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

Brendan Panici[1], Hosei Nakajima[1], Colleen M. Carlston[2], Hakan Ozadam[1], Can Cenik[1,3], Elif Sarinay Cenik[1,3]

1 Department of Molecular Biosciences, University of Texas at Austin, Austin, USA

2 School of Medicine, University of San Francisco, San Francisco, CA

3 co-corresponding authors

ABSTRACT

Variants that perturb splicing are major contributors to disease. Recent evidence implicate non-canonical intronic variants as a poorly characterized yet highly prevalent class of alterations associated with Mendelian disorders. Here, we report the first detailed RNA expression and splicing analysis from a large family with an intronic *RPL11* variant associated with Diamond Blackfan Anemia (DBA). DBA manifests with incomplete penetrance and partial expressivity yet the mechanism of this variability remains enigmatic. Our analysis 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 and an antisense exon-exon junction variant present only in carriers. We observed that *RPL11* transcript abundance is comparable among carriers regardless of symptom severity. In contrast, both the small and large ribosomal subunit transcripts were significantly overexpressed in individuals with more prominent DBA symptoms. 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.

INTRODUCTION

Many genetic diseases are incompletely penetrant. The nature of the causative genetic mutation, epistastic interaction with genetic background, or 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 and variable penetrance. DBA can lead to severe 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 in ~19 ribosomal protein genes, a ribosome maturation factor, *TSR2,* and two factors important for red blood cell development, *GATA1,* and *MYSM1* (Lipton & Ellis, 2010; Ulirsch et al., 2018; Khajuria et al., 2018; Sankaran et al., 2012). Despite being a dominant genetic disease due to haploinsufficiency, DBA is characterized by unpredictable penetrance [reviewed in (Farrar and Dahl, 2011)]. In a DBA related cohort, maternal inheritance was observed in 74% of the cases, suggesting the presence of silent carriers (Moetter et al., 2011). Moreover a study of DBA in a large family indicated an unexpectedly high frequency of carriers only with a mild thumb eminence whereas two members are affected by more classic DBA symptoms(Carlston et al., 2017). The same splicing variant had been observed in an individual with bilateral triphalangeal thumbs, heart defects, short stature and anemia who came to medical attention as an adult for fatigue and pallor (Narla et al., 2016).

While initial characterizations of DBA patients focused on nonsense and missense coding sequence mutations, recent evidence indicates that extended splice site variants are under-appreciated and common in DBA patients (Ulirsch et al., 2018). Splicing variants may disrupt a splicing site, activate cryptic or alternative splicing, or introduce a novel splicing pattern with potential consequences for the open reading frame of the corresponding gene (Anna & Monika, 2018a). Increased application of genome and exome sequencing in patients has led to the identification of intronic variants that might alter splicing. A rare, synonymous variant in *GFI1B*–a transcription regulator expressed in hematopoietic lineage– is responsible for variation in blood traits across humans and suppresses the formation of a GFI1B splicing isoform that preferentially promotes platelet formation and megakaryocyte differentiation(Polfus et al., 2016). In a disease

context, despite the presence of a large number of splice site variants in DBA patients –~20% predicted splicing defects– (Ulirsch et al., 2018), to the best of our knowledge, there has never been an experimental characterization of the RNA expression and splicing changes in patient samples harboring these variants.

Here, we report the first detailed RNA expression and splicing analysis in a large family affected by DBA caused by the ribosomal protein gene *RPL11*. Eight family members spanning four generations were found to have the *RPL11* splicing variant. Of the eight only two family members had clinically significant anemia, suggesting an unexpectedly low –25%– penetrance (Carlston et al., 2017). The proband in this study additionally had short stature, congenital heart defects, and required surgical intervention for intestinal volvulus, severe hypospadias, and a major thumb malformation, and his cousin required surgical repair of a cleft palate. Less-affected family members had small thumbs or thenar eminences that did not require surgical intervention, and one carrier of the familial variant was totally asymptomatic (Carlston et al., 2017). One individual was diagnosed with breast cancer at age 36, and an increased incidence of neoplasia including breast cancer has been observed in the DBA registry (Vlachos et al., 2012). For easier annotation throughout the manuscript, we will refer less affected carriers members as having mild DBA symptoms and the two patients who are significantly affected by DBA as having prominent DBA symptoms.

The low penetrance of anemia and other classic manifestations of DBA in most members of this family led us to question whether this was related to the nature of the splicing variant or differences in cellular responses to *RPL11* haploinsufficiency. *RPL11* variants are more associated with cleft palate and thumbs anomalies compared to other DBA genes such as *RPS19* (Draptchinskaia et al., 1999; Gazda et al., 2008). Lack of RPL11 is associated with impaired red blood cell development, reduced P53, increased c-MYC in mouse (Morgado-Palacin et al., 2015), and in contradiction increased P53 and apoptosis in zebrafish (Chakraborty et al., 2009). *RPL11* is a rare variant of DBA (Gazda et al., 2008): *RPL11* mutations affect 0.5 % of DBA patients, while the most common cause variant is *RPS19,* affecting ~39% of the patients in a cohort with 370 patients (Ulirsch et al., 2018). Among the genome sequences available through the Genome Aggregation database, loss of function single nucleotide variation (SNV) in *RPL11* and

RPS19 have been observed zero times whereas the expected loss of function SNV is 10 and 8.1 respectively suggesting both of these genes are similarly haploinsufficient (Lek et al., 2016).

The RNA expression analysis that we have done of this extended family with incomplete and variable penetrance of DBA in the context of a *RPL11* splice site variant, enabled us to address the following questions: How does the pattern of *RPL11* splicing in 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 afflicted with prominent DBA symptoms compared to *RPL11* variant carriers with few to no symptoms (will be referred as mild DBA)?

We find that the most abundant *RPL11* variant transcript is potentially translated via a late stop codon read through and encodes a 76% N terminal identical, C terminal variant RPL11 protein. We report a common gene expression signature in individuals who carry the *RPL11* variant compared to noncarrier family members including overexpression of *GATA2* and *CDK11A* mRNAs. Despite observing a similar significant decrease in *RPL11* transcript levels in carriers with both mild and prominent DBA, we observe a significant overexpression of small and other large ribosomal subunit transcripts only in carriers with prominent DBA compared to other carriers, suggesting a difference in response to decreased *RPL11* mRNA levels. Finally, we used differential proportionality analysis to ask whether the varying expression of a particular gene is in a relation of proportionality to the *RPL11* transcript. We surprisingly observed that an existing proportionality between transcripts of *RPL11* and mitochondrial components is significantly lost in the presence of *RPL11* variant. We don't observe a particular trend of up- or down-regulation, but rather there is significantly more variation in gene expression of mitochondrial components in all carrier individuals regardless of their symptoms. Overall, our results suggest a difference in response to the decreased levels of *RPL11* transcript in *RPL11* splicing variant carriers with prominent or mild DBA symptoms. Finally, we suggest a coordination between expression of *RPL11* and mitochondrial component transcripts, which is lost in the *RPL11* variant carriers regardless of their symptoms.

RESULTS:

RNA sequencing analysis of a large family with Diamond Blackfan Anemia with incomplete and variable penetrance

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.* Several carriers in the family are affected by Diamond Blackfan Anemia and display incomplete penetrance (**Figure-1**) (Carlston et al., 2017). 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 same samples at different times to ensure we don't observe bias due to batch effects.

Diamond Blackfan Anemia (DBA) in this family is caused by splice variant (c.396 3A>3G) in the intron 4 of the *RPL11* gene. We confirmed the presence of the intronic variant for each carrier (**Supplemental Figure-1**). After mapping the RNA sequencing data to the human genome, we calculated the gene-specific expression values for all samples (see Methods). 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 (**Supplemental Figure-2,** 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 (**Supplemental Figure-3**). The overall mean correlation coefficient was 0.97 between remaining samples given the outlier samples (**Supplemental Figure-2**).

After read depth normalization of all the samples (median of ratios based on DESeq2), we didn't observe any bias for total normalized read counts per individual samples (**Supplemental Figure-4A**). 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 (**Supplemental Figure-4B**). Given the large number of samples, library preparation was

Figure-1

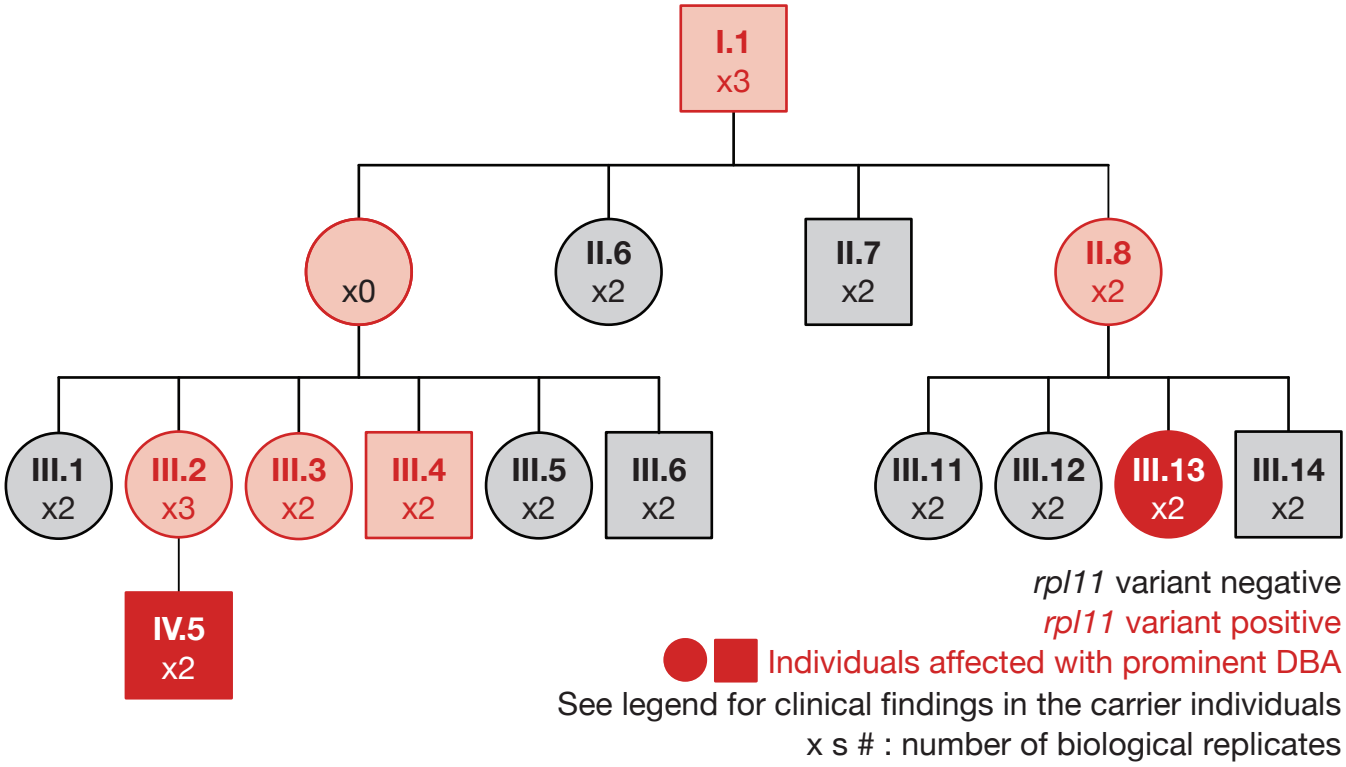


Figure-1: Four generation family pedigree depicts individuals who carry the pathogenic *RPL11* variant and their respective findings [adapted from (Carlston et al., 2017)]. Light and dark red circles and squares represent individuals who carry the *RPL11* variant. Clinical findings in the carrier individuals with prominent DBA symptoms (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. Less-affected family members carrying *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 totally asymptomatic.

carried out across several batches as described in **Supplemental Table-1**. Principal component analysis (PCA) suggested that batch variables do not explain a significant component of the variation observed in read counts (**Supplemental Figure-4C**). Taken together, 39 samples passed our strict quality control metrics and were used for all subsequent analyses.

The most abundant RPL11 alternate splicing isoform results in a frame shift and a 3'UTR read through

Mutations in splicing consensus sequences lead to altered recognition of introns and exons and this results in an aberrant splicing transcript (reviewed in (Anna & Monika, 2018b)). These mutations are found both in exons and introns which disrupt the dominant splice isoform, cause activation of cryptic splice sites generating new splice isoforms or lead to exon skipping. Conserved splicing acceptor and donor sites in the exon-intron boundaries are recognized by the splicing machinery at the 5' splice site (5'SS, exon:CAG/intron:GUAAGU) and 3' splice site (3'SS, exon:NYAG/intron:G) . Mutations in these sequences have been widely established to impact splicing events (reviewed in (Anna & Monika, 2018b)). Splicing alterations may consequently alter the encoded protein products by introducing frameshifts and premature stop codons. Such events typically lead to mRNA degradation through nonsense mediated decay (NMD) when the introduced stop codon precedes an exon-exon junction (reviewed in (Hug, Longman, & Cáceres, 2016)). NMD is a protective mechanism that limits the accumulation of aberrant proteins in cells, yet consequent loss of the functional protein product from the variant carrying gene copy can have deleterious effects (Morrill and Amon, 2019).

Diamond Blackfan Anemia (DBA) in this family is caused by an *RPL11* 5'splice site (5'SS) variant (c.396 intron4: 3A>3G) (**Figure-2A, Supplementary Figure-1**). Previous analyses of samples carrying this variant suggested the presence of a mis-splicing event leading to the inclusion of an alternative 5'SS at the +66 position relative to the canonical 5'SS (Carlston et al., 2017). This report relied on PCR amplification of RPL11 cDNA and Sanger sequencing of band isolated from a DNA gel.

