HMG Advance Access published February 13, 2013

Human Molecular Genetics, 2013 **1–10**
doi:10.1093/hmg/ddt035

Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders

Lam S. Nguyen[1], Hyung-Goo Kim[2,5], Jill A. Rosenfeld[3], Yiping Shen[2], James F. Gusella[2], Yves Lacassie[4], Lawrence C. Layman[5], Lisa G. Shaffer[3,†] and Jozef Gécz[1,6,*]

[1]School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, SA 5006, Australia, [2]Centre for Human Genetic Research, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, Boston, MA 02114, USA, [3]Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA 99207, USA, [4]Division of Genetics, Department of Paediatrics, Louisiana State University, Health Science Center and Children's Hospital, New Orleans, LA 70118, USA, [5]Department of Obstetrics and Gynaecology, Georgia Regents University, Augusta, GA 30912, USA and [6]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, SA 5006, Australia

Received August 15, 2012; Revised January 21, 2013; Accepted January 28, 2013

The nonsense-mediated mRNA decay (NMD) pathway functions not only to degrade transcripts containing premature termination codons (PTC), but also to regulate the transcriptome. *UPF3B* and *RBM8A*, important components of NMD, have been implicated in various forms of intellectual disability (ID) and Thrombocytopenia with Absent Radius (TAR) syndrome, which is also associated with ID. To gauge the contribution of other NMD factors to ID, we performed a comprehensive search for copy number variants (CNVs) of 18 NMD genes among individuals with ID and/or congenital anomalies. We identified 11 cases with heterozygous deletions of the genomic region encompassing *UPF2*, which encodes for a direct interacting protein of UPF3B. Using RNA-Seq, we showed that the genome-wide consequence of reduced expression of *UPF2* is similar to that seen in patients with *UPF3B* mutations. Out of the 1009 genes found deregulated in patients with *UPF2* deletions by at least 2-fold, majority (95%) were deregulated similarly in patients with *UPF3B* mutations. This supports the major role of deletion of *UPF2* in ID. Furthermore, we found that four other NMD genes, *UPF3A*, *SMG6*, *EIF4A3* and *RNPS1* are frequently deleted and/or duplicated in the patients. We postulate that dosage imbalances of these NMD genes are likely to be the causes or act as predisposing factors for neuro-developmental disorders. Our findings further emphasize the importance of NMD pathway(s) in learning and memory.

INTRODUCTION

ID is a broad term encompassing a clinically and genetically highly heterogeneous group of disorders characterized primarily by impaired learning and memory (1). ID affects as much as 3% of the population of resource rich countries and represents a major health burden (1). ID can be caused by point mutations, translocations, and structural variations, which contribute to 10–20% of cases (2). Recently, it has been shown that many ID genes encode for members of interacting proteins belonging to a small number of molecular networks, suggesting that many ID conditions share common underlying etiology (3, 4).

Two members of the NMD pathway, *UPF3B* and *RBM8A*, have been linked to ID. Complete loss of function mutations in *UPF3B* (Mendelian Inheritance in Man, MIM 300298), an X-linked gene, cause various forms of ID, autism, schizophrenia and attention deficit hyperactivity disorder (MIM 300676)

*To whom correspondence should be addressed at: Neurogenetics Laboratory, The University of Adelaide at the Women's and Children's Hospital, 72 King William Road, North Adelaide, South Australia 5006, Australia. Tel: +61 8 8161 6339; Fax: + 61 8 8161 7342; Email: jozef.gecz@adelaide.edu.au
†Current address: Paw Print Genetics, Genetic Veterinary Sciences, Inc., Spokane, WA 99202, USA

(5). Inter- as well as intra-familial variability in clinical expressivity is emerging as one of the main features of the loss of *UPF3B* function (5, 6). More recently, TAR syndrome (MIM 274000), has been described to be due primarily to compound heterozygous null (deletion) mutation and a regulatory single nucleotide polymorphism (SNP) of the *RBM8A* gene (MIM 605313) (7). About 7% of TAR cases are associated with ID in which cerebral dysgenesis might be the underlying cause (8). Micro-deletions and micro-duplications of the proximal 1q21.1 region, which contains *RBM8A*, are also frequently associated with ID, brain malformations and congenital anomalies without hallmarks of the TAR syndrome (8).

NMD functions to degrade transcripts containing PTCs and to regulate the transcriptome. In metazoan, the UPF3B protein is associated with the exon junction complex (EJC), which contains RBM8A and is deposited at the exon-exon junctions of the mRNA during splicing. When the EJC bound mRNA is exported to the cytoplasm, it acquires UPF2 (MIM 605529) via direct interaction with UPF3B. In the pioneer round of translation, if there is a PTC upstream of the last EJC, the ribosome stalls at the PTC; this triggers the formation of downstream complexes, which includes UPF1 (MIM 601430), and recruits exo- and endonucleases to degrade the PTC containing transcript. NMD acts to ensure that potentially deleterious truncated proteins are not made from erroneous transcription (9).

Strong evidence also supports a major role of NMD in regulating the expression of normal transcripts. Up to 5% of the transcriptome was reported to be deregulated in HeLa cell line depleted of UPF1, UPF2 or UPF3 proteins (10–13) and in lymphoblastoid cell lines (LCLs) of patients with complete loss of function mutations in *UPF3B* (14). Among the deregulated genes, direct physiological targets of NMD were suggested to contain general features that may introduce a stop codon into the open reading frame in a context that is sufficient to initiate NMD (15). These features include alternative splicing that causes frameshift and introduces PTC (16, 17), upstream open reading frame (18), intron in the 3' untranslated region (19), leaky translation due to differential usage of internal ribosomal entry sites (20), amino acid encoding for selenium (21) and long 3' untranslated region (22). It is thought that NMD acts to regulate expression of the transcripts necessary for cellular biogenesis in a tissue dependent and temporal specific manner (23, 24).

The function and development of the brain is particularly sensitive to dosage of NMD and EJC factors. Perhaps the most significant supporting evidence comes from the aforementioned patients with *UPF3B* mutations, and to a lesser extent, TAR patients carrying mutations in *RBM8A* and those with micro-duplications and deletions of the 1q21.1 proximal region. Many of the genes found deregulated in lymphoblastoid cell lines (LCLs) of patients with *UPF3B* mutations have important neuronal functions (14). It has been reported that UPF3A, a protein paralog of UPF3B, was stabilized in the patients with loss of UPF3B function and able to rescue some NMD function. Importantly, high level of UPF3A was suggested to reduce the severity of the neurological phenotypes associated with loss of *UPF3B* (12, 14). Studies using animal models have shown that depletion of NMD and EJC factors resulted in brain malformation or early embryonic lethality. In fruit flies, deletion of NMD genes *smg1, upf2* or *smg6* disrupted the formation of the neuromuscular junction synapse structure, reduced neurotransmission responses, and reduced synaptic vesicle cycling (25). In zebrafish, down regulating either one of these three genes resulted in aberrant eye and brain patterning and substantial increase in faulty somitogenesis and mortality rates (26). Haploinsufficiency of *Magoh*, an EJC component, caused microcephaly in mice (27). Depletion of the EJC factor eIF4A3 enhanced the synaptic strength of neurons in rat (28). Targeted disruption of either of *Upf1* or *Upf2* leads to embryonic lethality in mice (29, 30). Taken together, it is plausible that mutations in other NMD members in man could result in ID and other overlapping clinical features as found in patients with mutations in *UPF3B* and *RBM8A*.

