# Analysis of fibroblasts from patients with cblC and cblG genetic defects of cobalamin metabolism reveals global dysregulation of alternative splicing

Charif Rashka, Sébastien Hergalant, Natacha Dreumont, Abderrahim Oussalah, Jean-Michel Camadro, Virginie Marchand, Ziad Hassan, Matthias Baumgartner, David Rosenblatt, François Feillet, et al.

# Analysis of fibroblasts from patients with *cblC* and *cblG* genetic defects of cobalamin metabolism reveals global dysregulation of alternative splicing

Charif Rashka[1], Sébastien Hergalant[1], Natacha Dreumont[1], Abderrahim Oussalah[1,2], Jean-Michel Camadro[3], Virginie Marchand[4], Ziad Hassan[1], Matthias R. Baumgartner[5], David S. Rosenblatt[6], François Feillet[1,2], Jean-Louis Guéant[1,2], David Coelho[1,2,*]

1 Inserm UMRS 1256 NGERE – Nutrition, Genetics, and Environmental Risk Exposure, University of Lorraine, Nancy, F-54000, France

2 National Center of Inborn Errors of Metabolism, University Regional Hospital Center of Nancy, Nancy, F-54000, France

3 Université Denis Diderot – Paris 7, CNRS, UMR 7592, Paris, F-75205, France

4 University of Lorraine, CNRS, INSERM, UMS2008, IBSLor, Epitranscriptomics and RNA Sequencing Core Facility, Nancy, F-54000, France

5 radiz–Rare Disease Initiative Zürich, Clinical Research Priority Program for Rare Diseases, University of Zürich, Switzerland

6 Department of Human Genetics, McGill University, Montreal, Canada

***To whom correspondence should be addressed at:** UMRS 1256 Inserm-Université de Lorraine, NGERE, Faculté de Médecine, 9 Avenue de la Forêt de Haye - BP50184, 54505 Vandœuvre-lès-Nancy, France. Tel:+33 3 72 74 61 50; Fax: 03.72.74.61.31; Email: david.coelho@inserm.fr

**Abstract**

Vitamin B12 or cobalamin (Cbl) metabolism can be affected by genetic defects leading to defective activity of either methylmalonyl-CoA mutase or methionine synthase or both enzymes. Patients usually present with a wide spectrum of pathologies suggesting that various cellular processes could be affected by modifications in gene expression. We have previously demonstrated that these genetic defects are associated with subcellular mislocalization of RNA binding proteins and subsequent altered nucleo-cytoplasmic shuttling of mRNAs. In order to characterize the possible changes of gene expression in these diseases, we have investigated global gene expression in fibroblasts from patients with *cblC* and *cblG* inherited disorders by RNA-seq. The most differentially expressed genes are strongly associated with developmental processes, neurological, ophthalmologic and cardiovascular diseases. These associations are consistent with the clinical presentation of *cblC* and *cblG* disorders. Multivariate analysis of transcript processing revealed splicing alterations that led to dramatic changes in cytoskeleton organization, response to stress, methylation of macromolecules and RNA binding. The RNA motifs associated with this differential splicing reflected a potential role of RNA binding proteins such as HuR and HNRNPL. Proteomic analysis confirmed that mRNA processing was significantly disturbed. This study reports a dramatic alteration of gene expression in fibroblasts of patients with *cblC* and *cblG* disorders, which resulted partly from disturbed function of RNA binding proteins. These data suggest to evaluate the rescue of the mislocalization of RNA binding proteins as a potential strategy in the treatment of severe cases who are resistant to classical treatments with co-enzyme supplements.

## Introduction

In man, vitamin B12 or cobalamin (Cbl) metabolism is responsible for the synthesis of adenosylcobalamin (AdoCbl) and methylcobalamin (MeCbl), the two cofactors required for the mitochondrial methylmalonyl-CoA mutase and the cytosolic methionine synthase, respectively. These two biosynthetic pathways can be affected by genetic defects that are inherited in an autosomal recessive manner and were historically defined by complementation groups and names *cblA-G*, *cblJ* and *mut* [1]. The most common of these inherited disorders is the cobalamin C (*cblC*) defect with more than 500 patients diagnosed. This complementation group is defined by defects in the *MMACHC* gene (MIM#609831), affecting both AdoCbl and MeCbl metabolism. The MMACHC protein has been associated to different functions related to early processing of Cbl in the cells from cytosolic chaperone, removal of the upper-axial ligands of dietary Cbl: decyanation of cyanocobalamin and dealkylation of alkylcobalamins, to presentation of Cbl in the base-off state required by methionine synthase. The *cblG* defect is defined by mutations in the *MTR* gene coding for methionine synthase (MIM 250940) and approximately 50 patients have been diagnosed.

Patients with the *cblC* defect have methylmalonic acidemia combined with hyperhomocysteinemia while *cblG* patients have isolated hyperhomocysteinemia. Neurological, ophthalmological and hematological symptoms are common in patients from both groups with some specific symptoms [2, 3]. Patients with the *cblC* defect usually present with feeding difficulties, hypotonia, and lethargy and the clinical course includes mainly hematologic and neurologic symptoms. Renal failure, hepatic dysfunction, cardiomyopathy and visual impairment are also frequent. Almost all patients with the *cblG* defect present with megaloblastic anemia and neurologic disorders including lethargy, failure to thrive, developmental delay, nystagmus, hypotonia or hypertonia, ataxia, seizures, and blindness. Despite some common features, clinical presentation and severity of these two inborn errors can vary considerably [2, 3].

The mechanisms at the origin of the high variability in the severity and in the nature of the pathologies associated with these disorders are far from being understood but recent findings suggest that alterations of gene expression could contribute to these pathomechanisms. Indeed, it has been reported in fibroblasts from *cblC* patients, a relatively moderate overexpression of pro-apoptotic and anti-apoptotic genes [4].

While the *cblC* defect leads to reduced synthesis of the two Cbl co-enzymes, we have recently shown that the lack of MMACHC could also result in decreased activity of methionine synthase, probably through a weaker interaction with its partner methionine synthase reductase [5, 6]. Thus, a reduced activity of methionine synthase is a common feature of the two complementation groups *cblC* and *cblG*. This would result in two biochemical consequences that could alter regulation of gene expression. First, impairment of this enzymatic reaction will lead to the accumulation of the two substrates homocysteine and methyltetrahydrofolate, disturbing the folate cycle which is crucial for important cellular pathways including amino acid metabolism, but also biosynthesis of purine and pyrimidines. Methionine is the precursor of S-adenosyl-methionine (SAM), the universal methyl donor required for the methylation of nucleic acids and proteins. Consequently, the reduced synthesis of methionine and SAM will affect the methylation of DNA, RNA and proteins involved in the regulation of gene expression, including histones, transcription factors and their co-activators and RNA binding proteins. We have recently reported that a defective Cbl metabolism could also alter the subcellular localization of RNA binding proteins (RBP) such as ELAVL1/HuR by altering their methylation and phosphorylation status, resulting in abnormal mRNA nucleocytoplasmic transport, in a transgenic cellular mouse model and fibroblasts from patients [7]. Moreover, this study also revealed that this mislocalization of RBP was associated with dramatic transcriptomic changes.

The subcellular mislocalization of RNA binding proteins (RBP) observed in transgenic mouse and cell models and patient fibroblasts with impaired cellular availability of Cbl led to subsequent global dysregulation of mRNA trafficking and dramatic transcriptomic changes related to metabolic processing of RNA, cellular response to stress, regulation of cell cycle, neurogenesis and neuron differentiation, plasticity and development [7, 8]. These results suggested that transcriptomic

dysregulation could be an important contributor to the pathogenesis of the inherited disorders of Cbl metabolism through mechanisms that remain to be dissected.

The altered RBP trafficking may influence alternative splicing. For example, mutations in genes coding for RBP produce aberrant splicing linked to amyotrophic lateral sclerosis and autism [9, 10]. Since, HuR and other RBP are mislocalized in fibroblasts from patients with *cblC* and *cblG* disorders, we aimed to evaluate the consequences of these genetic defects on gene expression, including gene splicing, that could contribute to the spectrum of clinical manifestations reported in these disorders.

To address this issue, we performed a deep RNA-Seq analysis of gene expression and a proteomic analysis of fibroblasts from 3 *cblC* and 4 *cblG* patients, with mutations in the *MMACHC* and *MTR* genes respectively. Our data revealed for the first time in *cblC/G* cells altered gene expression and abnormal alternative splicing and highlights the contribution of RNA binding proteins in this process. The most differentially expressed genes are associated with developmental processes and with cardiovascular, neurological and visual manifestations of these disorders. Analysis of differential alternative splicing revealed that cytoskeleton organization, response to cellular stress and RNA binding are the most affected biological functions. Further analysis of RNA motifs significantly involved in the differential splicing events led to the identification of the corresponding RBP. Furthermore, proteomic analysis of the fibroblasts confirmed that mRNA processing was significantly affected and highlighted the role of HNRNPL in the changes in gene expression.

## Results

### RNA-Seq landscape and Identification of differentially expressed genes (DEG)

The transcriptome of fibroblasts from 3 *cblC* and 4 *cblG* patients was analyzed by RNA sequencing and compared to a control cell line with 3 biological replicates of each cell line. Description of the cell lines are presented in Supplementary Table S1. In *cblC* cells, differential expression analysis (false discovery rate or FDR<0.05 and fold change or FC>1.5) identified 84 genes with significantly increased expression and 90 genes with significantly decreased expression. In *cblG* cells, 110 genes

were upregulated and 73 genes were down-regulated. We identified a subset of 51 genes that was also up-regulated in *cblC* and in *cblG* cells, while 36 genes were found down-regulated in both groups (Fig. 1a). The ten DEG with the higher positive fold change and the ten with the higher negative fold change are presented in Table 1 with the associated p-value, FDR and a short summary of their function and the known associated pathologies. It is interesting to note that these genes are either down or over-expressed in both cell types with dramatic values of fold change. Several genes are associated with cardiac dysfunction (*TLL1*, *TRPV2*, *DUSP4*, *PPARG*) and nervous system development or function (*EPHA5*, *ISL2*, *SAMD5*, *CDON*). The full list of differentially expressed genes in both groups is presented in Supplementary Table S2.

**Functional analysis of DEG**

In order to functionally characterize the DEG, we carried out an annotation analysis using Panther with Gene Ontology (GO) and we sorted the GO terms by decreasing fold enrichment (Fig. 1b). In *cblC* cells, the analysis revealed a statistically significant down regulation of genes involved in the following pathways: muscle development (GO: 0014706) and differentiation (GO: 0051146) skeletal system development (GO: 0001501) but also neurological processes such as retinal ganglion cell axon guidance (GO: 0031290) and forebrain development (GO: 0030900). In *cblG* cells, the significant annotations referred mainly to decreased expression of genes associated with anatomical structure development (GO: 0048856), organ development (GO: 0048513) and actin cytoskeleton organization (GO: 0030036). Over-expressed genes in cells from both genetic defects were associated with regulation of localization (GO: 0032879). The finding that cytoskeleton of patient fibroblasts could be affected is consistent with the very high proportion of patient cells with an abnormal morphology compared to control (Supplementary Fig. S1).

**Experimental validation of RNA-Seq data**

In order to validate the transcriptomic data obtained by RNA-Seq, 28 genes were selected on the basis of their fold change, differential expression and potential relevance in the pathophysiology of the disease. Real-time quantitative PCR (RT-qPCR) of this set of genes was performed with mRNA from

the same cell lines used in the RNA-Seq study (Fig. 2a). Most RT-qPCR results confirm the transcriptomic data with 25 genes validated in *cblC* cells (89%) and 21 genes in *cblG* group (75%) (Fig. 2a). The relationship between RNA-Seq and RT-qPCR results was characterized by the calculation of Spearman's rho coefficient (ρ). The results indicate a strong linear correlation of the expression levels obtained by RNA-Seq analysis versus RT-qPCR. (*cblC*: Spearman's rho, ρ =0.8878; P-value = 8.57E-7 *cblG*; Spearman's rho, ρ =0.5978, P-value = 7.81E-4*;* Supplementary Fig. S2*)*. We also selected two other genes that are differentially expressed in *cblC* but not in *cblG* cells, *DMD* (FC *cblC*: -11.55) and *MYH1* (FC cblC: -4), to be validated by western blot and immunofluorescence in one cell line from each group (Fig. 2b). Our results confirm that these two genes are strongly under-expressed in *cblC* cells only.

