X-linked dystonia parkinsonism syndrome (XDP, lubag): disease-specific sequence change DSC3 in \(TAF1/DYT3\) affects genes in vesicular transport and dopamine metabolism

Thilo Herzfeld\(^1\), Dagmar Nolte\(^1\), Maria Grznarova\(^1\), Andrea Hofmann\(^2\), Joachim L. Schultze\(^2\) and Ulrich Müller\(^1\,*\)

\(^1\)Institute of Human Genetics, University of Giessen, Giessen, Germany and \(^2\)LIMES (Life and Medical Sciences) Institute, University of Bonn, Bonn, Germany

Received October 16, 2012; Revised and Accepted November 21, 2012

X-chromosomal dystonia parkinsonism syndrome (XDP, ‘lubag’) is associated with sequence changes within the \(TAF1/DYT3\) multiple transcript system. Although most sequence changes are intronic, one, disease-specific single-nucleotide change 3 (DSC3), is located within an exon (d4). Transcribed exon d4 occurs as part of multiple splice variants. These variants include exons d3 and d4 spliced to exons of \(TAF1\), and an independent transcript composed of exons d2–d4. Location of DSC3 in exon d4 and utilization of this exon in multiple splice variants suggest an important role of DSC3 in the XDP pathogenesis. To test this hypothesis, we transfected neuroblastoma cells with four expression constructs, including exons d2–d4 [d2–d4/wild-type (wt) and d2–d4/DSC3] and d3–d4 (d3–d4/wt and d3–d4/DSC3). Expression profiling revealed a dramatic effect of DSC3 on overall gene expression. Three hundred and sixty-two genes differed between cells containing d2–d4/wt and d2–d4/DSC3. Annotation clustering revealed enrichment of genes related to vesicular transport, dopamine metabolism, synapse function, \(Ca^{2+}\) metabolism and oxidative stress. Two hundred and eleven genes were differentially expressed in d3–d4/wt versus d3–d4/DSC3. Annotation clustering highlighted genes in signal transduction and cell–cell interaction. The data show an important role of physiologically occurring transcript d2–d4 in normal brain function. Interference with this role by DSC3 is a likely pathological mechanism in XDP. Disturbance of dopamine function and of \(Ca^{2+}\) metabolism can explain abnormal movement; loss of protection against reactive oxygen species may account for the neurodegenerative changes in XDP. Although d3–d4 also affect genes potentially related to neurodegenerative processes, their physiologic role as splice variants of \(TAF1\) awaits further exploration.

INTRODUCTION

\(TAF1\) encodes the TAF1 RNA polymerase II, TATA box-binding protein-associated factor, 250 kDa (TAF1) (1,2). The gene is located in the proximal long arm of the human X chromosome (Xq13.1) and is composed of 38 evolutionarily highly conserved exons. Exons 1–38 are part of the \(TAF1/DYT3\) multiple transcript system that includes five additional exons downstream of exon 38 (3–5). These five exons are referred to as exons d1–d5. Exons d1–d5 are evolutionarily conserved in primates only. There is a high degree of differential transcription and splicing within \(TAF1/DYT3\). Combinations of exons 1–38 can result in various isoforms of TAF1; multiple highly polymorphic variants can be generated by splicing exons d3 and d4 to various combinations of exons 1–37 including additional exons in \(TAF1\), e.g. exons 32′, 34′ and 35′ (3). Finally, exons d2–d4 can also be transcribed separately, regulated by a specific promoter (4).

The function of exons 1–38 is well established. By coding for TAF1 they are required for formation of the multi-protein complex that makes up transcription factor TFIID (1,2,6,7). TFIID is essential for regulation of transcription in...
combination with other transcription factors and RNA polymerase II. Evolutionary conservation of exons 1–38 is in line with the essential role of TAF1 in transcription in eukaryotes. In contrast, the function of the considerably less conserved exons d1–d5 is poorly understood.