In this study, we investigated the contribution of CNVs encompassing 18 known NMD and EJC genes (31, 32) in a large cohort of patients with various neuro-developmental disorders. We identified CNVs encompassing *UPF2*, *UPF3A* (MIM 605530), *SMG6* (MIM 610963), *EIF4A3* (MIM 608546) and *RNPS1* (MIM 606447), to be highly associated with these conditions. Using RNA-Seq, we provide further evidence that NMD is compromised in LCLs of patients with *UPF2* deletions in a similar manner to that of patients with *UPF3B* mutations. This overlapping expression profile implicates CNVs of *UPF2* in disease pathology and also points to a common molecular mechanism of NMD gene associated neuro-developmental disorders.

RESULTS

Enrichment of CNVs encompassing NMD and EJC factors in patients with neuro-developmental disorders

We assessed CNVs encompassing 18 known NMD and EJC genes to ID. To achieve this, we mined available CNV data in 57 356 patients with neuro-developmental disorders and 20 474 controls in large collated databases (see Materials and Methods; Supplementary Material, Table S1 and S2). We found that copy number losses of *UPF2* ($n = 11$), *UPF3A* ($n = 40$), *RBM8A* ($n = 50$) and copy number gains of *UPF2* ($n = 13$), *SMG6* ($n = 31$), *RBM8A* ($n = 101$), *EIF4A3* ($n = 22$) and *RNPS1* ($n = 25$) were significantly enriched in the patient cohort ($P < 0.05$ by Fisher's exact test, Table 1). In 89 cases where the mode of inheritance could be tested, the CNVs were mostly *de novo* (54 cases) or were transmitted from an apparently normal parent (35 cases) (Supplementary Material, Table S3). The CNVs were variable in size and contained other genes, which could also contribute to the patients' phenotypes. Patients carrying these CNVs presented with various clinical phenotypes which all included ID, brain malformations and/or developmental anomalies. CNVs encompassing *SMG5* (MIM 610962, 9 gains), *SMG6* (23 losses), *SMG7* (MIM 610964, 4 gains, 9 loses) *SMG8* (MIM 613175, 11 gains, 2 losses), *SMG9* (MIM 613176, 3 gains, 3 losses), *MAGOH* (MIM 602603, 2 gains, 2 losses), and *CASC3* (MIM 606504, 1 gain, 1 loss) were also more frequently observed in the patients than controls; however, these differences were not statistically significant (Table 1). Our findings suggested that both duplications and deletions of NMD and EJC factor genes could contribute to disease pathology.

Table 1. Dosage imbalances of NMD genes are frequently associated with ID. Genomic coordinates are in HG19. *P*-values were calculated using Fisher's exact test and are only displayed for genes with significant enrichment in the patient cohorts ($P < 0.05$).

Group	Gene	Chr	Start	Stop	Cytoband	Patients (n = 57,365)		Controls (n= 20,474)		P Value		Associated Syndrome
						Gain	Loss	Gain	Loss	Gain	Loss	
NMD Factors	*UPF1*	chr19	18 942 744	18 979 039	19p13.2–19p13.11	1	0	3	0			
	UPF2	chr10	11 962 021	12 085 023	10p14-p13	13	11	0	0	0.018903	0.034805	DiGeorge syndrome 2
	UPF3B	chrX	118 967 989	118 986 968	Xq25-Xq26	4	5	0	1			Various forms of ID
	UPF3A	chr13	115 047 059	115 071 291	13q34	20	41	5	1	0.323101	0.000043	Neural tube defect
	SMG1	chr16	18 816 176	18 937 726	16p12.3	3	3	5	9			
	SMG5	chr1	156 219 015	156 252 620	1q21.2	9	0	0	1			
	SMG6	chr17	1 963 133	2 207 069	17p13.3	31	23	2	5	0.003373	0.215482	17p13.3 microduplication syndrome
	SMG7	chr1	183 441 634	183 523 326	1q25	4	9	0	0			
	SMG8	chr17	57 287 371	57 292 611	17q22	11	2	0	0			
	SMG9	chr19	44 235 301	44 259 142	19q13.31	3	3	0	0			
	NBAS	chr2	15 307 032	15 701 454	2p24	8	5	7	2			
	DHX34	chr19	47 852 546	47 885 961	19q13.3	4	3	2	2			
EJC factors	*RBM8A*	chr1	145 507 638	145 511 444	1q21.1	101	50	2	1	0.000001	0.000655	TAR, proximal 1q21 microduplication and deletion syndrome
	MAGOH	chr1	53 692 564	53 704 207	1p32.3	2	2	1	0			
	CASC3	chr17	38 296 507	38 328 431	17q11-q21.3	1	1	0	0			
	EIF4A3	chr17	78 109 017	78 120 938	17q25.3	22	0	0	1	0.001210	1.000000	Not known
	RNPS1	chr16	2 303 122	2 317 797	16p13.3	25	4	1	4	0.003671	0.966198	16p13.3 duplication syndrome
	WIBG/ PYM	chr12	56 295 197	56 321 697	12q13.2	2	1	0	0			

Deletions of *UPF2* result in reduced *UPF2* expression

To investigate the molecular consequences of at least some CNVs identified, we examined the effects of heterozygous deletions of *UPF2*. Eight females and three males with *UPF2* deletions showed developmental delay with various dysmorphic features, speech and language problems, and deafness (Supplementary Material, Table S3). In 5 of these 11 deletions, where the inheritance could be tested, the deletions were *de novo*. *UPF2* deletions were not seen in controls (Table 1). Three of the deletion cases DGDP061, DGDP151, and DGAP195 are described in greater detail in the Supplementary clinical information and pictured in Figure 1. We did not detect any mutation by Sanger sequencing on the remaining *UPF2* allele in these three cases, suggesting UPF2 function was likely to be partially compromised. The minimal overlapping region deleted in all 11 *UPF2* patients was 234 kb. This region contains full coding region of *C10ORF47,* part of the coding region of *ECHDC3,* and *UPF2* (Fig. 2A). The functions of the *C10ORF47* or *ECHDC3* genes have not yet been determined, however they have been excluded because neither gene is expressed in the brain as investigated by RNA-Seq (BrainSpan: Atlas of the Developing Human Brain, http://developinghumanbrain.org).

We postulated at least three reasons why the *UPF2* gene is likely involved in ID: i) *UPF2* is ubiquitously expressed in all tissues including the brain (33); ii) UPF2 acts as a bridge between UPF3B (known to cause ID, see reference (5)) and the main NMD protein complex (34); and iii) UPF2 is essential for NMD function and cell development as demonstrated by both *in-vitro* and *in-vivo* experiments (10, 30, 35, 36). In lymphoblastoid cell lines (LCLs) derived from two patients with *UPF2* deletions (DGAP195 and GM06936) whom we could get access to, the *UPF2* mRNA levels were at 50% (Fig. 2B), while the UPF2 protein levels were at 50–75% with respect to the controls (Fig. 2C).