**Analysis of alternative splicing (AS) events**

We have previously demonstrated that impaired Cbl metabolism, including the *cblC* and *cblG* genetic defects, are associated with abnormal mRNA trafficking and at least aberrant splicing of *IRF3* [7, 8]. Thus, in order to further investigate the consequences of these inherited disorders on alternative splicing, we analyzed our RNASeq data for changes in alternative splicing events (AS) using rMATS software (Multivariate Analysis of Transcript Splicing) [15]. This software allows the quantification of 5 types of AS events: skipped exons (SE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), mutually exclusive exon (MXE) and retained intron (RI). The distribution of the density of genes significantly affected (difference in percentage of spliced index Delta PSI > 5% and FDR <0.01), differs according to the categories of AS events (SE, RI, MXE, A5SS and A3SS) in *cblC* and in *cblG* cells compared to control (Fig. 3a). A difference is also observed both groups with the genes affected by MXE and A5SS being more heterogeneous in the *cblG* cells (Fig. 3a). In both cell lines, more than 70% of the genes were affected by skipped exon while the proportion of other genes affected by A3SS, A5SS, MXE and RI events is relatively the same between the lines *cblC* and *cblG* versus control (Fig. 3b). Compared to control, we identified in *cblC* cells, 850 differential splicing events (Delta PSI> 5%, FDR <0.01) within 719 genes and in *cblG* cells, we found 1206 differential splicing events within 1007 genes (Table 2). A subset of 314 genes were affected in both groups.

Among the genes affected by differential AS events, a very small number is also differentially expressed (Table S2). Table 3 shows the 20 genes most significantly differentially spliced with their p-values, FDR and ranked by Delta PSI (Delta PSI> 5% and FDR <0.01). The full list of differentially spliced genes in both groups can be found in Supplementary Table S3.

**Functional analysis of differential AS events**

In order to gain insights into the potential consequences of the differential AS events, we performed a functional analysis of the affected genes. In *cblC* cells, gene ontology (GO) analysis showed that the top biological processes affected are related to metabolic processes, DNA repair, response to DNA damage and cellular stress, cytoskeleton organization but also modification of macromolecules including methylation and histone modification (Fig. 4). More importantly, the main molecular functions affected by the altered alternative splicing are functions related to the contribution of one carbon metabolism in the regulation of gene expression such as protein-lysine N-methyltransferase, SAM-dependent methyltransferase and also RNA binding activities. In *cblG* cells, the GO analysis also showed an involvement of differentially expressed genes in protein modification processes such as H3-K9 histones, repair of DNA damage, cytoskeleton organization, gene expression and RNA binding.

**Experimental validation of alternative splicing by RT-PCR**

In order to corroborate the rMATS analysis, we selected 4 genes affected by skipped exon (*XIAP, EEF1AKMT1* and *PRMT2)* or by alternative 5' splice site (*MBD5*) in order to validate the AS events by RT-PCR (Supplementary Table S4). Primers were designed in neighboring exons of AS events (Supplementary Table S5) and were tested on one *cblC* and one *cblG* cell line. Our results confirm a difference in the alternative splicing of *XIAP*, *MBD5*, *EEF1AKMT1* and *PRMT2* genes in the *cblC* and *cblG* cells compared to the control (Fig. 5). In the *cblC* line, the results of the RT-PCR confirmed the difference of exon inclusion for *PRMT2*. A difference in *XIAP* and *MBD5* splicing was also observed but the results differed from the rMAPS analysis. In the *cblG* line, we confirmed the rMATS results

for *XIAP*, and *EEF1AKMT1* genes and we detected a difference in *MBD5* splicing that differs from the rMATS data.

**Identification of the RBP involved in differential AS events**

In order to investigate further our hypothesis of the contribution of the RBP in the dysregulation of mRNA trafficking and alternative splicing detected in *cblC* and in *cblG* cells, we examined the regions surrounding the differential alternative splicing events previously identified, to identify the motifs or binding sites of RBP that could be significantly over-represented. Thus, we submitted the rMATS outputs with the rMAPS software using the default parameters, to generate RNA-maps for the identification of consensus sequences. In total, we identified 66 motifs around alternatively spliced exons in *cblC* and 49 motifs in *cblG* cells. The number of RBP recognizing these patterns are 43 in *cblC* and 44 in *cblG* (Fig. 6a). Among the 25 RBP shared by *cblC* and *cblG* cells, 14 were supposed to contribute to SE events. The full list of RBP associated with each consensus sequence ranked by decreasing enrichment (pattern> 2 and log (p)> 2.8) can be found in Supplementary Table S6. In order to identify the RBP that are particularly involved in the AS events found in patient cells, we have calculated the enrichment of their targets in the total number of genes affected by AS events using the reference Starbase V3 data for the RBP whose information about target genes are available. Our data indicate a significant enrichment for 2 RBP in the *cblC* cells only (PCBP2 and SRSF7), 6 RBP in the *cblG* group (HNRNPL, KHDRBS1, MSI1, RBM5, SRSF10 and SRSF3) and 13 RBP in the two genetic defects (ELAV1/HuR, FMR1, HNRNPC, HNRNPK, IGF2BP2, IGF2BP3, LIN28A, PTBP1, RBM37, SRSF1, SRSF9, TARDBP, and TIA1) (Fig. 6b).

**Subcellular localization of RBP**

We have previously discovered that abnormal mRNA metabolism associated with impaired Cbl metabolism resulted from mislocalization of RBP [7, 8]. Thus, we decided to investigate the subcellular localization of a subset of RBP whom contribution in differential AS events was highlighted by the rMAPS analysis by analyzing the subcellular localization of TDP43, FMR1, SRSF7, TIA1, HNRNPK and HNRNPL by immunofluorescence and confocal microscopy in control,

*cblC* and *cblG* fibroblasts (Supplementary Figures S3 and S4). In contrast to ELAV1/HuR that was previously found mislocalized in patient fibroblasts, we did not detect any difference in the subcellular localization of those RBP between patient and control cells.

**Proteomic analysis and expression level of HNRNPL**

Since the subcellular localization of these RBP was not altered in patient cells, we next analyzed their expression level. A proteomic analysis was performed on the same set of *cblC* and *cblG* cells by nano-LC/MSMS coupled to an Orbitrap Fusion ETD mass spectrometer. Functional annotation of the differentially expressed proteins (FC>1.25; pvalue<0.05; Supplementary Tables S7 and S8) revealed that mRNA processing (GO:0006397) was significantly disturbed in *cblC* (pvalue<0.0005) but not in *cblG* cells. However, we found that HNRNPL, one of the RBP identified in our rMAPS analysis, was significantly over-expressed in *cblC* (FC 4.28; pvalue 0.006) and in *cblG* cells (FC 3.36; pvalue 0.013). The overexpression of HNRNPL in cells with both defects was further confirmed by western blotting and quantification of immunostaining (Fig. 7).

**Role of HuR and HNRNPL in the transcriptomic and proteomic modifications in patient cells**

Our previous data have highlighted the major role of HuR in the transcriptomic changes associated with impaired Cbl metabolism [7]. Moreover, we have also shown using *cblC* and *cblG* cells that this effect resulted from its mislocalization and not from a modified expression level [7, 8]. However, our data also underscore the contribution of HNRNPL in the alterations of alternative splicing while its subcellular localization remains unchanged in contrast to an increase expression. Thus, in order to investigate the contributions of these two RBP in the transcriptomic and proteomic modifications associated with the *cblC* and *cblG* defects, we calculated the enrichments of their respective targets in the genes that we found differentially transcribed, alternatively spliced and expressed (Table 4). Significant values for HuR were found for the differential alternative splicing events (*cblC*: 1.62 (pvalue 3.85 10E-11); *cblG*: 1.63 (6.86 10E-23)) and protein expression (*cblC*: 1.51 (5.51 10E-4); *cblG*: 1.55 (1.02 10E-3)) for both genetic defects. Similar values were found for HNRNPL for alternative splicing (*cblC*: 1.74 (7.96 10E-2); *cblG*: 1.83 (1 10E-20)) but not for protein expression

(*cblC*: 1.22 (0.51); *cblG*: 1.53 (0,17)). No significant enrichment for the two RBP was found in the lists of the differentially transcribed genes in the two cell types.

**Discussion**

By analyzing the whole transcriptome of *cblC* and *cblG* cells, our study aimed at two main goals. First, we sought to characterize the gene expressions disrupted by these genetic defects that could contribute to the pathomechanisms of their clinical outcomes. Second, we wanted to investigate further the hypothesis based on our previous findings that the altered function of RBP associated with *cblC* and *cblG* disorders could affect gene expression by disturbing normal splicing. In order to achieve these goals, we analyzed the RNA sequences of the transcriptome and mRNA splicing in fibroblasts from 4 different *cblG* patients carrying mutations in the *MTR* gene and from 3 *cblC* patients with a defect in the *MMACHC* gene. We also performed a proteomic analysis of the same cells in order to confirm our findings about the contribution of RBP in the consequences of the transcriptomic deregulation.

A small number of studies have already described some modifications in gene expression in patient cells with genetic defects of intracellular metabolism of cobalamin. Fibroblasts from patients with mutations in *HCFC1* (*cblX* defect) or in *THAP11* gene shared some similarities in their transcriptome with a cluster of 46 differentially expressed genes compared to the controls [16]. Modifications of protein expression in patient cells with *cblD-MMA* defect have also been reported in an analysis of mitochondria, that lead to the identification of proteins mainly involved in apoptosis, oxidative stress and cell metabolism [4]. Two other studies focused on *cblC* cells identified changes of protein expression involved mainly in cytoskeleton organization, nervous system signaling, energy metabolism and cellular detoxification [17, 18].

Our results revealed dramatic transcriptomic changes in *cblC* and in *cblG* patient fibroblasts affecting developmental processes of muscles, bones and nervous system, in agreement with the manifestations of early onset of the disease [2, 3]. Among the most differentially expressed genes (DEG), several code for transcription factors that trigger cascades of dysregulation. Indeed, we found a massive

decrease in the expression of the transcription factors *ISL2* and *FOXE1* with a fold change in *cblC* cells reaching 58 and 128 and in *cblG* cells 30 and 145, respectively.

Interestingly, the functional analysis of DEG also highlighted a significant enrichment of genes involved in cytoskeleton organization, confirming a previous proteomic analysis of *cblC* cells [18]. Many studies indicate that consequences of cytoskeleton abnormalities could contribute to the development of neurological and cardiac dysfunctions. Indeed, we have previously shown in a cellular model of folate deficiency that disorganization of the cytoskeleton could lead to impair vesicular transport and synaptic functions [19]. It has also been shown that alterations of cytoskeleton could affect cell motility and axon guidance in patients with amyotrophic lateral sclerosis [20, 21]. Furthermore, our results indicate that among the differentially expressed genes identified in our study, many of them code for cytoskeletal proteins that have been directly associated with cardiac and neurological disorders.

