

Original Article

Loss of coordinated expression between ribosomal and mitochondrial genes revealed by comprehensive characterization of a large family with a rare Mendelian disorder



Brendan Panici ^{a,1}, Hosei Nakajima ^a, Colleen M. Carlston ^{b,2}, Hakan Ozadam ^a, Can Cenik ^a, Elif Sarinay Cenik ^{a,*}

^a Department of Molecular Biosciences, University of Texas at Austin, Austin, USA

^b University of Utah Health, Salt Lake City, UT, USA

A B S T R A C T

Non-canonical intronic variants are a poorly characterized yet highly prevalent class of alterations associated with Mendelian disorders. Here, we report the first RNA expression and splicing analysis from a family whose members carry a non-canonical splice variant in an intron of *RPL11* (c.396 +3A>G). This mutation is causative for Diamond Blackfan Anemia (DBA) in this family despite incomplete penetrance and variable expressivity. Our analyses revealed a complex pattern of disruptions with many novel junctions of *RPL11*. These include an *RPL11* transcript that is translated with a late stop codon in the 3' untranslated region (3'UTR) of the main isoform. We observed that *RPL11* transcript abundance is comparable among carriers regardless of symptom severity. Interestingly, both the small and large ribosomal subunit transcripts were significantly overexpressed in individuals with a history of anemia in addition to congenital abnormalities. Finally, we discovered that coordinated expression between mitochondrial components and *RPL11* was lost in all carriers, which may lead to variable expressivity. Overall, this study highlights the importance of RNA splicing and expression analyses in families for molecular characterization of Mendelian diseases.

1. Introduction

Many genetic diseases are incompletely penetrant and exhibit variable expressivity. The nature of the causative genetic mutation, epistatic interactions, and environmental factors can accentuate, mitigate, or sometimes completely mask a potentially pathogenic genetic variant. Diamond Blackfan Anemia (DBA) is an example of a monoallelic, genetic disease exhibiting incomplete penetrance and variable expressivity (reviewed in [1]). DBA can lead to developmental and hematological defects that often manifest in the first year of infancy and is caused by haploinsufficiency in the protein synthesis machinery due to loss of function or missense mutations. DBA mutations have been identified in a ribosome maturation factor, *TSR2*, two transcriptional regulators, *GATA1*, and *MYSM1*, and ~19 ribosomal protein genes including *RPL11* [2–5].

RPL11 mutations in DBA patients are associated with a spectrum of phenotypes. Compared to other DBA genes such as *RPS19*, individuals with *RPL11* mutations have an increased incidence of cleft palate and

thumb anomalies [6,7]. Loss of one functional copy of *RPL11* leads to impaired red blood cell development, reduced P53, increased c-MYC in mouse [8], and in contradiction increased P53 and apoptosis in zebrafish [9]. *RPL11* mutations are relatively common in DBA patients affecting ~7%. In comparison, *RPS19* mutations are the most prevalent cause (~30–38% of patients) [5].

While initial characterizations of DBA patients focused on nonsense and missense coding sequence mutations in ribosomal proteins, recent evidence indicates that extended splice site variants are underappreciated and highly prevalent [5]. Splicing variants may disrupt a splicing site, activate a cryptic/alternative splice isoform, or introduce novel junctions with adverse consequences for the open reading frame of the corresponding gene [10]. Genome and exome sequencing in patients revealed many potentially pathogenic intronic variants that may alter splicing. For example, a rare, synonymous variant in *GFI1B*—a transcription regulator expressed in hematopoietic lineage—is responsible for variation in blood traits across humans [11]. Despite the high prevalence of splice site variants in DBA patients (~18.5% [5]), no

* Corresponding author.

E-mail addresses: Brendan.Panici@bcm.edu (B. Panici), hakan.ozadam@utexas.edu (H. Ozadam), ccenik@austin.utexas.edu (C. Cenik), esarinay@utexas.edu (E.S. Cenik).

¹ Present address: Baylor College of Medicine, Houston, USA

² Present address: University of California, San Francisco School of Medicine, San Francisco, CA

<https://doi.org/10.1016/j.ygeno.2021.04.020>

Received 11 January 2021; Received in revised form 30 March 2021; Accepted 11 April 2021

Available online 20 April 2021

0888-7543/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

previous study has characterized the RNA expression and splicing changes in patient samples harboring such splice variants.

Here, we report the first detailed RNA expression and splicing analysis in a large family with a noncanonical intronic splice variant of the ribosomal protein gene, *RPL11*. Eight family members spanning four generations harbor the particular non-canonical *RPL11* splicing variant (c.396 +3A>G). In this family, two carriers have prominent congenital symptoms (congenital heart defect & short stature) and the other carriers are either asymptomatic or have small thumbs that does not require any surgical intervention [12]. This particular *RPL11* splicing variant had also been observed in an individual with bilateral triphalangeal thumbs, heart defects, short stature and anemia [13].

In this study, we address the following questions: How does the pattern of *RPL11* splicing differ in carriers compare to noncarrier relatives? Is there a shared gene expression signature among *RPL11* splicing variant carriers? Do we observe any significant changes in gene expression patterns in carrier individuals with a history of anemia compared to *RPL11* variant carriers with few to no symptoms?

Our analyses revealed a complex pattern of splicing disruptions including an *RPL11* transcript that is translated with a late stop codon in the 3' untranslated region (3'UTR) of the main isoform. *RPL11* transcript abundance is comparable among carriers regardless of symptom severity. Interestingly, both the small and large ribosomal subunit transcripts were significantly overexpressed in individuals with a history of anemia in addition to congenital abnormalities. Finally, we observed that coordinated expression between mitochondrial components and *RPL11* was lost in all carriers, which may lead to variable expressivity. Overall, this study highlights the importance of RNA splicing and expression analyses in families for molecular characterization of Mendelian diseases.

2. Results

2.1. RNA sequencing analysis of a large family with a noncanonical splice variant in *RPL11*

We performed whole blood RNA expression and splicing analysis from members of a family spanning four generations that harbor an intronic splicing variant in ribosomal protein gene *RPL11* (Fig. 1) [12]. Family members carrying the *RPL11* variant (light red circles/squares) that are not affected by anemia had small thumbs or thenar eminences that did not require surgical intervention, and one carrier of the familial variant was asymptomatic. Clinical findings in the carrier individuals

with a history of anemia (dark red circle and square) are as follows: Individual IV.5 was affected with bilateral thumb defects, anemia, short stature, congenital heart defects (ASD and double superior vena cava), peno-scrotal transposition hypospadias and volvulus. Individual III.13 has cleft palate, anemia, celiac-like gastrointestinal issues and slender but not hypoplastic thumbs. At the time samples were taken for RNA extraction, both individuals had mild anemia, since that time individual IV.5's hemoglobin continued to rise to now be within the lower range of normal. Detailed clinical characterization of this family was previously described in [12].

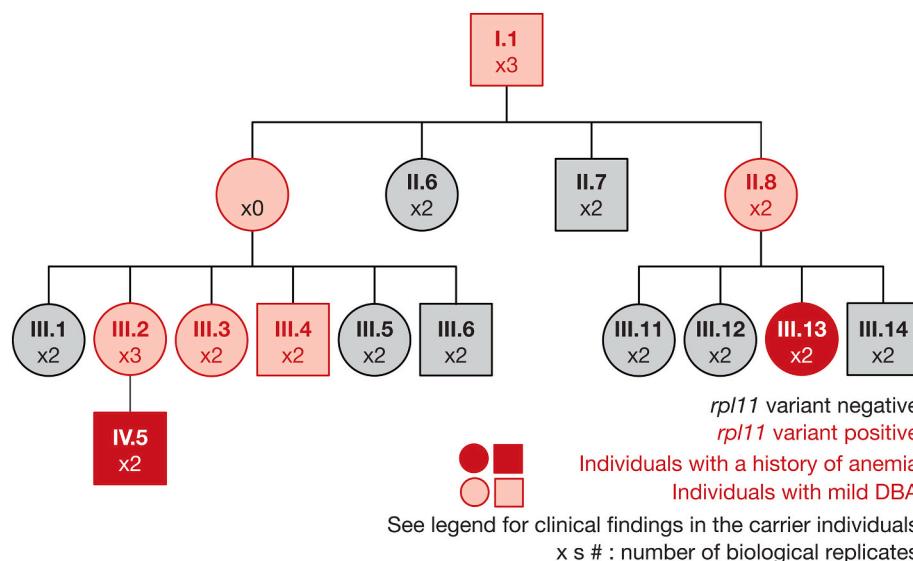
Using our RNA sequencing data, we first confirmed the presence of the intronic variant (c.396 +3A>G) for each carrier (S1 Figure). Given that all samples are generated from the same biological material source and family, we expected to observe global similarity in expression profiles across the samples. We identified five samples that behaved differently compared to the rest of the samples (S2 Figure, mean spearman correlation of outlier samples <0.9). These outlier samples predominantly had low percentage of genome mapping reads and were removed from further analyses (S3 Figure). The overall mean correlation coefficient was 0.97 between remaining samples (S2 Figure).

After sequencing depth normalization (median of ratios based on DESeq2), we didn't observe any bias for total normalized read counts per sample (S4A Figure). Principal component analyses revealed that RNA-Seq read counts from samples of the same individual tended to cluster suggesting that inter-individual variance in gene expression dominates technical variability in our data (S4B Figure). Given the large number of samples, library preparation was carried out across several batches as described in S1 Table. Principal component analysis (PCA) suggested that batch variables do not explain a significant component of the variation observed in read counts (S4C Figure). Taken together, 39 samples passed our strict quality control metrics and were used for all subsequent analyses.