Transcriptome-wide effects of reduced *UPF2* expression

We hypothesized that loss of one allele of *UPF2* would compromise NMD function similarly to the *UPF3B* mutations (5, 14). We sequenced polyA+ RNA extracted from LCLs of patients with *UPF2* deletions ($n = 2$) using RNA-Seq (Illumina GAII) (Fig. 3A). We generated 10 to 20 million reads per sample. We aligned the reads to the human genome (HG19) with CASAVA V1.6 (Illumina) and calculated gene expression as Reads Per Kilobase per Million mapped reads (RPKM) as previously described (14, 37). Since the patients with *UPF2* deletions are females and the controls and patients with *UPF3B* mutations used in subsequent comparisons are males, we excluded from the analysis all sex specific genes such as *Xist* and Y-linked genes; we also excluded genes known to be differently expressed between sexes in LCLs (38). Compared to the controls ($n = 2$), we identified 1009 differently expressed genes (DEGs) which were deregulated by at least 2-fold in the patients with *UPF2* deletions (40% of which were down regulated, 60% were upregulated; Supplementary Material, Table S4). This represents 10% of all expressed genes in LCL (expression threshold RPKM > 3). Thirty-eight percent of these DEGs (379 genes) overlapped with the DEGs deregulated also by 2-fold in the patients with *UPF3B* mutations using the same platform (RNA-Seq) and assessed against the same controls (Fig. 3A) (14). More importantly, ninety-five percent of

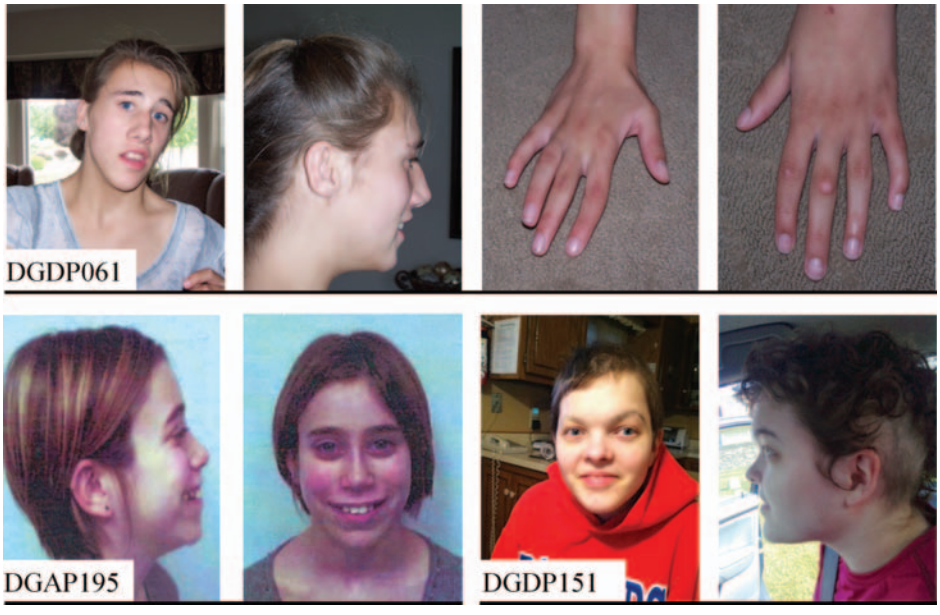

Figure 1. Clinical features of individuals with copy number loss leading to reduced expression of *UPF2*. Subject DGDP061 at 15 years; please note her mild facial dysmorphism and cupped ears, which were posteriorly rotated and slightly low set; she has clinodactyly of both 5th fingers. Subject DGAP195 at 12 years; she has developmental delay, ADHD, learning disabilities, major depression, and sensory problems. Subject DGDP151 at 15 years; she has a generalized triangular shape to her face and thin downslanting palpebral fissures; patches of missing hair are seen below behind her ear due to her chronic hair picking (trichotillomania).

all DEGs identified in either *UPF2* or *UPF3B* patients share the same trend of deregulation, i.e. either up or down regulation, with a high concordance level (Pearson's correlation efficiency $R = 0.79$ and 0.77, Fig. 3B and C, respectively).

To test whether the association in gene expression changes in patients with *UPF2* deletion and *UPF3B* mutations was random, we permutated lists containing a similar number of genes to the reported DEGs in the two patient groups. We then calculated the Pearson's correlation efficiency of gene expression differences, or fold changes with respect to the controls, between these two patient groups for each simulation (Supplementary Material, Fig. S1). After 100 simulations, the resulted R values (maximum 0.64, minimum 0.39 and average values 0.52) were much lower than the observed R values (0.79 and 0.77, Fig. 3B and C), suggesting that the observed correlation was true.

We selected 13 genes for further validation using reverse-transcribed quantitative real time PCR (RT-qPCR). Half of these genes were randomly selected, while the other half was chosen because they have been reported to be upregulated in patients with *UPF3B* mutations (5, 14). To test for reproducibility of the RNA-Seq results, we used RNAs extracted from different and extended sets of LCLs from the patients and controls (Fig. 4A). We showed that the majority of these genes (11 out of 13), including *GADD45B* and *ARHGAP24* isoform 1, which are known to be upregulated upon inhibition of NMD (5, 11), were similarly upregulated in both patient groups. Subsequently, we selected 3 genes *ROBO1*, *HNF1B* and *SERINC2* for validation using protein analyses (Fig. 4B and C). All three genes were shown to be upregulated by RNA-Seq (Fig. 4A and Supplementary Materials, Table S4). Western blot analyses were not as sensitive as RNA-Seq, but overall we were able to validate the trend of upregulation of ROBO1 and SERINC2 in *UPF2* and *UPF3B* patient groups with respect to the controls (Fig. 4C). We could only validate the upregulation of HNF1B in *UPF3B* patient group, but not *UPF2* patient group (Fig. 4C). In this respect, the gene expression changes identified by RNA-Seq in our study broadly reflected the changes of these proteins. These results were in agreement with a recent study which demonstrated that RNA-Seq and proteomic analysis had better concordance (than microarray and proteomic analysis), wherein 75% of genes and proteins were identified to have similar trend of expression changes (39). Taken together, our results led us to conclude that NMD was at least partially deficient in the patients with *UPF2* deletions, and that the transcriptome-wide consequences of compromised NMD in the patients with *UPF2* deletions and *UPF3B* mutations were very similar.

We hypothesized that the clinical presentations of ID in the patients with *UPF2* and *UPF3B* alterations were similar as a consequence of shared deregulated genes with known neuronal functions. Using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (40), we identified DEGs with known neuronal functions (Supplementary Material, Table S5). We only selected DEGs that were deregulated in both patient groups in the same trend, i.e. both up or down regulated, as identified by RNA-Seq (this study) and exon array from our previous study (14). These DEGs represent the most consistently deregulated genes when NMD is compromised and might be considered a priority in follow up functional studies in neuronal cell lines.

DISCUSSION

We have identified CNVs encompassing five NMD factors to be significantly associated with neuro-developmental disorders.