Figure-2

A

B

```
  rpl11          1 MAQDQGEKENPMRELRIRKLCLNICVGESGDRLTRAAKVLEQLTGQTPVF  50
* rpl11 variant   1 MAQDQGEKENPMRELRIRKLCLNICVGESGDRLTRAAKVLEQLTGQTPVF  50

  rpl11         51 SKARYTVRSFGIRRNEKIAVHCTVRGAKAEEILEKGLKVREYELRKNNFS  100
  rpl11 variant 51 SKARYTVRSFGIRRNEKIAVHCTVRGAKAEEILEKGLKVREYELRKNNFS  100

  rpl11        101 DTGNFGFGIQEHIDLGIKYDPSIGIYGLDFYVV -- LGRPGFS———————  140
  rpl11 variant 101 DTGNFGFGIQEHIDLGIKYDPSIGIYGLDFYVVCWVGQVSASQTRSAGQA  150

  rpl11        141 ———I ADKKRRTGCIGA———————————— KHRISKEEAMRWFQQKYDGII 174
  rpl11 variant 151 ALGP NTESAKRRPCAGSSRSMMGSSFLANKFPFLSKRA I K SFQ———— 193

  rpl11        175 LPG                                                  177
```

Figure-2: *RPL11* variant c.396 3A>G is in the intronic region and affects the splicing pattern of *RPL11,* with the most abundant variant transcript resulting in frameshift with a late stop in 3' untranslated region. 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 exon4 and exon 5, height of the boxes represent read abundance excluding functional splicing abundance. Start and end positions are given in comparison to splicing that produces functional RPL11 (start, end:0). The numbers of reads mapping to each junction in each individual are in Supplemental Table-2. The most abundant splicing variant splices 2 nucleotides downstream of exon 4 and results in a frame-shift and 3'UTR ribosome read-through. **B** Alignment of the potential protein encoded by the most abundant variant to RPL11, aminoacids are colored based on similarity.

Here, we took advantage of a high throughput sequencing approach to decipher the extent of mis-splicing events that were attributable to the presence of this variant. To this end, we analyzed RNA sequencing reads that map to the RPL11 gene in an unbiased fashion. We identified all alternative splice isoforms impacting the junction between exons 4 and 5 in the canonical splice variant. We classified these events with respect to 5'splice site position (5'SS) relative to donor site; hence the 5'SS of the canonical RPL11 isoform is denoted as +0 (**Figure 2A**). Complete data with the number of splice junctions counts per sample that map to exon-4–exon-5 junction is provided in **Supplemental Table-2**.

Surprisingly, the most abundant isoform in carriers is an alternative 5' splice site at +2 position and not the +66 isoform reported previously (total 5.7% compared to 0.7%, **Figure 2A**). In the earlier study by Carlston et al. 2017, the 2 nucleotides difference between the canonical and variant isoform would have been impossible to differentiate with a DNA gel analysis. Interestingly, the mRNA containing the 5'SS at +2 does not include any early stop codons, and instead results in a frame-shift with a stop codon in the 3'UTR of the RPL11 canonical transcript. Thus, this variant mRNA is unlikely to be an NMD substrate due to an absence of early stop codon before any intron. The hypothetical RPL11 protein variant which is translated from the 5'SS at +2 transcript, shares the first 134 amino acids in the N-terminal with wild-type RPL11 (**Figure 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 protein instability ((Arribere et al., 2016), **Figure 2B**).

In addition to the most abundantly observed +2 5'SS variant (5.7% of total exon-4–exon-5 junction reads), we identified ~80 nucleotide antisense exon-4–exon-5 junction reads from the +2 5'SS (0.7 % of total exon-exon junction reads). These antisense reads constitute the second most abundant alternative group among exon-4–exon-5 junction reads. Strikingly, these reads are specifically observed in carriers but not in noncarrier individuals (**Figure 2A**). Importantly, every single sample carrying the RPL11 variant expressed the antisense exon-exon junction reads. The alternative antisense exon-exon junctions are observed in multiple uniquely mapped reads with distinct end coordinates for the paired end reads in ~20 different RNA-Seq libraries from *RPL11*

variant carrier individuals, and thus are not attributable to alignment or library preparation artifacts. Furthermore, extensive sequence similarity searches with the antisense reads did not reveal any sequences in the human genome with similarity. 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 in any of the samples. Taken together, these findings reveal an intriguing antisense RNA product that is specifically produced from the RPL11 splice variant containing allele.

The remaining minor splice isoforms observed in carrier individuals included the complete skipping of exon 4 (exon-4 skipped, 0.6%), and cryptic 5' splice sites at -50 (0.6%), and -26 (0.2%) relative to the canonical *RPL11* 5'SS (**Figure 2A, Supplemental Table-2)**. All of the type of alternative splice junctions and their relative abundances were similar between carrier individuals with mild or prominent DBA symptoms. These results suggest that the severity of DBA is not associated with the altered frequency or type aberrant splicing event.

In noncarrier individuals, alternative isoforms constitute only 0.05% of observed exon-exon junction reads while the rest were from the canonical isoform suggesting the noncanonical isoforms are spliced at low frequency even in the absence of the variant. In contrast, carrier individuals with the *RPL11* splice site variant had ~91% of their RPL11 exon-4–exon-5 junction reads originating from the major isoform of RPL11 (**Figure-2A**).

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

To analyze gene expression changes among carriers with mild and prominent DBA symptoms relative to noncarriers, we performed two complementary analyses (limma-dream and DESeq2; see Methods). Limma-dream (differential expression for repeated measures) analyses adopt a linear mixed effects model to reduce false positive differentially expressed genes in a repeated measures experimental design by extending the limma-voom workflow((Law, Chen, Shi, & Smyth, 2014), (Ritchie et al., 2015)). DESeq2 analysis, on the other hand, derives regularized fold change and dispersion

estimated and can accommodate any generalized linear model for experimental design (Love, Huber, & Anders, 2014). Using these two approaches, we first compared RNA expression between noncarriers and all carriers and identified nine genes with RNA expression differences (DESeq2 adjusted p-value < 0.01; limma-dream p-value < 0.05, **Figure-3A**). Normalized counts for all of these 9 genes across individuals display the over and under expression patterns in carrier and noncarrier individuals (**Figure-3B**).

Interestingly, when we compared RNA expression between noncarriers and carriers with mild DBA symptoms, we identified 23 genes with RNA expression differences (DESeq2 adjusted p-value < 0.01; limma-dream p-value < 0.05, **Figure-S5A**). Genes that are differentially expressed between carriers with mild symptoms and noncarriers were not differentially expressed in individuals with prominent DBA symptoms (**Figure-S5B**). Complete list of genes that had significant expression differences are provided in **Supplemental Table-3.** Normalized values (DESeq2 normalization) for all genes are provided in **Supplemental Table-4.**

RPL11 is 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). Among the overexpressed genes in carriers **(Supplemental Table-3)**, *TMEM158* is a Ras induced senescence protein which has been suggested to be involved in proliferation arrest and its overexpression is linked to ovarian cancer ((Cheng et al., 2015)). Cyclin dependent kinase 11a(*CDK11A*), is a serine/threonine-protein kinase that is involved in cell cycle control(Malumbres, 2014). *AKAP12* is a tumor suppressor, inhibitor of cell migration and regulates cell-cycle progression through Src-mediated oncogenic signaling and cytoskeletal pathway (Hu et al., 2019; Akakura & Gelman, 2012). 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(Vlachos et al., 2012).

Among the under expressed genes in carriers, *EPHB3* is an ephrin receptor that is expressed in early embryos and important for palate formation (Risley et al., 2009;

Figure-3

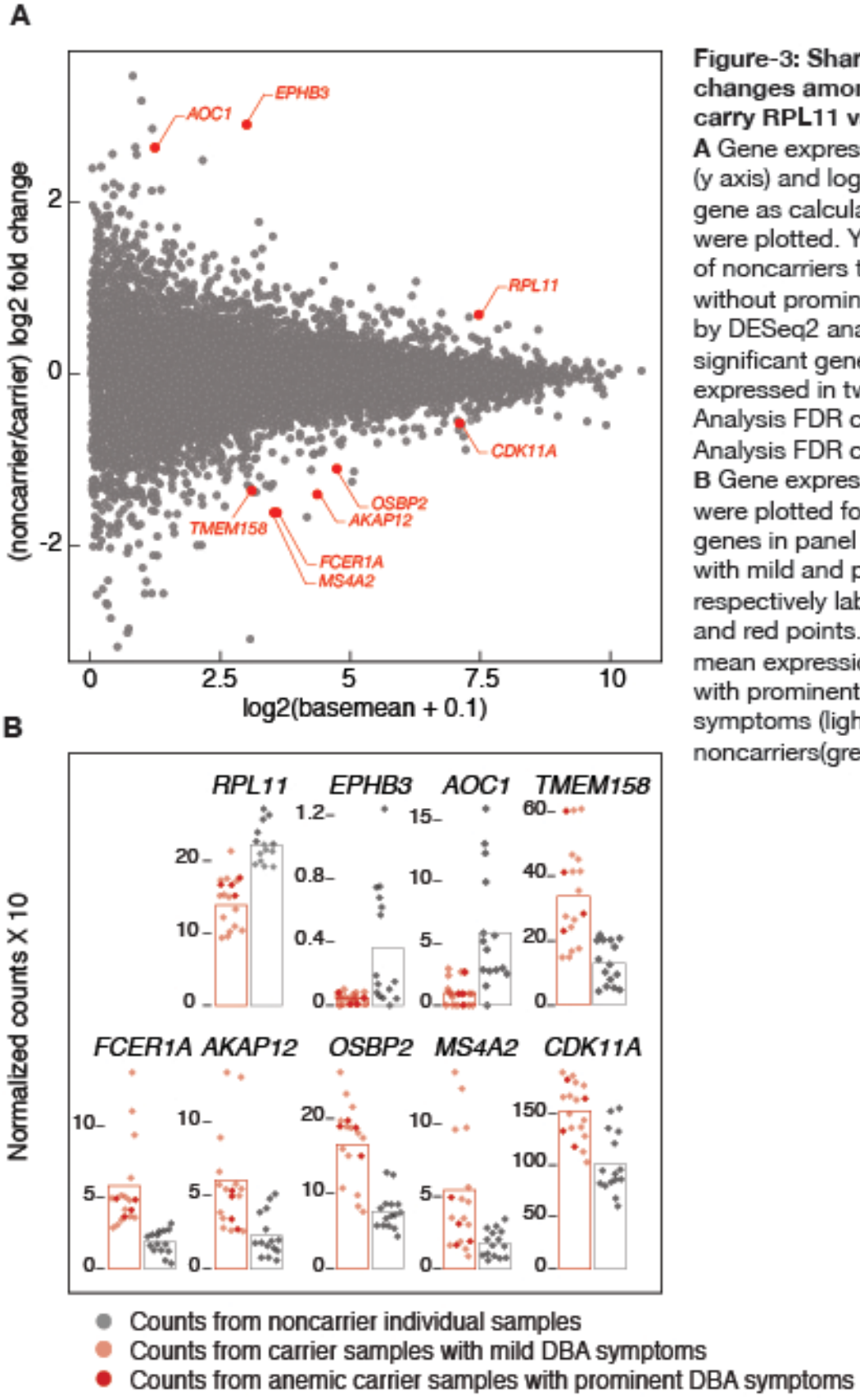


Figure-3: Shared gene expression changes among individuals that carry RPL11 variant c.396 3A>G.
A Gene expression log2 fold changes (y axis) and log2 base mean values per gene as calculated by Deseq2 analysis were plotted. Y axis is log2 fold change of noncarriers to carriers with and without prominent DBA as calculated by DESeq2 analysis. Red points are significant genes that are differentially expressed in two analyses. (R, Limma Analysis FDR cutoff<0.05, DESeq2 Analysis FDR cutoff<0.01).
B Gene expression values per individual were plotted for differentially expressed genes in panel A. Noncarriers, carriers with mild and prominent DBA were respectively labeled with grey, light red and red points. The bars represent mean expression among all carriers with prominent and mild DBA symptoms (light red) and among noncarriers(grey).

Baker, Vanderboom, Bell, & Antin, 2001). In this family among the two individuals with prominent DBA symptoms, cleft palate is observed in one, and congenital abnormalities are observed in the other individual. The observed differences in *EPHB3* transcript may have some connection to these symptoms.

These findings suggest that individuals with the *RPL11* variant express the canonical isoform at a reduced expression level with the c.396 3A>G variant producing about 14% of the major splice isoform generated from 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.

Reduced expression of small and large ribosomal protein gene expression in carriers with mild DBA symptoms that mirror the lower abundance of *RPL11* transcript

Ribosome assembly is a complex process that involves ribosomal RNAs, ~79 ribosomal proteins and ~300 trans-acting factors (Shajani et al., 2011, Woolford and Baserga, 2013). Thus, expression of ribosomal components is stochiometric (Taggart and Li, 2018; Amirbeigiarab et al., 2019), as loss of such coordination and over-production of ribosomal proteins would lead to unproductive energy consumption and may have proteotoxic consequences (Tye et al., 2019) and furthermore enhance cell proliferation, tumorigenesis and cancer metastasis [examples: RPS3A: (Naora et al., 1998), RPL5: (Fancello et al., 2017), RPL15: (Ebright et al., 2020), Guimaraes and Zavolan, 2016; (Ajore et al., 2017;reviewed in (Goudarzi and Lindström, 2016)].

We next determined the impact of reduced *RPL11* transcript expression on other ribosomal proteins given that their stoichiometry is tightly regulated (Taggart and Li, 2018; Amirbeigiarab et al., 2019). We reasoned that there might be a feedback regulation among other ribosomal proteins in response to the changes in *RPL11* transcript. While individual ribosomal proteins were not identified as differentially expressed, we reasoned that a feedback regulation would lead to a global regulation of the entire ribosomal subunits. Indeed, we observed that the expression of other ribosomal protein genes were reduced modestly but consistently for both the small and large ribosomal subunit proteins in carrier individuals with mild DBA symptoms. We

plotted the moderated log2-fold change (DESeq2) of all ribosomal proteins onto the human ribosome structure (PDB ID: 4UG0, (Khatter, et al., 2015)) to investigate any structural patterns to the feedback regulation (**Figure-4A**). We observed that reduced expression of the components to be widely distributed across the entire ribosome structure suggesting the coordinated regulation to be shared among almost all large and small ribosomal proteins with a few exceptions: Mean levels of *RPL27A, RPL28, RPL39* and *RPL26* transcripts in carriers with mild DBA are similar to noncarrier mean expression levels, whereas *RPS29*, *RPS17*, and *RPS26* transcript levels are slightly higher. When all ribosomal protein gene transcripts are considered together, the magnitude of the difference was small (rotating gene set analysis –ROAST– p value=0.1, example plots are in **Figure-4B**). This observation suggests an overall feedback regulation among other ribosomal protein genes in carrier individuals with mild DBA to compensate for the lower levels of *RPL11*.