Our RNAseq shows also DEG of genes potentially involved in cardiomyopathy manifestations presented by *cblC* and *cblG* patients. *MYH10* gene is strongly downregulated in *cblC* and *cblG* cells with a fold change (FC) close to 6 in both cell lines. Mutations in the *MYH10* gene, that encodes a non-muscle myosin, have been associated with developmental defects in brain and heart [22]. Moreover, *MYH10* mutations were also associated in mice with severe cardiac defects including ventricular septal defect [23]. The protein encoded by *PDLIM3*, a gene down-regulated in *cblC* and in *cblG* cells with a fold change reaching -19 and -69 respectively, is supposed to be involved in cytoskeletal assembly and plays an important role in mice cardiac muscles. Indeed, *PDLIM3* deficient mice lead to right ventricular chamber dilatation and dysfunction suggesting a role in the onset of cardiomyopathy [24]. The *TLL1* gene codes for a metalloprotease that is crucial for the development of mammalian heart and was downregulated in both cell types (FC *cblC* -19; FC *cblG* -68.6) [25]. The role of this gene was further confirmed by the identification of deleterious mutations of *TLL1* in four non-related patients with atrial septal defect [26]. *DMD* is another important gene strongly under expressed in *cblC* cells. It encodes dystrophin, a protein that connects the filamentous actin cytoskeleton to the extracellular matrix. Since the role of this protein is to stabilize the plasma

membrane, a functional deficiency of dystrophin leads to progressive muscle weakness leading to loss of ambulation, respiratory complications and heart failure. Indeed, dilated cardiomyopathy is considered as a major cause of death in patients with *DMD* mutations [27]. The precise molecular mechanisms causing the cellular damage triggered by the lack of dystrophin are still poorly understood. However, the main hypotheses suggest a direct mechanism, dysregulation of calcium pathways and an increased oxidative stress [28]. In addition to the F-actin structures, the dystrophin protein complex is also associated with the microtubules and intermediate filaments and the role of the cytoskeleton in heart failure is well documented [29].

Several genes up-regulated in *cblC* and in *cblG* cells is also associated with cardiac disorders. The overexpression of *DUSP4*, which encodes a dual-specificity protein phosphatase that can inactivate MAP kinases, has been observed in dilated cardiomyopathy [30]lynx1

. Moreover, specific cardiac overexpression of *DUSP4* is sufficient to cause cardiomyopathy in transgenic mice [31]. *PPARG* is another gene up-regulated in *cblC* and *cblG* fibroblasts. It encodes the peroxisome proliferator-activated receptor-gamma (PPARγ) a ligand-activated transcription factor that regulates tissue triglyceride accumulation. Its overexpression in mice has been associated with dilated cardiomyopathy [32]. PPARγ expression is repressed by SIRT1 [33]. We found an upregulation of PPARγ concurrently with a decreased expression of PPARα and SIRT1 in rats exposed to the fetal programming of the deficiency in vitamin B12 and folate and a decreased expression of SIRT1 in fibroblasts from patients with inherited disorders [34, 8]. Thus, the upregulation of *PPARG* reported in our patient cells is likely to result from the decreased activity of SIRT1 associated with impaired vitamin B12 metabolism [35, 8].

Our RNAseq shows differential expression of genes potentially involved in the neurologic disorders presented by *cblC* and *cblG* patients. *EPHA5* is a gene that was found to be strongly down-regulated in both *cblC* and *cblG* cells (FC *cblC:* 211; FC *cblG:* 13) and it has been found to be crucial for synaptogenesis during hippocampus development. This protein coded by this gene is a member of the ephrin receptors that participate in various cellular functions such as immune cell development and neuronal cell migration, including retinal ganglion cell axon guidance [36]. Another gene strongly

down-regulated in patient fibroblasts that has been associated with neurological and visual problems is *ISL2* (FC *cblC:* 58; FC *cblG:* 30), that codes for a transcription factor important for motoneurons in the embryonic spinal cord and for the pathfinding of retinal ganglion cell axons [37, 38].

Some of the up-regulated genes found in our cells could also contribute to the development of neurological disorders. Indeed, we reported an over-expression of *LYNX1* (FC *cblC:* 10.9; FC *cblG:* 8.4), a gene coding for a membrane protein known to modulate the function of nicotinic acetylcholine receptors, which contribute to a wide array of functions, including learning and memory. Interestingly, the increased expression of *LYNX1* impairs plasticity in the primary visual cortex of mice while its absence enhanced learning and memory abilities [39, 40]. We also found an up-regulation of *DAPK1* (FC *cblC:* 10.5; FC *cblG:* 14.6), which codes for a Ser/Thr kinase that plays important roles in neural development. DAPK1 contributes to the pathogenesis of many neurological diseases including epilepsy, Alzheimer and Parkinson diseases. Indeed, overexpression of *DAPK1* has been observed in the hippocampus of patients with Alzheimer's disease and is sufficient to produce locomotor abnormalities and death of dopaminergic neurons in a mouse model of Parkinson's disease [41, 42].

It is well known that disruptions in RNA metabolism, including aberrant alternative mRNA splicing, are associated with many pathologies, such neurological diseases, dilated cardiomyopathy, retinitis pigmentosa and many types of cancer [43]. Thus, in addition to investigate quantitative changes in the transcriptome of *cblC* and *cblG* cells, we also aimed at identifying the potential alterations of alternative splicing. Our results reveal that the number of differentially spliced genes (719 in *cblC* group and 1007 in *cblG* group) is much higher than the number of genes differentially expressed (174 genes in *cblC* group and 183 in *cblG* group; FDR<0.05; FC>1.5). Analysis of these differentially spliced genes indicated a significant enrichment in functions related to cellular response to stress, cytoskeleton organization and regulation of gene expression by methylation. We confirmed the alteration of alternative splicing by RT-PCR in *MBD5*, which encodes a protein of the methyl-CpG-binding domain (MBD) family that also includes *MECP2*, a gene associated with the Rett syndrome. Alterations of *MBD5* have been associated with epileptic encephalopathies, intellectual disability and autism spectrum disorder [44, 45]. A study performed with a mouse model of Rett syndrome has

revealed that the loss of MECP2 function was not only associated with widespread aberrations in gene expression but also with anomalous patterns of alternative splicing following neuronal stimulation [46].

We have recently demonstrated using *cblC* and *cblG* patient fibroblasts, that these genetic defects are associated with global mRNA mislocalization and abnormal subcellular localization of the RNA binding-proteins HuR, HnRNPA1 and SRSF1 [7, 8]. Our analysis of RNA motifs suggests that other RBPs play also important roles in the aberrant splicing detected in *cblC* and *cblG* fibroblasts: PCBP2 and SRSF7 in *cblC* cells, HNRNPL, KHDRBS1, MSI1, RBM5, SRSF10 and SRSF3 in *cblG* cells and ELAV1/HuR, FMR1, HNRNPC, HNRNPK, IGF2BP2, IGF2BP3, LIN28A, PTBP1, RBM37, SRSF1, SRSF9, TARDBP, and TIA1 in both cell types. Even though most of these RBP were not found to be mislocalized in patient cells, our data indicate that HNRPL is indeed over-expressed in *cblC* and in *cblG* fibroblasts. We also noticed that *HNRNPL* mRNA, as well as many of the most relevant genes differentially expressed in our study are known targets of HuR (*DMD, MYH10, PPARG, DUSP4, DAPK1*). Thus, our findings suggest that changes in HNRNPL expression could result from altered HuR subcellular localization and confirm our previous reports about the crucial role of HuR in the consequences of impaired metabolism of cobalamin [7, 8].

In summary, our data clearly show that the *cblC* and *cblG* defects lead to altered expression and splicing of genes involved in developmental processes and in cardiac and neurological pathomechanisms, which are consistent with the manifestations observed in patients. The finding that many of these dysregulated genes are directly involved in known pathologies confirm previous proteomic studies made with *cblC* cells [17, 18]. Moreover, we found that abnormal alternative splicing associated with both genetic defects also affects important biological processes such as cytoskeleton organization, modification of proteins involved in gene expression and RNA binding. Even though the analysis of our transcriptomic data at a quantitative level did not highlight the role of a specific RBP, our study reveals that these molecules play a major role in the alterations of alternative splicing and protein expression associated with the *cblC* and *cblG* defects. The characterization of the

specific contribution of each RBP on the transcriptome and on the proteome will require further experiments and will raise novel perspectives in the understanding of the molecular mechanisms underlying the clinical manifestations presented by patients with *cblC* and *cblG* defects of cobalamin metabolism. Moreover, our study also suggests that the rescue of the mislocalization of RBP such as HuR, could constitute a potential novel strategy in the treatment of severe cases who are resistant to classical treatments with co-enzyme supplements.

## Materials and Methods

### Patient fibroblasts, cell culture and RNA collection

Patient fibroblasts (3 cblC, 4 cblG and a control line) were obtained from the Reference Center for Inherited Metabolic Diseases of the Great East Inter-region, the University Hospital of Nancy, the University Hospital for Children in Zurich and the McGill University. The description of each cell line is available in supplementary Table S1. All cell lines are from Caucasian origin. Cells were used with passages between 10 and 15 and were free from mycoplasma infection. Cells were maintained in culture in Dulbecco's Modified Eagle Medium (DMEM) Glucose 4500 mg/mL (Sigma Alrich, D5796) supplemented with 10% fetal calf serum (Eurobio, CVFSVF0001), 1% sodium pyruvate (Sigma Aldrich, S8636), 1% penicillin-streptomycin (Sigma Aldrich, P4333) and incubated at 37°C, 5% CO2 until confluent. The culture medium was changed every 3 days. Then, cells were trypsinized and total RNA was isolated from the cell pellet using the RNAeasy mini extraction kit (Qiagen) and treated with DNase I (Ambion, AM1906). RNA quality was assessed by using the 2100 Bioanalyzer (Agilent).

### RNA-Seq library construction and Transcriptome sequencing

3-5 ug of each RNA DNAseI-treated was used for polyA isolation using NEBNext® Poly(A) mRNA Magnetic Isolation Module (ref#E7490S) following the manufacturer's instructions. Enrichment of polyA+ fraction was checked by capillary electrophoresis using a PicoRNA chip on Bioanalyzer 2100 (Agilent Technologies). PolyA+ RNAs were converted to library with the Scriptseq V2 RNA-Seq kit

from Epicentre using the manufacturer's instructions. After 15 cycles of PCR amplification, the libraries were purified using the Agencourt AMPure XP beads (ref#A63880, Beckman Coulter) at a ratio of 0.9x. Library quality was assessed using a High Sensivity DNA chip on a Bioanalyzer 2100, while quantification was done using a fluorometer (Qubit 3.0 fluorometer, Invitrogen). Libraries were multiplexed and subjected for high-throughput sequencing using an Illumina HiSeq 1000 instrument with 2*101 bp paired-end read runs.

**Bioinformatic analysis of RNA-Seq data**

The quality control of nucleotide sequences was performed using FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Following removal of low quality sequences, mapping of the reads was performed by HISAT2 v2.0.4 - [11] on a version of the hg38 reference genome which accounts for splicing donor and acceptor sites and "normal" SNP polymorphism (https://cloud.biohpc.swmed.edu/index.php/s/grch38_snp/download). Spliced isoforms were assembled from the reconstructed transcripts obtained with Stringtie v1.3.3b [12], allowing for the identification of new isoforms in addition to those overlapping known transcripts from the Ensembl official annotations (Homo_sapiens.GRCh38.89.gtf). Data were normalized with the trimmed means of M-values (TMM) standardization method [13] and then transformed in logarithm base 2 (log2-) and counts per million (cpm-), as previously described [14]. Paired-end RNA sequencing of the 24 samples generated an average of 110.2 million reads per sample. The overall alignment rate was 90.55% (sd +/- 1.86%), with an average quality of mapped read at 99.97% (PHRED score 37.2) and an average insert size of 362.5 bp (sd +/- 621.4). About 10.94% (from 9.7% to 11.4%) of reads mapped to intronic regions. 55204 transcripts were reconstructed during the next step, which led to a transcriptome of 50818 unique transcripts and 16808 unique genes after filtering for very low expression against the 3 control samples (mean cpm value > 0.5 was considered proof of expression in the control group). Linear modelling with empirical Bayes was then applied to assess differential expression at gene and at transcript level between both control and each patient cell line.

**Statistical analysis, differentially expressed genes and functional annotations**

Batch effects between the 3 replicate runs were corrected using the ComBat function available in the R sva package. Linear modelling with empirical Bayes was applied to assess differential expression at gene and at transcript level between both control and each patient cell line. Differential expression p-values were obtained by using a moderated Student t-test and adjusted for the false discovery rate (FDR) with the Benjamini-Hochberg procedure. Differentially expressed genes (DEGs) were identified with FDR < 0.05 and a fold-change (FC) above 1.5. Gene ontology analysis for significant biological processes and molecular functions (FDR < 0.05) for DEG were performed with Panther database.