2.2. The most abundant *RPL11* alternate splicing isoform results in a frame shift and a 3'UTR readthrough

Diamond Blackfan Anemia (DBA) in this family is caused by an *RPL11* 5'splice site (5'SS) variant (c.396 +3A>G) (Fig. 2A, S1 Figure). This variant is not in the canonical splice donor site, and thus we first characterized the consequent perturbation in the splicing of the *RPL11* gene. Specifically, we hypothesized that differences in splicing patterns of *RPL11* transcript may underlie differences in symptom among carriers.

Fig. 1. Four generation family pedigree depicts individuals who carry the pathogenic *RPL11* variant c.396 +3A>G [adapted and simplified from (Carlston et al., 2017)]. Light red circles and squares represent individuals who carry the *RPL11* variant. Clinical findings in the carrier individuals with a history of anemia (dark red circle and square) are as follows: Individual IV.5 was affected with bilateral thumb defects, anemia, short stature, congenital heart defects (ASD and double superior vena cava), peno-scrotal transposition hypospadias and volvulus. Individual III.13 has cleft palate, anemia, celiac-like gastrointestinal issues and slender but not hypoplastic thumbs. Other family members carrying the *RPL11* variant (light red circles/squares) had small thumbs or thenar eminences that did not require surgical intervention, and one carrier of the familial variant was asymptomatic (annotated as individuals with mild DBA).



A

Fig. 2. *RPL11* variant c.396 + 3A > G alters the splicing pattern of *RPL11*. A Diagram depicts exon 4-exon 5 junction RNA reads with percentages from carriers and noncarriers of the pathogenic *RPL11* variant. Left, green boxes represent exon 4 and exon 5, height of the boxes represent read abundance unless noted otherwise. Start and end positions are given in comparison to splicing isoform that produces functional canonical *RPL11* protein (start, end:0). The numbers of reads mapping to each junction per individual are listed in Table S2. The most abundant splicing variant (marked with *) results in a frame-shift and a 3'UTR ribosome readthrough. B Alignment of canonical *RPL11* and the protein encoded by the most abundant alternate isoform was shown where amino acids are colored based on similarity.

B

RPL11	1 MAQDQGEKENPMRELIRKLCLNICVGESGDRRLTRAALKVLEQLTGQTPVF	50
* RPL11 variant	1 MAQDQGEKENPMRELIRKLCLNICVGESGDRRLTRAALKVLEQLTGQTPVF	50
RPL11	51 SKARYTVRSFGIRRNEKIAVHCTVRGAKEEILEKGLKVREYELRKNNFS	100
RPL11 variant	51 SKARYTVRSFGIRRNEKIAVHCTVRGAKEEILEKGLKVREYELRKNNFS	100
RPL11	101 DTGNFGFGIQEHIDLGIKYDPSIGIYGLDFYVV — LGRPGFS —	140
RPL11 variant	101 DTGNFGFGIQEHIDLGIKYDPSIGIYGLDFYVV CWVGQVSASQTRSAGQA	150
RPL11	141 — IADKKRRTGCIGA — KHRISKEEAMRWFQQKYDGII	174
RPL11 variant	151 ALGP NTESAKRRPCAGSSRSMMGSSFLANKFPFLSKRA I K SFQ —	193
RPL11	175 LPG	177

We classified all exon junction reads proximal to the *RPL11* variant with respect to the 5'splice site position (5'SS) denoting the donor site of the canonical *RPL11* isoform as +0 (Fig. 2A; S2 Table). We found that both the types of aberrant splice junctions and their relative abundances were very similar among all carrier individuals regardless of symptom severity, suggesting that the severity of symptoms is not associated with the frequency or type of aberrant splicing event.

The most abundant minor isoform in carriers has an alternative 5' splice site at the +2 position (5.7% in carriers compared to 0.02% in noncarriers, Fig. 2A). A previous study of samples carrying the same variant had revealed a mis-splicing event with an alternative 5'SS at the +66 position relative to the canonical 5'SS [12]. This report relied on PCR amplification of *RPL11* cDNA and Sanger sequencing of a band isolated from a DNA gel thus likely missed the most abundant +2 5'SS variant. Importantly, the isoform containing the 5'SS at +2 results in a frame-shift with a stop codon in the 3'UTR of the *RPL11* canonical transcript. Thus, this isoform is likely a poor substrate for nonsense-mediated RNA decay due to an absence of stop codon upstream of an intron [14]. This alternative *RPL11* isoform when translated shares the first 134 amino acids with the canonical *RPL11* protein (Fig. 2B). We currently do not know whether this aberrant splice variant is translated into a stable protein as 3'UTR ribosome readthrough can lead to mRNA degradation via no-go decay (NGD; reviewed in [15] or protein

instability [16], Fig. 2B).

In addition to the most abundantly observed +2 5'SS variant, we detected ~80 nucleotide antisense exon-4-exon-5 junction reads from the +2 5'SS (1.2% of total exon-exon junction reads). Strikingly, these reads are specifically observed in carriers but not in noncarrier individuals (Fig. 2A). Importantly, every single sample carrying the *RPL11* variant expressed the antisense exon-exon junction reads (S2 Table). The alternative antisense exon-exon junctions are observed in multiple uniquely mapped reads with distinct end coordinates in ~20 different RNA-Seq libraries from *RPL11* variant carriers, and thus are not attributable to alignment or library preparation artifacts. Furthermore, no other regions in the human genome have sequence similarity to the antisense reads. We wondered whether these junction reads could originate from a spliced antisense transcript; however, we did not observe any reads besides those mapping to the exon-exon junction on the antisense strand overlapping the *RPL11* locus. Taken together, these findings reveal an intriguing antisense RNA product that is specifically produced from the *RPL11* splice variant containing allele.

2.3. Shared gene expression changes among individuals that carry the *RPL11* variant c.396 + 3A > G

We next sought to determine differential gene expression patterns

among carriers compared to noncarriers. For this purpose, we used two complementary approaches (DESeq2 and Limma-dream; see Methods). *RPL11* was among the set of differentially expressed genes with an estimated abundance of ~%57 in carriers compared to noncarriers (DESeq2 log-fold change is 0.8, noncarriers to carriers). We hypothesized that *RPL11* levels might be different among carriers with mild symptoms versus those with a history of anemia but failed to observe any significant difference in *RPL11* expression. Hence, *RPL11* transcript abundance differences in carriers are unlikely to explain DBA symptom severity in these individuals.

We identified nine genes with RNA expression differences among carriers compared to noncarriers. (DESeq2 adjusted *p*-value <0.01; limma-dream *p*-value <0.05, Fig. 3A and B). Interestingly, when we restricted the RNA expression comparison to only individuals with mild DBA symptoms compared to noncarriers, we identified 23 genes with RNA expression differences (DESeq2 adjusted *p*-value <0.01; limma-dream *p*-value <0.05, Fig. S5A). Genes that are differentially expressed between carriers with mild symptoms and noncarriers were surprisingly not differentially expressed in individuals with a history of anemia (Fig. S5B; S3 Table and S4 Table). This result suggests the existence of additional compensatory changes only in individuals with mild or no DBA symptoms (small thumb or asymptomatic).

Among the genes overexpressed in carriers (S3 Table), we identified three implicated in cancer biology: *TMEM158*, *CDK11A* and *AKAP12*. *TMEM158* is a Ras induced senescence protein which is involved in proliferation arrest and its overexpression is linked to ovarian cancer [17]. Cyclin dependent kinase 11a (*CDK11A*), is a serine/threonine-protein kinase that controls cell cycle [18], and *AKAP12* is a tumor suppressor, inhibitor of cell migration and regulates cell-cycle progression through Src-mediated oncogenic signaling and cytoskeletal pathways [19,20]. An individual in this family with mild DBA symptoms was diagnosed with breast cancer at age 36, and there is an increased observed incidence of neoplasia including breast in the DBA registry. Thus, the observed significant overexpression of tumor suppressor genes; *TMEM158*, *CDK11A* and *AKAP12*, could be valuable understanding manifestation of increased incidence of neoplasia in DBA patients [21].

Among the under expressed genes in carriers, we identified *EPHB3*, an ephrin receptor that is expressed in early embryos and important for palate formation [22,23]. Among two family members with a history of anemia, cleft palate is observed in one, and congenital abnormalities are observed in the other. The observed differences in *EPHB3* transcript may relate to these symptoms.

Taken together, our findings suggest that individuals with the *RPL11* variant express the canonical isoform at a reduced expression level. We estimate that the c.396 +3A>G variant produces ~14% of the major splice isoform compared to the wild-type allele. Furthermore, this reduction in *RPL11* transcript leads to consistent but modest changes in RNA expression in a small number of genes that are involved in immunity, cell cycle, and cellular migration.

2.4. Reduced expression of all ribosomal protein genes in carriers with mild DBA symptoms mirror the lower abundance of *RPL11*

Expression of ribosomal components is stoichiometric [24,25]. Loss of expression coordination leads to unproductive energy consumption, proteotoxic stress [26], impaired cell proliferation, tumorigenesis and metastasis [examples: *RPS3A* [27], *RPL5* [28], *RPL15* [29–32]].