Small and large ribosomal protein genes are significantly overexpressed in individuals with prominent DBA compared to carriers with mild or no DBA symptoms

Given that carriers with mild DBA symptoms (one individual with no symptoms) have lower ribosomal protein gene expression compared to noncarriers (**Figure-4**), we expected to observe a similar trend of lower expression of ribosomal protein gene expression in carriers regardless of symptom severity. While carriers with mild DBA have overall lower expression of large and small ribosomal protein genes compared to noncarriers, surprisingly, we didn't observe a similar trend for the carrier individuals with prominent DBA symptoms. In contrast, both small and large ribosomal proteins are significantly overexpressed in individuals with prominent DBA compared to carrier individuals with mild symptoms (**Figure-5A and B**, ROAST, p-value=0.04; see Methods). All of the large subunit ribosomal protein genes are overexpressed in carriers with prominent DBA except *RPL11* (**Figure-5A)**. All of the small subunit proteins are also overexpressed in carriers with prominent DBA in comparison to carriers with mild symptoms. In contrast, among the small subunit ribosomal protein genes in carriers with mild DBA, only three genes; *RPS29*, *RPS17,* and *RPS26* deviate from this pattern. The rest of the ribosomal protein genes are modestly lower or similar

Figure-4

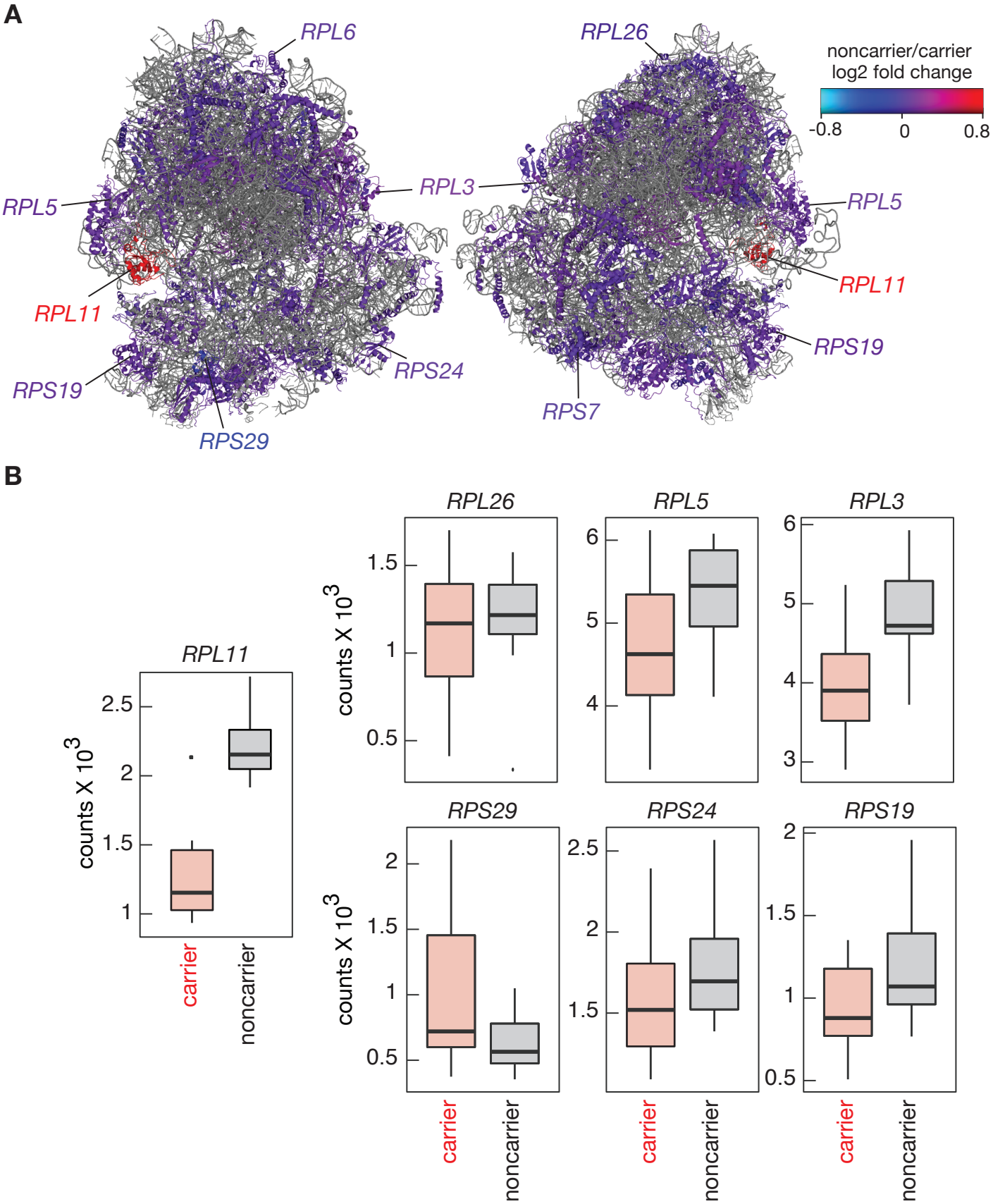


Figure-4: Effect of *RPL11* variant c.396 3A>G on the expression of both large and small ribosomal subunits. A Log2 shrinkage values generated by DESeq2 analysis for each ribosomal protein gene were plotted in color scale on the human ribosome structure (PDB ID: 4UG0). **B** Distribution of normalized *RPL11* expression in carriers –both with mild and prominent DBA symptoms– (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.

Figure-5

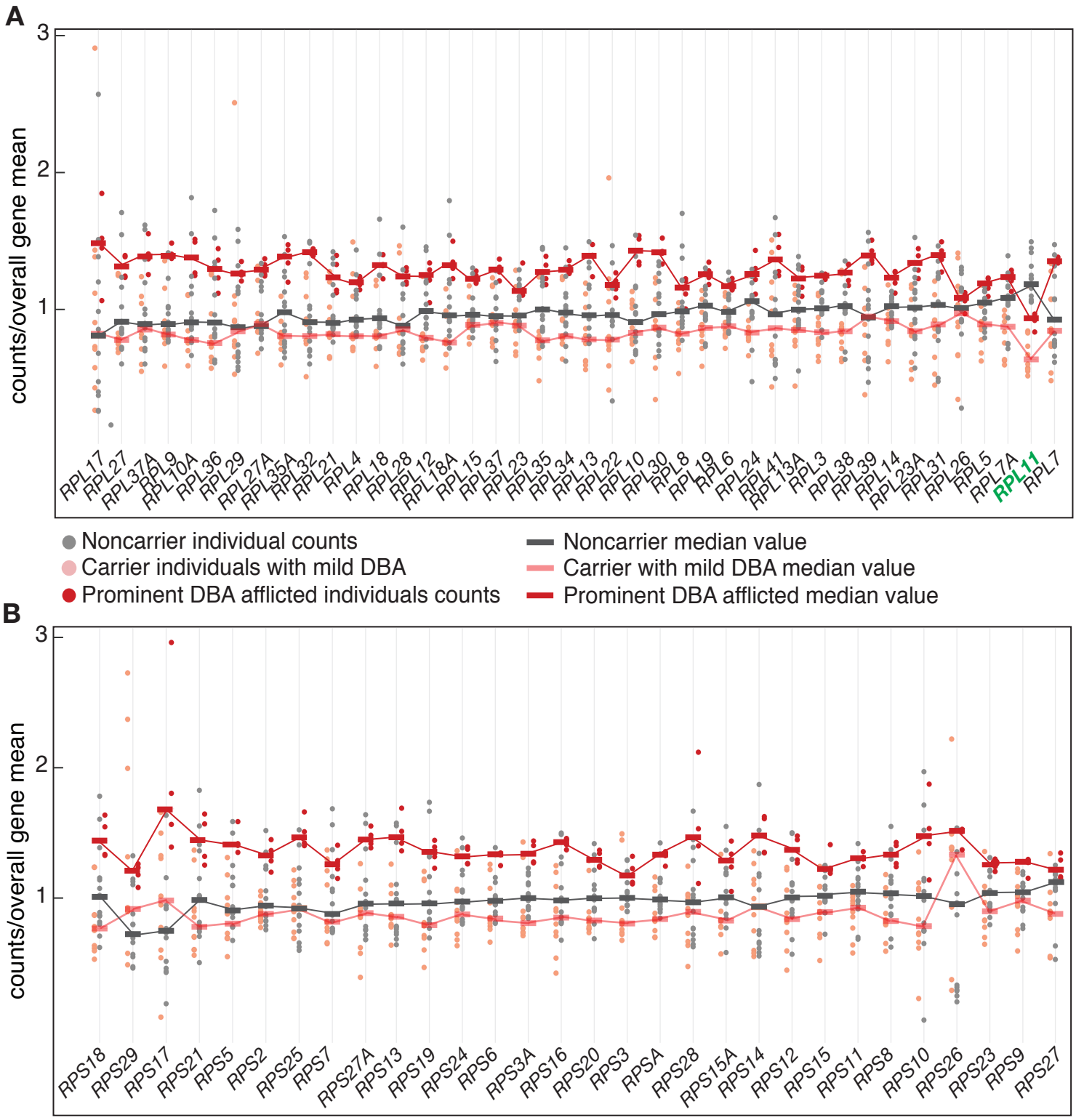


Figure-5: Large and subunit ribosomal protein mRNAs are differentially overexpressed in carrier individuals with prominent DBA in comparison to other carriers. Counts per each ribosomal protein gene in each individual were divided to the overall gene mean of the cohort and were plotted in noncarriers (grey), carriers with minimal DBA symptoms (light red), and carriers with more prominent DBA symptoms (dark red). Counts that map to each large (top plot) and small (bottom plot) ribosomal protein transcripts were divided to the overall mean of 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 (carriers with mild DBA–light red, carriers with prominent DBA–dark red, noncarriers: grey). Large and small subunit ribosomal protein mRNAs except *RPL11* (highlighted in green) are significantly overexpressed in carriers with prominent DBA compared to carriers with mild DBA (ROAST–rotating gene set analysis, p value= 0.04). The difference between carriers without DBA and noncarriers for ribosomal protein mRNA expression is not significant (ROAST, p value=0.1)

in mild DBA carriers compared with noncarriers. Taken together, 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 DBA compared to carriers with mild DBA.

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

Gene expression is coordinated at multiple levels including co-regulation through the action of transcription factors and chromatin remodelers, coupling of RNA processing with export and post-transcriptional processes [reviewed in (Komili & Silver, 2008; Moore, 2005)]. Moreover, organelles such as mitochondria have their own genomes and the expression of these components need to be coordinated with the majority of their components that are nuclear. Hence, signaling between the organelle genome and the nuclear genome ensures vital processes like electron transport chain function are robustly encoded [reviewed in (Woodson and Chory, 2008)].

Ribosomes are a central component of the cellular metabolic processes as gate-keepers of protein production. Consequently, we wondered whether the carriers lost coordination between *RPL11* and any other genes. To address this question, we took advantage of the recently formalized concept of emergent proportionality (Erb, Quinn, Lovell, & Notredame, 2017). Emergent proportionality utilizes the variance of log-ratios of read counts between a given pair of genes to derive the so called 'theta E value', which enables comparison between the coordinated expression of a pair genes across a biological variable interest. A given gene pair is expressed coordinately if the ratio of their read counts is maintained regardless of their expression level among individuals.

Proportionality analyses have numerous advantages over normalization dependent approaches that are predominantly utilized in RNA sequencing analyses. In particular gene expression measurements as mapped read counts are by definition relative abundance measures that are better analyzed using compositional data analysis methods. Importantly, distance metrics based on proportionality have recently been shown to significantly improve cluster identification in very high-dimensional data,

demonstrating the improved biological inferences when more appropriate statistical analyses are employed (Skinnider et al., 2019).

We utilized emergent proportionality as a method of identifying whether a set of genes lose coordination of expression due to haploinsufficiency of *RPL11.* Specifically, the variance of log-ratios of read counts was calculated for all genes and *RPL11* (**Supplemental Figure-6A**). For a given gene pair, if either group (carrier or noncarriers) contributes the majority of the total variance of log-ratios, this pair would have a low theta E value hence likely a candidate for loss of coordination between *RPL11* and the other gene. **(Supplemental Figure-6A and 6B).**

The overall direction of variance for all genes suggested that carrier individuals had higher variance of gene expression ratios with respect to *RPL11* ($p < 2.2e-16$ binomial test). Of the top results (theta E < 0.15), carrier individuals dominated the variance for all 357 pairs of genes ($p < 2.2e-16$; binomial test), suggesting an overall loss of coordination of gene expression in relation to *RPL11* in carriers **(Figure-6A)**. As a control to *RPL11*, we have conducted the same analysis with CDK11, a gene we found to be differentially increased in carriers (**Figure-3A)**. No significant results for directionality or gene ontology were obtained suggesting our results are not a by-product of differential expression or non-relating inherent properties of the carrier and noncarrier groups **(Figure-6A)**.

To illustrate the expression patterns, we plotted two exemplary genes; one with similar and one with different variance of log-ratios in comparison of carriers and noncarriers (**Figure-6B, C**). In particular *RPL4 to RPL11* (theta E value=0.91) ratio among carriers and noncarriers had similar variance suggesting that their expression coordination was similar in both groups (**Figure-6B, C**; left panels). On the other hand, *FIBP to RPL11* (theta E value=0.01) ratios were much more variable in carriers (**Figure-6B, C**; right panels). FIBP (FGF1 intracellular binding protein) interacts with aFGF and could be involved with its mitogenic action (Kolpakova et al., 1998). Moreover, FIBP depleted breast cancer cells display decreased proliferation and migration suggesting its role in cell growth and migration (Xu et al., 2014).