**RT-PCR and western blotting**

Total RNA was isolated from the cell pellet of 3 cblC, 4 cblG and 1 control (WT). We used custom plates containing 28 pairs of lyophilized primers, designed and pre-validated by Biorad to study gene expression by RT-qPCR. Two reference genes (*AASDH* and *ALAS1*) were used to normalize expression levels. Gene expression levels were pre-analyzed using StepOnePlus Real-Time PCR System (Thermofischer). A Student test was established on RT-qPCR data. Fold changes were calculated using the 2-deltadelta (CT) method and transformed into log2 (FC). Spearman correlation (ρ) was calculated to establish a relationship between RNA-Seq and RT-qPCR data. Western blot of MYH1, DMD and HNRNPL proteins were performed with the following antibodies (DMD, Abcam, ab15277; MYH1, Abcam ab91506; B-actin, Cell Signaling 4970; HNRNPL, Santa Cruz Sc-46673) as described [7].

**Alternative splicing of cblC and cblG patient fibroblasts and functional annotation**

RNA-Seq data was uploaded to the rMATS online software (version 3.0, 8) [15] to identify differential alternating splice (AS) events between two groups control and cblC/cblG cell lines, corresponding to all five basic types of AS patterns: skipped exon (SE), alternative 5' spliced site (A5SS), alternative 3' spliced site (A3SS), mutually exclusive exon (MXE) and intron retention (RI). For each AS event, we used both the reads mapped to the splice junctions and the reads mapped to the exon body as the input for rMATS. Parameters used to detect significant AS events are FDR <1% and

delta PSI>5%. Functional annotation of significant differentially spliced genes was performed using Panther database.

**Experimental validation of alternative splicing events**

Five differentially spliced genes, *XIAP*, *MBD5*, *EEF1AKMT1* and *PRMT2* were selected to validate the rMATS results by RT-PCR. PCR primers were designed using the Primer-BLAST (V.3) software (Supplementary Table S4).

**Motif enrichment analysis and identification of RNA Binding Protein (RBP)**

rMATS outputs were uploaded to the online rMAPS portal (http://rmaps.cecsresearch.org/MTool/) and the motif map tool was used to test common RNA motif enrichments within nearby splicing regions, according to the 5 considered alternative splicing events. Default settings were used apart for the FDR threshold for the rMATS results that was lowered to 1%. Thresholds were applied on both p-value and z-score RMAPS outputs to assess statistical significance for over and under-represented consensus motifs (RBP recognition patterns; p-value < 1e-3 and z-score > 2-fold evaluated background). The complete list of recognition patterns for RBPs is available in Supplementary Table S5. From the Starbase V3 Clip-Seq experiment database, we compiled lists of RBP targets for all RBP available, with a stringency of at least 3 whenever possible. We only considered targets that are mRNAs with protein coding as Ensembl biotype (20244 genes). The enrichment of each RBP target in the number of AS events was calculated as follow: (number of genes affected by AS events that are also known target of RBP)/number of genes affected by AS events)/(number of genes known to be target of RBP)/total number of genes.

**LC-MS/MS acquisition**

Protein extracts were analyzed using an Orbitrap Fusion Tribrid equipped with an EASY-Spray nanoelectrospray ion source and coupled to an Easy nano-LC Proxeon 1000 system (Thermo Fisher Scientific, San Jose, CA). Chromatographic separation of peptides was performed with the following parameters: Acclaim PepMap100 C18 pre-column (2 cm, 75 µm i.d., 3 µm, 100 Å), Pepmap-RSLC

Proxeon C18 column (50 cm, 75 µm i.d., 2 µm, 100 Å), 300 nl/min flow, gradient rising from 95 % solvent A (water, 0.1% formic acid) to 35% solvent B (100 % acetonitrile, 0.1% formic acid) in 120 minutes followed by column regeneration for 50 minutes. Peptides were analyzed in the Orbitrap cell, in full ion scan mode, at a resolution of 120,000 (at m/z 200), with a mass range of m/z 350-1550 and an AGC target of 4x105. Fragments were obtained by higher-energy C-trap dissociation (HCD) activation with a collisional energy of 30%, and a quadrupole isolation window of 1.6 Da. MS/MS data were acquired in the linear ion trap in a top-speed mode, with a total cycle of 3 seconds with an AGC target of 1x104. The maximum ion accumulation times were set to 100 ms for MS acquisition and 35 ms for MS/MS acquisition in parallelization mode. All MS/MS data were processed with an in-house Mascot search server (Matrix Science, Boston, MA; version 2.4.1) using Proteome Discoverer 2.1 (Thermo Scientific). The mass tolerance was set to 6 ppm for precursor ions and 0.5 Da for fragments. The following modifications were used in variable modifications: oxidation (M), phosphorylation (STY), acetylation (Protein N-term, K). The maximum number of missed cleavages by trypsin was limited to two for all proteases used. MS/MS data were searched against the UniProtKB RefSeq Homo sapiens database.

**Quantitative analysis in label-free experiments**

For quantitative proteomic analyses, the data of proteomic experiments are the results of three independent samples. Cell extracts were prepared as described above, and the protein content of the samples was measured in order to quantify the amount of proteins to be precipitated with acetone. Four times volume of cooled acetone (-20°C) were added to sample volume containing 50 µg of protein extracts. Vortexed tubes were incubated for 60 minutes at -20°C then centrifuged for 10 minutes at 13,000-15,000 x g. The protein pellet was heated for 20 minutes at 95°C and then cooled on ice for 20 minutes before performing, in triplicate, digestion overnight at 37°C by sequencing grade trypsin (12.5 µg/ml; Promega, Madison, Wi, USA) in 30 µl of 25 mmol/L NH4HCO3. LC-MS/MS acquisition with a 2-hour gradient. Label-free quantification in between subject analysis was performed on raw data with Progenesis-Qi software 4.1 (Nonlinear Dynamics Ltd, Newcastle, U.K.) using the following procedure: (i) chromatograms alignment, (ii) peptide abundances normalization,

(iii) statistical analyses of features, and (iv) peptides identification using the Mascot server through Proteome Discoverer 2.1 (Thermo Scientific). A decoy search was performed and the significance threshold was fixed to 0.01. The resulting files were imported into Progenesis-LC software. Normalized abundances of proteins from trypsin digests with similar normalized abundance variations (ANOVA p value < 0.05) were classified together by the AutoClass Bayesian clustering system and visualized with Javatreeview (http://jtreeview.sourceforge.net/).

**Acknowledgments**

This work was supported by Fondation Jérôme Lejeune, CHRU Nancy Hospital and Région Lorraine.

*Conflict of Interest statement.* None declared.

**Author Contributions**

C.R., S.H., N.D., A.O., J.M.C., V.M. and Z.H. provided input on methodology, validation, investigation, formal analysis and data curation; M.R.B. and D.S.R. provided input on conceptualization, methodology and resources; F.F., J.L.G. and D.C. provided input on conceptualization, writing and supervision.

**References**


[1] Coelho, D., Kim, J.C., Miousse, I.R., Fung, S., du Moulin, M., Buers, I., Suormala, T., Burda, P., Frapolli, M., Stucki, M., et al. (2012) Mutations in ABCD4 cause a new inborn error of vitamin B12 metabolism. *Nat. Genet.*, **10**, 1152-1155.

[2] Fischer, S., Huemer, M., Baumgartner, M., Deodato, F., Ballhausen, D., Boneh, A., Burlina, A.B., Cerone, R., Garcia, P., Gökçay, G., et al. (2014) Clinical presentation and outcome in a series of 88 patients with the cblC defect. *J. Inherit. Metab. Dis*., **37**, 831-840.

[3] Huemer, M., Bürer, C., Ješina, P., Kožich, V., Landolt, M.A., Suormala, T., Fowler, B., Augoustides-Savvopoulou, P., Blair, E., Brennerova, K., et al. (2015) Clinical onset and course, response to treatment and outcome in 24 patients with the cblE or cblG remethylation defect complemented by genetic and in vitro enzyme study data. *J. Inherit. Metab. Dis.*, **38**, 957-967

[4] Richard, E., Monteoliva, L., Juarez, S., Pérez, B., Desviat, L.R., Ugarte, M., Albar, J.P. (2006) Quantitative analysis of mitochondrial protein expression in methylmalonic acidemia by two-dimensional difference gel electrophoresis. *J. Proteome Res*., **5**, 1602-1610.

[5] Fofou-Caillierez, M.B., Mrabet, N.T., Chéry, C., Dreumont, N., Flayac, J., Pupavac, M., Paoli, J., Alberto, J.M., Coelho, D., Camadro, J.M., et al. (2013). Interaction between methionine synthase isoforms and MMACHC: characterization in cblG-variant, cblG and cblC inherited causes of megaloblastic anaemia. *Hum. Mol. Genet*., **22**, 4591-4601.

[6] Bassila, C., Ghemrawi, R., Flayac, J., Froese, D.S., Baumgartner, M.R., Guéant, J.L., Coelho, D. (2017) Methionine synthase and methionine synthase reductase interact with MMACHC and with MMADHC. *Biochem. Biophys. Acta Mol. Basis Dis*., **1863**, 103-112.

[7] Battaglia-Hsu, S.F., Ghemrawi, R., Coelho, D., Dreumont, N., Mosca, P., Hergalant, S., Gauchotte, G., Sequeira, J.M., Ndiongue, M., Houlgatte, R., et al. (2018) Inherited disorders of cobalamin metabolism disrupt nucleocytoplasmic transport of mRNA through impaired methylation/phosphorylation of ELAVL1/HuR. *Nucleic Acids Res*., **46**, 7844-7857.

[8] Ghemrawi, R., Arnold, C., Battaglia-Hsu, S.F., Pourié, G., Trinh, I., Bassila, C., Charif, R., Wiedemann, A., Flayac, J., Robert, A., et al. (2019) SIRT1 activation rescues the mislocalization of RNA-binding proteins and cognitive defects induced by inherited cobalamin disorders. *Metabolism*, **101**, 153992

[9] Dredge, B.K., Polydorides, A.D., Darnell, R.B. (2001) The splice of life: alternative splicing and neurological disease. *Nat. Rev. Neurosci*., **2**, 43-50.

[10] Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandal, M.J., Hartl, C., Leppa, V., Ubieta, L.T., Huang, J., et al. (2016) Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*, **540**, 423-427.

[11] Kim, D., Langmead, B., Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357-360.

[12] Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol*., **33**, 290-295.

[13] Robinson, M.D., Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*., **11**, R25.

[14] Law, C., Paquet, M., Kania, A. (2014) Emergence of motor circuit activity. *PLoS One*, e93836.

[15] Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., Xing, Y. (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E5593-5601.

[16] Quintana, A.M., Yu, H.C., Brebner, A., Pupavac, M., Geiger, E.A., Watson, A., Castro, V.L., Cheung, W., Chen, S.H., Watkins, D., et al. (2017) Mutations in THAP11 cause an inborn error of cobalamin metabolism and developmental abnormalities. *Hum. Mol. Genet.*, **26**, 2838-2849.

[17] Caterino, M., Pastore, A., Strozziero, M.G., Di Giovamberardino, G., Imperlini, E., Scolamiero, E., Ingenito, L., Boenzi, S., Ceravolo, F., Martinelli, D., et al. (2015) The proteome of cblC defect: in vivo elucidation of altered cellular pathways in humans. *J. Inherit. Metab. Dis.*, **38**, 969-979.

[18] Hannibal, L., DiBello, P.M., Yu, M., Miller, A., Wang, S., Willard, B., Rosenblatt, D.S., Jacobsen, D.W. (2011) The MMACHC proteome: hallmarks of functional cobalamin deficiency in humans. *Mol. Genet. Metab.* **103**, 226-239.

[19] Akchiche, N., Bossenmeyer-Pourié, C., Kerek, R., Martin N., Pourié, G., Koziel, V., Helle, D., Alberto, J.M., Ortiou, S., Camadro, J.M., et al. (2012) Homocysteinylation of neuronal proteins contributes to folate deficiency-associated alterations of differentiation, vesicular transport, and plasticity in hippocampal neuronal cells. *FASEB. J.*, **26**, 3980-3992.