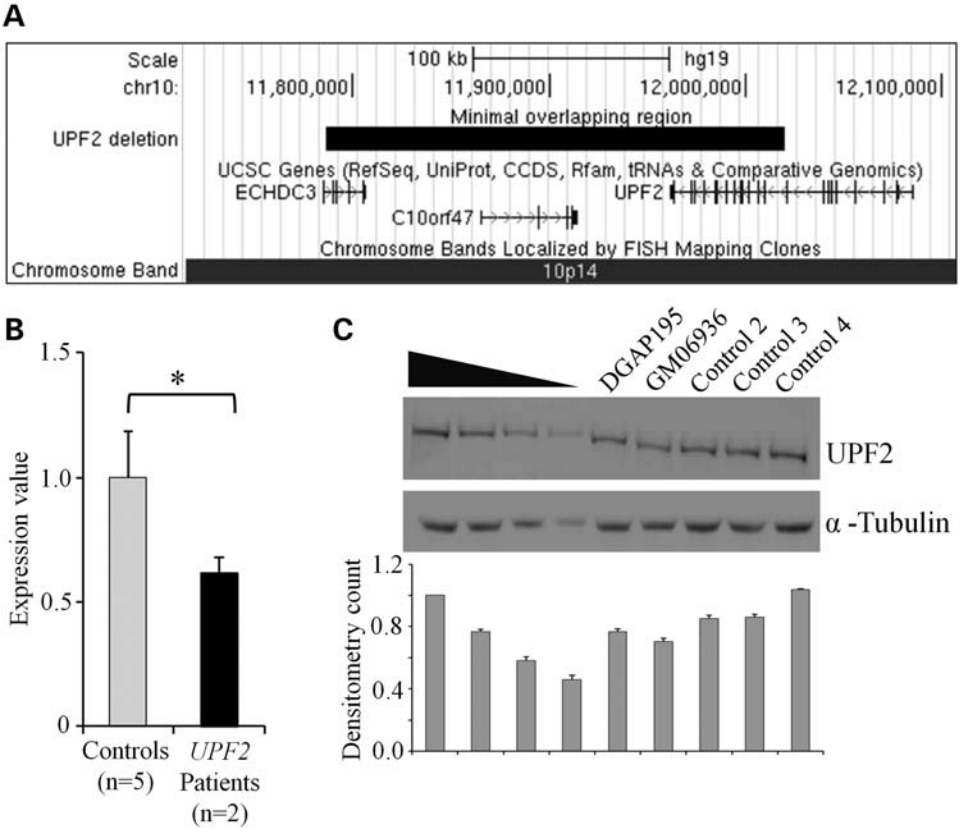

Figure 2. Deletions of *UPF2* result in reduced *UPF2* expression. (**A**) Minimal overlapping region deleted at 10p14 in 11 *UPF2* patients (snap shot from UCSC). (**B**) RT-qPCR expression of *UPF2* mRNA in the patients ($n = 2$) and controls ($n = 5$). Expression was measured using relative standard curve method and normalized against *ACTB* mRNA expression in the same sample. Results are representative of two independent experiments with the same results. $^{*}P < 0.05$ by Student's two-tailed t-test. (**C**) Western blot analysis of UPF2. Top panels: protein lysates from two patients and four controls are shown. Lysates from control 1 was loaded in decreasing amount in lanes 1–4 (100%, 75%, 50% and 25%, respectively) for comparison. Lysates from all other lanes were loaded as 100%. UPF2 expression is variable in the controls and is at 50–75% of normal level in the patients. α-Tubulin was used as loading control. These blots are representative of two independent experiments with similar results. Bottom panel: respective densitometry count showing relative UPF2 protein level in each sample. Densitometry ($\pm$ SD) was performed directly on the images captured at low exposure time points.

These CNVs likely drive or contribute to these conditions, including three known syndromes: *UPF3A*/neural tube defects (41), *SMG6*/17p13.3 microduplication syndrome (MIM 613215) and *RNPS1*/16p13.3 duplication syndrome (MIM 613458); and two novel pathogenic regions involving *UPF2*, which lies outside the critical region of DiGeorge Syndrome 2 (MIM 601362), and *EIF4A3* (Table 1). Interestingly, four of these gene loci are close to the subtelomeric regions (except *UPF3A* which lies within this region), which are rich in repeats and prone to rearrangements. The breakpoints of CNVs encompassing these genes also appear to be frequently within these regions (*UPF2* CNVs, 15 out of 24 events; *EIF4A3* CNVs, 16 out of 22 events) (Supplementary Materials, Table S3). This may offer an explanation for the relatively high frequency of CNVs reported here. Indeed, it is estimated that cryptic subtelomeric rearrangements collectively contribute to up to 10% of all ID cases (42, 43).

We further investigated the role of *UPF2* deletion in ID using RNA-Seq. We show that loss of one allele of *UPF2* resulted in deregulation of the transcriptome similarly to that of previously studied patients with ID and loss of *UPF3B* function (Fig. 3). Generally, the extent of transcriptome-wide effect is higher in the patients with single *UPF2* allele deletions than in those with *UPF3B* mutations. This could be attributed to the fact that UPF2 also has a role in maintaining the telomere integrity (44), and is involved in at least two branches of NMD, the classical NMD branch (11) and the UPF2/RNPS1 dependent branch (45). Alternatively, this could also be due to an effect of other contiguously deleted gene(s) from the *UPF2* region, or simply due to the small number of samples tested. Remarkably though, the expression profiles of the patients with single *UPF2* allele deletions and *UPF3B* mutations were similar, pointing to a common molecular pathway, i.e. NMD, being deregulated (Fig. 3). This also suggested that the deletion of a single *UPF2* allele had significant effects to the overall deregulation of the transcriptome.

The patients with *UPF2* deletions showed a broader phenotypic spectrum than those with *UPF3B* mutations. The hallmarks of the *UPF2* deletion associated phenotypes include significant speech delay, hearing impairment, joint contractures and reduction deformity of legs (no digits) (Supplementary Material, Table S3). The differences in clinical consequences of *UPF3B* and *UPF2* mutations may also be due to the fact that unlike for *UPF3B*, where the autosomal UPF3A paralog