Hence, we hypothesized that other ribosomal proteins may respond in a coordinated manner to reduced *RPL11* transcript expression. While individual ribosomal proteins were not identified as differentially expressed, we reasoned that a feedback regulation may lead to a coordinated regulation of the ribosomal subunits. Indeed, we observed that the expression of ribosomal protein genes was reduced modestly but consistently for both the small and large ribosomal subunit proteins in carrier individuals with mild DBA symptoms. We plotted the moderated

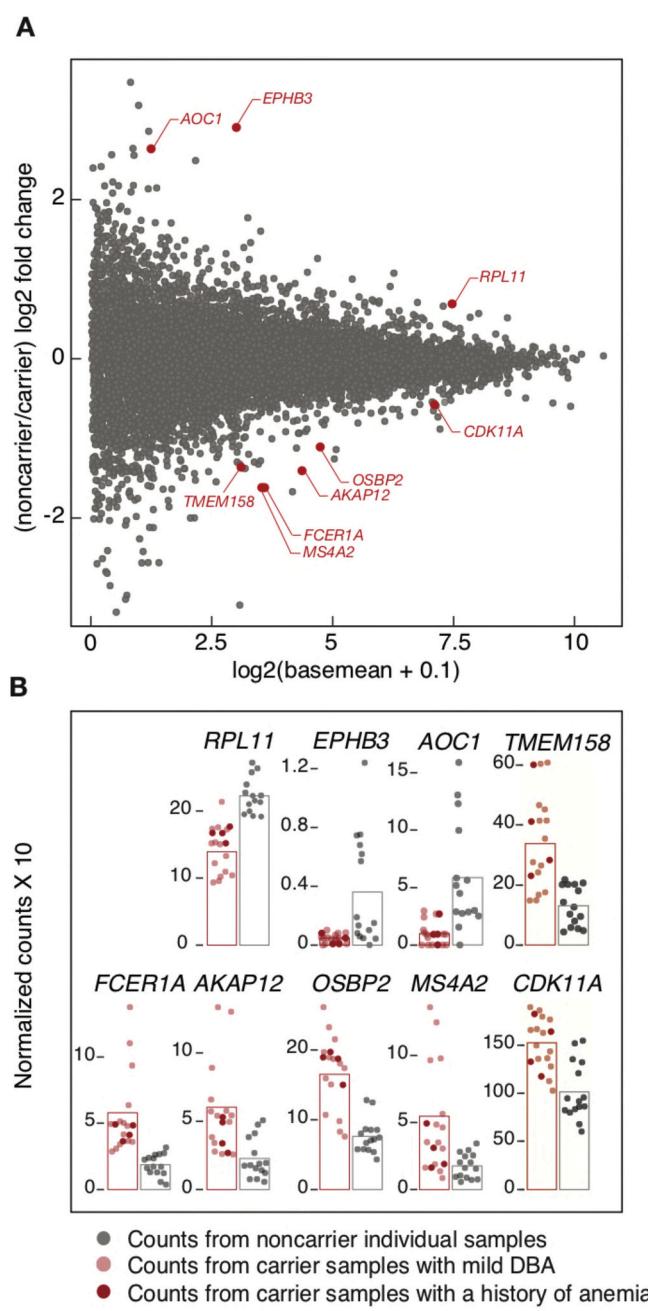


Fig. 3. Shared gene expression changes among individuals that carry *RPL11* variant c.396 +3A>G. Gene expression log₂ fold changes (y-axis) and log₂ base mean values per gene as calculated by DESeq2 were plotted. Y-axis represents log₂ fold change of noncarriers compared to carriers. Red points are genes that are differentially expressed in two complementary approaches. (see Methods; Limma FDR cutoff <0.05, DESeq2 FDR cutoff <0.01). B Gene expression values per sample were plotted for differentially expressed genes in panel A. Noncarriers were labeled with grey. Carriers were shown in red and samples from individuals with a history of anemia are indicated with darker shade of red. The bars represent mean expression among all carriers (light red) and among noncarriers(grey).

log₂-fold change (DESeq2) of all ribosomal proteins onto the human ribosome structure (PDB ID: 4UG0 [33]) to investigate any structural patterns to the feedback regulation (Fig. 4A). We observed widely distributed reduced expression suggesting shared regulation among large and small ribosomal proteins with few exceptions (*RPL27A*, *RPL28*, *RPL39* and *RPL26*, *RPS29*, *RPS17*, and *RPS26*; example plots are

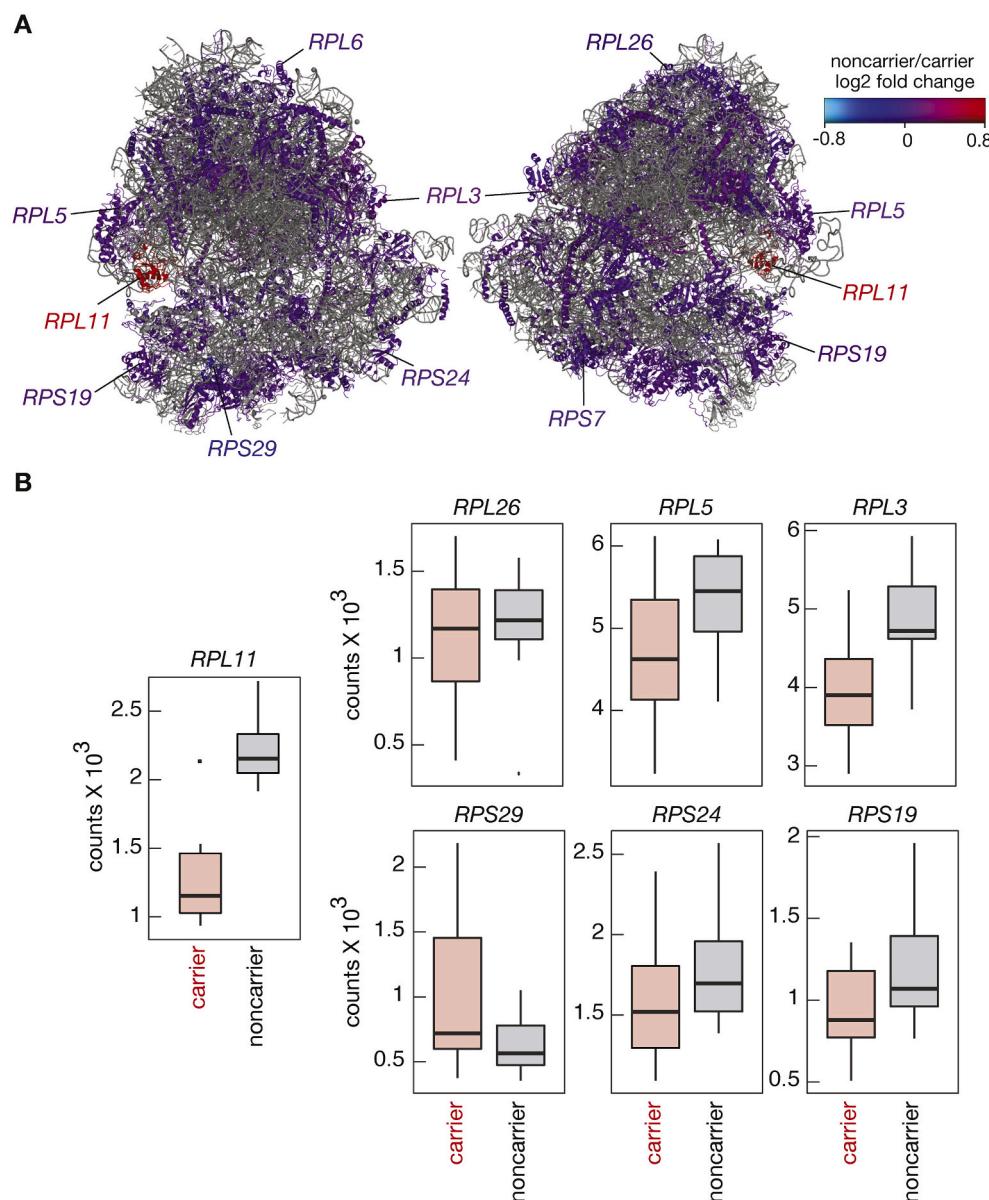


Fig. 4. Effect of *RPL11* variant c.396 +3A>G on the expression of both large and small ribosomal subunits. A Log2 shrinkage values (DESeq2) for each ribosomal protein gene were plotted in color scale on the human ribosome structure (PDB ID: 4UG0). B Distribution of normalized *RPL11* expression in carriers (light red) and noncarriers (grey) were plotted in the left panel. Right six panels display the distribution of three exemplary large (top) and three small (bottom) subunit ribosomal protein RNAs in carrier (red) and noncarrier (grey) individuals.

shown in Fig. 4B). This observation reveals a coordinated feedback regulation among other ribosomal protein genes in carrier individuals with mild DBA to compensate for the lower levels of *RPL11*.

2.5. Small and large ribosomal protein genes are significantly overexpressed in individuals with a history of anemia compared to carriers with mild or no symptoms

Given that carriers with mild DBA symptoms (one individual with no symptoms) have lower ribosomal protein gene expression compared to noncarriers (Fig. 4), we expected to observe a similar trend of lower expression of ribosomal protein gene expression in carriers regardless of symptom severity. Surprisingly, we observed a different trend for the carrier individuals with a history of anemia: Both small and large ribosomal proteins are significantly overexpressed in these individuals compared to carrier individuals with mild symptoms (Fig. 5A and B, ROAST, *p*-value = 0.04; see Methods). Specifically, all ribosomal protein genes except four (*RPL11*, *RPS29*, *RPS17*, and *RPS26*) are overexpressed in carriers with a history of anemia (Fig. 5). These results suggest that the compensation at the gene expression level for other ribosomal

proteins in response to lower levels of *RPL11* transcript might be different in individuals with prominent symptoms compared to carriers with mild DBA.