Evidence for a function of these exons in the brain of primates comes from findings in the X-linked dystonia parkinsonism syndrome (XDP). In this disorder several disease-specific single-nucleotide changes (DSCs) and a small deletion are found within the region spanning exons d1–d5 (3). These sequence changes and the deletion were detected by sequencing the XDP critical region as determined by the analysis of linkage and of allelic association. DSCs found in the region were then fine-mapped by the study of allelic association in 46 XDP patients and in 629 control X chromosomes. Five DSCs within the TAF1/DYT3 multiple transcript system were only observed in X chromosomes of patients but in none of the controls. They are thought to alter function of TAF1/DYT3. Other coding genes and exons in the region which were not known at the time have been identified but shown not to cause disease (8,9).

XDP is a primary dystonia characterized by generalized adult-onset dystonia and concurrent parkinsonism in ~50% of cases (11,12). Severe neurodegeneration occurs in patients’ brains, mainly affecting the striatum (13,14). Presently, the molecular mechanisms causing XDP are not understood. There is some evidence that a SVA (SINE-VNTR-ALU) transposable element in intron 31 of TAF1 in patients might affect expression of a variant of the TAF1/DYT3 multiple transcript system (10). This variant (TA14-391) includes exon 34 and its expression appears to be reduced in the caudate of XDP patients when compared with caudate of healthy controls. Antibodies against a peptide encoded by exon 34 identified the expression of this protein in the rat striatum (15). However, a role of this variant in neurodegeneration has not been shown. Composition of the complete transcript of variant TA14-391 has not been studied and it is not known whether it includes exons d3 and d4. The latter is particularly important, since exon d4 harbours the only disease-specific exonic mutation (referred to as DSC3) detected in XDP patients.

To explore the possibility of a role of DSC3 in the pathogenesis of XDP, we studied the effect of this mutation on overall gene expression in a neuronal cell line. We transfected SK-N-AS neuroblastoma cells with various expression constructs. One contained exons d2–d4 that encode an independent transcript, driven by a TATA box-less promoter regulated by transcription factor Ikaros (4). A second construct included only exons d3–d4 that are spliced to various exons of TAF1.

RESULTS

Figure 1A depicts the TAF1/DYT3 multiple transcript system. Note that exons d2–d4 are transcribed as an independent

Figure 1. Schematic of the TAF1/DYT3 locus. (A) depicts exons 26–38 of TAF1 and downstream exons d1–d5. Molecular changes in XDP within this transcript unit are indicated above. OGT denotes the adjacent gene coding for UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase. >indicates a specific TATA box-free promoter. Accession number of transcript d2–d4 is given. Note that exons d3–d4 are only transcribed as splice variants of TAF1 exons. (B) Exons d2–d4 and d3–d4 used for cloning in expression vector pcDNA3.1. Cloning site, promoter and relevant resistance genes of pcDNA3.1 are shown in (C).
transcript (3) driven by a specific promoter (4). The processed d2–d4 transcript is 1999 bp in length. Exons d3–d4 are only transcribed as variants spliced to various exons of TAF1. Exon 38 is never included in these variants (3).

Analysis of d exons

We started this investigation by studying the expression pattern of d2–d3, d3–d4 and TAF1 in the brain in comparison with housekeeping genes GAPDH and RPS18. RPS18 expression was set at 1. As shown in Fig. 2 relative expression of d2–d3 as part of the d2–d4 transcript is less than that of exons d3–d4 that occur as part of TAF1 exons. Transcription of TAF1 exons was higher than that of both d2–d4 and d3–d4. Yet TAF1 was markedly less expressed than GAPDH. We then examined the expression of d2–d3 and of d3–d4 in various regions of the human brain. Figure 3A and B shows that both d2–d3 and d3–d4 are preferentially expressed in the brain. Expression of both transcripts is significantly higher in the cerebral cortex and in the nucleus accumbens than in the kidney, liver and skeletal muscle. Particularly high expression of d3–d4 transcripts was found in the putamen.