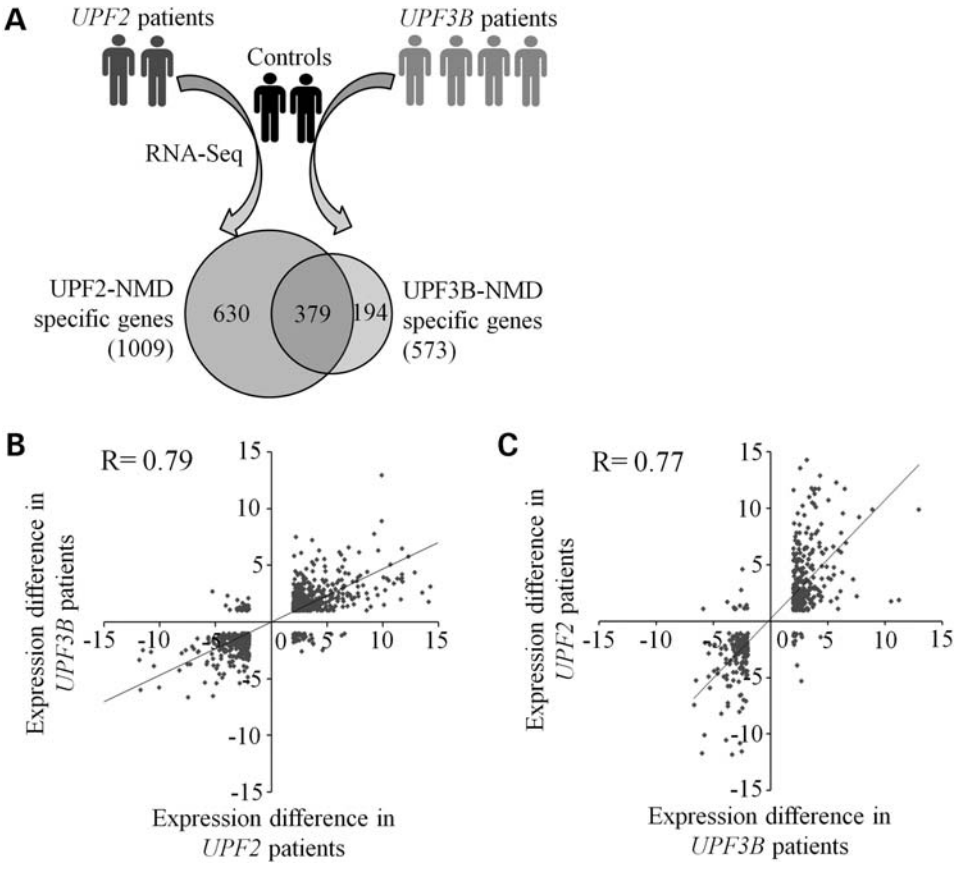

Figure 3. Transcriptome deregulation of patients with *UPF2* deletions is similar to that of patients with *UPF3B* mutations. (**A**) Overall design and outcome of the comparison. Compared to controls, 379 genes were deregulated by at least 2-fold in both patient groups. (**B**) Gene expression differences (or expression fold change with respect to the controls) of DEGs identified in the patients with *UPF2* deletions (1009 genes) and their corresponding values in the patients with *UPF3B* mutations. (**C**) Gene expression differences (or expression fold change with respect to the controls) of DEGs identified in the *UPF3B* patients (573 genes) and their corresponding values in the *UPF2* patients.

is able to rescue the loss of *UPF3B*-NMD function (14), there is no UPF2 paralog in human or other species. Interestingly, the mouse model with heterozygous deletion of *Upf2* is phenotypically normal with no apparent neuronal or learning and memory deficits (30) (Bo Porse, personal communication). In this respect, it is also important to note that the majority of NMD targets do not seem to be conserved between human and the *Upf2* knock out mouse tissues (14). This adds additional complexity when studying NMD across species.

We validated the association of CNVs encompassing *RBM8A* (both gains and losses) with neuro-developmental disorders in this study using an expanded cohort of patients and a different cohort of controls with regard to the previous report (8) (Table 1). There are some overlapping clinical features in these patients with TAR syndrome, including thinning of the corpus callosum and cysts (Supplementary Material, Table S3). RBM8A and other EJC constituents are in fact dispensable for NMD function. In fruit flies and nematodes, NMD does not require EJC (46–49). In man, NMD branch independent of the EJC factors RBM8A, MAGOH, CASC3 and EIF4A3 has also been described (45), although this branch was thought to be a minor NMD branch (50). Perhaps this is why partial loss of *RBM8A* may only act as a pre-disposing factor and requires a second hit (the regulatory SNP) for fully manifested TAR syndrome.

Copy number gains of *UPF2, SMG6, RBM8A, EIF4A3* and *RNPS1* were significantly enriched in ID patients (Table 1). This suggests that enhancement of NMD function may also be deleterious for normal neuronal development and function. In fact, overexpression of the NMD gene *SMG1* enhanced UPF1 phosphorylation and accelerated the decay rate of target transcripts (51); and the varying amount of RNPS1 in different HeLa strains has been shown to affect NMD efficiency (52). However, increasing gene dosage may not necessarily lead to enhanced NMD function, since limiting proteins in protein complexes may be titrated out by the over expressed factors, resulting in attenuation of NMD. Specifically, overexpression of upf1 in yeast abolished the requirement for upf2, indicating that high level of upf1 is sufficient to maintain NMD in the active state (53). In HeLa cells, NMD efficiency was enhanced upon over expression of *SMG5*, *SMG6* and *SMG7*, but not *UPF1*, *UPF2*, *UPF3A*, *UPF3B* and *SMG7* (23). Notably, the genes deregulated in HeLa cells depleted of NMD factors UPF1, UPF2 or UPF3 did not overlap much with those reported in LCLs of patients with UPF3B mutations (14). As such, the consequences of copy number gains of NMD and EJC factors in LCLs might be different and require further investigation. Taken together, our results suggest that CNVs encompassing NMD and EJC factor genes affect NMD function

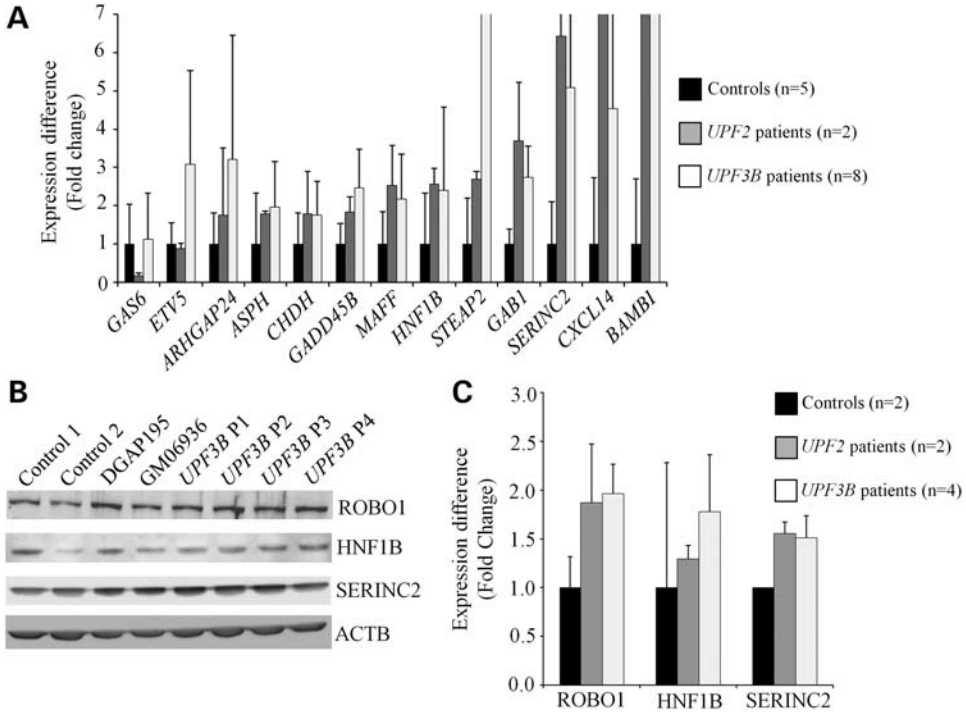

Figure 4. Validation of RNA-Seq analysis. (**A**) Validation of mRNA expression differences of 13 selected genes by RT-qPCR. Mean expression (±SD) of *GAS6, ETV5, ARHGAP24, ASPH, CHDH, GADD45B, MAFF, HNF1B, STEAP2, GAB1, SERINC2, CXCL14* and *BAMBI* in controls (black bar, $n = 5$), patients with *UPF2* deletions (gray bar, $n = 2$) and patients with *UPF3B* mutations (white bars, $n = 8$). Values were normalized to *ACTB* mRNA level in the same samples using a relative standard curve method and converted into fold change. Results are averaged from two independent RT-qPCRs. Table S6 contains the sequences of all primers used in this study. (**B**) Western blot analysis of ROBO1, HNF1B and SERINC2, which were all reported to be upregulated in *UPF2* and *UPF3B* patients by RNA-Seq. Expression of the proteins were detected using commercially available antibodies: ROBO1 (Santa Cruz, sc-25672), HNF1B (sc-7411) and SERINC2 (sc-138678). ACTB was used as a loading control. Image at low exposure time points were captured directly from the membrane using camera-based quantitative biomolecular imager (ImageQuant LAS 4000, Fujifilm). The blots shown are representative of two independent experiments. (**C**) Densitometry analysis of the western blots. Densitometry (±SD) was performed on images shown in (B) at two low exposure time points using Multi Gauge V.3 (Fujifilm). Values were normalized against signal obtained from ACTB bands from the same blot, and normalized against the controls to convert to fold change.

and result in a range of neurodevelopmental clinical presentations.