2.6. Coordination between expression of mitochondrial components and *RPL11* are lost in all carriers

As previously implicated and shown in this study, feedback loops connect the expression of functionally related genes such as ribosomal proteins [24,25]. A novel hypothesis regarding variable expressivity of a phenotype is loss of gene expression coordination. Specifically, we aimed to identify clusters of genes that are expressed proportionally to *RPL11* yet lose this co-expression pattern once one copy of *RPL11* is altered. This type of relationship cannot be detected through conventional differential expression analyses as *RPL11* loss would be predicted to lead to increased variance in the expression of the related genes and not their simple down- or over-expression.

To investigate this hypothesis, we used the variance of log-ratios of read counts between a given pair of genes as a function of *RPL11* carrier status. This metric, called emergent proportionality ('theta E value'),

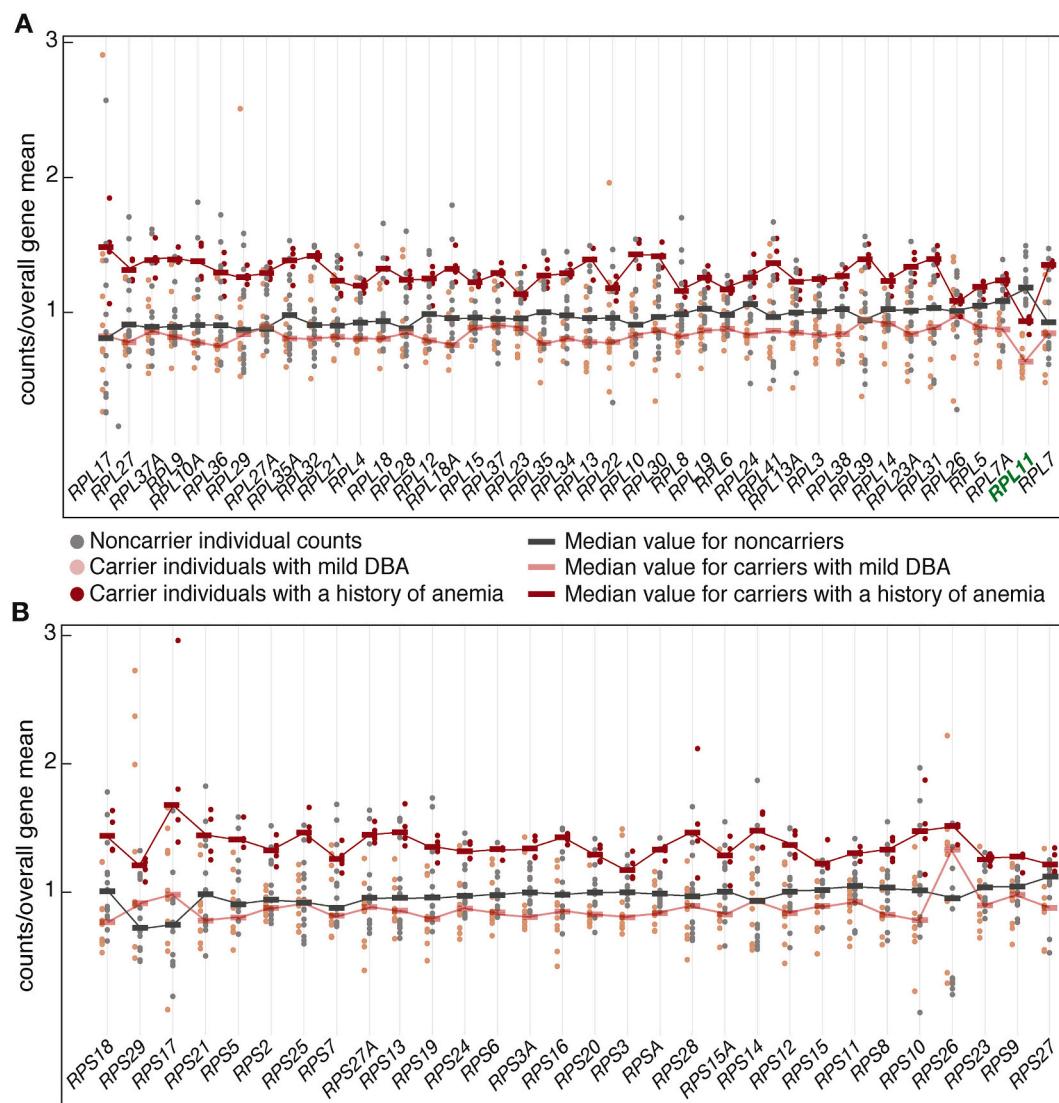


Fig. 5. Large and subunit ribosomal protein mRNAs are differentially overexpressed in carrier individuals with a history of anemia compared to other carriers. Read counts for all ribosomal protein genes in each sample were divided to the mean of the cohort and were plotted for noncarriers (grey), and carriers. Carrier individuals with (dark red) or without (light red) a history of anemia were separated. Read counts that map to each large (top plot) and small (bottom plot) ribosomal protein transcript were divided by the mean for the corresponding ribosomal protein gene (y-axis). The thick color bars on the plot represent the median value for each gene, these bars are connected with a fine line to represent the overall pattern for each group. Large and small subunit ribosomal protein mRNAs except *RPL11* (highlighted in green) are significantly overexpressed in carriers with a history of anemia compared to carriers with mild DNA (ROAST–rotating gene set analysis, p value = 0.04). The difference between carriers without a history of anemia and noncarriers is not significant for ribosomal protein mRNA expression (ROAST, p value = 0.1).

enables discovery of gene pairs that are coordinately expressed if the ratio of their respective read counts is maintained among individuals regardless of their absolute expression level and carrier status [34,35]. Proportionality analyses formalize compositional data analysis methods for RNA expression and robustly infer underlying biological variables [35,36].

We identified 357 genes with higher variance of expression ratios compared to *RPL11* between carrier and noncarrier individuals ($\theta_E < 0.15$; S6 Figure; Methods). Remarkably, carrier individuals dominated the variance for every single one of these top 357 genes ($p < 2.2e-16$; binomial test), suggesting an overall loss of coordination of gene expression in relation to *RPL11* in carriers (Fig. 6A). As a control, we conducted the same analysis with *CDK11*, a gene we found to be differentially increased in carriers (Fig. 3A). No significant results for directionality were obtained revealing that our results are not a simple consequence of differential expression (Fig. 6A).

To better illustrate these expression patterns, we plotted one gene

with similar and another with different variance of log-ratios (Fig. 6B, C). *RPL4* to *RPL11* (θ_E value = 0.91) ratio among carriers and noncarriers had similar variance suggesting that their expression coordination was similar in both groups (Fig. 6B, C; left panels). On the other hand, *FIBP* to *RPL11* (θ_E value = 0.01) ratios were much more variable in carriers (Fig. 6B, C; right panels).

We next determined the functional categories of genes that lost expression coordination with respect to *RPL11*. We found a very strong enrichment for genes related to mitochondria components and function (Fig. 7A and B; S7 Figure). In particular, there was a 45.4-fold enrichment for genes related to mitochondrial large ribosomal subunit, and a 10.6-fold enrichment for genes associated with the mitochondrial membrane (S5 Table; adjusted p -value < 0.05). We validated the robustness of these results to potential outliers via sub-sampling to calculate false discovery rates (S6B Figure). Overall, these results reveal a loss of coordination between *RPL11* and mitochondrial components among carriers compared to noncarriers.