[20] Prudencio, M., Belzil, V.V., Batra, R., Ross, C.A., Gendron, T.F., Pregent, L.J., Murray, M.E., Overstreet, K.K., Piazza-Johnston, A.E., Desaro, P. (2015) Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat. Neurosci.,* **18**, 1175-1182.

[21] D'Erchia, A.M., Gallo, A., Manzari, C., Raho, S., Horner, D.S., Chiara, M., Valletti, A., Aiello, I., Mastropasqua, F., Ciaccia, L., et al. (2017) Massive transcriptome sequencing of human spinal cord tissues provides new insights into motor neuron degeneration in ALS. *Sci. Rep.*, **7**, 10046.

[22] Ridge, L.A., Mitchell, K., Al-Anbaki, A., Shaikh Qureshi, W.M., Stephen, L.A., Tenin, G., Lu, Y., Lupu, I.E., Clowes, C., Robertson, A., et al. (2017) Non-muscle myosin IIB (Myh10) is required for epicardial function and coronary vessel formation during mammalian development. *PLoS Genet.*, **13**, e1007068.

[23] Ma, X., Adelstein, R.S. (2014) A point mutation in Myh10 causes major defects in heart development and body wall closure. *Circ. Cardiovasc. Genet*., **7**, 257-265.

[24] Pashmforoush, M., Pomiès, P., Peterson, K.L., Kubalak, S., Ross J. Jr, Hefti, A., Aebi, U., Beckerle, M.C., Chien, K.R. (2001) Adult mice deficient in actinin-associated LIM-domain protein reveal a developmental pathway for right ventricular cardiomyopathy. *Nat. Med*., **7**, 591-597.

[25] Clark, T.G., Conway, S.J., Scott, I.C., Labosky, P.A., Winnier, G., Bundy, J., Hogan, B.L., Greenspan, D.S. (1999) The mammalian Tolloid-like 1 gene, Tll1, is necessary for normal septation and positioning of the heart. *Development,* **126**, 2631-2642.

[26] Stańczak, P., Witecka, J., Szydło, A., Gutmajster, E., Lisik, M., Auguściak-Duma, A., Tarnowski, M., Czekaj, T., Czekaj, H., Sieroń, A.L. (2009) Mutations in mammalian tolloid-like 1 gene detected in adult patients with ASD. *Eur. J. Hum. Genet*., **17**, 344-351.

[27] Kamdar F., Garry D.J. (2016) Dystrophin-Deficient Cardiomyopathy. *J. Am. Coll. Cardiol*., **67**, 2533-2546.

[28] Allen, D.G., Whitehead, N.P., Froehner, S.C. (2016) Absence of Dystrophin Disrupts Skeletal Muscle Signaling: Roles of Ca2+, Reactive Oxygen Species, and Nitric Oxide in the Development of Muscular Dystrophy. *Physiol. Rev*., **96**, 253-305.

[29] Hein, S., Kostin, S., Heling, A., Maeno, Y., Schaper, J. (2000) The role of the cytoskeleton in heart failure. *Cardiovasc. Res*., **45**,273-278.

[30] Communal, C., Colucci, W.S., Remondino, A., Sawyer, D.B., Port, J.D., Wichman, S.E., Bristow M.R., Singh, K. (2002) Reciprocal modulation of mitogen-activated protein kinases and mitogen-activated protein kinase phosphatase 1 and 2 in failing human myocardium. *J. Card. Fail.*, **8**, 86-92.

[31] Choi, J.C., Wu, W., Muchir, A., Iwata, S., Homma, S., Worman, H.J. (2012) Dual specificity phosphatase 4 mediates cardiomyopathy caused by lamin A/C (LMNA) gene mutation. *J. Biol. Chem*., **287**, 40513-40524.

[32] Son, N.H., Park, T.S., Yamashita, H., Yokoyama, M., Huggins, L.A., Okajima, K., Homma, S., Szabolcs, M.J., Huang, L.S., Goldberg, I.J. (2007) Cardiomyocyte expression of PPARgamma leads to cardiac dysfunction in mice. *J. Clin. Invest*., **117**, 2791-2801.

[33] Picard, F., Kurtev, M., Chung, N., Topark-Ngarm, A., Senawong, T., Machado De Oliveira, R., Leid, M., McBurney, M.W., Guarente, L. (2004) Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR-gamma. *Nature*, **429**, 771-776.

[34] Garcia, M.M., Guéant-Rodriguez, R.M., Pooya, S., Brachet, P., Alberto, J.M., Jeannesson, E., Maskali, F., Gueguen, N., Marie, P.Y., Lacolley, P., et al. (2011) Methyl donor deficiency induces cardiomyopathy through altered methylation/acetylation of PGC-1α by PRMT1 and SIRT1. *J. Pathol*., **225**, 324-335.

[35] Guéant, J.L., Namour, F., Guéant-Rodriguez, R.M., Daval J.L. (2013) Folate and fetal programming: a play in epigenomics? *Trends Endocrinol. Metab*., **24**, 279-289.

[36] Murai, K.K., Pasquale, E.B. (2004) Eph receptors, ephrins, and synaptic function. *Neuroscientist*., *10*, 304-314.

[37] Thaler, J.P., Koo, S.J., Kania, A., Lettier, K., Andrews, S., Cox, C., Jessell, T.M., Pfaff, S.L. (2004) A postmitotic role for Isl-class LIM homeodomain proteins in the assignment of visceral spinal motor neuron identity. *Neuron.*, **54,** 337-350.

[38] Pak, W., Hindges, R., Lim, Y.S., Pfaff, S.L., O'Leary, D.D. (2004) Magnitude of binocular vision controlled by islet-2 repression of a genetic program that specifies laterality of retinal axon pathfinding. *Cell*. **119**, 567-578.

[39] Miwa, J.M., Stevens, T.R., King, S.L., Caldarone, B.J., Ibanez-Tallon, I., Xiao, C., Fitzsimonds, R.M., Pavlides, C., Lester, H.A., Picciotto, M.R., et al. (2006) The prototoxin lynx1 acts on nicotinic acetylcholine receptors to balance neuronal activity and survival in vivo. *Neuron*, **51**, 587-600.

[40] Morishita, H., Miwa, J.M., Heintz, N., Hensch, T.K. (2010) Lynx1, a cholinergic brake, limits plasticity in adult visual cortex. *Science*, **330**,1238-1240.

[41] You, M.H., Kim, B.M., Chen, C.H., Begley, M.J., Cantley, L.C., Lee, T.H. (2017) Death-associated protein kinase 1 phosphorylates NDRG2 and induces neuronal cell death. *Cell Death Differ*. **24**, 238-250.

[42] Su, Y., Deng, M.F., Xiong, W., Xie, A.J., Guo, J., Liang, Z.H., Hu, B., Chen, J.G., Zhu, X., Man, H.Y., et al. (2019) MicroRNA-26a/Death-Associated Protein Kinase 1 Signaling Induces Synucleinopathy and Dopaminergic Neuron Degeneration in Parkinson's Disease. *Biol. Psychiatry.*, **85**, 769-781.

[43] Scotti M.M., Swanson M.S. (2016) RNA mis-splicing in disease. *Nat. Rev. Genet*., **17**, 19-32.

[44] Carvill, G.L., Heavin, S.B., Yendle, S.C., McMahon, J.M., O'Roak, B.J., Cook J., Khan, A., Dorschner, M.O., Weaver, M., Calvert, S., et al. (2013) Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat. Genet.*, **45**, 825-830.

[45] Talkowski, M.E., Mullegama, S.V., Rosenfeld, J.A., van Bon, B.W., Shen, Y., Repnikova, E.A., Gastier-Foster J., Thrush, D.L., Kathiresan, S., Ruderfer, D.M., et al. (2011) Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. *Am. J. Hum. Genet.,* **89**, 551-563.

[46] Osenberg S., Karten A., Sun J., Li J., Charkowick S., Felice C.A., Kritzer M., Nguyen M.V.C., Yu, P., Ballas N. (2018) Activity-dependent aberrations in gene expression and alternative splicing in a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E5363-E5372.

Legends to Figures

**Fig. 1** Identification and analysis of differentially expressed genes (DEGs) identified by RNA sequencing. **a** Venn diagram of differentially expressed genes in *cblC* and in *cblG* cells. **b** Gene ontology (GO) analysis of DEGs for significant biological process and molecular functions

Fig. 1

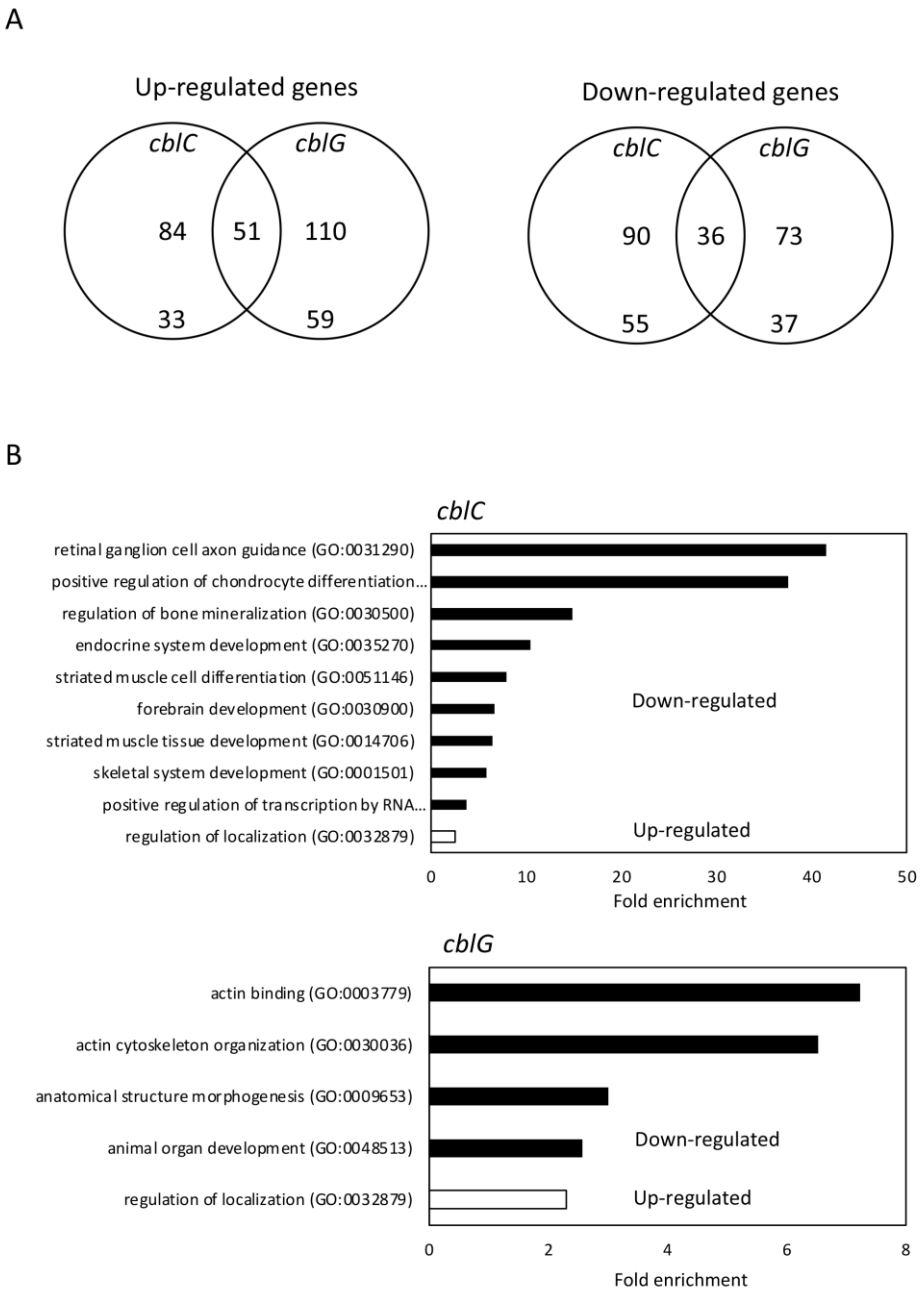

**Fig. 2** RT-qPCR validation of the DEGs. **a** Comparison of the fold change obtained by RNA-Seq and by RT-qPCR for a selection of 28 genes. **b** Expression of DMD and MYH1 assessed by ewtern blot and by immunostaining visualized by confocal microscopy.