CNVs of some of the NMD gene loci were also found in apparently normal controls, but at a much lower frequency (Table 1). In some cases, the NMD CNVs were inherited from an apparently normal parent (Supplementary Material, Table S3), which is not unusual (54). However, the explanations for this apparent 'non-penetrance' may lie in the extent of individual's innate NMD efficiency, which may not go below the 'clinically relevant' threshold. Cumulative evidence points towards considerably variable NMD efficiency in the populations, which could be influenced by expression quantitative trait loci (55), competitive pathway (56) and feedback regulation (23, 57). It is also possible that some CNVs are simply non-pathogenic. Larger datasets and further investigations are required to address this complexity. Our results also demonstrate that NMD efficiency can be influenced by copy numbers as well as point mutations in the NMD and EJC genes. Such genetic variations can help to explain variable NMD efficiency which has been shown to modulate not only the phenotypic outcomes of genetic disorders involving NMD (58–60), but also patients' response to PTC read-through drugs (61).

Our work demonstrates a broader than expected involvement of NMD in normal brain development and function. We hope it will trigger future investigations into the role of NMD genes in a broad spectrum of neuro-developmental and other congenital disorders. Better understanding of the inter-individual NMD variation will also help to interpret the complexity of the human phenotype.

MATERIALS AND METHODS

CNV screening methodology by the Signature Genomics Laboratory

From March 2004 through November 2010, 45 744 probands were referred to the Signature Genomic Laboratories for clinical aCGH testing, with the most common clinical presentations being ID, developmental delay, and/or multiple congenital anomalies. Patients' DNAs were subjected to aCGH using either bacterial artificial chromosome (BAC) microarray (Signature Chip) or a whole genome oligonucleotide-based microarray (SignatureChip Oligo Solution). All abnormalities detected by aCGH were visualized by metaphase or interphase fluorescence *in situ* hybridization (FISH) using one or more BAC clones determined to be in the abnormal region by aCGH (62). The six versions of the BAC microarray had increasing coverage of the genome, and the array version used depended on the date of sample submission. Versions 1.0 through 4.0, used from March 2004 until November 2007, had targeted coverage

of the genome and were used to test 15 529 samples. The BAC-based SignatureChip WholeGenome (WG) versions 1.0 and 2.0, were used to test 8,475 and 732 samples, respectively, from November 2007 to December 2009. A comparison of the contents of versions 1.0 through 4.0 and WG version 1.0 can be found at http://www.signaturegenomics.com/clone_list.html. For all BAC platforms, DNA extraction, labeling, and microarray analysis were performed as previously described (63, 64). Oligonucleotide-based microarray analysis was performed using either a 105K-feature WG microarray (SignatureChip Oligo Solution version 1, custom-designed by Signature Genomics; manufactured by Agilent Technologies, Santa Clara, CA) or a 135K-feature WG microarray (SignatureChip Oligo Solution version 2, custom-designed by Signature Genomics; manufactured by Roche Nimblegen, Madison, WI) as previously described (64, 65). These oligo-based platforms were used to test 21,008 probands from February 2008 to November 2010.

Bioinformatic analysis of collated CNV data from multiple databases

For the patient cohorts, we utilized the resources of five different databases: The Developmental Gene Anatomy Project (DGAP, 200 cases) at Harvard Medical School (MA, USA), The Developmental Gene Discovery Project (DGDP, 120 cases) at Georgia Regents University (GA, USA), The National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository (NJ, USA, 45 cases), Signature Genomic Laboratories' Genoglyphix Chromosome Aberration Database (GCAD, Spokane, WA, 45 744 cases) (Supplementary Material, Table S1) and The Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER, 11 247 cases). These databases contain high quality curated genome-wide CNV information obtained from diagnostic cases. To assess the frequency of these NMD CNVs in controls, we extracted data from the Database of Genomic Variants (DGV) (66). Due to redundancy of the representation of the HapMap samples, we counted the controls in only one study as representative (67–71). The total number of controls accounted for in DGV was 12,145. To increase the number of controls for comparison, we used CNV information from a recently published dataset of 8,329 cases (72). In total, we have surveyed CNVs encompassing known NMD and EJC genes in 57 356 patients and 20 474 controls (Table 1; Supplementary Material, Table S2).

General methods

Maintaining LCLs, RNA extraction, RT-PCR, RT-qPCR and western blot were performed using protocols previously described (14).

RNA-Seq analysis

Approximately twenty μg of RNA was processed and sequenced by GeneWorks (Australia) on the Illumina GAII platform. Reads were aligned by CASAVA v1.6 using the exact same protocols as previously described (14). Raw and processed RNA-Seq data have been deposited to the Gene Expression Omnibus repository under accession number GSE35420.

Statistical analysis

Where appropriate, Student's t-test, Fisher's exact test or Pearson's correlation test was employed to aid analysis when comparing difference between two groups.

AUTHOR CONTRIBUTIONS

LSN performed most the data analysis; HK, YL, JFG & LCL discovered the original deletions encompassing *UPF2* in ID patients and provided the clinical information; YS performed aCGH in 3 patients; JAR & LGS contributed the CNV data from Signature's database; LSN and JG designed the project; JG supervised the project; the manuscript was written by LSN and JG. We would like to thank the patients and their families for the support and ongoing contribution to the study.

This study makes use of data generated by the DECIPHER Consortium. A full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

WEB RESOURCES

The URLs for data presented herein are as follows:

Database of Developmental Gene Anatomy Project, http://www.bwhpathology.org/dgap/

The National Institute of General Medical Sciences Human Genetic Cell Repository, http://ccr.coriell.org/Sections/Collections/NIGMS/?SsId=8

Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER), http://decipher.sanger.ac.uk/about

Database of Genomic Variants (DGV), http://projects.tcag.ca/variation/

BrainSpan: Atlas of the Developing Human Brain, http://developinghumanbrain.org

Database for Annotation, Visualization and Integrated Discovery (DAVID), http://david.abcc.ncifcrf.gov/home.jsp

Online Medelian Inheritance in Man, http://www.ncbi.nlm.nih.gov/omim/

UCSC Genome Browser, http://genome.ucsc.edu/

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. Jill A. Rosenfeld is an employee of Signature Genomics, a subsidiary of PerkinElmer, Inc. The other authors declare no conflict of interest.