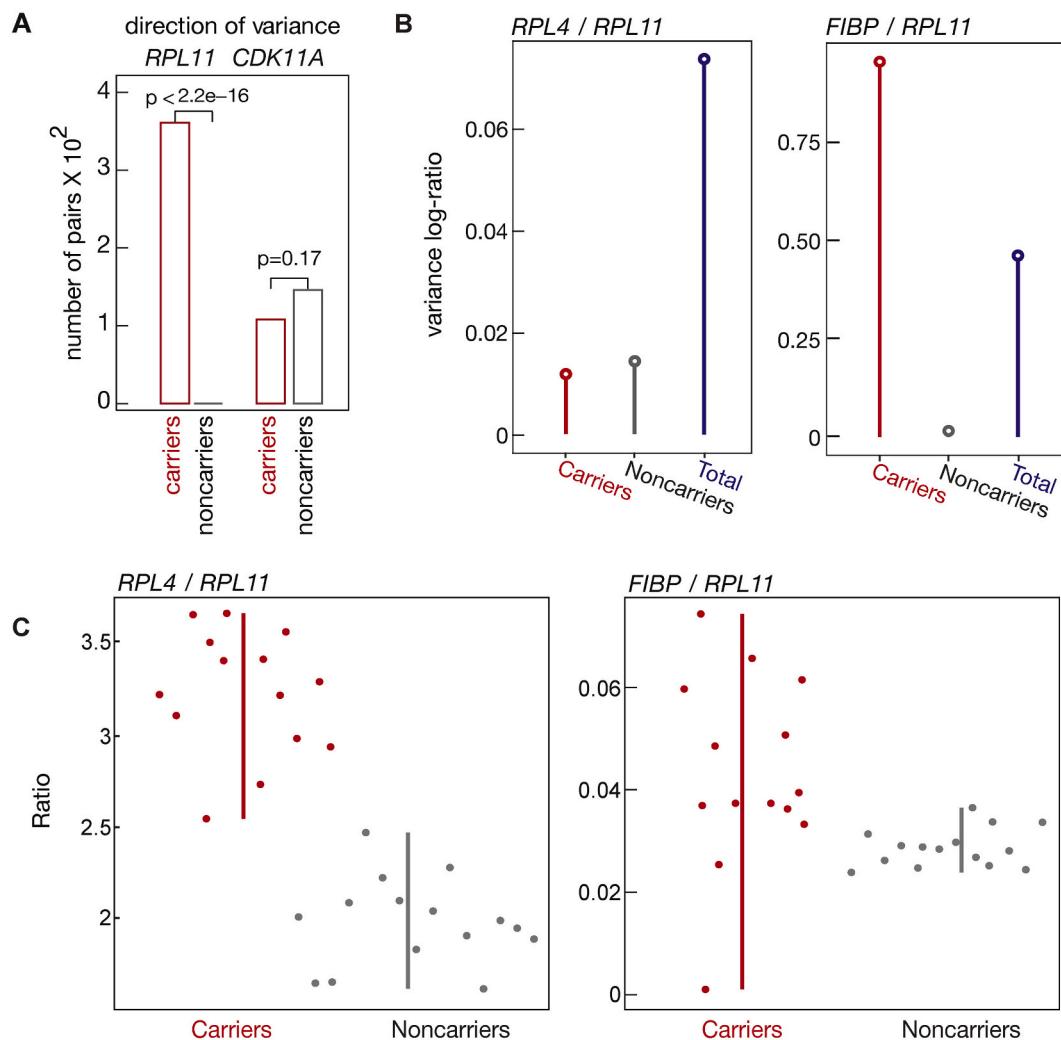


Fig. 6. Emergent proportionality highlights loss of coordination between gene pairs. A The log-ratio of read counts of *RPL11* or *CDK11A* compared to all other genes were calculated. Variance of the log-ratios were used to calculate emergent proportionality (theta E) comparing carriers to non-carriers [34]. For each gene pair with theta E < 0.15—for proportionality to *RPL11*—or with theta E < 0.25—for proportionality to *CDK11A*—, we determined the group (carrier or noncarriers) that dominates the total group variance. The significance of the bias towards either group was determined using a binomial test. B The total variance log-ratio (VLR) is plotted along with the VLR for noncarriers and carriers. Left panel: No genotype group dominates the total VLR, therefore *RPL4* and *RPL11* have a high theta E value (0.91). Right panel: VLR is higher in carrier samples for *FIBP* / *RPL11* (theta E = 0.01). C The ratio of two gene pairs for a sample is plotted and colored by *RPL11* status. These ratios are ordered by their sample index. Left panel: The ratio of *RPL4* and *RPL11* has similar variance by genotype, although the mean of the ratios differs. Right panel: For the ratio of *FIBP* and *RPL11*, carrier samples are spread out while noncarriers are distinctly more clustered.

3. Discussion

Variants in splicing regulatory sequences are major causative factors in a wide range of diseases including DBA [5,37]. Variants causing improper splicing can alter the encoded protein sequence and thus be detrimental to protein function. Here, we carried out an RNA sequencing study of a family spanning four generations with an inherited intronic variant affecting one of the two copies of *RPL11*. This noncanonical splice variant (c.396 +3A>G) is causative for Diamond Blackfan Anemia (DBA) in this family with incomplete penetrance and variable expressivity.

We first determined the frequencies of aberrant splicing patterns attributable to this intronic variant. A previously missed splice junction with a 5' splice donor site at +2 position relative to the canonical *RPL11* isoform has the highest abundance. Interestingly, this isoform encodes a protein that is 75% identical to *RPL11* at its N-terminus and includes a late stop in the 3'UTR of the canonical isoform. Given the frameshifted polypeptide sequence in its C-terminal, this isoform may have proteotoxic consequences. In particular, SYO1, the dedicated chaperone for

assembling *RPL5* and *RPL11* into 4S RNP complex, binds to the N-terminal of *RPL11* [38]. Thus, the aberrant *RPL11* isoform may sequester SYO1 and abrogate ribosome biogenesis. Future work analyzing protein samples from this family for the variable C terminus of the *RPL11* variant would further clarify whether this alternate isoform is stable.

The expression of *RPL11* in carrier individuals is ~57% of noncarrier individuals. If we assume that 50% of *RPL11* mRNA is transcribed from the variant *RPL11* locus, then we can deduce that ~43% *RPL11* transcripts originating from the aberrantly spliced variant are degraded. This leaves only ~7%—or lower if there is dosage compensation—of *RPL11* isoforms originating from the variant allele among the pool of *RPL11* transcripts. Importantly, all carriers express the same set of aberrant splice isoforms with similar relative abundance. Therefore, symptom severity in this family is not attributable to difference in the splicing pattern among the carriers.

Among splice isoforms encoded by the *RPL11* variant, we were surprised that the second most abundant splice variant was anti-sense to the *RPL11* gene. These antisense junction reads are not present in any of the noncarriers, but surprisingly all of the carriers express them. Antisense

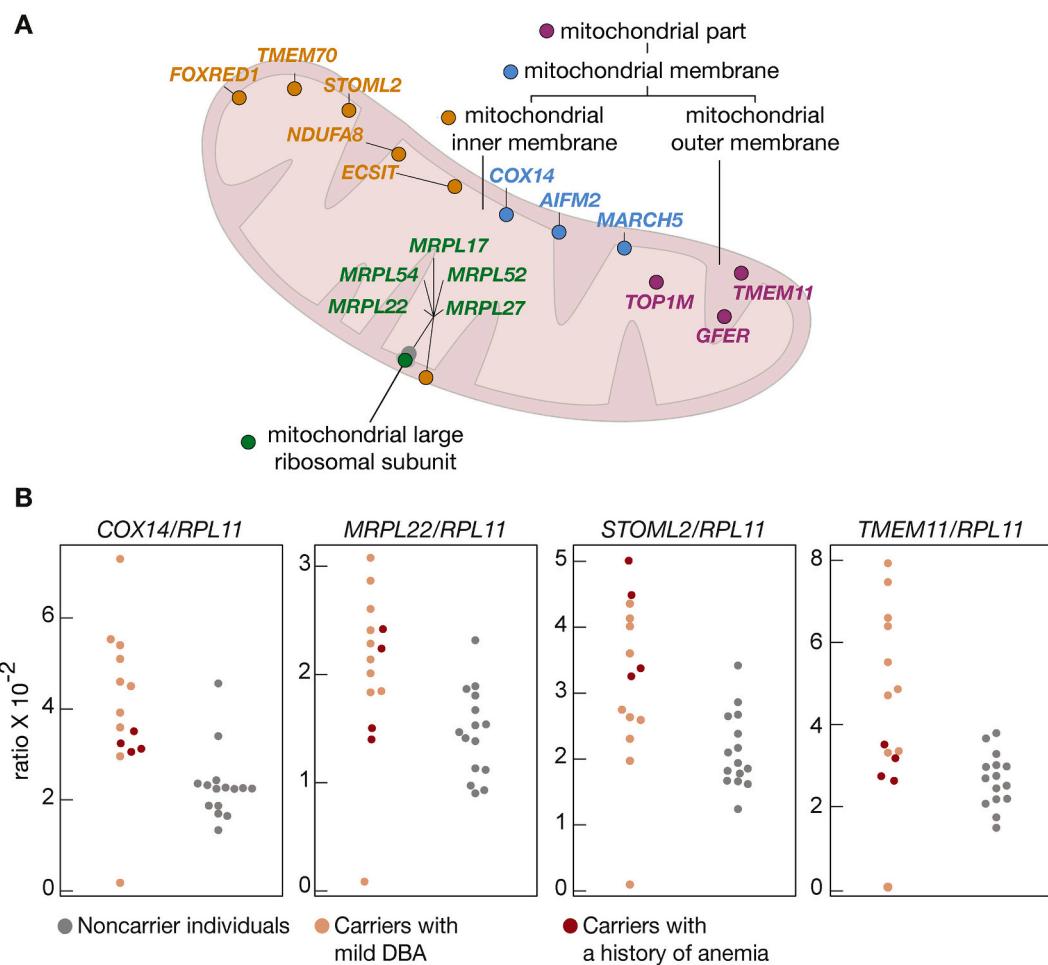


Fig. 7. Expression of mitochondrial genes becomes uncoordinated with *RPL11* expression in carriers compared to noncarriers. A Genes with the lowest theta E values were highly enriched for mitochondrial localization as shown in the diagram [GO category significant enrichment analysis, adjusted *p*-value < 0.05, FuncAssociate 3.0 [64], Table S5]. The only GO category that is not shown in this diagram belongs to cytoplasmic large ribosomal subunit components. B For selected mitochondrial genes, the ratio of read counts were plotted for the two groups.

exon-exon junctions have not been studied extensively in humans but were previously reported in *Plasmodium falciparum* [39] where they were detected in many transcripts. One hypothesis is that the antisense exon-exon junction reads result from splicing of an antisense transcript. The antisense RNA isoform we identified is only ~80–100 nucleotides long, and we did not detect any other antisense read that maps to elsewhere in the *RPL11* coding region suggesting that antisense transcription of the *RPL11* gene would have to be highly unstable. We think that antisense transcription spanning only the exon4-exon5 region of the *RPL11* gene remains a more plausible alternative. Antisense oligonucleotides can be used as a tool to skip an exon or restore a cryptic splicing event [40]. However, the function of the detected antisense exon-exon junction remains unknown.

The family carrying *RPL11* variant fortunately display low penetrance for DBA symptoms (25%) compared to other DBA patients (~75%). The two most affected patients have anemia in addition to congenital defects. The low penetrance might be specific to the ribosomal protein gene *RPL11*. Alternatively, the nature of the variant itself may be responsible. However, we observed highly similar splicing patterns for in all carrier individuals suggesting that any cis-acting splicing effects are unlikely to be explanatory.