Fig. 2

A

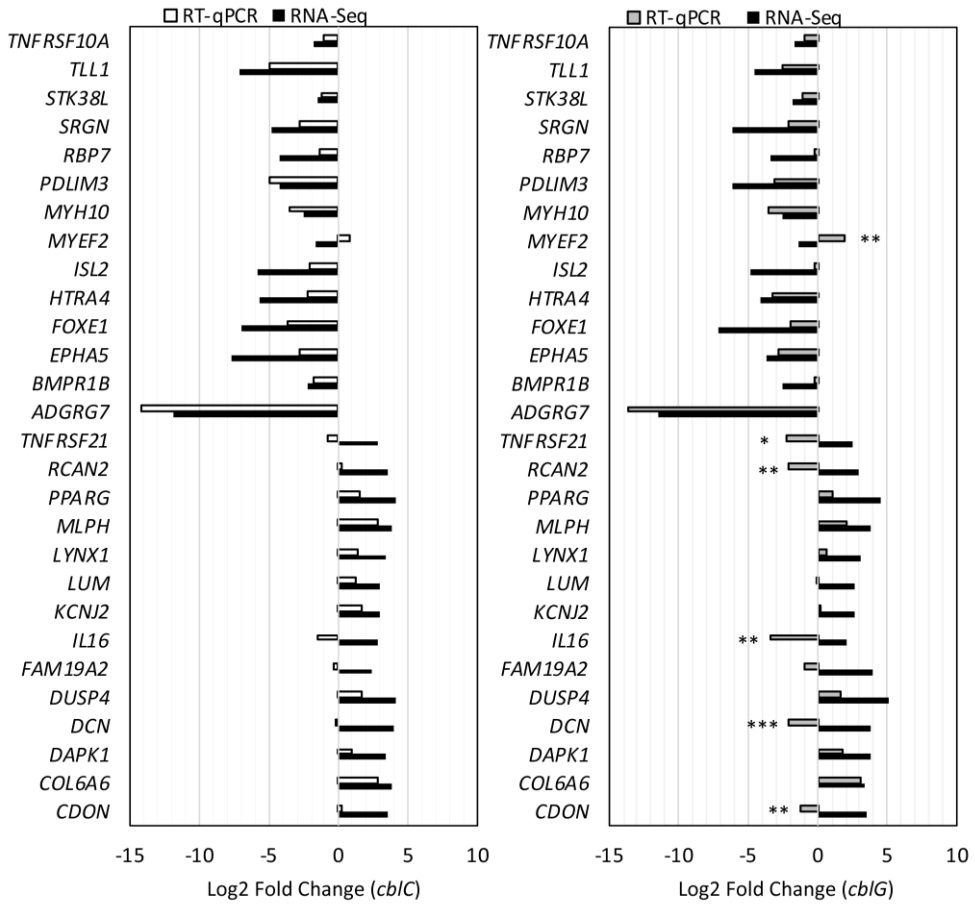

Fig. 2

B

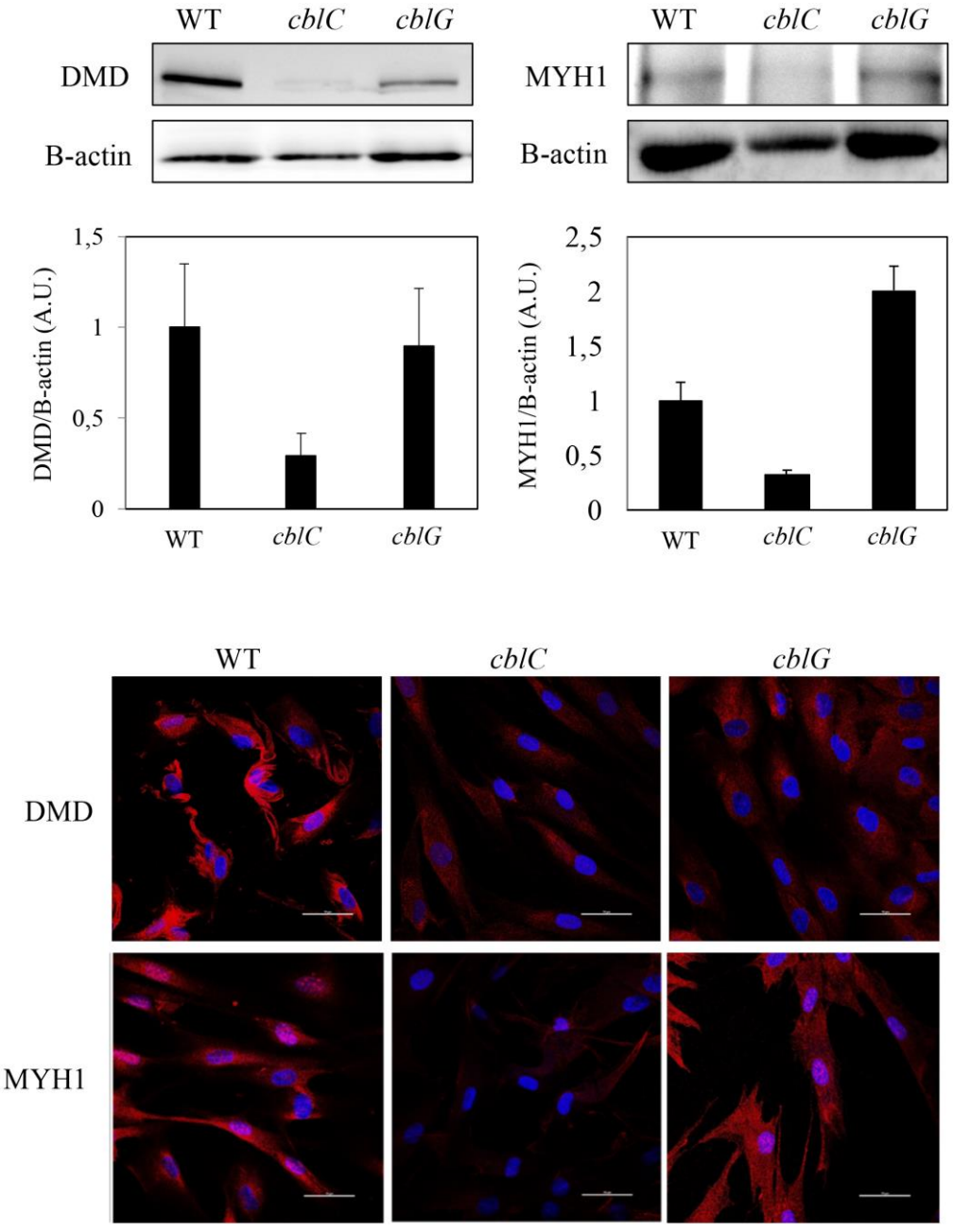

**Fig. 3** Alternative splicing (AS) analysis in *cblC* and *cblG* patient fibroblasts. **a** Violin plots representing the density of differential alternative splicing events found in *cblC* (N=3x3) and *cblG* (N=4x3) replicate samples compared to control (N=3) (rMATS analysis, FDR < 1% and delta-PSI > 5%). **b** Number and relative proportions of differential AS events found in patient cells compared to controls (rMATS analysis, Delta PSI >5%, FDR<0,01)

Fig. 3

A

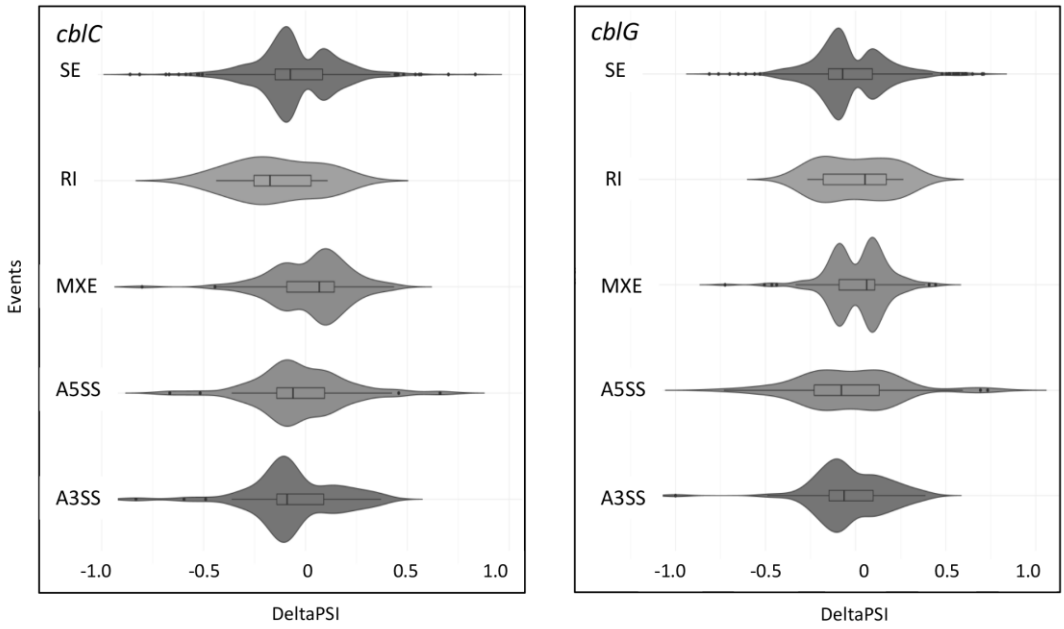

Fig. 3

B

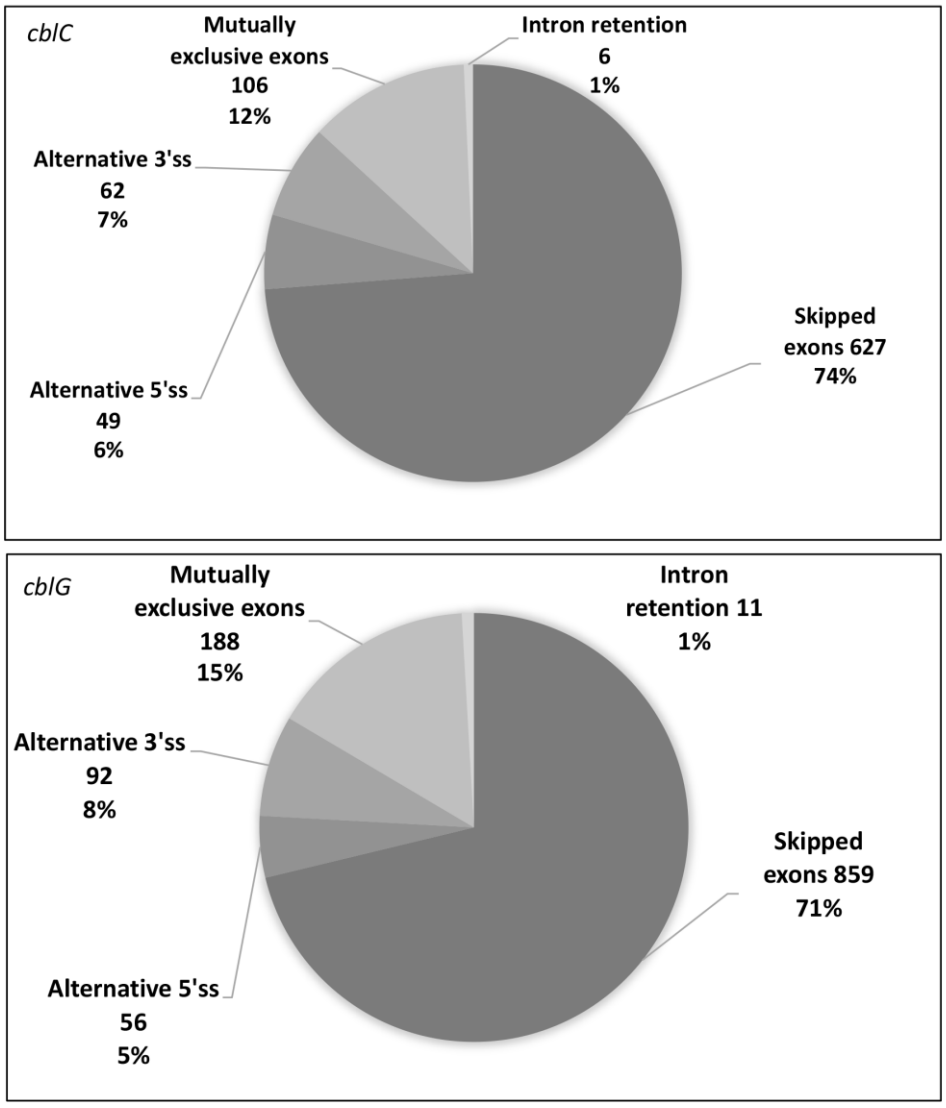

**Fig. 4** Functional annotations with relative fold-enrichment of the genes affected by differential alternative splicing (Panther analysis).