FUNDING

This work was supported by grants from the Australian NH&MRC Project Grant APP628952 to J.G. J.G. is supported by NH&MRC Principal Research Fellowship APP1041920. L.S.N.'s PhD Scholarship is supported by the MS McLeod Foundation.

REFERENCES

1. Gecz, J., Shoubridge, C. and Corbett, M. (2009) The genetic landscape of intellectual disability arising from chromosome X. *Trends Genet.*, **25**, 308–316.
2. Sagoo, G.S., Butterworth, A.S., Sanderson, S., Shaw-Smith, C., Higgins, J.P. and Burton, H. (2009) Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. *Genet. Med.*, **11**, 139–146.
3. Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P. *et al.* (2011) Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*, **478**, 57–63.
4. Ramakers, G.J. (2002) Rho proteins, mental retardation and the cellular basis of cognition. *Trends Neurosci.*, **25**, 191–199.
5. Tarpey, P.S., Nguyen, L.S., Raymond, F.L., Rodriguez, J., Hackett, A., Vandeleur, L., Smith, R., Shoubridge, C., Edkins, S., Stevens, C. *et al.* (2007) Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nat. Genet.*, **39**, 1127–1133.
6. Laumonnier, F., Shoubridge, C., Antar, C., Nguyen, L.S., Van Esch, H., Kleefstra, T., Briault, S., Fryns, J.P., Hamel, B., Chelly, J. *et al.* (2010) Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism. *Mol. Psychiatry*, **15**, 767–776.
7. Albers, C.A., Paul, D.S., Schulze, H., Freson, K., Stephens, J.C., Smethurst, P.A., Jolley, J.D., Cvejic, A., Kostadima, M., Bertone, P. *et al.* (2012) Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat. Genet.*, **44**, 435–439.
8. Rosenfeld, J.A., Traylor, R.N., Schaefer, G.B., McPherson, E.W., Ballif, B.C., Klopocki, E., Mundlos, S., Shaffer, L.G. and Aylsworth, A.S. (2012) Proximal microdeletions and microduplications of 1q21.1 contribute to variable abnormal phenotypes. *Eur. J. Hum. Genet.*
9. Nicholson, P., Yepiskoposyan, H., Metze, S., Zamudio Orozco, R., Kleinschmidt, N. and Muhlemann, O. (2010) Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell Mol. Life Sci.*, **67**, 677–700.
10. Wittmann, J., Hol, E.M. and Jack, H.M. (2006) hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol. Cell Biol.*, **26**, 1272–1287.
11. Chan, W.K., Huang, L., Gudikote, J.P., Chang, Y.F., Imam, J.S., MacLean, J.A. 2nd and Wilkinson, M.F. (2007) An alternative branch of the nonsense-mediated decay pathway. *EMBO J.*, **26**, 1820–1830.
12. Chan, W.K., Bhalla, A.D., Le Hir, H., Nguyen, L.S., Huang, L., Gecz, J. and Wilkinson, M.F. (2009) A UPF3-mediated regulatory switch that maintains RNA surveillance. *Nat. Struct. Mol. Biol.*, **16**, 747–753.
13. Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F. and Dietz, H.C. (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.*, **36**, 1073–1078.
14. Nguyen, L.S., Jolly, L., Shoubridge, C., Chan, W.K., Huang, L., Laumonnier, F., Raynaud, M., Hackett, A., Field, M., Rodriguez, J. *et al.* (2011) Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability. *Mol. Psychiatry.*
15. Nagy, E. and Maquat, L.E. (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.*, **23**, 198–199.
16. Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E. *et al.* (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.*, **26**, 163–175.
17. Wang, J., Vock, V.M., Li, S., Olivas, O.R. and Wilkinson, M.F. (2002) A quality control pathway that down-regulates aberrant T-cell receptor (TCR) transcripts by a mechanism requiring UPF2 and translation. *J. Biol. Chem.*, **277**, 18489–18493.
18. Ruiz-Echevarria, M.J. and Peltz, S.W. (2000) The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. *Cell*, **101**, 741–751.
19. He, F., Peltz, S.W., Donahue, J.L., Rosbash, M. and Jacobson, A. (1993) Stabilization and ribosome association of unspliced pre-mRNAs in a yeast upf1- mutant. *Proc. Natl. Acad. Sci. USA.*, **90**, 7034–7038.
20. Welch, E.M. and Jacobson, A. (1999) An internal open reading frame triggers nonsense-mediated decay of the yeast SPT10 mRNA. *EMBO J.*, **18**, 6134–6145.
21. Moriarty, P.M., Reddy, C.C. and Maquat, L.E. (1998) Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA. *Mol. Cell Biol.*, **18**, 2932–2939.
22. Kebaara, B.W. and Atkin, A.L. (2009) Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **37**, 2771–2778.
23. Huang, L., Lou, C.H., Chan, W., Shum, E.Y., Shao, A., Stone, E., Karam, R., Song, H.W. and Wilkinson, M.F. (2011) RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol. Cell*, **43**, 950–961.
24. Bruno, I.G., Karam, R., Huang, L., Bhardwaj, A., Lou, C.H., Shum, E.Y., Song, H.W., Corbett, M.A., Gifford, W.D., Gecz, J. *et al.* (2011) Identification of a MicroRNA that Activates Gene Expression by Repressing Nonsense-Mediated RNA Decay. *Mol. Cell*, **42**, 500–510.
25. Long, A.A., Mahapatra, C.T., Woodruff, E.A. 3rd, Rohrbough, J., Leung, H.T., Shino, S., An, L., Doerge, R.W., Metzstein, M.M., Pak, W.L. *et al.* (2010) The nonsense-mediated decay pathway maintains synapse architecture and synaptic vesicle cycle efficacy. *J. Cell Sci.*, **123**, 3303–3315.
26. Wittkopp, N., Huntzinger, E., Weiler, C., Sauliere, J., Schmidt, S., Sonawane, M. and Izaurralde, E. (2009) Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. *Mol. Cell Biol.*, **29**, 3517–3528.
27. Silver, D.L., Watkins-Chow, D.E., Schreck, K.C., Pierfelice, T.J., Larson, D.M., Burnetti, A.J., Liaw, H.J., Myung, K., Walsh, C.A., Gaiano, N. *et al.* (2010) The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat. Neurosci.*, **13**, 551–558.
28. Giorgi, C., Yeo, G.W., Stone, M.E., Katz, D.B., Burge, C., Turrigiano, G. and Moore, M.J. (2007) The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell*, **130**, 179–191.
29. Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M. and Dietz, H.C. (2001) Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.*, **10**, 99–105.
30. Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Monch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E., Nerlov, C. and Porse, B.T. (2008) NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.*, **22**, 1381–1396.
31. Vicente-Crespo, M. and Palacios, I.M. (2010) Nonsense-mediated mRNA decay and development: shoot the messenger to survive? *Biochem. Soc. Trans.*, **38**, 1500–1505.
32. Le Hir, H. and Andersen, G.R. (2008) Structural insights into the exon junction complex. *Curr. Opin. Struct. Biol.*, **18**, 112–119.
33. Zetoune, A.B., Fontaniere, S., Magnin, D., Anczukow, O., Buisson, M., Zhang, C.X. and Mazoyer, S. (2008) Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues. *BMC Genet.*, **9**, 83.
34. Chamieh, H., Ballut, L., Bonneau, F. and Le Hir, H. (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.*, **15**, 85–93.
35. He, F., Li, X., Spatrick, P., Casillo, R., Dong, S. and Jacobson, A. (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol. Cell*, **12**, 1439–1452.
36. Rehwinkel, J., Letunic, I., Raes, J., Bork, P. and Izaurralde, E. (2005) Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA*, **11**, 1530–1544.
37. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
38. Zhang, W., Bleibel, W.K., Roe, C.A., Cox, N.J. and Eileen Dolan, M. (2007) Gender-specific differences in expression in human lymphoblastoid cell lines. *Pharmacogenet. Genomics*, **17**, 447–450.
39. Ning, K., Fermin, D. and Nesvizhskii, A.I. (2012) Comparative analysis of different label-free mass spectrometry based protein abundance estimates