To understand causes of incomplete penetrance and variable expressivity of DBA in this family, we investigated differential gene expression patterns in carriers with mild symptoms versus those with a history of anemia. Interestingly, carrier individuals with no history of anemia modestly downregulate other large and small ribosomal proteins

suggesting a compensation mechanism for the lower expression of *RPL11*. On the other hand, carriers with a history of anemia significantly overexpress the majority of small and large ribosomal protein transcripts compared to other carriers. This could either be a response to symptom severity or directly be causative of the manifestations of DBA.

A lack of compensation in the expression of other ribosomal protein genes could lead to higher levels of proteotoxicity due to the increased abundance of unassembled ribosomal proteins. Majority of the overexpressed large subunit ribosomal protein mRNAs, if translated, would need to be degraded by the proteasome as they are usually unstable if not assembled into mature ribosomal subunits [41]. Such toxicity hasn't been documented in DBA patients before but would be an interesting avenue for further investigation.

Loss of coordination between ribosomal proteins and mitochondria may be yet another explanation for variable expressivity. When *RPL11* is reduced, expression of mitochondrial genes is not simply going up or down. Instead, mitochondrial genes are expressed with much higher variance compared to ribosomal protein genes. Among the genes that are significantly variable across *RPL11* carriers, *PMI* and *TMEM11* control mitochondrial morphogenesis in *Drosophila melanogaster* [42]. Another variably expressed gene, *STOML2* was suggested to alter mitophagy and mitochondrial apoptosis pathways through PINK and MEK/ERK pathways respectively [43,44].

The increased variance in expression of mitochondrial components may lead to variable expressivity of DBA symptoms. For instance, the heme biosynthetic pathway which is important for red blood cell

function originates in mitochondria [45]. Interestingly, patients with Pearson syndrome who harbor large mitochondrial lesions can frequently be misdiagnosed as having Diamond Blackfan anemia since Pearson syndrome shares many symptoms with DBA including early onset of anemia with occasional spontaneous improvement and sporadic genetic occurrence [46,47]. Occasional observation of congenital heart disease in DBA patients could also be related to the increased variance among mitochondria genes given the connection between mitochondrial biology and vascular function. Thus, manifestation of DBA and variable expressivity could potentially be linked to the lack of coordination in mitochondrial components.

The variable expressivity of DBA symptoms could also be related with non-mitochondria related genes that have increased variance across *RPL11* variant carriers. Nonsense variants in FGF1 intracellular binding protein (FIBP), the gene with the lowest theta-E value, has been associated with overgrowth such as large thumbs, and variable congenital abnormalities in the heart and facial morphology [48,49]. Thus, FIBP gene is a potential candidate underlying variable congenital abnormalities observed in DBA patients. Another significantly variable gene across *RPL11* variant carriers, *LIMS2*, was identified in two siblings with severe cardiomyopathy [50].

Taken together, our study highlights the importance of unbiased systematic RNA expression analyses in families with non-canonical splice site variants that are causative for a Mendelian disorder.

4. Materials and methods

4.1. Sample collection and human subjects

Sample names, corresponding batches and index sequences for the RNA libraries are provided in Table S1. We generated RNA-Seq libraries from at least two separate blood samples from 8 non-carriers, 5 carriers and 2 patients afflicted by anemia. In total, 43 RNA-Seq samples were generated and extensive quality control metrics were inspected prior to further analyses. A few of the RNAseq libraries were prepared from the same samples at different times to ensure we don't observe bias due to batch effects. IRB approval as indicated in Carlston et al. 2017.

4.2. RNA extraction and library preparation

RNA was extracted using Maxwell RSC instrument (Promega) from 50 μ l of whole blood specimens with a white cell count greater than 50,000/ μ l. For depletion of ribosomal RNAs from the sample, 100 ng total RNA was mixed with 250 ng mixed pools of anti-sense rDNA oligos (Table S6). RNA-DNA mix was incubated at 95 °C for 2 min, then the temperature was ramped down –0.1 °C per second to 45 °C. 1 μ l Hybridase Thermostable RNase H (Lucigen, H39500) and 1 μ l 10 X digestion buffer (500 mM Tris-HCl, 1 M NaCl, 200 mM MgCl₂) was preheated to 45 °C and added to the RNA-DNA mix. After RNase H incubation for 1 h at 45 °C, the samples were moved to 37 °C, and incubated for 1 h in 50 μ l DNase reaction with 3 μ l Turbo DNase (ThermoFisher, AM2238) and 1× Turbo DNase buffer. 300 μ l solution containing 0.2%SDS, 300 mM NaAcetate and 2 μ l glycoblue was used to stop the reaction followed by acid-phenol extraction and ethanol precipitation. All of the precipitated RNA was converted into a sequencing library using SMARTer® smRNA-Seq Kit for Illumina (Thermo Fisher/Takara 635,032). The kit protocol was used with the following specifics: RNA was initially hydrolysed for 4 min at 94 °C. The recommended additional AmpureXP cDNA purification was applied. The libraries were sequenced using 150 cycles on a HiSeq 2500. RNA libraries were prepared in four technical batches. Samples from individuals III.1, III.2, and III.4 were included in all of the batches to ensure the differences that are observed are not due to any batch biases. Sample names, corresponding batches and index sequences are provided in Table S1.

4.3. Preprocessing and quality control of RNA sequencing data

Sequencing reads were first trimmed to remove adapter sequences and low quality bases using cutadapt [51]. Specifically, we used the following parameters: “-a GATCGGAAGAGCACACGTCTGAAGTCAGTCACNNNNNATCTCGTATGCCGTTCTGCTTG -A GATCGGAAGAGC ACACGTCTGAACTCCAGTCACNNNNNATCTCGTATGCCGTTCTGCTTG -A CAGATCGGAAGAGCGTCGTAGGGAAAGAGTGTAGATC TCGGTGGTC -u 6 -U 6 -q 22 -m 22”. Individual clipped fastq files were analyzed for quality control statistics using fastqc (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were aligned to the human genome using hisat2 [52] with default parameters. Alignment files were sorted and indexed using samtools [53]. Reads that map to protein-coding genes were counted using Rsubread's featureCounts function for pair-end sequencing [54]. Transcript annotations were retrieved from Ensembl's v31 basic gene annotation file (06.2019) [55] and were filtered to include only genes with the \protein coding tag. Ensembl ids were converted to HGNC symbols through biomaRt's *hsapiens_gene_ensembl* database and the getBM function with the following parameters: “attributes = c(“ensembl_gene_id”, “hgnc_symbol”), filter = “ensembl_gene_id”, values = ids, mart = ensemble_database”.

4.4. Quality control measures For RNA-seq data

The number of reads mapping to each protein-coding gene was compared for all pairwise combinations of samples. Samples were removed from further analyses if their mean Spearman correlation below 0.90 when compared to all others. (S2 Figure). This filter resulted in the elimination of five samples. Next, we compared the annotated gender for each sample with that inferred from the data. Specifically, we profiled the gender of each individual based on their expression of chrY-linked genes and matched to the annotated gender. Our results indicated one sample displayed an apparent gender discrepancy compared to the annotation, as well as the inferred gender of other samples belonging to the same individual. Hence, this sample was removed. For the remaining samples, the distribution of counts was plotted, and no systematic bias as a function phenotype was observed (S4A Figure). Counts were transformed with DESeq2's vst function and examined using principal component analysis (PCA) to cluster samples (S4B Figure) (Love, Huber, & Anders, 2014). PCA analyses indicated that batch variables did not correlate with the first 4 principle components suggesting limited confounding technical variability to due batch effects (S4C Figure).

4.5. Splicing patterns in *RPL11*

For splicing pattern analysis of *RPL11*, samtools-1.9 was used to extract all sequencing reads that align to the *RPL11* gene locus (chr1:23,689,806–23,698,835) with MAPQ greater than 10. Integrative Genomics Viewer v2.6.0 (IGV) was used for visualization. All BAM files were converted to junction BED files in IGV by using the “export features” function while the entire *RPL11* gene is shown in IGV, to better quantify junction reads. To select junctions spanning exon 4 and 5 from the junction BED files, junctions starting between 23,693,000–23,694,999 and ending between 23,695,000–23,695,999 were selected. This strategy captures all junction reads that arise due to skipping exon 4. Because the start and end positions in the junction BED files are adjusted by the max read length on the junction, max read length on the start positions were added to the start positions and max read length on the end positions were subtracted from the end positions. The corrected junction start and end positions, read count, and the strand of junction were aggregated into a table in R version 3.6.3 (Table S2). SplitterAlv1.2.1 was used to reconfirm the main effect of the variant in *RPL11* [56]. The effect of the most abundant splicing variant on translation was visualized in ExPASy (<https://www.expasy.org/>).

4.6. Antisense junction read analyses

Genes transcribed from two strands were split into two gtf files using gencode's v31 basic gene annotation (06.2019) file [55]. bedtools v2.27.1 was used to intersect sequencing reads with genes in the strand-separated gtf files, generating two files for each sample, one for genes transcribed from each strand. Samtools-1.9 was used to filter for reads with mapping quality (MAPQ) greater than 10. awk was used to select antisense junction reads from the alignment files. R version 3.6.3 was used to quantify each antisense junction read and filter for junctions that had more than 3 reads in at least 5 samples. The selected junction reads were visualized in Integrative Genomics Viewer v2.6.0.

4.7. Splice region variants in the human genome

Splice variant data from gnomAD v3 was used for analyzing splice region variants. Loss-of-function (LoF) metrics data in gnomAD v2.1.1 [57] was used to filter genes by probability of loss-of-function intolerance (pLI), with thresholds of pLI greater than 0.96 or less than 0.04. The value 0.96 was taken from the pLI of *RPL11*. Python 3.7.4 was used to download and aggregate the data for ribosomal protein genes and genes with matching pLI from the last step, from the gnomAD API. The table was read into R version 3.6.3 and analyzed.

