Kruse JP, Gu W. Modes of p53 regulation. *Cell*. 2009; 137: 609-22.
65. Manfredi JJ. The Mdm2-p53 relationship evolves: Mdm2 swings both ways as an oncogene and a tumor suppressor. *Genes Dev*. 2010; 24: 1580-9.
66. Marine JC, Lozano G. Mdm2-mediated ubiquitylation: p53 and beyond. *Cell Death Differ*. 2010; 17: 93-102.
67. Marechal V, Elenbaas B, Piette J, et al. The ribosomal L5 protein is associated with mdm-2 and mdm-2-p53 complexes. *Mol Cell Biol*. 1994; 14: 7414-20.
68. Cui D, Li L, Lou H, et al. The ribosomal protein S26 regulates p53 activity in response to DNA damage. *Oncogene*. 2014; 33: 2225-35.
69. Zhou X, Hao Q, Liao J, et al. Ribosomal protein S14 unties the MDM2-p53 loop upon ribosomal stress. *Oncogene* 2013; 32: 388-96.
70. Yadavilli S, Mayo LD, Higgins M, et al. Ribosomal protein S3: A multi-functional protein that interacts with both p53 and MDM2 through its KH domain. *DNA Repair (Amst)*. 2009; 8: 1215-24.
71. Zhu Y, Poyurovsky MV, Li Y, et al. Ribosomal protein S7 is both a regulator and a substrate of MDM2. *Mol Cell*. 2009; 35: 316-26.
72. Ofir-Rosenfeld Y, Boggs K, Michael D, et al. Mdm2 regulates p53 mRNA translation through inhibitory interactions with ribosomal protein L26. *Mol Cell*. 2008; 32: 180-9.
73. Zhang Y, Shi Y, Li X, et al. Inhibition of the p53-MDM2 interaction by adenovirus delivery of ribosomal protein L23 stabilizes p53 and induces cell cycle arrest and apoptosis in gastric cancer. *J Gene Med*. 2010; 12: 147-56.

74. Lindstrom MS, Jin A, Deisenroth C, et al. Cancer-Associated Mutations in the MDM2 Zinc Finger Domain Disrupt Ribosomal Protein Interaction and Attenuate MDM2-Induced p53 Degradation. *Mol Cell Biol.* 2007; 27: 1056-1068.
75. Gilkes DM, Chen L, Chen J. MDMX regulation of p53 response to ribosomal stress. *EMBO J.* 2006; 25: 5614-5625.
76. Kreiner G, Bierhoff H, Armentano M, et al. A neuroprotective phase precedes striatal degeneration upon nucleolar stress. *Cell Death Differ.* 2013; 20: 1455-64.
77. Deisenroth C, Zhang Y. The Ribosomal Protein-Mdm2-p53 Pathway and Energy Metabolism: Bridging the Gap between Feast and Famine. *Genes Cancer.* 2011; 2: 392-403.
78. Donati G, Bertoni S, Brighenti E, et al. The balance between rRNA and ribosomal protein synthesis up- and downregulates the tumour suppressor p53 in mammalian cells. *Oncogene.* 2011; 30: 3274-88.
79. Yuan X, Zhou Y, Casanova E, et al. Genetic inactivation of the transcription factor TIF-IA leads to nucleolar disruption, cell cycle arrest, and p53-mediated apoptosis. *Mol Cell.* 2005; 19: 77-87.
80. Parlato R, Kreiner G, Erdmann G, et al. Activation of an endogenous suicide response after perturbation of rRNA synthesis leads to neurodegeneration in mice. *J Neurosci.* 2008; 28: 12759-64.
81. Finkbeiner E, Haindl M, Muller S. The SUMO system controls nucleolar partitioning of a novel mammalian ribosome biogenesis complex. *EMBO J.* 2011; 30: 1067-78.
82. Pestov DG, Strezoska Z, Lau LF. Evidence of p53-dependent cross-talk between ribosome biogenesis and the cell cycle: effects of nucleolar protein Bop1 on G(1)/S transition. *Mol Cell Biol.* 2001; 21: 4246-55.
83. Holzel M, Rohrmoser M, Schlee M, et al. Mammalian WDR12 is a novel member of the Pes1-Bop1 complex and is required for ribosome biogenesis and cell proliferation. *J Cell Biol.* 2005; 170: 367-78.
84. Donati G, Montanaro L, Derenzini M. Ribosome biogenesis and control of cell proliferation: p53 is not alone. *Cancer Res.* 2012; 72: 1602-7.