Figure-6

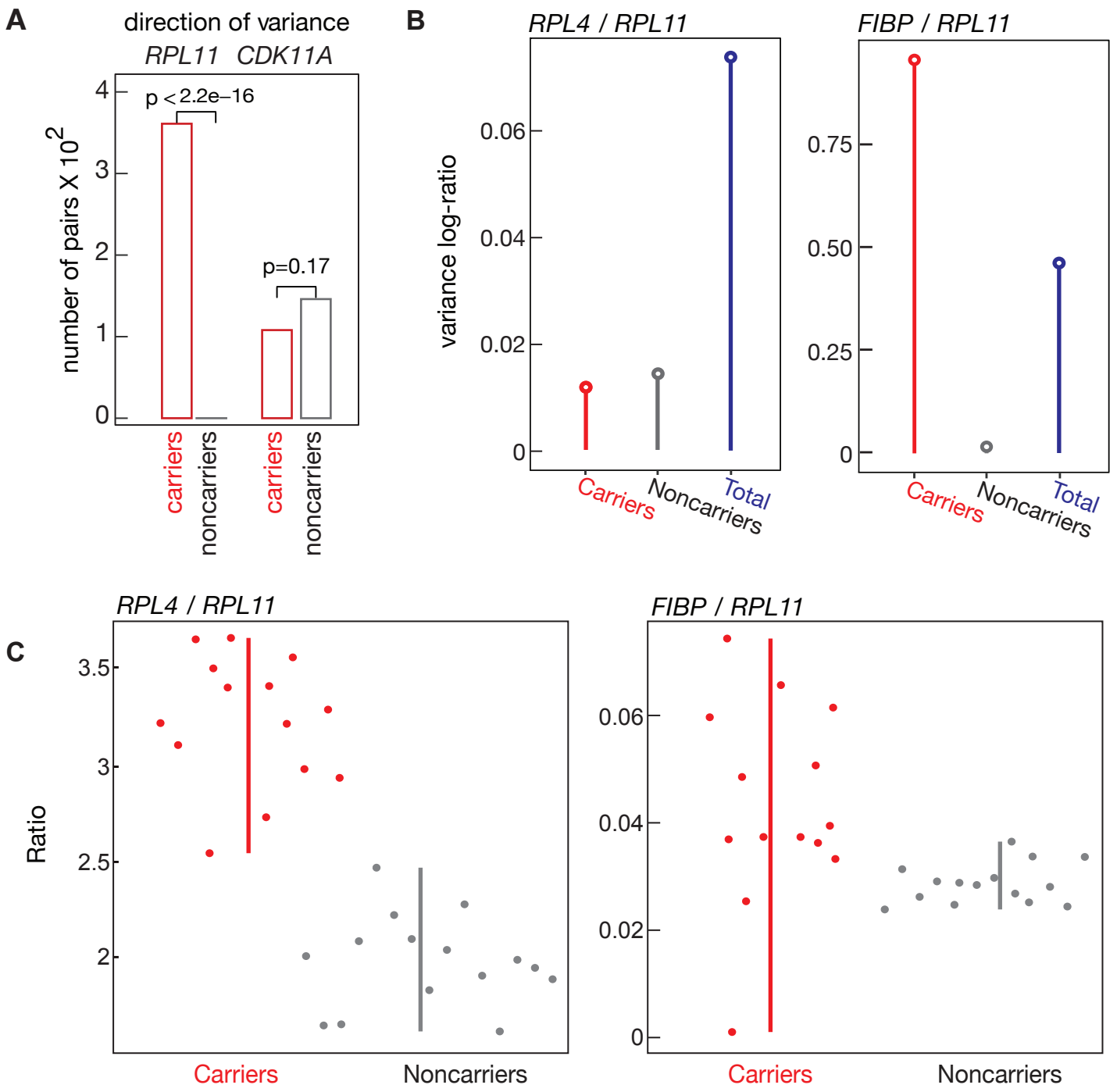


Figure-6: Emergent proportionality highlights loss of coordination between gene pairs. **A** The log-ratio of read counts of *RPL11* or *CDK11A* compared to all genes were calculated. Variance of the log-ratios were used to calculate emergent proportionality (theta E) as previously described comparing carriers to non-carriers (Erb et al., 2017). For each gene pair with theta E <0.025 –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 noncarrier and carrier samples. Left panel: No genotype group dominates the total VLR, therefore *RPL4* and *RPL11* are not in emergent proportionality and have a high theta E value (0.91). Right panel: VLR is higher in carriers samples for *FIBP* / *RPL11* and thus these two genes are in emergent proportionality (theta E = 0.01). **C** The ratio of two gene pairs for a sample is plotted and colored by *RPL11* status. Left panel: The ratio of *RPL4* and *RPL11* has a similar variance by genotype, although the mean of their ratios differs. Right panel: For the ratio of *FIBP* and *RPL11*, carrier samples are spread out while noncarriers are distinctly more clustered.

Gene ontology (GO) analysis was performed for the genes with a theta E less than 0.15 (corresponding to 0.3 FDR) with respect to *RPL11* using FuncAssociate 3.0 (Berriz, et.al., 2009). All significant GO terms with an adjusted p-value below 0.05 were related to mitochondria components and function including: mitochondrial large ribosomal subunit, inner membrane, and membrane components, such as a mitochondrial topoisomerase gene, *TOP1M* (**Supplementary Table-5**). Overall, of the top 57 genes with a theta E less than 0.15, 16 are related to mitochondria **(Figure-7A)**. Individual pairwise read count ratios of a few of example mitochondrial genes (*COX14*, *MRPL22*, *STOML2, TMEM11*) display the increased variance in carriers with mild and prominent DBA **(Figure-7B)**. Pairwise read count ratios of all these mitochondrial components in carrier and noncarrier samples are provided (**Supplementary Figure-7)**. One outlier sample with low counts to mitochondrial genes is likely not the main driver for the detected loss of coordination, as we have performed rigorous sub-sampling to calculate false discovery rates (**Supplementary Figure-6B**). These results overall reveal a loss of coordination between the cytoplasmic large ribosomal subunit or *RPL11* and mitochondrial large ribosomal subunit, and other mitochondrial components such as mitochondrial membrane proteins in *RPL11* variant carriers.

DISCUSSION:

Variants in splicing regulatory sequences have recently been implicated as major causative factors in a wide range of diseases including DBA (Ulirsch et al., 2018). Splicing regulatory elements can be found in both exons and introns. Variants causing improper splicing can alter the encoded protein sequences and thus be detrimental for proper protein function given that precise pre-mRNA splicing is a prerequisite for correct protein translation. 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* exon4-exon5 junction, which is causative for Diamond Blackfan Anemia (DBA).

Using this strategy, we first determined the frequencies of aberrant splicing patterns produced due to this intronic variant. A previously missed *RPL11* splice variant (5′ splice site at position +2 relative to major *RPL11* isoform) was found to have the highest abundance. Interestingly, this variant is likely a poor NMD substrate given that the

Figure-7

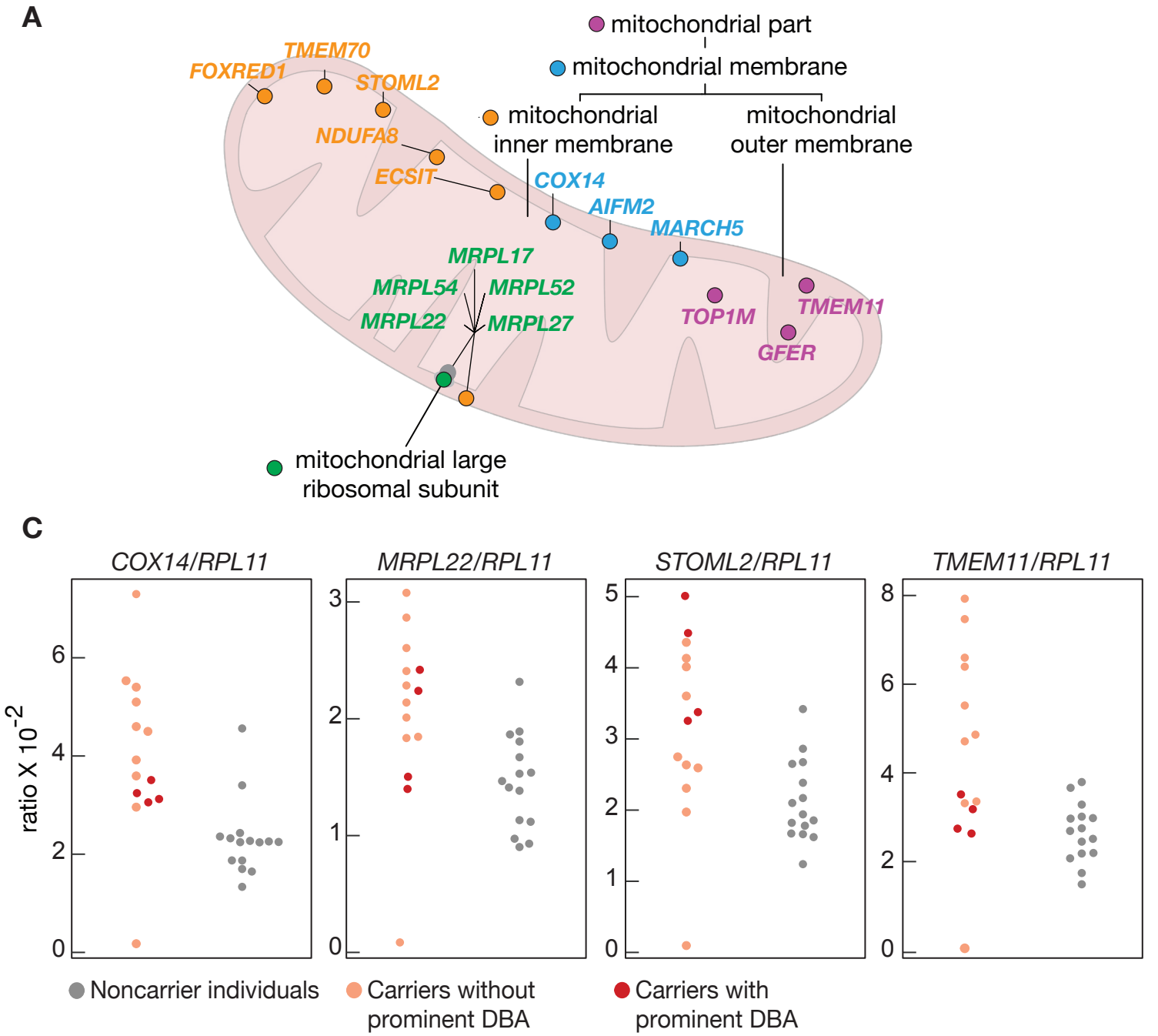


Figure-7: Expression of mitochondrial genes becomes uncoordinated with RPL11 expression in carriers compared to noncarriers. A Genes with the highest theta E value were highly enriched for mitochondrial localization as shown in the diagram [GO significant enrichment analysis, p-value< 0.05, FuncAssociate 3.0, Supplementary Table-5, (Berriz et al. 2009)]. The only GO category that is not shown in this diagram belongs to large ribosomal subunit components **B** For selected mitochondrial genes, the ratio of read counts were plotted for the two groups. Each dot represents a sample from carriers without and with prominent DBA symptoms as well as noncarriers –labeled with dark red, light red and grey respectively.

inclusion of a late stop in the 3'UTR of the canonical isoform. Since this isoform encodes a protein that is 75% identical to RPL11 at its N terminus, it might have proteotoxic consequences. SYO1, the dedicated chaperone for assembling RPL5 and RPL11 into 5S RNP complex, binds to the N terminal section of RPL11 (Calviño et al., 2015). Thus, the aberrant RPL11 might sequester SYO1 and further abrogate ribosome biogenesis.

The expression of *RPL11* in carrier individuals is ~57% of noncarrier individuals (more details in the next results section). If we assume 50% of *RPL11* mRNA is transcribed from the variant *RPL11* locus, then ~43% of aberrantly spliced variant *RPL11* transcripts must be degraded by NMD. This leaves only ~7%–or lower if there is dosage compensation– of aberrantly spliced *RPL11* among the pool of *RPL11* transcripts in carriers. The most abundantly detected +2 5'SS variant is probably not degraded by NMD and the absence of NMD might explain the higher abundance of this particular variant among all the detected *RPL11* transcript variants.

While examining other splice isoforms encoded by the *RPL11* variant, we were surprised that the second most abundant exon4-exon5 junction splice variant was specifically present in all carriers 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 exon-exon junctions have not been studied extensively in humans, but were previously reported in *Plasmodium falciparum* (Sorber, Dimon, & DeRisi, 2011) 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 elsewhere to the *RPL11* coding region suggesting that antisense transcription of the *RPL11* gene would have to be highly unstable to explain this observation. We think that antisense transcription only spanning the exon4-exon5 region of the *RPL11* gene remains a more plausible alternative.

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 carrier individuals with and without prominent DBA symptoms suggesting that any cis-acting effects are unlikely to be explanatory.

To understand causes and consequences of a low and variable penetrance in this family with DBA, we further investigated whether there is any evidence of differential gene expression patterns in carriers with mild or prominent DBA symptoms. Interestingly, carrier individuals with mild DBA modestly downregulate other large and small ribosomal proteins suggesting that there might be a compensation mechanism for the lower expression of *RPL11*. On the other hand, carriers with prominent DBA significantly overexpress the majority of small and large ribosomal protein transcripts. This could either be a response to symptom severity in more severely afflicted individuals or directly 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 large amount of unassembled ribosomal proteins. Gene expression range of haploinsufficient genes are small and trigger toxicity with increase copy number of expression (Morrill and Amon, 2019). Such toxicity hasn't been documented in DBA patients before but would be an interesting avenue for further investigation.

The difference in expression of ribosomal protein genes in carriers with mild and prominent DBA could be triggered by alterations in inflammatory mechanisms, p38 MAPK, TNF-α and NLK (Nemo-like kinase) which are observed to be elevated or activated in DBA (Bibikova et al., 2014; Pesciotta et al., 2015; Wilkes et al., 2020). Moreover, severity of DBA symptoms are also associated with decreased levels of HSP70 (Gastou et al., 2017). We haven't observed gene expression differences for these groups of genes but our detection may be (1) limited by the number of individuals with prominent DBA symptoms in the family or (2) these changes may be triggered post-transcriptionally.