Previously, we had described several small open reading frames (ORFs) in d2-d4. The largest open reading frame (ORF) of 447 bp in d2 was used for the generation of a fusion protein (see Materials and Methods). Immunization of rabbits resulted in antisera that recognized the fusion protein on western blots, but did not reveal specific bands in any portion of the brain studied including the frontal cortex, putamen, caudate nucleus, thalamus and substantia nigra (not shown).

Expression constructs

We prepared four different expression constructs: d2–d4/wt (construct 1a) and d2–d4/DSC3 (construct 2a) containing wild-type (wt) and DSC3 variant exons d2–d4, respectively; d3–d4/wt (construct 1b) and d3–d4/DSC3 (construct 2b) containing wt and DSC3 variant exons d3–d4. Figure 1A shows parts of the TAF1/DYT3 multiple transcript system including the location of all changes observed in XDP. DSC3 is highlighted. The exons used for the preparation of expression constructs are depicted in Fig. 1B. Figure 1C shows expression vector pcDNA3.1+ in which the inserts were cloned. Expression of the inserts was quantified by quantitative PCR (qPCR) after transfection of SK-N-AS cells using the various constructs. As shown in Fig. 4, RNA encoded by both wt and DSC3-containing exons d2–d4 and d3–d4 is \( \sim 10^4 \)–\( 10^5 \) fold more abundant in transfected SK-N-AS cells than in cells transfected with vector only (vo).

Expression analysis

We examined the effect of over-expression of the four cloned inserts (d2–d4/wt and d2–d4/DSC3 as well as d3–d4/wt and d3–d4/DSC3) on the overall level of gene expression of the transfected SK-N-AS cells. As controls we studied overall gene expression in cells transfected with vo and in untransfected cells.

We compiled all genes whose expression level differed at least two-fold between the various transfected cell lines.
Expression of 362 genes differs between cells containing d2–d4/wt and cells containing d2–d4/DSC3. This difference is even higher when comparing RNA from cells over-expressing d2–d4/wt to those containing vo. This comparison revealed 485 differentially expressed genes. Conversely, the difference is lower (67 genes) in RNA from cells containing d2–d4/DSC3 versus vo. Comparable results were obtained with constructs including exons d3 and d4 only. The expression level of 211 genes differed between d3–d4/wt and d3–d4/DSC3 transfected cells. At 446 genes, the difference was highest when d3–d4/wt containing cells were compared with those containing vo. Only 78 genes were found to be dysregulated when comparing RNA from d3–d4/DSC3 over-expressing cells with RNA from cells containing vo. The comparatively low number of genes differentially regulated in both d2–d4/DSC3 and d3–d4/DSC3 compared with vo demonstrates that DSC3 dramatically compromises the effect of d2–d4/wt and d3–d4/wt on overall gene expression. Figure 5B gives hierarchical clustering of the 362 genes differentially expressed in cell lines transfected with d2–d4/wt versus d2–d4/DSC3. Up-regulated genes are given in red, down-regulated genes in blue. Similarity of the expression level of these genes is also shown in vo and in untransfected cells. Expression of two-thirds of the genes differentially expressed between d2–d4/wt and d2–d4/DSC3 are up-regulated by d2–d4/wt. Figure 5C shows hierarchical clustering of the 211 genes differentially expressed in cell lines transfected with d3–d4/wt versus d3–d4/DSC3. Similar to constructs d2–d4, about two-thirds of the 211 differentially expressed genes are up-regulated by d3–d4/wt.

Supplementary Material, Table S1a lists all 362 genes expression of which differs at least two-fold in cells over-expressing d2–d4/wt versus d2–d4/DSC3. The 211 genes differentially expressed in d3–d4/wt versus d3–d4/DSC3 are given in Supplementary Material, Table S1b. Although differences in gene expression are quite striking using d2–d4 and d3–d4 constructs (both wt and DSC3 variant), the differentially regulated genes differ between the constructs.