85. Mu-Shui Dai, Hua Lu. Crosstalk between c-Myc and ribosome in ribosomal biogenesis and cancer. *J Cell Biochem.* 2008; 105: 670-677.
86. Bachmann M, Moroy T. (2005). The serine/threonine kinase Pim-1. *Int J Biochem Cell Biol.* 2005; 37: 726-30.
87. Soeiro I, Mohamedali A, Romanska HM, et al. p27Kip1 and p130 cooperate to regulate hematopoietic cell proliferation in vivo. *Mol Cell Biol.* 2006; 26: 6170-84.
88. Iadevaia V, Caldarola S, Biondini L, et al. PIM1 kinase is destabilized by ribosomal stress causing inhibition of cell cycle progression. *Oncogene.* 2010; 29: 5490-9.
89. Morscio J, Dierickx D, Tousseyn T. Molecular pathogenesis of B-cell posttransplant lymphoproliferative disorder: what do we know so far? *Clin Dev Immunol.* 2013; 2013: 150835.
90. Narla A, Ebert BL. Ribosomopathies: human disorders of ribosome dysfunction. *Blood.* 2010; 115: 3196-3205.
91. Valdez BC, Henning D, So RB, et al. The Treacher Collins syndrome (TCOF1) gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc Natl Acad Sci USA.* 2004; 101: 10709-10714.
92. Heiss NS, Knight SW, Vulliamy TJ, et al. X-linked dyskeratosiscongenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet.* 1998; 19: 32-38.
93. Welting TJ, van Venrooij WJ, Pruijn GJ. Mutual interactions between subunits of the human RNase MRP ribonucleoprotein complex. *Nucleic Acids Res.* 2004; 32: 2138-2146.
94. Ganapathi KA, Austin KM, Lee CS, et al. The human Schwachman-Diamond syndrome protein, SBDS, associates with ribosomal RNA. *Blood.* 2007; 110: 1458-1465.
95. Ebert BL, Pretz J, Bosco J, et al. Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature.* 2008; 451: 335-339.
96. Cretien A, Proust A, Delaunay J, et al. Genetic variants in the noncoding region of RPS19 gene in Diamond-Blackfan anemia: potential implications for phenotypic heterogeneity. *Am J Hematol.* 2010; 85:111-6.

97. Willig TN, Draptchinskaia N, Dian-zani I, et al. Mutations in ribosomal protein S19 gene and Diamond-Blackfan anemia: wide variations in phenotypic expression. *Blood*. 1999; 94: 4294-306.
98. Angelini M, Cannata S, Mercaldo V, et al. Missense mutations associated with Diamond-Blackfan anemia affect the assembly of ribosomal protein S19 into the ribosome. *Hum Mol Genet*. 2007; 16: 1720-7.
99. Chatr-Aryamontri A, Angelini M, Garelli E, et al. Nonsense-mediated and nonstop decay of ribosomal protein S19 mRNA in Diamond-Blackfan anemia. *Hum Mutat*. 2004; 24: 526-33.
100. Gregory LA, Aguisa-Touré AH, Pinaud N, et al. Molecular basis of Diamond-Blackfan anemia: structure and function analysis of RPS19. *Nucleic Acids Res*. 2007; 35: 5913-21.
101. Choessel V, Bacqueville D, Rouquette J, et al. Impaired ribosome biogenesis in Diamond-Blackfan anemia. *Blood*. 2007; 109: 1275-83.
102. Choessel V, Fribourg S, Aguisa-Toure AH, et al. Mutation of ribosomal protein RPS24 in Diamond-Blackfan anemia results in a ribosome biogenesis disorder. *Hum Mol Genet*. 2008; 17: 1253-1263
103. Robledo S, Idol RA, Crimmins DL, et al. The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA*. 2008; 14: 1918-29.
104. Farrar JE, Nater M, Caywood E, et al. Abnormalities of the large ribosomal subunit protein, Rpl35a, in Diamond-Blackfan anemia. *Blood*. 2008; 112: 1582-92.
105. Zhang F, Hamanaka RB, Bobrovnikova-Marjon E, et al. Ribosomal stress couples the unfolded protein response to p53-dependent cell cycle arrest. *J Biol Chem*. 2006; 281: 30036-45.
106. Chen D, Zhang Z, Li M, et al. Ribosomal protein S7 as a novel modulator of p53-MDM2 interaction: binding to MDM2, stabilization of p53 protein, and activation of p53 function. *Oncogene* 2007; 26: 5029-37.
107. Ofir-Rosenfeld Y, Boggs K, Michael D, et al. Mdm2 regulates p53 mRNA translation through inhibitory interactions with ribosomal protein L26. *Mol Cell*. 2008; 32: 180-9.