Loss of coordination between ribosomal proteins and mitochondria may be yet another explanation for the partial penetrance. When the coordination breaks down upon reduced abundance of *RPL11*, expression of mitochondrial genes is not simply going down or up. Instead, they behave with much higher variance compared to ribosomal

protein genes. 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 (Sano et al., 1959). Furthermore hematological abnormalities are also observed in mitochondrial disorders (Fontenay, Cathelin, Amiot, Gyan, & Solary, 2006); and thus manifestation of DBA and thus variable expressivity of symptoms could be linked to the lack of coordination in mitochondrial components.

The implications of losing coordination of mitochondrial components to *RPL11* are unknown but mitochondrial malfunction due to loss of coordination among mitochondrial components upon ribosomal haploinsufficiency might in turn result in further mitochondrial genome deletions. A surprisingly unexpected frequency of large mitochondrial lesions had been reported in Diamond Blackfan anemia (DBA) patients; 4.6% of a DBA patient cohort tested contained large mtDNA deletions (Gagne et al., 2014). Knockdown of *RPS19* in zebrafish , the most common DBA allele, induces S6 kinase phosphorylation and autophagy through reactive oxygen species which also suggests the involvement of altered mitochondrial function (Heijnen et al., 2014). Impairment of mitochondrial function also affects nuclear genome stability due to mitochondrial ROS, dNTP production, metal ion detoxification or disturbances in apoptotic signaling [reviewed in (Kaniak-Golik and Skoneczna, 2015)]. Thus, increased cancer incidence among DBA patients could also be potentially related to mitochondrial dysfunction in addition to the alterations that are observed in P53 levels and AKT pathway (Antunes et al., 2015). Taken together, our results reveal the crucial role of detailed RNA expression characterization in a Mendelian disease and lead to a novel hypothesis about loss of gene expression coordination that may underlie partial peneterance.

Acknowledgements

This work was supported in part by the Texas Rising Star Award [to E.S.C.] and National Institutes of Health [1R35GM138340-01 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 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.

MATERIALS AND METHODS:

Sample Collection and Human Subjects

Sample names, corresponding batches and index sequences for the RNA libraries are provided in **Supplementary Table-1**. IRB approval as indicated in Carlston et al. 2017.

RNA extraction and library preparation

RNA was extracted using Maxwell RSC instrument (Promega) from 50 ul of whole blood specimens with a white cell count greater than 50 000/μl. For depletion of ribosomal RNAs from the sample, 100 ng total RNA was mixed with 250 ng mixed pools of anti-sense rDNA oligos (**Supplemental Table-6**). RNA-DNA mix was incubated at 95°C for 2 min, then the temperature was ramped down -0.1^{o}C per second to 45°C. 1 ul Hybridase Thermostable RNase H (Lucigen, H39500) and 1 ul 10 X digestion buffer (500 mM Tris-HCl, 1M NaCl, 200 mM $MgCl_2$) was preheated to 45°C and added to the RNA-DNA mix. After RNAse H incubation for 1 hour at 45°C, the samples were moved to 37°C, and incubated for 1 hour in 50 μl DNAse reaction with 3 μl Turbo DNAse (ThermoFisher, AM2238) and 1X Turbo DNAse buffer. 300 μl solution containing 0.2 %SDS, 300 mM NaAcetate and 2 μl 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 635032). The kit protocol was used with the following specifics: RNA was initially hydrolysed for 4 minutes 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 **Supplementary Table-1**.

Preprocessing and Quality Control of RNA Sequencing Data

Sequencing reads were first trimmed to remove adapter sequences and low quality bases using cutadapt(Martin, 2011). Specifically, we used the following parameters: "-a

GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG -A GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG -A CAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTC -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 (Kim, Paggi, Park, Bennett, & Salzberg, 2019) with default parameters. Alignment files were sorted and indexed using samtools (Li et al., 2009). Reads that map to protein-coding genes were counted using Rsubread's featureCounts function for pair-end sequencing (Liao, Smyth, & Shi, 2014). Transcript annotations were retrieved from Gencode's v31 basic gene annotation file (06.2019) (Frankish et al., 2019) 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)".

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. **(Supplemental Figure-2)**. 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 **(Supplemental Figure-4A)**. Counts were transformed with DESeq2's vst function and examined using principal component analysis (PCA) to cluster samples **(Supplemental Figure-4B)** (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 **(Supplemental Figure-4C).**

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 23693000-23694999 and ending between 23695000-23695999 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 **(Supplemental Table-2**). SpliceAIv1.2.1 was used to reconfirm the main effect of the variant in *RPL11* (Jaganathan et al., 2019). The effect of the most abundant splicing variant on translation was visualized in ExPASy (https:/www.expasy.org/).

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 (Frankish et al., 2019). 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.

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 (Karczewski et al., 2020) 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.

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 Love, Anders, Kim, & Huber, 2016. Genes with an adjusted p-value of less than equal to 0.01 were deemed significant (Love et al., 2014). 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 (Hoffman & Schadt, 2016; Hoffman & Roussos, 2020; Law et al., 2014). 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 (Hoffman & Roussos, 2020). 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 (Wu et al., 2010). Comparisons with p-values under 0.05 were deemed significant.

Differential Proportionality

For differential proportionality, the definition established for emergent proportionality by Erb *et al.* 2017, was used (Erb, Quinn, Lovell, & Notredame, 2017; Lovell, Pawlowsky-Glahn, Egozcue, Marguerat, & Bähler, 2015). 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 **(Supplemental-Figure-6A)**. 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 **(Supplemental-Figure-6B)**. 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 (Berriz *et al.*, 2009).

Figure Legends

Figure-1: Four generation family pedigree depicts individuals who carry the pathogenic *RPL11* variant and their respective findings [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 prominent DBA symptoms (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. Less-affected family members carrying *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 totally asymptomatic.

Figure-2: *RPL11* variant c.396 3A>G is in the intronic region and affects the splicing pattern of *RPL11*, with the most abundant variant transcript resulting in frameshift with a late stop in 3' untranslated region. 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 exon4 and exon 5, height of the boxes represent read abundance excluding functional splicing abundance. Start and end positions are given in comparison to splicing that produces functional RPL11 (start, end:0). The numbers of reads mapping to each junction in each individual are in Supplemental Table-2. The most abundant splicing variant splices 2 nucleotides downstream of exon 4 and results in a frame-shift and 3'UTR ribosome read-through. **B** Alignment of the potential protein encoded by the most abundant variant to RPL11, aminoacids are colored based on similarity.

Figure-3: Shared gene expression changes among individuals that carry *RPL11* variant c.396 3A>G.A Gene expression log2 fold changes (y axis) and log2 base mean values per gene as calculated by Deseq2 analysis were plotted. Y axis is log2 fold change of noncarriers to carriers with and without prominent DBA as calculated by DESeq2 analysis. Red points are significant genes that are differentially expressed in two analyses. (R, Limma Analysis FDR cutoff<0.05, DESeq2 Analysis FDR cutoff<0.01). **B** Gene expression values per individual were plotted for differentially expressed genes in panel A. Noncarriers, carriers with mild and prominent DBA were respectively labeled with grey, light red and red points. The bars represent mean expression among all carriers –with prominent and mild DBA symptoms (light red) and among noncarriers(grey).

Figure-4: Effect of *RPL11* variant c.396 3A>G on the expression of both large and small ribosomal subunits. A Log2 shrinkage values generated by DESeq2 analysis 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 –both with mild and prominent DBA symptoms– (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.

Figure-5: Large and subunit ribosomal protein mRNAs are differentially overexpressed in carrier individuals with prominent DBA in comparison to other

carriers. Counts per each ribosomal protein gene in each individual were divided to the overall gene mean of the cohort and were plotted in noncarriers (grey), carriers with minimal DBA symptoms (light red), and carriers with more prominent DBA symptoms (dark red). Counts that map to each large (top plot) and small (bottom plot) ribosomal protein transcripts were divided to the overall mean of 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 (carriers with mild DBA–light red, carriers with prominent DBA–dark red, noncarriers: grey). Large and small subunit ribosomal protein mRNAs except *RPL11* (highlighted in green) are significantly overexpressed in carriers with prominent DBA compared to carriers with mild DBA (ROAST–rotating gene set analysis, p value= 0.04). The difference between carriers without DBA and noncarriers for ribosomal protein mRNA expression is not significant (ROAST, p value=0.1)

Figure-6: Emergent proportionality highlights loss of coordination between gene pairs. **A** The log-ratio of read counts of *RPL11* or *CDK11A* compared to all genes were calculated. Variance of the log-ratios were used to calculate emergent proportionality (theta E) as previously described comparing carriers to non-carriers (Erb et al., 2017). For each gene pair with theta E <0.025 –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 noncarrier and carrier samples. Left panel: No genotype group dominates the total VLR, therefore *RPL4* and *RPL11* are not in emergent proportionality and have a high theta E value (0.91). Right panel: VLR is higher in carrier samples for *FIBP* / *RPL11* and thus these two genes are in emergent proportionality (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 a similar variance by genotype, although the mean of their ratios differs. Right panel: For the ratio of *FIBP* and *RPL11*, carrier samples are spread out while noncarriers are distinctly more clustered

Figure-7: Expression of mitochondrial genes becomes uncoordinated with *RPL11* expression in carriers compared to noncarriers. A Genes with the highest theta E value were highly enriched for mitochondrial localization as shown in the diagram [GO category significant enrichment analysis, p-value< 0.05, FuncAssociate 3.0, Supplementary Table-5, (Berriz et al. 2009)]. The only GO category that is not shown in this diagram belongs to large ribosomal subunit components. **B** For selected mitochondrial genes, the ratio of read counts were plotted for the two groups. Each dot represents a sample from carriers without and with prominent DBA symptoms as well as noncarriers –labeled with dark red, light red and grey respectively.

Supplemental Figure-1: An exemplary of IGV snapshots of *RPL11* gene at exon4-intron4 region from example carrier and non-carrier individuals. Few reads that map to *RPL11* variant c.396 3A>G was used as a confirmation of the carrier status.

Supplemental Figure-2: Plot of spearman correlation across RNA samples and individuals. 5 samples with lower spearman correlation values (<0.9) were discarded from further RNA expression analysis. These 5 samples were further investigated in Figure S3. The mean correlation value across samples after the sample removal is 0.97.

Supplemental Figure-3: Samples with <0.9 Spearmen correlation from Figure S1 generally have lower percentage of genome mapping reads. Samples II.8_s2, III.3_s1_t3, III.4_s1_t4, which were removed from further analysis, generally have low counts and percentages of mapped reads in comparison to the other example samples; II.6_s2 and III.1_s1_t3.

Supplemental Figure-4: A Normalized assigned counts were plotted for each sample. Grey, light red and dark red points represent noncarriers, carriers with few manifestations of DBA and carriers with prominent DBA findings respectively. **B** PCA analysis plot is shown for each sample annotated with individual identifier from Figure 1, followed by a number that stands for sample number for the particular individual. RNA expression libraries were prepared in four batches and three samples were included in all of the batches. **C** Libraries that were prepared in each batch was colored accordingly (batch 1–purple, 2–blue, 3–green, 4–orange) in the two PCA plots; first two variance components are plotted left, third and fourth components are plotted on the

right. Note that PCA plots in B and C are the same, where each sample were labeled with a sample name, or a batch number respectively.

Supplemental Figure-5: Shared gene expression changes among carrier individuals with few manifestations of DBA compared to noncarriers. A Gene expression log fold changes (y axis) and log base mean values per gene were plotted on the left large plot. Red labeled points are significant genes that are differentially expressed (R, Limma Analysis FDR cutoff<0.05, Deseq2 Analysis FDR cutoff<0.01). **B** Mean gene expression of carriers mild and prominent DBA (y axis on the left and right plots respectively) are compared to noncarriers (x axis). Red points on the plot represent significantly over and under expressed genes in carriers with mild DBA symptoms in comparison to noncarriers.

Supplemental Figure-6: Implementation of individual sampling allows for more robust emergent proportionality results. **A** Equal group sizes are formed by randomly sampling noncarriers individuals equal to the number of carrier individuals with mild and prominent DBA (n=7). One sample from each individual in the pool is then selected. Theta E is calculated with this reduced pool. This process is repeated for 100 permutations and the results are averaged. **B** False discovery rates (FDR) are calculated in a similar manner, where samples are first randomly mixed. The number of gene pairs with a theta E below cutoffs is then calculated. Finally, the expected results are compared to observed data, yielding an FDR of 0.3 for a stringent theta E cutoff of 0.025.

Supplemental Figure-7: The loss of coordination of mitochondrial component genes to RPL11 in carriers are plotted for each gene. The 16 mitochondrial related genes that are (theta E > 0.15) in emergent proportionality with respect to RPL11 are plotted by their ratio to RPL11 and are colored by phenotype.

Supplemental Table-1: Details about the individual samples and RNA libraries are provided.

Supplemental Table-2: Related to Figure 2. Raw number of reads per sample are provided for each of the splice junctions that are plotted in Figure-2.

Supplemental Table-3: Log fold changes, mean expression values as calculated by Dream and DESeq Analysis are provided with P values. First tab provides the list for comparison of noncarriers to all carriers with and without prominent DBA. Second tab provides the list of comparison of noncarriers to carriers without prominent DNA only, this list is larger than the first list.

Supplemental Table-4: DESeq normalized reads per sample are provided. Please check family diagram in Figure 1 for the status of each individual.

Supplemental Table-5: Gene ontology (GO) analysis was performed for the genes with a theta E less than 0.15 (corresponding to 0.3 FDR) with respect to *RPL11* using FuncAssociate 3.0 (Berriz, et.al., 2009). The list is provided for all significant GO terms with an adjusted p-value below 0.05.

Supplemental Table-6: Antisense ribosomal DNA oligos used for the rRNA depletion for RNA-seq library preparation.

REFERENCES:

Ajore, R., Raiser, D., McConkey, M., Jöud, M., Boidol, B., Mar, B., . . . Nilsson, B. (2017). Deletion of ribosomal protein genes is a common vulnerability in human cancer, especially in concert with TP53 mutations. *EMBO Mol Med*, *9*(4), 498-507.