Functional annotation clustering

Annotation categories were assigned to the 362 genes differentially expressed in cells transfected with d2–d4/wt versus d2–d4/DSC3 (Table 1). We identified the three most significant annotation categories. They are referred to as annotation clusters 1a–3a. Annotation cluster 1a includes 30 genes related to vesicular and synapse function. At 3.32 the enrichment score is highest for this group of genes. Twenty-six genes related to Ca$^{2+}$ metabolism and function are assigned to cluster 2a (enrichment score 2.53). Genes responsive to stress, mainly mediated by reactive oxygen species (ROS), are grouped in annotation cluster 3a (enrichment score 2.22, 18 genes). Supplementary Material, Table 2a lists the genes assigned to the three annotation clusters 1a–3a.

Annotation categories were also assigned to the 211 genes differentially expressed in d3–d4/wt versus d3–d4/DSC3 expressing SK-N-AS cells. The three categories of genes found to be most significantly enriched are designated clusters 1b–3b (Table 2). One hundred and four genes related to signal transduction were assigned to annotation cluster 1b. The enrichment score for this group of genes is 14.04. Annotation cluster 2b includes genes related to differentiation of vasculature and cell movement (enrichment score 5.97). Genes related to cell–cell interaction were assigned to cluster 3b (enrichment score 5.71). Supplementary Material, Table 2b lists the genes assigned to the three annotation clusters 1b–3b.

Pathway analysis

Figure 6A depicts the interaction network of genes dysregulated in cells transfected with d2–d4/DSC3 when compared with those transfected with d2–d4/wt. A total of 105 gene products interacted with at least one other gene product. Gene products with many interaction partners included transcriptional regulators (EGR1), genes involved in the regulation of cell proliferation (GGC, PRKCA) and in neurotransmission and hormonal regulation (CALCB, CRH). Of note, SYP (synaptophysin), a synaptic vesicle protein involved in synaptic plasticity was also a central node in the network. Figure 6B shows the interaction network of genes dysregulated in cells transfected with d3–d4/DSC3 versus d3–d4/wt. These genes did not form pronounced networks. Only 38 genes interacted with at least one other gene. Three genes coding for members of the SERPIN family (serine protease inhibitors) were parts of the largest network of 13 interacting gene products. Neither network revealed genes interacting with known miRNAs.

DISCUSSION

The experiments show that exons d2–d4 dramatically affect gene expression. d2–d4 encode a distinct transcript regulated by a specific promoter (4). The disease-specific exonic sequence change DSC3 in exon d4 alters a normal pattern of
L-dopa and dopamine.

scripts generated by splicing to various exons of d4 are physiologically expressed as part of alternative transcription of cellular gene expression. Exons d3–d4 affect different genes than d2–d4. Furthermore, d3–d4 affect gene expression as well and DSC3 also appears to exert loss of function. However, exons d3–d4 alone affect gene expression and DSC3 interferes with d2–d4-mediated regulation of cellular gene expression by a loss of function mechanism. Exons d3–d4 affect different genes than d2–d4. Furthermore, d3–d4 are physiologically expressed as part of alternative transcripts generated by splicing to various exons of TAF1 [see Fig. 1; (3)]. The role of these two exons as parts of alternative transcripts of TAF1 or other genes is currently not understood.

Annotation clustering revealed that among the 362 genes dysregulated by DSC3 (comparison d2–d4/wt versus d2–d4/DSC3) many are involved in vesicular and dopamine function. SYT7 codes for synaptotagmin VII (22). Synaptotagmins are membrane-bound calcium sensors. They regulate Ca2+-dependent exocytosis of secretory vesicles and release of neurotransmitters. DLG4, also known as synapse-associated protein 90, may be required at the postsynaptic membrane for clustering of receptors, ion channels and signaling proteins (23). SNAP91 (synaptosomal-associated protein, 91 kDa homolog) is a synaptic protein that accumulates at nerve terminals and is important for vesicle recycling (24). DBH codes for dopamine β-dehydroxylase. It converts dopamine into norepinephrine and is located in synaptic vesicles (25).