Fig. 4

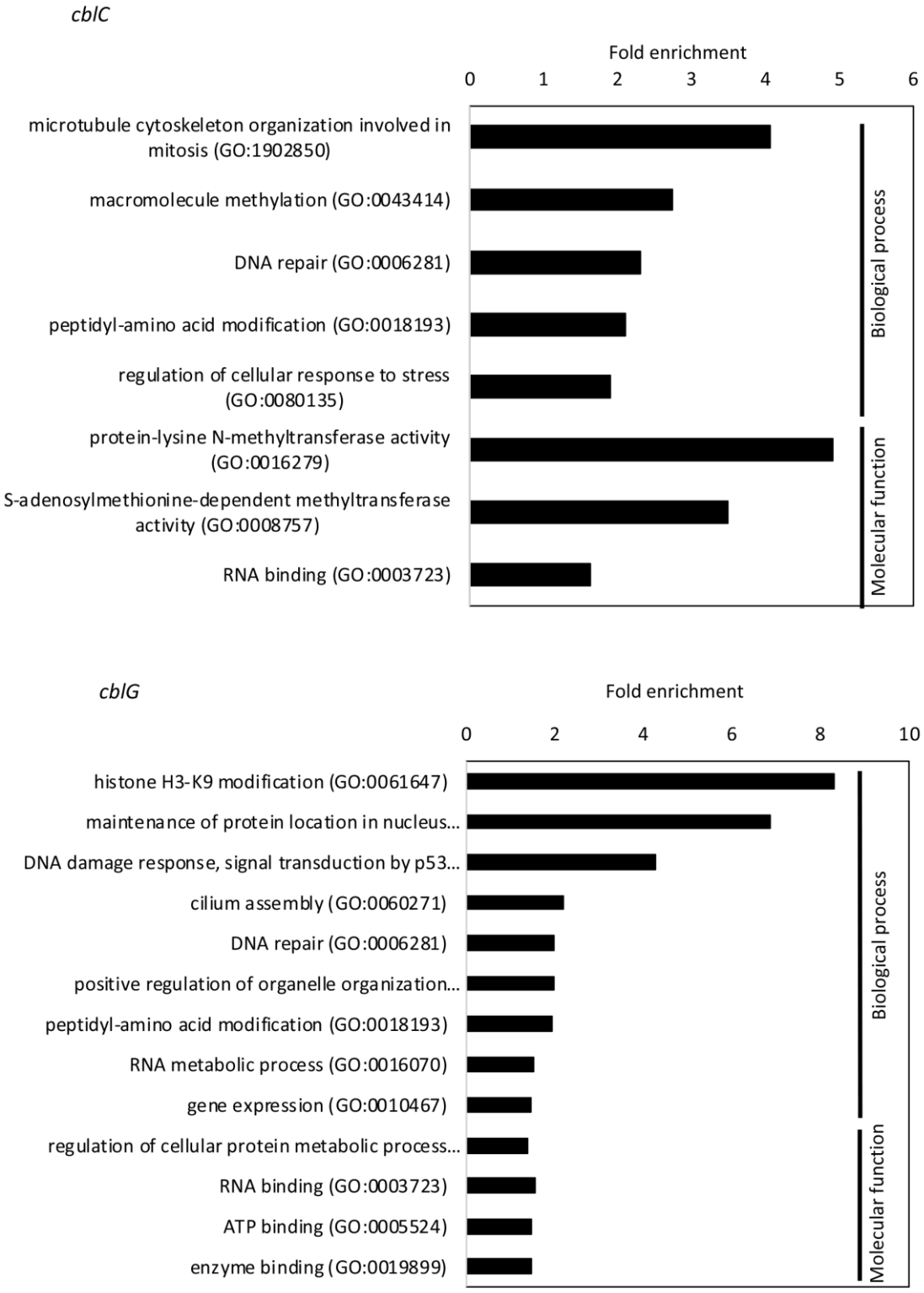

**Fig. 5** Experimental validation by RT-PCR of abnormal alternative splicing in *cblC* and *cblG* fibroblasts compared to control. The proportion of exon skipping was calculated by densitometric analysis of the bands (*: pvalue<0.1; **: pvalue<0.05; n=3).

Fig. 5

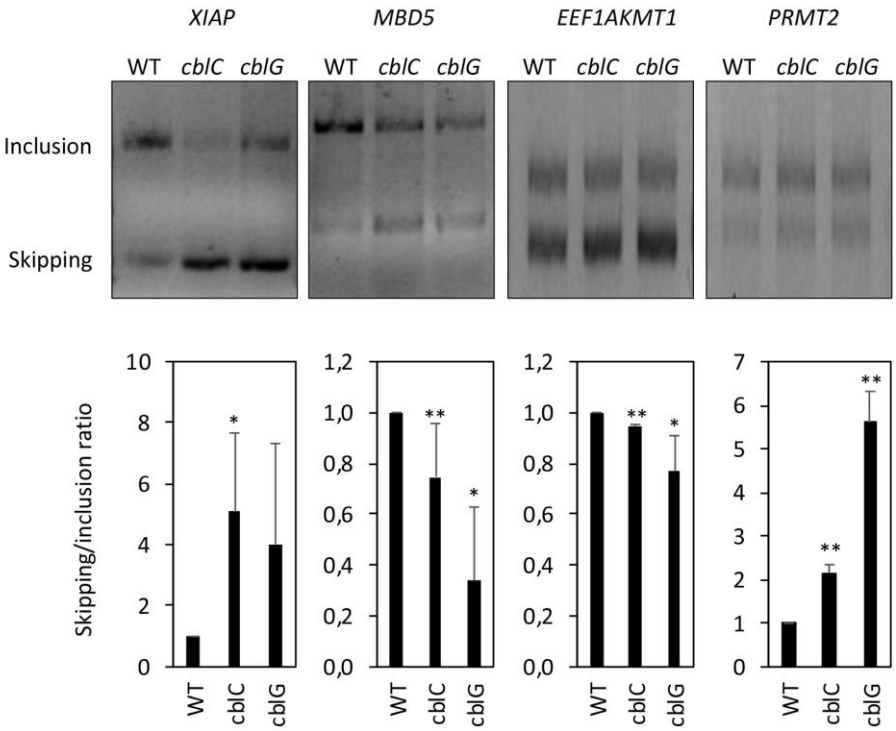

**Fig. 6** Identification of RNA Binding Proteins responsible for abnormal alternative splicing events in *cblC* and *cblG* fibroblasts. **a** Number of consensus sequences and associated RNA binding proteins significantly enriched in differential alternative splicing events identified in *cblC* and in *cblG* cells. **b** Enrichment of RBP targets identified in differential alternative splicing events.

Fig. 6

A

Number of consensus sequences and associated RBP
(log(p)>2,8; Enrichment >2)

| Events | *cblC* | | *cblG* | | Common *cblC-cblG* |
|---|---|---|---|---|---|
| | Consensus sequences | RBP | Consensus sequences | RBP | RBP |
| SE | 35 | 32 | 23 | 23 | 14 |
| A5SS | 4 | 3 | 7 | 7 | 0 |
| A3SS | 9 | 7 | 8 | 8 | 2 |
| MXE | 17 | 9 | 10 | 5 | 0 |
| RI | 0 | 0 | 1 | 1 | 0 |
| Total | 66 | 43 | 49 | 44 | 25 |

B

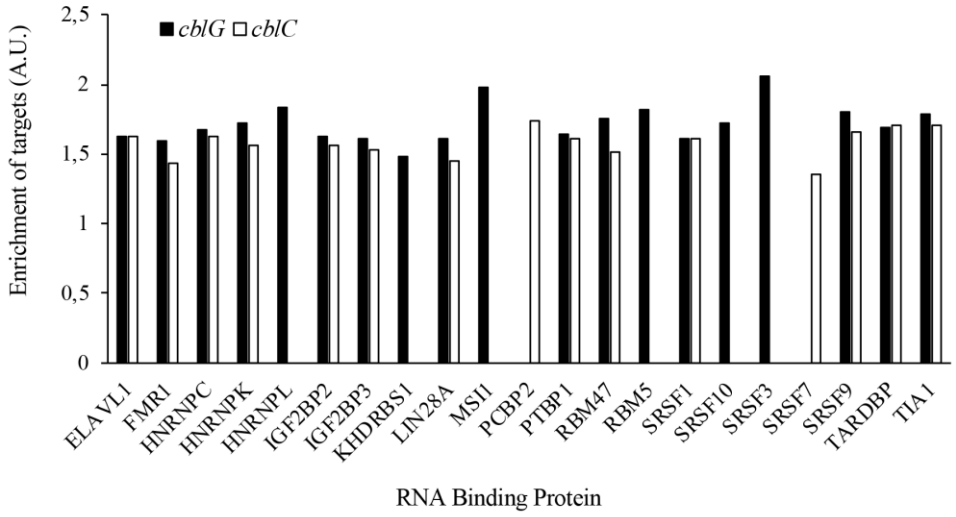

**Fig. 7** Expression of HNRNPL in *cblC* and *cblG* fibroblasts compared to control. **a** Western blot of HNRNPL and densitometric analysis expression (n=3). Beta-actin was used as an internal control of protein expression. **b** Immunofluorescence staining of HNRNPL. Fibroblasts were stained with anti-HNRNPL and DAPI and visualized by confocal microscopy. Intensity of the staining was measured using Image J freeware (n=3).