4.8. Differential expression analyses

Differential expression was conducted with the R packages DESeq2 and limma-dream. Carriers, including DBA patients were compared to noncarriers, with the only covariate for the model being *RPL11* variant status. For DESeq2, technical replicates were collapsed by summing the counts of that individual's technical replicates. Differentially expressed genes were identified using the standard pipeline in [58]. Genes with an adjusted *p*-value of less than equal to 0.01 were deemed significant [59]. As a complementary approach, Dream was utilized in conjunction with limma to accommodate the repeated measures in experimental design through a linear mixed-effects model [60–62]. The counts of technical replicates were summed for each sample. Independent samples of an individual were then modeled with a linear mixed-effects model along with *RPL11* status [62]. The results for both analyses were combined by considering genes with a *p*-value less than 0.05 in limma-dream and an adjusted *p*-value of less than 0.01 in DESeq2 as potentially differentially expressed between groups. Rotation gene set testing for linear models was employed through the FRY function in limma to test group differences in ribosomal genes [63]. Comparisons with *p*-values under 0.05 were deemed significant.

4.9. Differential proportionality

erbFor differential proportionality, the definition established for emergent proportionality by Erb et al. 2018, was used [34,35]. Modifications to the methods regard robust sampling, which randomly samples an equal number of individuals to represent each group delimited by *RPL11* status for 100 permutations. For each permutation, theta E, a measure of emergent proportionality, was calculated between *RPL11* and all other genes, using *RPL11* status as the biological variable of interest. The average of these permutations was utilized in further analysis (S6A Figure). The false discovery rate was estimated by randomly assigning *RPL11* status to all individuals, calculating theta E with the modified methods described above, repeating for 100 permutations, and comparing expected to observed values (S6B Figure). For the ratio of each gene to *RPL11*, one group will have a greater amount of variance than the other. The direction of variance is defined as which *RPL11* status possesses the dominating amount of variation in the ratio of two genes. With this classification of direction of variance, a binomial test was used to determine whether the direction of variance was significantly different between the group of interests with the assumption this

direction is equally distributed. Theta E less than 0.15, representing the most significant results, was used to identify gene-pairs where a single group contributes the majority of the variance. Gene ontology analysis was performed through FuncAssociate 3.0 using an ordered list of genes (Theta E < 0.15), and terms were kept with an adjusted *p*-value cutoff of 0.05 [64].

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2021.04.020>.

Acknowledgements

This work was supported in part by the Texas Rising Star Award [to E.S.C.] and National Institutes of Health [GM138340 to E.S.C.; CA204522 to C.C.], Cancer Prevention and Research Institute of Texas [RR180042 to C.C.] and Welch Foundation [F-2027-20200401 to C.C.]. The authors thank Arlen Johnson, and Andrew Fire for critical discussions. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing and storage resources that have contributed to the research results reported within this paper. URL: <http://www.tacc.utexas.edu>. We also acknowledge the family for their participation in this research.

References

- [1] J.E. Farrar, N. Dahl, Untangling the phenotypic heterogeneity of Diamond Blackfan anemia, *Semin. Hematol.* 48 (2011) 124–135.
- [2] J.M. Lipton, S.R. Ellis, Diamond Blackfan anemia 2008–2009: broadening the scope of ribosome biogenesis disorders, *Curr. Opin. Pediatr.* 22 (2010) 12–19.
- [3] R.K. Khajuria, M. Munschauer, J.C. Ulirsch, C. Fiorini, L.S. Ludwig, S.K. McFarland, et al., Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis, *Cell* 173 (2018) 90–103, e19.
- [4] V.G. Sankaran, R. Ghazvinian, R. Do, P. Thiru, J.-A. Vergilio, A.H. Beggs, et al., Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia, *J. Clin. Invest.* 122 (2012) 2439–2443.
- [5] J.C. Ulirsch, J.M. Verboon, S. Kazerounian, M.H. Guo, D. Yuan, L.S. Ludwig, et al., The genetic landscape of diamond-Blackfan Anemia, *Am. J. Hum. Genet.* 103 (2018) 930–947.
- [6] N. Drapchinskaya, P. Gustavsson, B. Andersson, M. Petterson, T.N. Willig, I. Dianzani, et al., The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia, *Nat. Genet.* 21 (1999) 169–175.
- [7] H.T. Gazda, M.R. Sheen, A. Vlachos, V. Choesmel, M.-F. O'Donohue, H. Schneider, 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.* 83 (2008) 769–780.
- [8] L. Morgado-Palacin, G. Varetti, S. Llanos, G. Gómez-López, D. Martínez, M. Serrano, Partial loss of Rpl11 in adult mice recapitulates diamond-Blackfan Anemia and promotes Lymphomagenesis, *Cell Rep.* 13 (2015) 712–722.
- [9] A. Chakraborty, T. Uechi, S. Higa, H. Torihara, N. Kenmochi, Loss of ribosomal protein L11 affects zebrafish embryonic development through a p53-dependent apoptotic response, *PLoS One* 4 (2009), e4152.
- [10] A. Anna, G. Monika, Splicing mutations in human genetic disorders: examples, detection, and confirmation, *J. Appl. Genet.* 59 (2018) 253–268.
- [11] L.M. Polfus, R.K. Khajuria, U.M. Schick, N. Pankratz, R. Pazoki, J.A. Brody, et al., Whole-exome sequencing identifies loci associated with blood cell traits and reveals a role for alternative GFI1B splice variants in human hematopoiesis, *Am. J. Hum. Genet.* 99 (2016) 785.
- [12] C.M. Carlson, Z.A. Afify, J.C. Palumbos, H. Bagley, C. Barbagelata, W. L. Woorderchak-Donahue, et al., Variable expressivity and incomplete penetrance in a large family with non-classical diamond-Blackfan anemia associated with ribosomal protein L11 splicing variant, *Am. J. Med. Genet. A* 173 (2017) 2622–2627.
- [13] A. Narla, D. Yuan, S. Kazerounian, C. LaVasseur, J.C. Ulirsch, J. Narla, et al., A novel pathogenic mutation in RPL11 identified in a patient diagnosed with diamond Blackfan anemia as a young adult, *Blood Cells Mol. Dis.* 61 (2016) 46–47.
- [14] Z. Yi, M. Sanjeev, G. Singh, The branched nature of the nonsense-mediated mRNA decay pathway, *Trends Genet.* (2020), <https://doi.org/10.1016/j.tig.2020.08.010>.
- [15] K.T. Powers, J.-Y.A. Szeto, C. Schaffitzel, New insights into no-go, non-stop and nonsense-mediated mRNA decay complexes, *Curr. Opin. Struct. Biol.* 65 (2020) 110–118.
- [16] J.A. Arribere, E.S. Cenik, N. Jain, G.T. Hess, C.H. Lee, M.C. Bassik, et al., Translation readthrough mitigation, *Nature* (2016), <https://doi.org/10.1038/nature18308> [cited 2 Jun 2016].
- [17] Z. Cheng, J. Guo, L. Chen, N. Luo, W. Yang, X. Qu, Overexpression of TMEM158 contributes to ovarian carcinogenesis, *J. Exp. Clin. Cancer Res.* 34 (2015) 75.
- [18] M. Malumbres, Cyclin-dependent kinases, *Genome Biol.* 15 (2014) 122.