Although the system used is artificial, the findings provide direct evidence for a possible molecular pathological mechanism underlying the XDP. DSC3 interferes with normal function of d2–d4 on the regulation of genes involved in dopamine processing and function (Fig. 7). The postulated dysfunction of dopamine could explain parkinsonism and possibly dystonia as well. Significantly, other genes mutated in some forms of monogenic dystonias cause depletion of dopamine. Heterozygous mutations of GCH1 result in insufficient synthesis of biopterin, an important co-factor of dopamine synthesis. The resulting autosomal dominant disorder is referred to as dopa-responsive dystonia (DRD, dystonia 5a) (26). Homozygous mutations in the gene coding for TH result in autosomal recessive DRD (dystonia 5b) also by interfering with dopamine synthesis (dystonia 5b) (27). Furthermore, loss of normal dopamine function is the major cause of Parkinson’s disease (PD) (28). Disturbed dopamine function in XDP is also supported by a moderate beneficial therapeutic effect of l-dopa in patients (29). l-Dopa mainly ameliorates the parkinsonian signs tremor and bradykinesia. The absence of dramatic therapeutic benefit of l-dopa is consistent with the observed neurodegeneration of the striatum in XDP.

SYP codes for synaptophysin. It is required for recycling of synaptic vesicles by binding to synaptofibrin and formation of a SNARE (soluble N-ethylmaleimide-sensitive-factor attachment protein receptor) complex (16). SNAP25 encodes the 25 kDa synaptosomal-associated protein, another important component of SNARE (17). RAB26, RAB27B and RAB8B encode Ras-related brain proteins (18). They are members of the Ras oncoprotein family, i.e. proteins of the cell membrane involved in vesicle trafficking. RAB26 might interact with GTP cyclohydrolase I (19). DDC encodes DDC which converts l-dopa into dopamine (20). GCH1 codes for GTP cyclohydrolase I the rate limiting enzyme in the synthesis of tetrahydrobiopterin, an important cofactor for the amino acid mono-oxygenases phenylalanine-, tryptophan- and tyrosine hydroxylase (TH) (21). Of these, TH converts tyrosine into l-dopa and dopamine. SLC18A1 encodes a protein of the solute carrier 18 family. These carriers transport cytoplasmic monoamines such as dopamine and other catecholamines to synaptic vesicles. SYT7 codes for synaptotagmin VII (22). Synaptotagmins are membrane-bound calcium sensors. They regulate Ca2+-dependent exocytosis of secretory vesicles and release of neurotransmitters. DLG4, also known as synapse-associated protein 90, may be required at the postsynaptic membrane for clustering of receptors, ion channels and signaling proteins (23). SNAP91 (synaptosomal-associated protein, 91 kDa homolog) is a synaptic protein that accumulates at nerve terminals and is important for vesicle recycling (24). DBH codes for dopamine β-dehydroxylase. It converts dopamine into norepinephrine and is located in synaptic vesicles (25).

Although the system used is artificial, the findings provide direct evidence for a possible molecular pathological mechanism underlying the XDP. DSC3 interferes with normal function of d2–d4 on the regulation of genes involved in dopamine processing and function (Fig. 7). The postulated dysfunction of dopamine could explain parkinsonism and possibly dystonia as well. Significantly, other genes mutated in some forms of monogenic dystonias cause depletion of dopamine. Heterozygous mutations of GCH1 result in insufficient synthesis of biopterin, an important co-factor of dopamine synthesis. The resulting autosomal dominant disorder is referred to as dopa-responsive dystonia (DRD, dystonia 5a) (26). Homozygous mutations in the gene coding for TH result in autosomal recessive DRD (dystonia 5b) also by interfering with dopamine synthesis (dystonia 5b) (27). Furthermore, loss of normal dopamine function is the major cause of Parkinson’s disease (PD) (28). Disturbed dopamine function in XDP is also supported by a moderate beneficial therapeutic effect of l-dopa in patients (29). l-Dopa mainly ameliorates the parkinsonian signs tremor and bradykinesia. The absence of dramatic therapeutic benefit of l-dopa is consistent with the observed neurodegeneration of the striatum in XDP. In contrast, no morphological changes are found in DRDs in which the administration of l-dopa can result in complete remission of signs and symptoms.