Akakura, S., & Gelman, I. H. (2012). Pivotal role of AKAP12 in the regulation of cellular adhesion dynamics: control of cytoskeletal architecture, cell migration, and mitogenic signaling. *Journal of signal transduction*, *2012*.

Amirbeigiarab, S., Kiani, P., Sanchez, A. V., Krisp, C., Kazantsev, A., Fester, L., . . . Ignatova, Z. (2019). Invariable stoichiometry of ribosomal proteins in mouse brain tissues with aging. *Proceedings of the National Academy of Sciences*, *116*(45), 22567-22572.

Anna, A., & Monika, G. (2018a). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*, *59*(3), 253-268.

Anna, A., & Monika, G. (2018b). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of applied genetics*, *59*(3), 253-268.

Antunes, A. T., Goos, Y. J., Pereboom, T. C., Hermkens, D., Wlodarski, M. W., Da Costa, L. et al. (2015). Ribosomal Protein Mutations Result in Constitutive p53 Protein Degradation through Impairment of the AKT Pathway. *PLoS Genet*, *11*(7), e1005326.

Arribere, J. A., Cenik, E. S., Jain, N., Hess, G. T., Lee, C. H., Bassik, M. C., & Fire, A. Z. (2016). Translation readthrough mitigation. *Nature*, *534*(7609), 719-723.

Baker, R. K., Vanderboom, A. K., Bell, G. W., & Antin, P. B. (2001). Expression of the receptor tyrosine kinase gene EphB3 during early stages of chick embryo development. *Mechanisms of Development*, *104*(1-2), 129-132.

Berriz, G. F., Beaver, J. E., Cenik, C., Tasan, M., & Roth, F. P. (2009). Next generation software for functional trend analysis. *Bioinformatics*, *25*(22), 3043-3044.

Bibikova, E., Youn, M. Y., Danilova, N., Ono-Uruga, Y., Konto-Ghiorghi, Y., Ochoa, R. et al. (2014). TNF-mediated inflammation represses GATA1 and activates p38 MAP kinase in RPS19-deficient hematopoietic progenitors. *Blood*, *124*(25), 3791-3798.

Calviño, F. R., Kharde, S., Ori, A., Hendricks, A., Wild, K., Kressler, D., . . . Sinning, I. (2015). Symportin 1 chaperones 5S RNP assembly during ribosome biogenesis by occupying an essential rRNA-binding site. *Nature communications*, *6*(1), 1-8.

Carlston, C. M., Afify, Z. A., Palumbos, J. C., Bagley, H., Barbagelata, C., Wooderchak-Donahue, W. L., . . . Carey, J. C. (2017). 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*(10), 2622-2627.

Chakraborty, A., Uechi, T., Higa, S., Torihara, H., & Kenmochi, N. (2009). Loss of ribosomal protein L11 affects zebrafish embryonic development through a p53-dependent apoptotic response. *PLoS One*, *4*(1), e4152.

Cheng, Z., Guo, J., Chen, L., Luo, N., Yang, W., & Qu, X. (2015). Overexpression of TMEM158 contributes to ovarian carcinogenesis. *Journal of Experimental & Clinical Cancer Research*, *34*(1), 1-9.

Da Costa, L., Willig, T.-N., Fixler, J., Mohandas, N., & Tchernia, G. (2001). Diamond-Blackfan anemia. *Current Opinion in Pediatrics*, *13*(1), 10-15.

Draptchinskaia, N., Gustavsson, P., Andersson, B., Pettersson, M., Willig, T. N., Dianzani, I. et al. (1999). The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat Genet*, *21*(2), 169-175.

Erb, I., Quinn, T., Lovell, D., & Notredame, C. (2017). *Differential proportionality-a normalization-free approach to differential gene expression*. The 7th International Workshop on Compositional Data, CoDaWork 2017: Proceedings book, 55-70.

Fancello, L., Kampen, K. R., Hofman, I. J., Verbeeck, J., & De Keersmaecker, K. (2017). The ribosomal protein gene RPL5 is a haploinsufficient tumor suppressor in multiple cancer types. *Oncotarget*, *8*(9), 14462-14478.

Farrar, J. E., & Dahl, N. (2011). Untangling the phenotypic heterogeneity of Diamond Blackfan anemia. *Semin Hematol*, *48*(2), 124-135.

Fontenay, M., Cathelin, S., Amiot, M., Gyan, E., & Solary, E. (2006). Mitochondria in hematopoiesis and hematological diseases. *Oncogene*, *25*(34), 4757-4767.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., . . . Armstrong, J. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research*, *47*(D1), D766-D773.

Gagne, K. E., Ghazvinian, R., Yuan, D., Zon, R. L., Storm, K., Mazur-Popinska, M., . . . Higman, M. A. (2014). Pearson marrow pancreas syndrome in patients suspected to have Diamond-Blackfan anemia. *Blood, The Journal of the American Society of Hematology*, *124*(3), 437-440.

Gastou, M., Rio, S., Dussiot, M., Karboul, N., Moniz, H., Leblanc, T. et al. (2017). The severe phenotype of Diamond-Blackfan anemia is modulated by heat shock protein 70. *Blood advances*, *1*(22), 1959-1976.

Gazda, H. T., Sheen, M. R., Vlachos, A., Choesmel, V., O'Donohue, M. F., Schneider, H., . . . Beggs, A. H. (2008). Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. *Am J Hum Genet*, *83*(6), 769-780.

Goudarzi, K. M., & Lindström, M. S. (2016). Role of ribosomal protein mutations in tumor development (Review). *Int J Oncol*, *48*(4), 1313-1324. doi:10.3892/ijo.2016.3387

Guimaraes, J. C., & Zavolan, M. (2016). Patterns of ribosomal protein expression specify normal and malignant human cells. *Genome Biol*, *17*(1), 236.

Heijnen, H. F., van Wijk, R., Pereboom, T. C., Goos, Y. J., Seinen, C. W., van Oirschot, B. A. et al. (2014). Ribosomal protein mutations induce autophagy through S6 kinase inhibition of the insulin pathway. *PLoS Genet*, *10*(5), e1004371.

Hoffman, G. E., & Roussos, P. (2020). dream: Powerful differential expression analysis for repeated measures designs. *Bioinformatics*. doi:10.1093/bioinformatics/btaa687
Hoffman, G. E., & Schadt, E. E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*, *17*(1), 483.

Hu, T., Wu, X., Li, K., Li, Y., He, P., Wu, Z., . . . Guan, M. (2019). AKAP12 Endogenous Transcripts Suppress The Proliferation, Migration And Invasion Of Colorectal Cancer Cells By Directly Targeting oncomiR-183-5p. *OncoTargets and therapy*, *12*, 8301.

Hug, N., Longman, D., & Cáceres, J. F. (2016). Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic acids research*, *44*(4), 1483-1495.

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., . . . Farh, K. K. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, *176*(3), 535-548.e24.

Jessica Moetter,M., Joerg Meerpohl, Alexandra Fischer, Thorsten Simon, Sandra Urbani ak, Shinsuke Hirabayashi, Friedrich Kapp, Peter Noellke, Christian Kratz, Charlotte M Niemeyer and Marcin W. Wlodarski. (2011). Analysis of Ribosomal Protein Genes Associated with Diamond Blackfan Anemia (DBA) In German DBA Patients and Their Relatives. *Blood* , 118:729.

Kaniak-Golik, A., & Skoneczna, A. (2015). Mitochondria–nucleus network for genome stability. *Free Radical Biology and Medicine*, *82*, 73-104.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434-443.

Khajuria, R. K., Munschauer, M., Ulirsch, J. C., Fiorini, C., Ludwig, L. S., McFarland, S. K., . . . Sankaran, V. G. (2018). Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. *Cell*, *173*(1), 90-103.e19.

Khatter, H., Myasnikov, A. G., Natchiar, S. K., & Klaholz, B. P. (2015). Structure of the human 80S ribosome. *Nature*, *520*(7549), 640-645.

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, *37*(8), 907-915.

Kolpakova, E., Wiedlocha, A., Stenmark, H., Klingenberg, O., Falnes, P. Ø., & Olsnes, S. (1998). Cloning of an intracellular protein that binds selectively to mitogenic acidic fibroblast growth factor. *Biochemical Journal*, *336*(1), 213-222.

Komili, S., & Silver, P. A. (2008). Coupling and coordination in gene expression processes: a systems biology view. *Nature Reviews Genetics*, *9*(1), 38-48.

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology*, *15*(2), R29.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Exome, A. C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285-291.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . 1000, G. P. D. P. S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923-930.

Lipton, J. M., & Ellis, S. R. (2010). Diamond Blackfan anemia 2008-2009: broadening the scope of ribosome biogenesis disorders. *Curr Opin Pediatr*, *22*(1), 12-19.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 550.

Lovell, D., Pawlowsky-Glahn, V., Egozcue, J. J., Marguerat, S., & Bähler, J. (2015). Proportionality: a valid alternative to correlation for relative data. *PLoS Comput Biol*, *11*(3), e1004075.

Malumbres, M. (2014). Cyclin-dependent kinases. *Genome biology*, *15*(6), 1-10.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), 10-12.

Moore, M. J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. *Science, 309*(5740), 1514-1518

Morgado-Palacin, L., Varetti, G., Llanos, S., Gómez-López, G., Martinez, D., & Serrano, M. (2015). Partial Loss of Rpl11 in Adult Mice Recapitulates Diamond-Blackfan Anemia and Promotes Lymphomagenesis. *Cell Rep*, *13*(4), 712-722.

Morrill, S. A., & Amon, A. (2019). Why haploinsufficiency persists. *Proc Natl Acad Sci U S A*, *116*(24), 11866-11871.

Narla, A., Yuan, D., Kazerounian, S., LaVasseur, C., Ulirsch, J. C., Narla, J., . . . Gazda, H. (2016). A novel pathogenic mutation in RPL11 identified in a patient diagnosed with

diamond Blackfan anemia as a young adult. *Blood cells, molecules & diseases*, *61*, 46. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27667165

Naora, H., Takai, I., Adachi, M., & Naora, H. (1998). 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*(3), 741-753.

Pesciotta, E. N., Lam, H. S., Kossenkov, A., Ge, J., Showe, L. C., Mason, P. J. et al. (2015). In-Depth, Label-Free Analysis of the Erythrocyte Cytoplasmic Proteome in Diamond Blackfan Anemia Identifies a Unique Inflammatory Signature. *PLoS One*, *10*, e0140036.

Polfus, L. M., Khajuria, R. K., Schick, U. M., Pankratz, N., Pazoki, R., Brody, J. A., . . . Huang, J. (2016). Whole-exome sequencing identifies loci associated with blood cell traits and reveals a role for alternative GFI1B splice variants in human hematopoiesis. *The American Journal of Human Genetics*, *99*(2), 481-488.

Risley, M., Garrod, D., Henkemeyer, M., & McLean, W. (2009). EphB2 and EphB3 forward signalling are required for palate development. *Mech Dev*, *126*(3-4), 230-239.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, *43*(7), e47.

Sankaran, V. G., Ghazvinian, R., Do, R., Thiru, P., Vergilio, J. A., Beggs, A. H., . . . Gazda, H. T. (2012). Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J Clin Invest*, *122*(7), 2439-2443.

Sano, S., Inoue, S., Tanabe, Y., Sumiya, C., & Koike, S. (1959). Significance of mitochondria for porphyrin and heme biosynthesis. *Science*, *129*(3344), 275-276.

Shajani, Z., Sykes, M. T., & Williamson, J. R. (2011). Assembly of bacterial ribosomes. *Annu Rev Biochem*, *80*, 501-526.

Skinnider, M. A., Squair, J. W., & Foster, L. J. (2019). Evaluating measures of association for single-cell transcriptomics. *Nature methods*, *16*(5), 381-386.

Sorber, K., Dimon, M. T., & DeRisi, J. L. (2011). RNA-Seq analysis of splicing in Plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nucleic acids research*, *39*(9), 3820-3835.

Taggart, J. C., & Li, G. W. (2018). Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes. *Cell Syst*, *7*(6), 580-589.e4.

Tye, B. W., Commins, N., Ryazanova, L. V., Wühr, M., Springer, M., Pincus, D., & Churchman, L. S. (2019). Proteotoxicity from aberrant ribosome biogenesis compromises cell fitness. *Elife*, *8*.

Ulirsch, J. C., Verboon, J. M., Kazerounian, S., Guo, M. H., Yuan, D., Ludwig, L. S., . . . Genovese, G. (2018). The genetic landscape of Diamond-Blackfan anemia. *The American Journal of Human Genetics*, *103*(6), 930-947.

Vlachos, A., Rosenberg, P. S., Atsidaftos, E., Alter, B. P., & Lipton, J. M. (2012). Incidence of neoplasia in Diamond Blackfan anemia: a report from the Diamond Blackfan Anemia Registry. *Blood, The Journal of the American Society of Hematology*, *119*(16), 3815-3819.

Wilkes, M. C., Siva, K., Chen, J., Varetti, G., Youn, M. Y., Chae, H. et al. (2020). Diamond Blackfan anemia is mediated by hyperactive Nemo-like kinase. *Nature communications*, *11*(1), 1-17.