Fig. 7

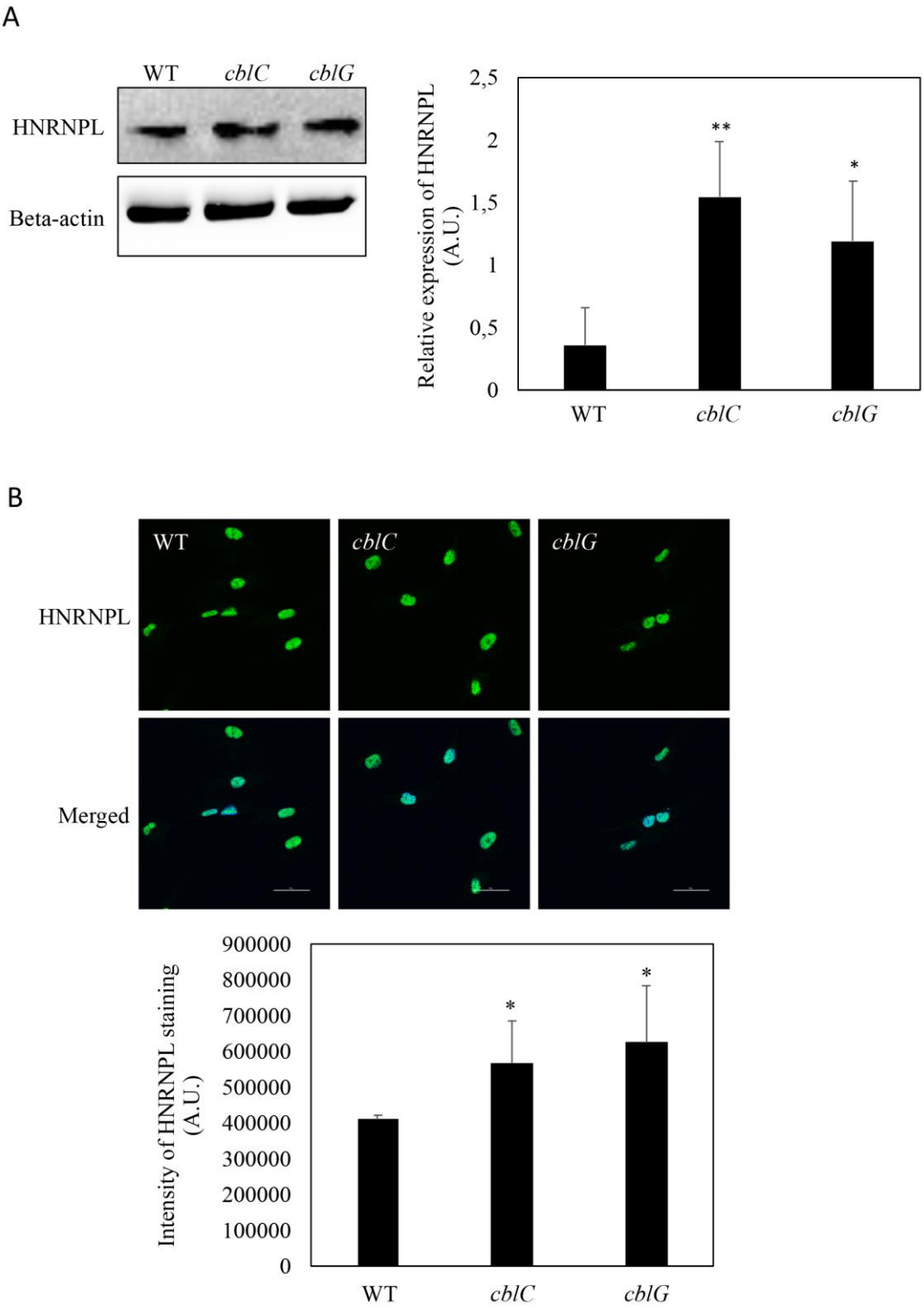

**Table 1** Top20 of under and over-expressed genes common to *cblC* and *cblG* groups, ranked by fold change (FC)

| Gene | Log2 FC | | p-value | | FDR | | |
|---|---|---|---|---|---|---|---|
| | *cblC* | *cblG* | *cblC* | *cblG* | *cblC* | *cblG* | Functions |
| *ADGRG7* | -11,95 | -11,48 | 5,45E-17 | 2,63E-17 | 9,17E-13 | 4,42E-13 | G protein-coupled receptor |
| *EPHA5* | -7,72 | -3,71 | 9,46E-06 | 4,86E-04 | 3,38E-03 | 4,23E-02 | Protein-tyrosine kinase (nervous system development) |
| *TLL1* | -7,1 | -4,52 | 2,61E-05 | 1,96E-04 | 7,41E-03 | 2,46E-02 | Zinc-dependent, metalloprotease (development of mammalian heart) |
| *FOXE1* | -7,00 | -7,18 | 1,14E-07 | 3,55E-08 | 1,60E-04 | 3,73E-05 | Transcription factor (cleft palate, thyroid morphogenesis) |
| *FAM9C* | -6,43 | -4,02 | 2,63E-06 | 2,81E-06 | 1,44E-03 | 1,21E-03 | Anti-apoptotic nuclear protein |
| *ISL2* | -5,87 | -4,9 | 1,48E-08 | 1,99E-08 | 2,49E-05 | 2,56E-05 | Transcription factor (nervous system ) |
| *HTRA4* | -5,67 | -4,06 | 5,29E-07 | 1,50E-06 | 4,88E-04 | 7,21E-04 | Serine protease (degradation of misfolded proteins) |
| *AP001999.1* | -5,5 | -3,84 | 7,37E-06 | 1,56E-05 | 3,03E-03 | 4,85E-03 | Long Intergenic Non-Protein Coding RNA |
| *CCDC144NL* | -5,37 | -5,2 | 8,09E-09 | 3,98E-09 | 1,51E-05 | 7,43E-06 | Coiled-coil domain-containing protein (diabetic retinopathy) |
| *SRGN* | -4,88 | -6,17 | 2,25E-05 | 3,42E-06 | 6,52E-03 | 1,44E-03 | Proteoglycan (inflammatory response) |
| *SAMD5* | 3,54 | 4,82 | 5,57E-06 | 1,59E-08 | 2,55E-03 | 2,22E-05 | Eph receptor binding protein (nervous system development) |
| *RCAN2* | 3,56 | 2,98 | 3,26E-04 | 2,67E-04 | 3,37E-02 | 2,95E-02 | Regulator of calcineurin 2 |
| *CDON* | 3,62 | 3,49 | 1,57E-04 | 3,91E-05 | 2,16E-02 | 9,25E-03 | Multifunctional co-receptor (myogenesis, neurogenesis) |
| *TRPV2* | 3,7 | 3,82 | 7,18E-05 | 2,91E-04 | 1,37E-02 | 3,13E-02 | Stretch-sensitive Ca2+-permeable channel (cardiomyopathy and muscular dystrophy) |
| *COL6A6* | 3,85 | 3,35 | 1,53E-04 | 1,23E-04 | 2,16E-02 | 1,93E-02 | Collagen Type VI Alpha 6 |
| *MLPH* | 3,86 | 3,79 | 2,01E-04 | 9,30E-05 | 2,59E-02 | 1,59E-02 | Melanophilin |
| *DCN* | 4,01 | 3,74 | 1,04E-06 | 5,99E-07 | 7,93E-04 | 3,59E-04 | Small leucine-rich proteoglycan (corneal dystrophy, myogenesis) |
| *DUSP4* | 4,08 | 5,05 | 8,54E-06 | 2,03E-07 | 3,26E-03 | 1,55E-04 | Dual specificity phosphatase (cardiomyopathy) |
| *PPARG* | 4,10 | 4,47 | 1,90E-07 | 6,53E-09 | 2,28E-04 | 9,98E-06 | Nuclear receptor, regulator of adipocyte differentiation (cardiac dysfunction) |
| *PRXL2A* | 4,49 | 4,61 | 2,95E-04 | 1,09E-05 | 3,23E-02 | 3,73E-03 | Peroxiredoxin Like (redox regulatory protein) |

**Table 2** Number of differential alternative splicing (AS) events (Delta PSI >5%; FDR<0,01) and affected genes (rMATS analysis)

| Events | Number of differential AS events | | Number of genes affected | | Genes in common |
|---|---|---|---|---|---|
| | *cblC* | *cblG* | *cblC* | *cblG* | |
| SE | 627 | 859 | 517 | 686 | 244 |
| A5SS | 49 | 56 | 48 | 55 | 16 |
| A3SS | 62 | 92 | 61 | 87 | 26 |
| MXE | 106 | 188 | 87 | 168 | 27 |
| RI | 6 | 11 | 6 | 11 | 1 |
| Total | 850 | 1206 | 719 | 1007 | 314 |

**Table 3** Top20 genes affected by events of alternative splicing in *cblC* and *cblG* groups (ranked by Delta-PSI)

| Event *cblC* | Gene | p-value | FDR | Delta PSI (WT-*cblC*) | Event *cblG* | Gene | p-value | FDR | Delta PSI (WT-*cblG*) |
|---|---|---|---|---|---|---|---|---|---|
| SE | *TVP23C-CDRT4* | 0,00E+00 | 0,00E+00 | 0,834 | A5SS | *ULK3* | 0,00E+00 | 0,00E+00 | 0,733 |
| A3SS | *XRCC4* | 0,00E+00 | 0,00E+00 | 0,832 | A5SS | *FAAP20* | 3,51E-07 | 5,08E-05 | 0,731 |
| MXE | *TVP23C-CDRT4* | 0,00E+00 | 0,00E+00 | 0,802 | MXE | *TVP23C-CDRT4* | 0,00E+00 | 0,00E+00 | 0,725 |
| SE | *DRAM2* | 0,00E+00 | 0,00E+00 | 0,702 | SE | *TVP23C-CDRT4* | 0,00E+00 | 0,00E+00 | 0,711 |
| A5SS | *C8orf59* | 0,00E+00 | 0,00E+00 | 0,666 | SE | *DRAM2* | 0,00E+00 | 0,00E+00 | 0,702 |
| A5SS | *ULK3* | 0,00E+00 | 0,00E+00 | 0,661 | A5SS | *TTLL3* | 2,20E-11 | 9,95E-09 | 0,693 |
| A3SS | *ZNF655* | 1,06E-10 | 1,37E-07 | 0,596 | SE | *NDUFB2* | 0,00E+00 | 0,00E+00 | 0,649 |
| SE | *USP53* | 0,00E+00 | 0,00E+00 | 0,568 | SE | *GUSBP11* | 2,48E-08 | 5,07E-06 | 0,615 |
| SE | *DRAM2* | 0,00E+00 | 0,00E+00 | 0,559 | SE | *USP28* | 1,52E-05 | 1,17E-03 | 0,611 |
| SE | *ALDH3A2* | 9,66E-06 | 8,27E-04 | 0,54 | SE | *AP000347.1* | 2,70E-08 | 5,46E-06 | 0,61 |
| SE | *TVP23C* | 0,00E+00 | 0,00E+00 | -0,861 | A3SS | *XRCC4* | 0,00E+00 | 0,00E+00 | -1 |
| SE | *TVP23C-CDRT4* | 0,00E+00 | 0,00E+00 | -0,814 | SE | *GHR* | 0,00E+00 | 0,00E+00 | -0,812 |
| SE | *ZNF561* | 4,26E-10 | 1,82E-07 | -0,685 | SE | *TVP23C* | 0,00E+00 | 0,00E+00 | -0,761 |
| SE | *DMKN* | 4,29E-14 | 3,77E-11 | -0,669 | SE | *TVP23C-CDRT4* | 0,00E+00 | 0,00E+00 | -0,699 |
| SE | *ECHDC2* | 3,53E-09 | 1,12E-06 | -0,622 | SE | *LIMS2* | 0,00E+00 | 0,00E+00 | -0,651 |
| SE | *CROCCP2* | 8,40E-10 | 3,17E-07 | -0,587 | SE | *AP000347.2* | 4,74E-06 | 4,57E-04 | -0,61 |
| SE | *DMKN* | 4,83E-11 | 2,66E-08 | -0,564 | SE | *CCDC74B* | 6,12E-05 | 3,45E-03 | -0,561 |
| SE | *ZNF45* | 6,32E-07 | 8,57E-05 | -0,53 | SE | *CCDC74B* | 6,81E-05 | 3,74E-03 | -0,531 |
| SE | *SNX11* | 1,11E-05 | 9,22E-04 | -0,524 | SE | *CROCCP2* | 6,17E-06 | 5,71E-04 | -0,515 |
| SE | *CROCCP2* | 8,52E-10 | 3,19E-07 | -0,523 | SE | *SENP7* | 9,35E-12 | 5,08E-09 | -0,505 |

**Table 4** Enrichment of HuR and HNRNPL in the genes differentially transcribed, spliced and expressed compared to controls (pvalues)

| | HuR | | HNRNPL | |
|---|---|---|---|---|
| | *cblC* | *cblG* | *cblC* | *cblG* |
| Transcription level | 1.08 (0.57) | 1.14 (0.70) | 1.06 (0.86) | 0.79 (0.53) |
| Alternative splicing | 1.62 (3.85 $10^{E}$-11) | 1.63 (6.86 $10^{E}$-23) | 1.74 (7.96 $10^{E}$-2) | 1.83 (1 $10^{E}$-20) |
| Protein expression | 1.51 (5.51 $10^{E}$-4) | 1.55 (1.02 $10^{E}$-3) | 1.22 (0.51) | 1.53 (0,17) |

Abbreviations: Cbl: cobalamin; RBP: RNA bonding protein; AdoCbl: adenosylcobalamin; MeCbl: methylcobalamin; DEG: differentially expressed gene; AS: alternative splicing