Other factors such as dysregulation of Ca2+ metabolism and function might also contribute to dystonia in XDP. For example, CACNA2D2 codes for the alpha-2/delta subunit of calcium channels, which play a role in synaptic plasticity and may contribute to the pathogenesis of dystonia (30). Additional studies are needed to fully understand the role of these and other genes in the pathogenesis of XDP and other monogenic dystonias.
a voltage-gated calcium channel. Calcium channels are required for the influx of Ca²⁺ into excitatory neurons after depolarization and for triggering vesicular neurotransmitter release (30,31). Given such channels are mutated in other neurodegenerative diseases (32,33), their disturbed regulation might also contribute to the phenotype of XDP.

DSC3 also affects d2–d4-mediated regulation of genes protective against ROS (annotation cluster 3a). This might explain at least partially neurodegeneration observed in XDP. Significantly, ROS appear to play a role in the pathogenesis of neurodegenerative diseases including amyotrophic lateral sclerosis, Alzheimer and PD (34–37).

Functional annotation clustering of genes regulated by d3–d4 shows enrichment of genes in signal transduction and post-translational modification (e.g. glycosylation) (Table 2). Genes required for blood vessel proliferation are also enriched.

### Table 1. Functional annotation clustering of 362 genes differentially expressed in cells transfected with d2–d4/wt versus d2–d4/DSC3

<table>
<thead>
<tr>
<th>Annotation Cluster 1a</th>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_CC_FAT GO:0031410</td>
<td>cytoplasmic vesicle</td>
<td>30</td>
<td>3.60E-006</td>
</tr>
<tr>
<td>GOTERM_CC_FAT GO:0031982</td>
<td>vesicle</td>
<td>13</td>
<td>1.41E-004</td>
</tr>
<tr>
<td>GOTERM_CC_FAT GO:0016023</td>
<td>cytoplasmic membrane-bounded vesicle</td>
<td>23</td>
<td>3.28E-004</td>
</tr>
<tr>
<td>GOTERM_CC_FAT GO:0030141</td>
<td>secretory granule</td>
<td>12</td>
<td>3.87E-004</td>
</tr>
<tr>
<td>GOTERM_CC_FAT GO:0031988</td>
<td>membrane-bounded vesicle</td>
<td>23</td>
<td>5.09E-004</td>
</tr>
<tr>
<td>GOTERM_CC_FAT GO:0060205</td>
<td>cytoplasmic membrane-bounded vesicle lumen</td>
<td>6</td>
<td>1.10E-003</td>
</tr>
<tr>
<td>GOTERM_CC_FAT GO:0031983</td>
<td>vesicle lumen</td>
<td>6</td>
<td>1.35E-003</td>
</tr>
<tr>
<td>GOTERM_CC_FAT GO:0031093</td>
<td>platelet alpha granule lumen</td>
<td>5</td>
<td>6.10E-003</td>
</tr>
</tbody>
</table>

### Table 2. Functional annotation clustering of 211 genes differentially expressed in cells transfected with d3–d4/wt versus d3–d4/DSC3