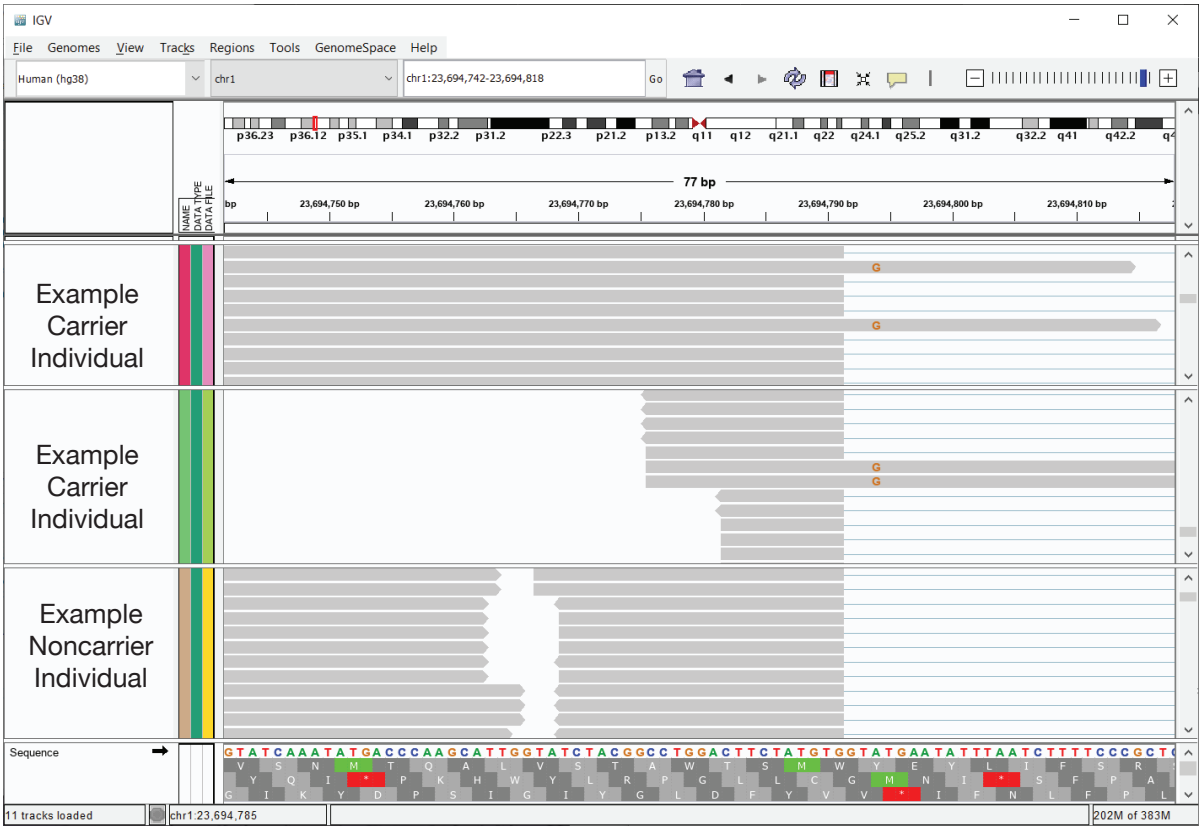
Woodson, J. D., & Chory, J. (2008). Coordination of gene expression between organellar and nuclear genomes. *Nature Reviews Genetics*, *9*(5), 383-395.

Woolford, J. L., & Baserga, S. J. (2013). Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics*, *195*(3), 643-681.

Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M. L., Visvader, J. E., & Smyth, G. K. (2010). ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics*, *26*(17), 2176-2182.

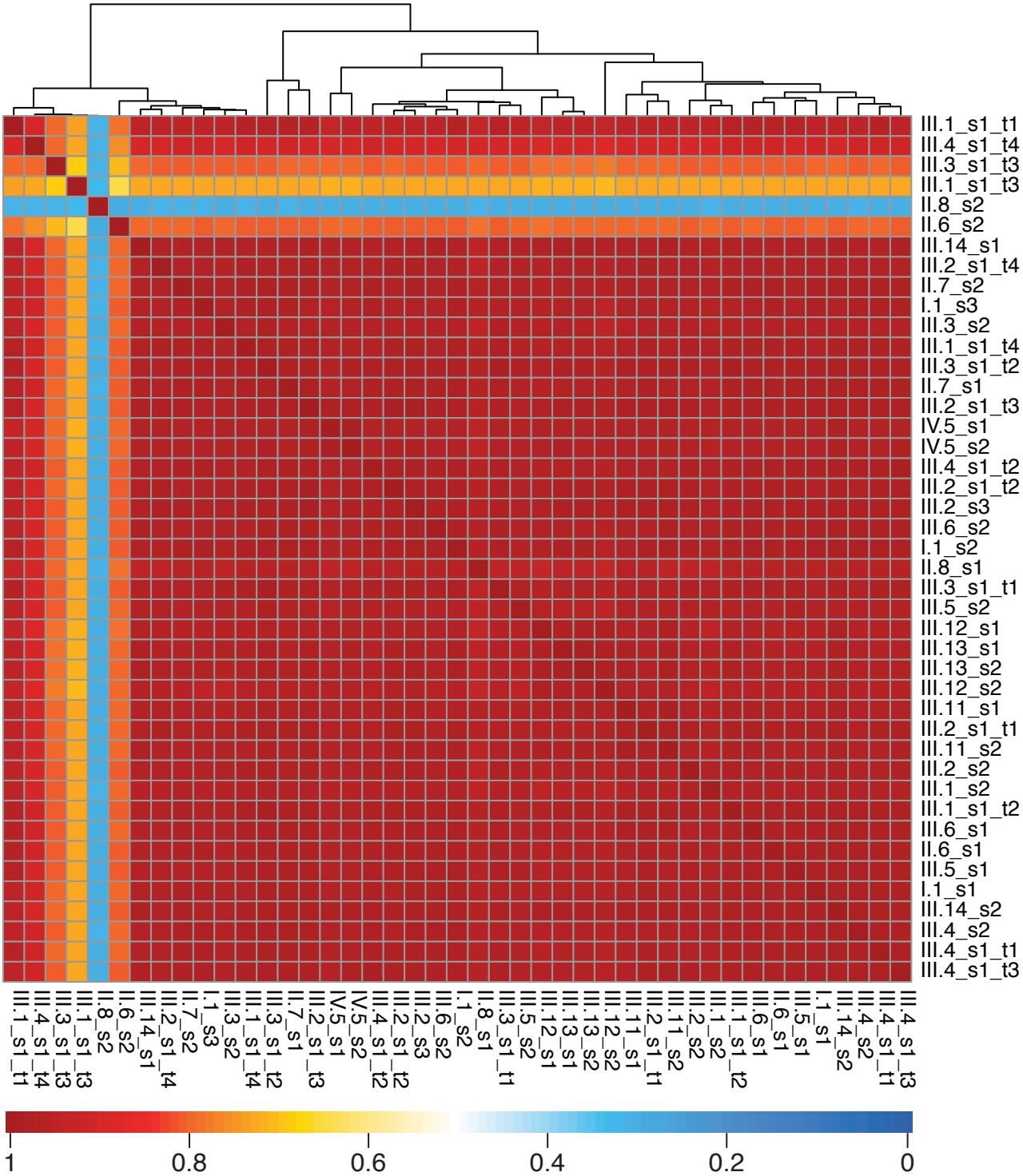
Xu, S., Li, X., Gong, Z., Wang, W., Li, Y., Nair, B. C., . . . Chen, J. (2014). Proteomic analysis of the human cyclin-dependent kinase family reveals a novel CDK5 complex involved in cell growth and migration. *Molecular & Cellular Proteomics*, *13*(11), 2986-3000.

Supplemental Figure-1



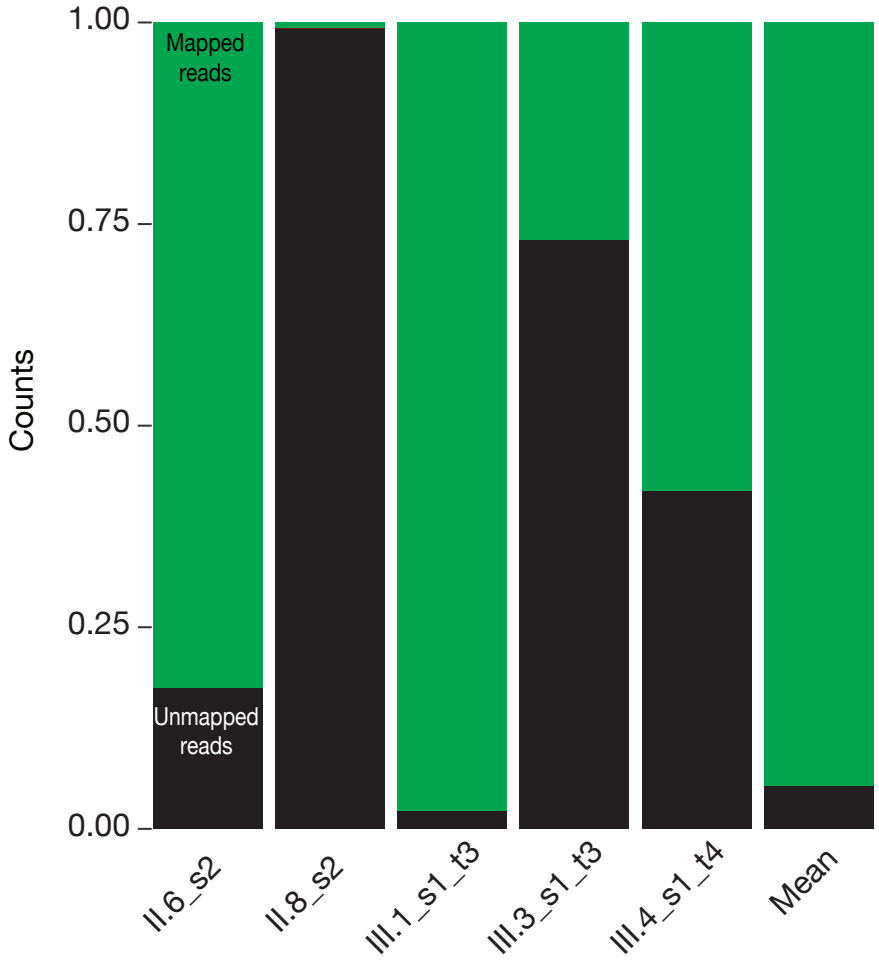
Supplemental Figure-1: An exemplary of IGV snapshot of *RPL11* gene at exon4-intron4 region from example carrier and non-carrier individuals. Few reads that map to *RPL11* variant c.396 3A>G was used as a confirmation of the carrier status.

Supplemental Figure-2



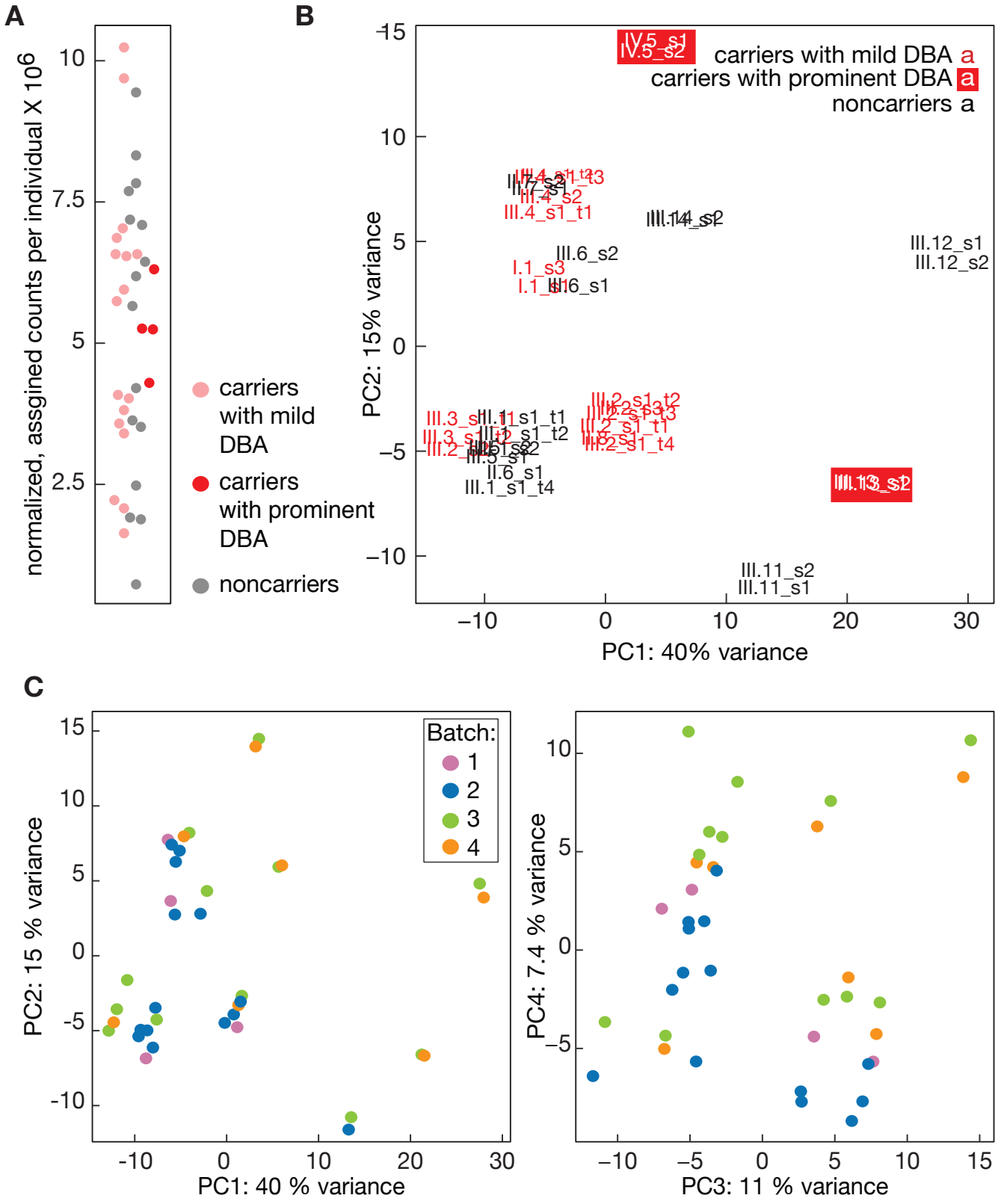
Supplemental Figure-2: Plot of spearman correlation across RNA samples and individuals. 5 samples with lower spearman correlation values (<0.9) were discarded from further RNA expression analysis. These 5 samples were further investigated in Figure S3. The mean correlation value across samples after the sample removal is 0.97. Across samples, example: III.4, refers to the individual; example: s1 refers to the biological sample taken, example: t1 refers to a technical replicate (RNAseq library was prepared from the same sample at different times). The libraries were prepared in four batches. A few of the samples were included and processed as technical replicates in all of the batches to ensure that a batch bias doesn't exist (Supplementary Table-1).