and their correlation with RNA-Seq gene expression data. *J. Proteome Res.*, **11**, 2261–2271.
40. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
41. Luo, J., Balkin, N., Stewart, J.F., Sarwark, J.F., Charrow, J. and Nye, J.S. (2000) Neural tube defects and the 13q deletion syndrome: evidence for a critical region in 13q33–34. *Am. J. Med. Genet.*, **91**, 227–230.
42. Rio, M., Molinari, F., Heuertz, S., Ozilou, C., Gosset, P., Raoul, O., Cormier-Daire, V., Amiel, J., Lyonnet, S., Le Merrer, M. *et al.* (2002) Automated fluorescent genotyping detects 10% of cryptic subtelomeric rearrangements in idiopathic syndromic mental retardation. *J. Med. Genet.*, **39**, 266–270.
43. Knight, S.J., Regan, R., Nicod, A., Horsley, S.W., Kearney, L., Homfray, T., Winter, R.M., Bolton, P. and Flint, J. (1999) Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet*, **354**, 1676–1681.
44. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, **318**, 798–801.
45. Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W. and Kulozik, A.E. (2005) Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol. Cell*, **20**, 65–75.
46. Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P. and Izaurralde, E. (2003) Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. *EMBO J.*, **22**, 3960–3970.
47. Metzstein, M.M. and Krasnow, M.A. (2006) Functions of the nonsense-mediated mRNA decay pathway in Drosophila development. *PLoS Genet.*, **2**, e180.
48. Palacios, I.M., Gatfield, D., St Johnston, D. and Izaurralde, E. (2004) An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature*, **427**, 753–757.
49. Longman, D., Plasterk, R.H., Johnstone, I.L. and Caceres, J.F. (2007) Mechanistic insights and identification of two novel factors in the C. *elegans NMD pathway.* Genes. Dev., **21**, 1075–1085.
50. Behm-Ansmant, I. and Izaurralde, E. (2006) Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. *Genes Dev.*, **20**, 391–398.
51. Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y. and Ohno, S. (2001) Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.*, **15**, 2215–2228.
52. Viegas, M.H., Gehring, N.H., Breit, S., Hentze, M.W. and Kulozik, A.E. (2007) The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway. *Nucleic Acids Res.*, **35**, 4542–4551.
53. Maderazo, A.B., He, F., Mangus, D.A. and Jacobson, A. (2000) Upf1p control of nonsense mRNA translation is regulated by Nmd2p and Upf3p. *Mol. Cell Biol.*, **20**, 4591–4603.
54. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. and Lee, C. (2004) Detection of large-scale variation in the human genome. *Nat. Genet.*, **36**, 949–951.
55. Seoighe, C. and Gehring, C. (2010) Heritability in the efficiency of nonsense-mediated mRNA decay in humans. *PLoS One*, **5**, e11657.
56. Gong, C., Kim, Y.K., Woeller, C.F., Tang, Y. and Maquat, L.E. (2009) SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs. *Genes Dev.*, **23**, 54–66.
57. Yepiskoposyan, H., Aeschimann, F., Nilsson, D., Okoniewski, M. and Muhlemann, O. (2011) Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA*, **17**, 2108–2118.
58. Resta, N., Susca, F.C., Di Giacomo, M.C., Stella, A., Bukvic, N., Bagnulo, R., Simone, C. and Guanti, G. (2006) A homozygous frameshift mutation in the ESCO2 gene: evidence of intertissue and interindividual variation in Nmd efficiency. *J. Cell Physiol.*, **209**, 67–73.
59. Kerr, T.P., Sewry, C.A., Robb, S.A. and Roberts, R.G. (2001) Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay? *Hum. Genet.*, **109**, 402–407.
60. Hutchinson, S., Furger, A., Halliday, D., Judge, D.P., Jefferson, A., Dietz, H.C., Firth, H. and Handford, P.A. (2003) Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype? *Hum. Mol. Genet.*, **12**, 2269–2276.
61. Linde, L. and Kerem, B. (2008) Introducing sense into nonsense in treatments of human genetic diseases. *Trends Genet.*, **24**, 552–563.
62. Traylor, R.N., Fan, Z., Hudson, B., Rosenfeld, J.A., Shaffer, L.G., Torchia, B.S. and Ballif, B.C. (2009) Microdeletion of 6q16.1 encompassing EPHA7 in a child with mild neurological abnormalities and dysmorphic features: case report. *Mol. Cytogenet.*, **2**, 17.
63. Ballif, B.C., Theisen, A., Coppinger, J., Gowans, G.C., Hersh, J.H., Madan-Khetarpal, S., Schmidt, K.R., Tervo, R., Escobar, L.F., Friedrich, C.A. *et al.* (2008) Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Mol. Cytogenet.*, **1**, 8.
64. Ballif, B.C., Theisen, A., McDonald-McGinn, D.M., Zackai, E.H., Hersh, J.H., Bejjani, B.A. and Shaffer, L.G. (2008) Identification of a previously unrecognized microdeletion syndrome of 16q11.2q12.2. *Clin. Genet.*, **74**, 469–475.
65. Duker, A.L., Ballif, B.C., Bawle, E.V., Person, R.E., Mahadevan, S., Alliman, S., Thompson, R., Traylor, R., Bejjani, B.A., Shaffer, L.G. *et al.* (2010) Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur. J. Hum. Genet.*, **18**, 1196–1201.
66. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. and Lee, C. (2004) Detection of large-scale variation in the human genome. *Nat Genet*, **36**, 949–951.
67. Locke, D.P., Sharp, A.J., McCarroll, S.A., McGrath, S.D., Newman, T.L., Cheng, Z., Schwartz, S., Albertson, D.G., Pinkel, D., Altshuler, D.M. *et al.* (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet*, **79**, 275–290.
68. McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S., Gabriel, S.B., Lee, C., Daly, M.J. *et al.* (2006) Common deletion polymorphisms in the human genome. *Nat Genet*, **38**, 86–92.
69. McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A. *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet*, **40**, 1166–1174.
70. Gusev, A., Lowe, J.K., Stoffel, M., Daly, M.J., Altshuler, D., Breslow, J.L., Friedman, J.M. and Pe'er, I. (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Res*, **19**, 318–326.
71. Simon-Sanchez, J., Scholz, S., Fung, H.C., Matarin, M., Hernandez, D., Gibbs, J.R., Britton, A., de Vrieze, F.W., Peckham, E., Gwinn-Hardy, K. *et al.* (2007) Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet*, **16**, p1–14.
72. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V. *et al.* (2011) A copy number variation morbidity map of developmental delay. *Nat Genet*, **43**, 838–846.