<table>
<thead>
<tr>
<th>Annotation Cluster 1b</th>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP_PIR_KEYWORDS</td>
<td>Signal</td>
<td>85</td>
<td>9.44E-021</td>
</tr>
<tr>
<td>UP_SEQ_FEATURE</td>
<td>Signal peptide</td>
<td>85</td>
<td>1.43E-020</td>
</tr>
<tr>
<td>SP_PIR_KEYWORDS</td>
<td>Glycoprotein</td>
<td>89</td>
<td>4.93E-015</td>
</tr>
<tr>
<td>UP_SEQ_FEATURE</td>
<td>Glycosylation site: N-linked (GlcNAc ... )</td>
<td>82</td>
<td>1.80E-012</td>
</tr>
<tr>
<td>UP_SEQ_FEATURE</td>
<td>Disulfide bond</td>
<td>64</td>
<td>9.41E-012</td>
</tr>
<tr>
<td>SP_PIR_KEYWORDS</td>
<td>Disulfide bond</td>
<td>65</td>
<td>1.12E-011</td>
</tr>
<tr>
<td>GOTERM_CC_FAT GO:0005576</td>
<td>extracellular region</td>
<td>56</td>
<td>8.12E-011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation Cluster 2b</th>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_BP_FAT GO:0001568</td>
<td>blood vessel development</td>
<td>16</td>
<td>2.15E-008</td>
</tr>
<tr>
<td>GOTERM_BP_FAT GO:0001944</td>
<td>vasculature development</td>
<td>16</td>
<td>2.97E-008</td>
</tr>
<tr>
<td>GOTERM_BP_FAT GO:0048514</td>
<td>blood vessel morphogenesis</td>
<td>14</td>
<td>1.84E-007</td>
</tr>
<tr>
<td>GOTERM_BP_FAT GO:0006928</td>
<td>cell motion</td>
<td>20</td>
<td>2.43E-007</td>
</tr>
<tr>
<td>GOTERM_BP_FAT GO:0016477</td>
<td>cell migration</td>
<td>14</td>
<td>3.81E-006</td>
</tr>
<tr>
<td>GOTERM_BP_FAT GO:0048870</td>
<td>cell motility</td>
<td>14</td>
<td>1.20E-005</td>
</tr>
<tr>
<td>GOTERM_BP_FAT GO:0051674</td>
<td>localization of cell</td>
<td>14</td>
<td>1.20E-005</td>
</tr>
<tr>
<td>GOTERM_BP_FAT GO:0001525</td>
<td>angiogenesis</td>
<td>9</td>
<td>1.17E-004</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation Cluster 3b</th>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_BP_FAT GO:0007155</td>
<td>cell adhesion</td>
<td>24</td>
<td>3.98E-007</td>
</tr>
<tr>
<td>GOTERM_BP_FAT GO:0022610</td>
<td>biological adhesion</td>
<td>24</td>
<td>4.08E-007</td>
</tr>
<tr>
<td>SP_PIR_KEYWORDS</td>
<td>Cell adhesion</td>
<td>15</td>
<td>4.54E-005</td>
</tr>
</tbody>
</table>
Interestingly, normal glycosylation is disturbed in many neurodegenerative diseases such as Alzheimer disease, amyotrophic lateral sclerosis, Parkinson and other rare neurological syndromes (38–43). Similarly, DSC3-mediated disturbed glycosylation might contribute to the neurodegenerative phenotype of XDP (44). An important role of DSC3-mediated altered function of d-exons in the pathogenesis of XDP is further supported by preferential expression of these exons in the human brain. Pathway analysis of d3–d4/wt versus d3–d4/DSC3 (Fig. 6B) revealed a network of several genes including three members of the SERPIN gene family (serine protease inhibitors) some of which have been implicated in neurodegenerative disorders (45,46). SERPINE1 is also a major node in d2–d4/wt versus d2–d4/DSC3 (Fig. 6A).

Apparent absence of translation of d2–d4-encoded RNA suggests a regulatory function of this RNA. In silico analysis indicates disturbance of a putative hair-pin structure of d3–d4 and of d4-encoded RNA by DSC3 (see Supplementary material online). This finding would be consistent with d2–d4 encoding a pri-miRNA. Alternatively, d2–d4 might act as a long regulatory RNA. Interestingly, an entry in the genome database (http://genome.ucsc.edu/, date last accessed December 7, 2012) suggests the existence of a gene encoding a small neural regulatory RNA (*BCYRN1*, RefSeq: NR_001568). The unprocessed transcript of this gene has a size of 518 928 bp, spans the X-chromosome from 70 430 035 to 70 948 962, and is transcribed from telomere to centromere. It cannot be excluded that DSC3 and/or other DSCs affect the function of this regulatory RNA. However, the experiments presented here clearly show that DSC3 affects transcription of various genes when transcribed from centromere to telomere, i.e. in the transcriptional direction of *TAF1/DYT3*.