- [19] S. Akakura, I.H. Gelman, Pivotal role of AKAP12 in the regulation of cellular adhesion dynamics: control of cytoskeletal architecture, cell migration, and Mitogenic signaling, *J Signal Transduct.* 2012 (2012) 529179.
- [20] T. Hu, X. Wu, K. Li, Y. Li, P. He, Z. Wu, et al., AKAP12 endogenous transcripts suppress the proliferation, migration and invasion of colorectal Cancer cells by directly targeting oncomiR-183-5p, *Oncotargets Ther.* 12 (2019) 8301–8310.
- [21] A. Vlachos, P.S. Rosenberg, E. Atsidaftos, B.P. Alter, J.M. Lipton, Incidence of neoplasia in diamond Blackfan anemia: a report from the diamond Blackfan Anemia registry, *Blood.* 119 (2012) 3815–3819.
- [22] M. Risley, D. Garrod, M. Henkemeyer, W. McLean, EphB2 and EphB3 forward signalling are required for palate development, *Mech. Dev.* 126 (2009) 230–239.
- [23] R.K. Baker, A.K. Vanderboom, G.W. Bell, P.B. Antin, Expression of the receptor tyrosine kinase gene EphB3 during early stages of chick embryo development, *Mech. Dev.* 104 (2001) 129–132.
- [24] S. Amirbeigiarab, P. Kiani, A. Velazquez Sanchez, C. Krisp, A. Kazantsev, L. Fester, et al., Invariable stoichiometry of ribosomal proteins in mouse brain tissues with aging, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 22567–22572.
- [25] J.C. Taggart, G.-W. Li, Production of protein-complex components is stoichiometric and lacks general feedback Regulation in eukaryotes, *Cell Systems* (2018) 580–589, e4, <https://doi.org/10.1016/j.cels.2018.11.003>.
- [26] B.W. Tye, N. Commins, L.V. Ryazanova, M. Wühr, M. Springer, D. Pincus, et al., Proteotoxicity from aberrant ribosome biogenesis compromises cell fitness, *Elife.* 8 (2019), <https://doi.org/10.7554/elife.43002>.
- [27] H. Naora, I. Takai, M. Adachi, H. Naora, Altered cellular responses by varying expression of a ribosomal protein gene: sequential coordination of enhancement and suppression of ribosomal protein S3a gene expression induces apoptosis, *J. Cell Biol.* 141 (1998) 741–753.
- [28] L. Fancello, K.R. Kampen, I.J.F. Hofman, J. Verbeeck, K. De Keersmaecker, The ribosomal protein gene RPL5 is a haploinsufficient tumor suppressor in multiple cancer types, *Oncotarget.* 8 (2017) 14462–14478.
- [29] J.C. Guimaraes, M. Zavolan, Patterns of ribosomal protein expression specify normal and malignant human cells, *Genome Biol.* (2016), <https://doi.org/10.1186/s13059-016-1104-z>.
- [30] R. Ajore, D. Raiser, M. McConkey, M. Jöud, B. Boidol, B. Mar, et al., Deletion of ribosomal protein genes is a common vulnerability in human cancer, especially in concert with TP53 mutations, *EMBO Mol Med.* (2017), <https://doi.org/10.15252/emmm.201606660>.
- [31] K.M. Goudarzi, M.S. Lindström, Role of ribosomal protein mutations in tumor development (review), *Int. J. Oncol.* 48 (2016) 1313–1324.
- [32] R.Y. Ebright, S. Lee, B.S. Wittner, K.L. Niederhoffer, B.T. Nicholson, A. Bardia, et al., Dereulation of ribosomal protein expression and translation promotes breast cancer metastasis, *Science.* 367 (2020) 1468–1473.
- [33] H. Khatter, A.G. Myasnikov, S.K. Natchiar, B.P. Klaholz, Structure of the human 80S ribosome, *Nature.* 520 (2015) 640–645.
- [34] T.P. Quinn, I. Erb, M.F. Richardson, T.M. Crowley, Understanding sequencing data as compositions: an outlook and review, *Bioinformatics.* 34 (2018) 2870–2878.
- [35] Erb I, Quinn T, Lovell D, Notredame C. Differential proportionality - a normalization-free approach to differential gene expression. doi:<https://doi.org/10.1101/134536>.
- [36] M.A. Skinnider, J.W. Squair, L.J. Foster, Evaluating measures of association for single-cell transcriptomics, *Nat. Methods* (2019), <https://doi.org/10.1038/s41592-019-0372-4>.
- [37] B.B. Cummings, J.L. Marshall, T. Tukiainen, M. Lek, S. Donkervoort, A.R. Foley, et al., Improving genetic diagnosis in Mendelian disease with transcriptome sequencing, *Sci. Transl. Med.* 9 (2017), <https://doi.org/10.1126/scitranslmed.aal5209>.
- [38] F.R. Calvino, S. Kharde, A. Ori, A. Hendricks, K. Wild, D. Kressler, et al., Symportin 1 chaperones 5S RNP assembly during ribosome biogenesis by occupying an essential rRNA-binding site, *Nat. Commun.* (2015), <https://doi.org/10.1038/ncomms7510>.
- [39] K. Sorber, M.T. Dimon, J.L. DeRisi, RNA-Seq analysis of splicing in *Plasmodium falciparum* uncovers new splice junctions, alternative splicing and splicing of antisense transcripts, *Nucleic Acids Res.* (2011) 3820–3835, <https://doi.org/10.1093/nar/gkq1223>.
- [40] A. Aartsma-Rus, G.-J.B. van Ommen, Antisense-mediated exon skipping: A versatile tool with therapeutic and research applications, *RNA* (2007) 1609–1624, <https://doi.org/10.1261/rna.653607>.
- [41] E. Maicas, F.G. Pluthero, J.D. Friesen, The accumulation of three yeast ribosomal proteins under conditions of excess mRNA is determined primarily by fast protein decay, *Mol. Cell. Biol.* (1988) 169–175, <https://doi.org/10.1128/mcb.8.1.169>.
- [42] M. Macchi, N. El Fissi, R. Tufi, M. Bentobji, J.-C. Liévens, L.M. Martins, J. Royet, T. Rival, The *Drosophila* inner-membrane protein PMI controls crista biogenesis and mitochondrial diameter, *J. Cell Sci.* 126 (2013) 814–824.
- [43] Y. Zheng, C. Huang, L. Lu, K. Yu, J. Zhao, M. Chen, L. Liu, Q. Sun, Z. Lin, J. Zheng, J. Chen, J. Zhang, STOML2 potentiates metastasis of hepatocellular carcinoma by promoting PINK1-mediated mitophagy and regulates sensitivity to lenvatinib, *J. Hematol. Oncol.* 14 (2021) 16.
- [44] G. Hu, J. Zhang, F. Xu, H. Deng, W. Zhang, S. Kang, W. Liang, Stomatin-like protein 2 inhibits cisplatin-induced apoptosis through MEK/ERK signaling and the mitochondrial apoptosis pathway in cervical cancer cells, *Cancer Sci.* 109 (2018) 1357–1368.
- [45] S. Sano, S. Inoue, Y. Tanabe, C. Sumiya, S. Koike, Significance of Mitochondria for Porphyrin and Heme Biosynthesis, *Science* (1959) 275–276, <https://doi.org/10.1126/science.129.3344.275>.
- [46] K.E. Gagne, R. Ghazvinian, D. Yuan, R.L. Zon, K. Storm, M. Mazur-Popinska, et al., Pearson marrow pancreas syndrome in patients suspected to have Diamond-Blackfan anemia, *Blood* (2014) 437–440, <https://doi.org/10.1182/blood-2014-01-545830>.
- [47] M. Fontenay, S. Cathelin, M. Amiot, E. Gyan, E. Solary, Mitochondria in hematopoiesis and hematological diseases, *Oncogene* (2006) 4757–4767, <https://doi.org/10.1038/sj.onc.1209606>.
- [48] N. Akawi, S. Ben-Salem, L. Lahti, J. Partanen, B.R. Ali, L. Al-Gazali, A recessive syndrome of intellectual disability, moderate overgrowth, and renal dysplasia predisposing to Wilms tumor is caused by a mutation in FIBP gene, *Am. J. Med. Genet. A* 170 (2016) 2111–2118.
- [49] C. Thauvin-Robinet, L. Duplomb-Jego, F. Limoge, D. Picot, A. Masurel, B. Terriat, C. Champilou, D. Minot, J. St-Onge, P. Kuentz, Y. Duffourd, J. Thevenon, J.-B. Rivière, L. Faivre, Homozygous FIBP nonsense variant responsible of syndromic overgrowth, with overgrowth, macrocephaly, retinal coloboma and learning disabilities, *Clin. Genet.* 89 (2016) e1–e4.
- [50] J.W. Chardon, A.C. Smith, J. Woulfe, E. Pena, K. Rakhrus, C. Dennie, C. Beaulieu, L. Huang, J. Schwartzentruber, C. Hawkins, M.B. Harms, S. Dojeiji, M. Zhang, FORGE Canada Consortium, J. Majewski, D.E. Bulman, K.M. Boycott, D.A. Dymet, LIMS2 mutations are associated with a novel muscular dystrophy, severe cardiomyopathy and triangular tongues, *Clin. Genet.* 88 (2015) 558–564.
- [51] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBNet.journal.* 17 (2011) 10–12.
- [52] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, *Nat. Biotechnol.* (2019) 907–915, <https://doi.org/10.1038/s41587-019-0201-4>.
- [53] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., The sequence alignment/map format and SAMtools, *Bioinformatics.* 25 (2009) 2078–2079.
- [54] Y. Liao, G.K. Smyth, W. Shi, Feature counts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics* (2014) 923–930, <https://doi.org/10.1093/bioinformatics/btt656>.
- [55] A. Frankish, M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, et al., GENCODE reference annotation for the human and mouse genomes, *Nucleic Acids Res.* 47 (2019) D766–D773.
- [56] K. Jagannathan, S.K. Panagiotopoulou, J.F. McRae, S.F. Darbandi, D. Knowles, Y. I. Li, et al., Predicting Splicing from Primary Sequence with Deep Learning, *Cell* (2019) 535–548, e24, <https://doi.org/10.1016/j.cell.2018.12.015>.
- [57] K.J. Karczewski, L.C. Francioli, D.G. MacArthur, The mutational constraint spectrum quantified from variation in 141,456 humans, *Yearbook of Paediatric Endocrinol.* (2020), <https://doi.org/10.1530/ey.17.14.3>.
- [58] M.I. Love, S. Anders, V. Kim, W. Huber, RNA-Seq workflow: gene-level exploratory analysis and differential expression, *F1000Research* (2016) 1070, <https://doi.org/10.12688/f1000research.7035.2>.
- [59] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* 15 (2014) 550.
- [60] C.W. Law, Y. Chen, W. Shi, G.K. Smyth, Voom: precision weights unlock linear model analysis tools for RNA-seq read counts, *Genome Biol.* 15 (2014) R29.
- [61] G.E. Hoffman, E.E. Schadt, Variancepartition: interpreting drivers of variation in complex gene expression studies, *BMC Bioinformatics* 17 (2016) 483.
- [62] G.E. Hoffman, P. Roussos, dream: Powerful differential expression analysis for repeated measures designs, *Bioinformatics* (2020), <https://doi.org/10.1093/bioinformatics/btaa687>.
- [63] D. Wu, E. Lim, F. Vaillant, M.-L. Asselin-Labat, J.E. Visvader, G.K. Smyth, ROAST: rotation gene set tests for complex microarray experiments, *Bioinformatics* (2010) 2176–2182, <https://doi.org/10.1093/bioinformatics/btq401>.
- [64] G.F. Berriz, J.E. Beaver, C. Cenik, M. Tasan, F.P. Roth, Next generation software for functional trend analysis, *Bioinformatics.* 25 (2009) 3043–3044.