Supplemental Figure-3



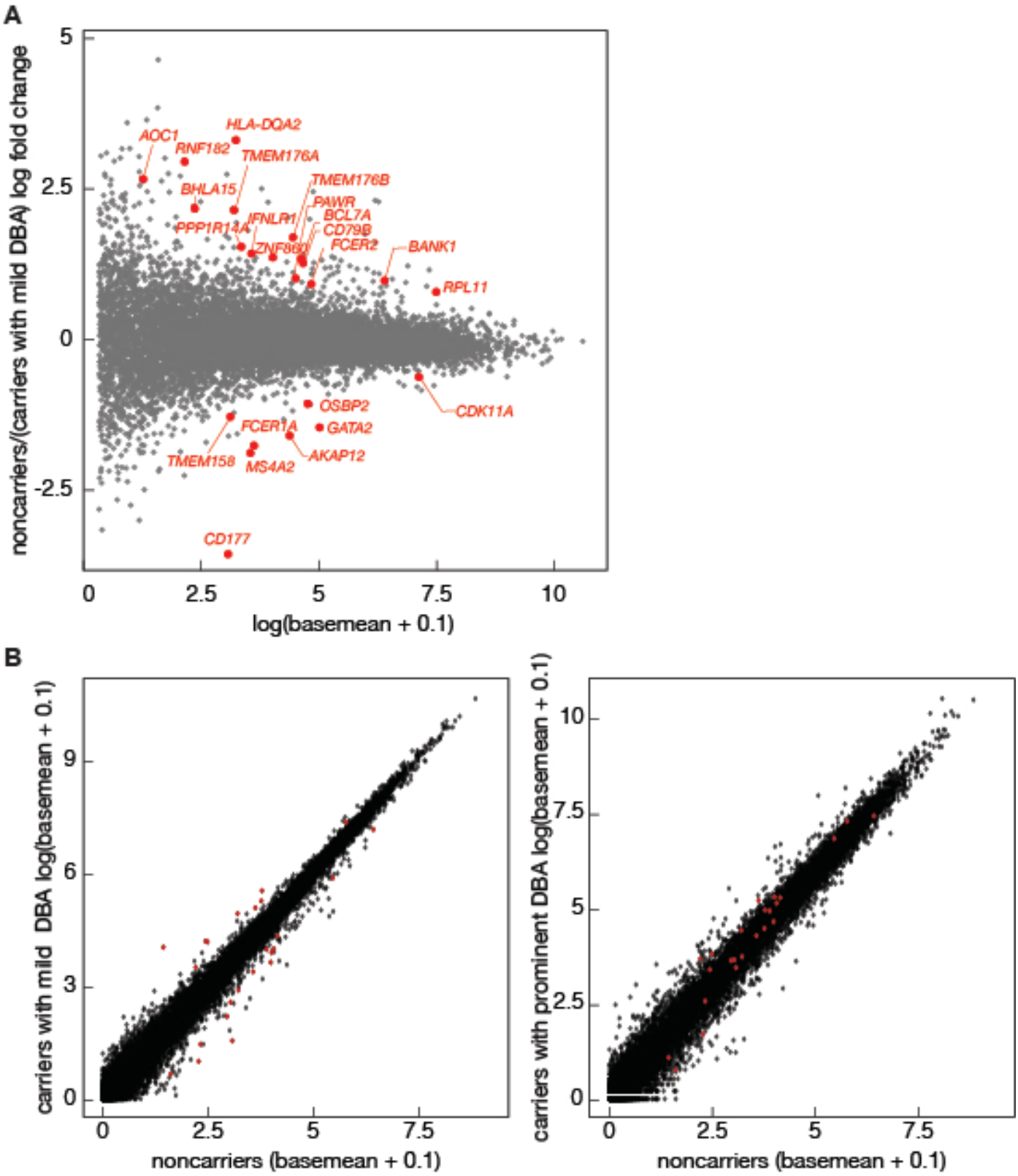
Supplemental Figure-3: Samples with <0.9 Spearmen correlation from Figure S1 generally have lower percentage of genome mapping reads. Samples II.8_s2, III.3_s1_t3, III.4_s1_t4, which were removed from further analysis, generally have low counts and percentages of mapped reads in comparison to the other example samples; II.6_s2 and III.1_s1_t3.

Supplemental Figure-4



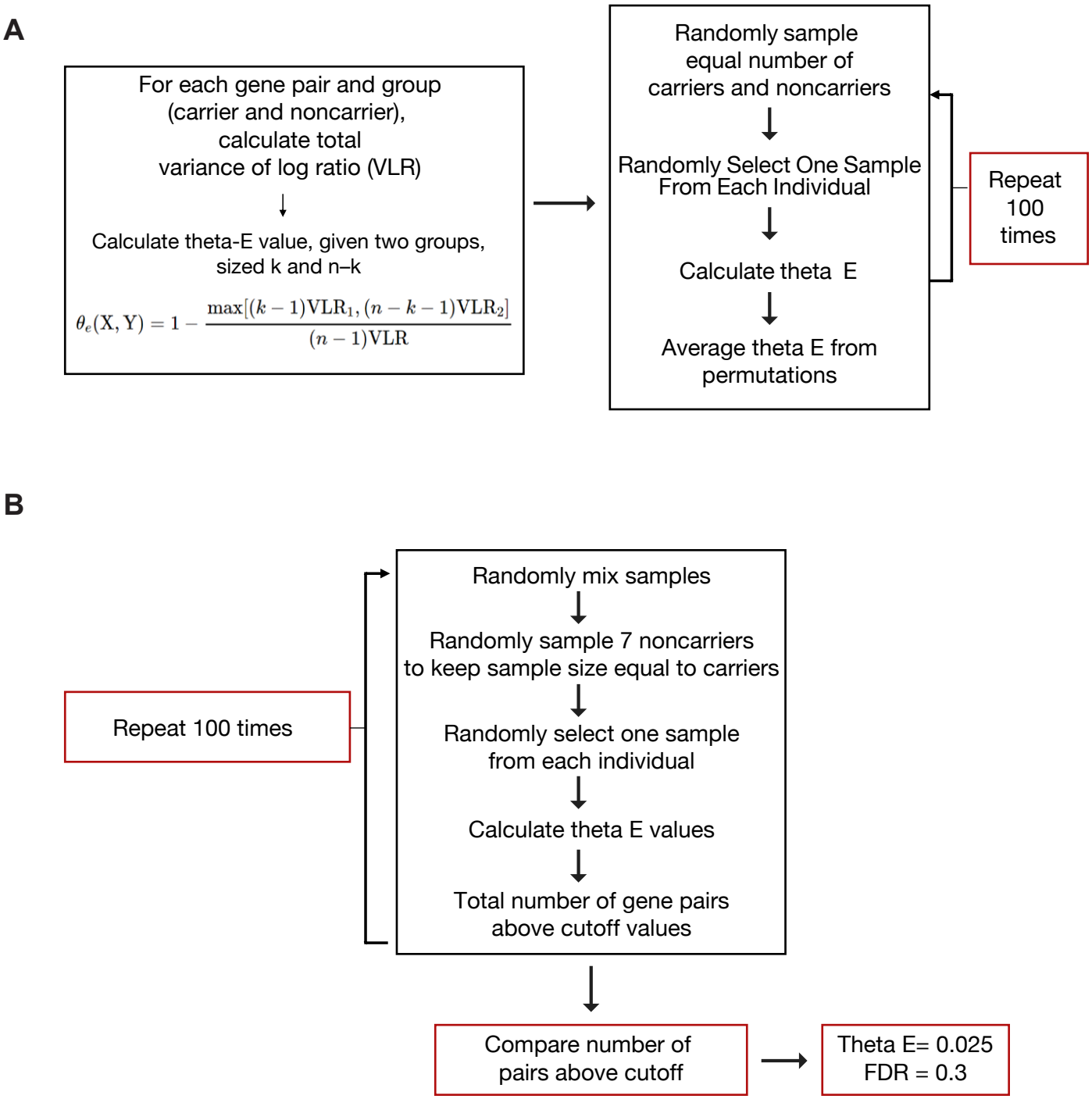
Supplemental Figure-4: A Normalized assigned counts were plotted for each sample. Grey, light red and dark red points represent noncarriers, carriers with few manifestations of DBA and carriers with prominent DBA findings respectively. **B** PCA analysis plot is shown for each sample annotated with individual identifier from Figure 1, followed by a number that stands for sample number for the particular individual. RNA expression libraries were prepared in four batches and three samples were included in all of the batches. **C** Libraries that were prepared in each batch was colored accordingly (batch 1–purple, 2–blue, 3–green, 4–orange) in the two PCA plots; first two variance components are plotted left, third and fourth components are plotted on the right. Note that PCA plots in B and C are the same, where each sample were labeled with a sample name, or a batch number respectively.

Supplemental Figure-5



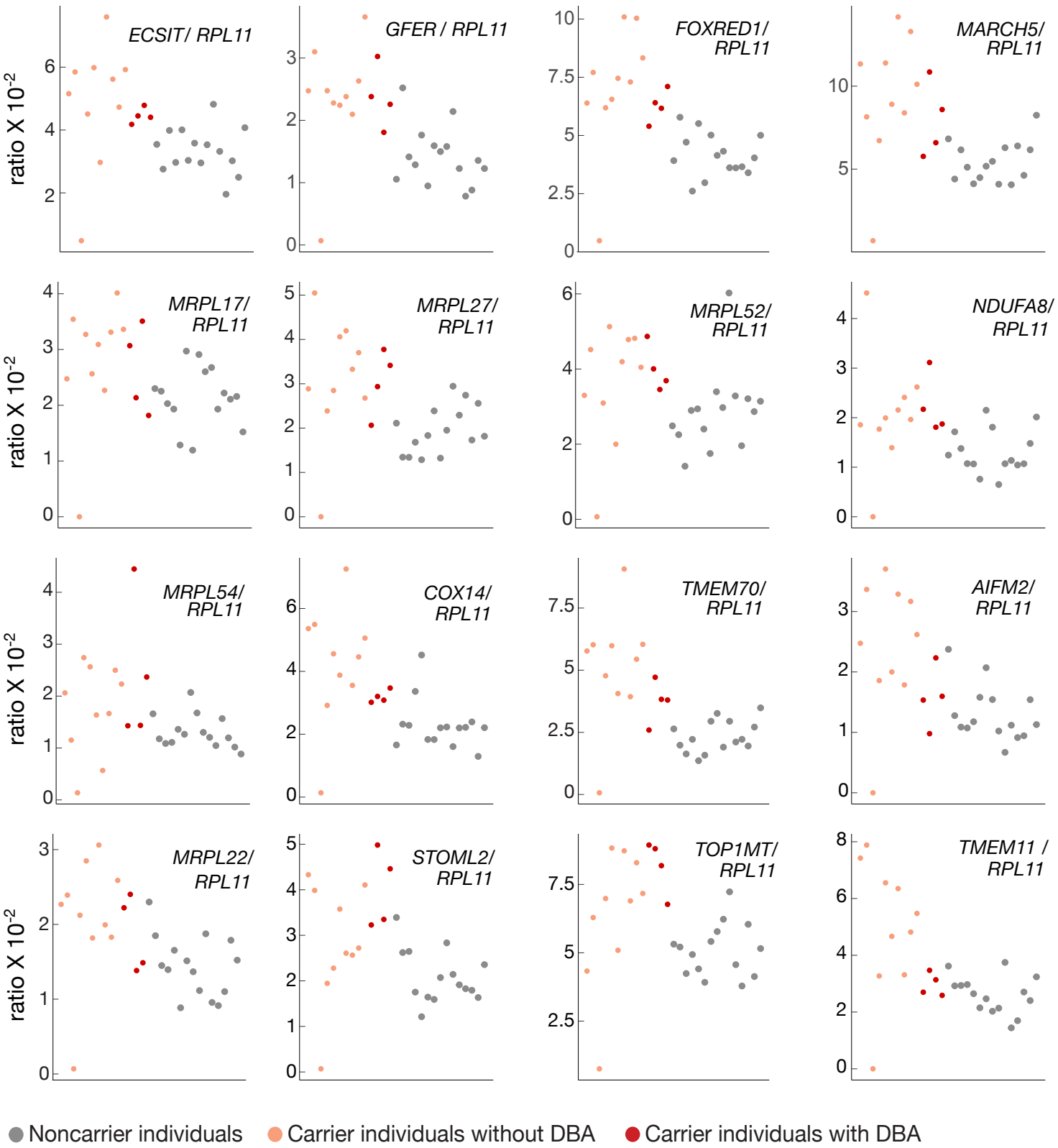
Supplemental Figure-5: Shared gene expression changes among carrier individuals with few manifestations of DBA compared to noncarriers. A Gene expression log fold changes (y axis) and log base mean values per gene were plotted on the left large plot. Red labeled points are significant genes that are differentially expressed (R, Limma Analysis FDR cutoff<0.05, Deseq2 Analysis FDR cutoff<0.01). **B** Mean gene expression of carriers mild and prominent DBA (y axis on the left and right plots respectively) are compared to noncarriers (x axis). Red points on the plot represent significantly over and under expressed genes in carriers with mild DBA symptoms in comparison to noncarriers.

Supplemental Figure-6



Supplemental Figure-6: Implementation of individual sampling allows for more robust emergent proportionality results. A Equal group sizes are formed by randomly sampling noncarriers individuals equal to the number of carrier individuals with mild and prominent DBA (n=7)**.** One sample from each individual in the pool is then selected. Theta E is calculated with this reduced pool. This process is repeated for 100 permutations and the results are averaged. **B** False discovery rates (FDR) are calculated in a similar manner, where samples are first randomly mixed. The number of gene pairs with a theta E below cutoffs is then calculated. Finally, the expected results are compared to observed data, yielding an FDR of 0.3 for a stringent theta E cutoff of 0.025.

Supplemental Figure-7



Supplemental Figure-7: The loss of coordination of mitochondrial component genes to *RPL11* in carriers are plotted for each gene. The 16 mitochondrial related genes that are (theta E > 0.15) in emergent proportionality with respect to *RPL11* are plotted by their ratio to *RPL11* and are colored by phenotype.