108. Badhai J, Frojmark AS, J Davey E, et al. Ribosomal protein S19 and S24 insufficiency cause distinct cell cycle defects in Diamond-Blackfan anemia. *Biochim Biophys Acta*. 2009; 1792: 1036-42.
109. Teng T, Mercer CA, Hexley P, et al. Loss of tumor suppressor RPL5/RPL11 does not induce cell cycle arrest but impedes proliferation due to reduced ribosome content and translation capacity. *Mol Cell Biol*. 2013; 33: 4660-71.
110. Jia Q, Zhang Q, Zhang Z, et al. Transcriptome analysis of the zebrafish model of Diamond-Blackfan anemia from RPS19 deficiency via p53-dependent and -independent pathways. *PLoS One*. 2013; 8: e71782.
111. Yadav GV, Chakraborty A, Uechi T, et al. Ribosomal protein deficiency causes Tp53-independent erythropoiesis failure in zebrafish. *Int J Biochem Cell Biol*. 2014; 49: 1-7.
112. Zhang Z, Jia H, Zhang Q, et al. Assessment of hematopoietic failure due to Rpl11 deficiency in a zebrafish model of Diamond-Blackfan anemia by deep sequencing. *BMC Genomics*. 2013; 14: 896.
113. Avondo F, Roncaglia P, Crescenzo N, et al. Fibroblasts from patients with Diamond-Blackfan anaemia show abnormal expression of genes involved in protein synthesis, amino acid metabolism and cancer. *BMC Genomics*. 2009; 10: 442.
114. Gazda HT, Kho AT, Sanoudou D, et al. Defective ribosomal protein gene expression alters transcription, translation, apoptosis, and oncogenic pathways in Diamond-Blackfan anemia. *Stem Cells*. 2006; 24: 2034-44.
115. Horos R1, Ijspeert H, Pospisilova D, et al. Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. *Blood*. 2012; 119: 262-72.
116. Cmejlova J, Dolezalova L, Pospisilova D, et al. Translational efficiency in patients with Diamond-Blackfan anemia. *Haematologica*. 2006; 91: 1456-64.
117. Flygare J, Karlsson S. Diamond-Blackfan anemia: erythropoiesis lost in translation. *Blood*. 2007; 109: 3152-4.

118. Freson, K., Devriendt, K., Matthijs, G., et al. Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood*. 2001; 98: 85-92.
119. Weiss MJ, Mason PJ, Bessler M. What's in a name? *J Clin Invest*. 2012; 122: 2346-9.
120. Vlachos A, Klein GW, Lipton JM. The Diamond Blackfan Anemia Registry: tool for investigating the epidemiology and biology of Diamond-Blackfan anemia. *J Pediatr Hematol Oncol*. 2001; 23: 377-382.
121. Vlachos A, Ball S, Dahl N, et al. Diagnosing and treating Diamond Blackfan anaemia: results of an international clinical consensus conference. *Br J Haematol*. 2008; 142: 859-76
122. Glader BE, Backer K. Elevated red cell adenosine deaminase activity: a marker of disordered erythropoiesis in Diamond-Blackfan anaemia and other haematologic diseases. *Br J Haematol*. 1988; 68: 165-168.
123. Boria I, Garelli E, Gazda HT, et al. The ribosomal basis of Diamond-Blackfan Anemia: mutation and database update. *Hum Mutat*. 2010; 31: 1269-79.
124. Quarello P, Garelli E, Carando A, et al. Diamond-Blackfan anemia: genotype-phenotype correlations in Italian patients with RPL5 and RPL11 mutations. *Haematologica*. 2010; 95: 206-13.
125. Zhang J, Harnpicharnchai P, Jakovljevic J, et al. Assembly factors Rpf2 and Rrs1 recruit 5S rRNA and ribosomal proteins rpL5 and rpL11 into nascent ribosomes. *Genes Dev*. 2007; 21: 2580-92.
126. Gazda HT, Sheen MR, Vlachos A, 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; 83: 769-80.
127. Horos R, von Lindern M. Molecular mechanisms of pathology and treatment in Diamond Blackfan Anaemia. *Br J Haematol*. 2012; 159: 514-27.
128. Singh G, Akcakanat A, Sharma C, et al. The effect of leucine restriction on Akt/mTOR signaling in breast cancer cell lines in vitro and in vivo. *Nutr Cancer* 2011; 63: 264-71.

129. Jaako P, Debnath S, Olsson K, et al. Dietary L-leucine improves the anemia in a mouse model for Diamond-Blackfan anemia. *Blood*. 2012; 120: 2225-8.
130. Pospisilova D, Cmejlova J, Hak J, et al. Successful treatment of a Diamond-Blackfan anemia patient with amino acid leucine. *Haematologica*. 2007; 92: e66-7.
131. Li H. and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*. 2009; 25: 1754-60.
132. Li H, Handsaker B, Wysoker A, et al. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. 2009; 25: 2078-9.
133. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38: e164.
134. Seelow D, Schwarz JM, Schuelke M. GeneDistiller--distilling candidate genes from linkage intervals. *PLoS One*. 2008; 3: e3874.
135. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013; Chapter 7: Unit7.20.
136. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013; 14: 178-92.
137. Li J, Lupat R, Amarasinghe KC, et al. CONTRA: copy number analysis for targeted resequencing. *Bioinformatics* 2012; 28: 1307-13.
138. Cmejla R, Cmejlova J, Handrkova H, et al. Identification of mutations in the ribosomal protein L5 (RPL5) and ribosomal protein L11 (RPL11) genes in Czech patients with Diamond-Blackfan anemia. *Hum Mutat*. 2009; 30: 321-7.
139. Boocock GR, Morrison JA, Popovic M, et al. Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat Genet*. 2003; 33: 97-101.
140. Calado RT, Graf SA, Wilkerson KL, et al. Mutations in the SBDS gene in acquired aplastic anemia. *Blood*. 2007; 110: 1141-6.
141. Gradi A, Imataka H, Svitkin YV, et al. A novel functional human eukaryotic translation initiation factor 4G. *Mol Cell Biol*. 1998; 18: 334-42.

142. Carron C, O'Donohue MF, Choesmel V, et al. Analysis of two human pre-ribosomal factors, bystin and hTsr1, highlights differences in evolution of ribosome biogenesis between yeast and mammals. *Nucleic Acids Res.* 2011; 39: 280-91.
143. Idol RA, Robledo S, Du HY, et al. Cells depleted for RPS19, a protein associated with Diamond Blackfan Anemia, show defects in 18S ribosomal RNA synthesis and small ribosomal subunit production. *Blood Cells Mol Dis.* 2007; 39: 35-43.
144. Fukuda MN, Miyoshi M, Nadano D. The role of bystin in embryo implantation and in ribosomal biogenesis. *Cell Mol Life Sci.* 2008; 65: 92-99.
145. Zhao M, Wang Q, Wang Q, et al. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics.* 2013; 14 Suppl 11:S1.
146. O'Rawe J, Jiang T, Sun G, et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med.* 2013; 5: 28.

## 9. Abbreviations

1000g	1000 genome database
30Q	phred quality 30
Bop1	Block of proliferation 1
BWA	Burrows-Wheeler Aligner
BYSL	bystin-like
c-MYC	V-myc avian myelocytomatosis viral oncogene homolog
cADA	Adenosine deaminase
DBA	Diamond-Blackfan anemia
dbSNP	The Single Nucleotide Polymorphism database
E2F	E2F transcription factor
eIF	Eukaryotic initiation factor
FANCx	Fanconi anemia (x refers to competent group)
G1 phase	Gap 1 phase
G2 phase	Gap 2 phase
GAIT	Gamma interferon inhibitor of translation element
GATA1	GATA binding protein 1
GTP	Guanosin triphosphate
HbF	Fetal hemoglobin
IFN $\gamma$	Interferon, gamma
IGV	Integrative Genomics Viewer
IRES	Internal ribosome entry site
MDM2	Mouse double minute 2 homolog
mTOR	Mammalian target of rapamycin
NF- $\kappa$ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
OMIM	Online Mendelian Inheritance in Man
p27 <sup>Kip1</sup>	Cyclin-dependent kinase inhibitor 1B
p53	Tumor protein p53

Pes1	Pescadillo ribosomal biogenesis factor 1
PIM1	Pim-1 oncogene
Pol I	Polymerase I
RMPR	RNA component of mitochondrial RNA processing endoribonuclease
RPLx	Ribosomal protein of large subunit (x refers to number)
RPSx	Ribosomal protein of small subunit (x refers to number)
PCR	Polymerase chain reaction
SBDS	Shwachman-Bodian-Diamond syndrome
SegDup	Segmental duplication databaseS phase
S	Synthetic phase
snoRNA	Small nucleolar RNA
TCOF1	Treacher Collins-Franceschetti syndrome 1
TIF-11A	TATA box binding protein (TBP)-associated factor, RNA
Pol I	polymerase I
VCF	Variant call format
Zn finger	Zinc finger

## Attachment

### Gene\_panel

RPSA, RPS2, RPS3, RPS3A, RPS4X, RPS4Y, RPS5, RPS6, RPS7, RPS8, RPS9, RPS10, RPS11, RPS12, RPS13, RPS14, RPS15, RPS15A, RPS16, RPS17, RPS18, RPS19, RPS20, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27, RPS27A, RPS28, RPS29, RPS30, RPL3, RPL4, RPL5, RPL6, RPL7, RPL7A, RPL8, RPL9, RPL10, RPL10A, RPL11, RPL12, RPL13, RPL13A, RPL14, RPL15, RPL17, RPL18, RPL18A, RPL19, RPL21, RPL22, RPL23, RPL23A, RP 24, RPL26, RPL27, RPL27A, RPL28, RPL29, RPL30, RPL31, RPL32, RPL34, RPL35, RPL35A, RPL36, RPL36A, RPL37, RPL37A, RPL38, RPL39, RPL40, RPL41, RPLP0, RPLP1, RPLP2, RPL3, GATA1, FANCB, FANCO, FANCL, FANCD2, FANCD, FACD, FAD, FANCE, FANCG, FANCC, FACC, FANCF, FANCD1, FANCM, FANCI, FANCP, XPF, FANCO, FANCN, FANCA, RAD51C, SBDS

## Gene\_panel.py

```
import csv
import os

listing = os.listdir('path/to/exome_summary/file/')
for f in listing:
    info=[]
    data=[]
    with open ('path/to/exome_summary/file/' + f, "rt") as csvfile:
        reader = csv.reader(csvfile, delimiter=',', quotechar='"')
        for row in reader:
            data.append(row)

    data1=[]
f1= open ('/path/to/Gene_panel')
for i in f1:
    for j in data[1:]:
        if i.strip() == j[1].split('(')[0] and j[2] != 'synonymous SNV' and float(j[27])>50:
            data1.append(j)

f3= csv.writer(open('/path/to/output/' + f, 'wt'))
info=[]
f3.writerow(data[0])
for i in data1:
    f3.writerow(i)
```

## **Clinvar.py**

```
import csv
import sys

rs=[]
info=[]
with open('/path/to/clinvar.csv', "rt") as csvfile:
    reader = csv.reader(csvfile, delimiter=',', quotechar="")
    for row in reader:
        rs.append(row[3])
        info.append(row)

result=[]
patient= open('/path/to/PX')
for i in patient:
    if i.strip().split()[8] in rs:
        result.append(i + info[rs.index(i.strip())])

f=csv.writer(open('/path/to/PX_clinvar.csv', 'wt'))
for i in result:
    f.writerow(i)
```