In conclusion, the experiments demonstrate a dramatic effect of DSC3 on gene expression. DSC3 is located in exon d4 of the *TAF1/DYT3* multiple transcript system and is the only exonic disease-specific change found in the XDP. DSC3 affects the expression of a number of genes regulated by transcript d2–d4. Many of these genes are required for normal brain function, including genes in dopamine metabolism, vesicle and synapse function, Ca\(^{2+}\) metabolism and protection against ROS. The DSC3-mediated dysregulation of these genes might be the molecular pathological mechanism in XDP.

**MATERIALS AND METHODS**

**Cell lines**

Neuroblastoma cell line SK-N-AS was purchased from ATCC (LGC Standards GmbH, Wesel, Germany) (Nr. CRL-2137).
Constructs

cDNA was prepared from human fetal brain RNA (Clontech). Exons d2–d4 and exons d3–d4 were amplified by PCR using 5′ phosphorylated primers (Ex2-F ACACTTCCTGCACCTCAGCAAAG, Ex3-F GTACCAATGAACAAGCCATTCAG, Ex4-R GTAGCAACCCACTCGGGTC). The amplification products were ligated to the EcoRV restriction site of expression vector pcDNA3.1 (Invitrogen/Life Technologies, Paisley, UK). DSC3 was introduced into the sequence by in vitro mutagenesis using the ‘QuickChange II Site-directed mutagenesis kit’ (Stratagene/Agilent, Santa Clara, CA, USA) and primers c.1259t (GAACCAAGGCTTTGAGTCTCTTTGTCTCATCTCG) and c.1259t_antisense (CGGAGATGAGAAAGAAGACTCAAGGCCTTGGTC). Constructs were used for transfection of competent E. coli TOP10 cells (Invitrogen/Life Technologies, Paisley, UK). Plasmids were isolated using the EndoFree Plasmid Maxi Kit (Qiagen, Hilden, Germany) and sequenced on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) to confirm intactness.

Cell culture and transfection

SK-N-AS cells were grown in DMEM (10% FCS; 37°C; 5% CO2). Effectene transfection reagent (Qiagen, Hilden, Germany) was used for transfection of cells using Scal-
linearized expression vector constructs. Five different Scal-
linearized plasmids were utilized: (i) Plasmids containing wt exons d2–d4 (d2–d4/wt); (ii) plasmids containing exons d2–d4 with DSC3 (d2–d4/DSC3); (iii) plasmids containing wt exons d3–d4 (d2–d3/wt); (iv) plasmids containing exons d3–d4 with DSC3 (d3–d4/DSC3); (iv) plasmids without insert. Each transfection was performed in triplicate. Positive clones were selected using geneticin (G418) at a final concentra-
tion of 400 μg/ml.

Quantitative PCR

The level of the expression of cloned inserts of the four con-
structs (d2–d4/wt; d2–d4/DSC3; d3–d4/wt; d3–d4/DSC3) and of transcripts in various tissues was determined by qPCR. Billiant II SYBR Green Master Mix (Stratagene/Agilent, Santa Clara, CA, USA) was used for amplification. cDNA was synthesized from 500 ng of total RNA that was either isolated from cells or purchased from Clontech, Mountain View, CA, USA applying the QuantiTect Rev Transcrip-
tion Kit from Qiagen (Hilden, Germany). Fluorescence data were compiled applying a MX3000P qPCR system. Annealing and extension were performed in one single step at 60°C for
1 min. PCR products obtained by qPCR were sequenced to confirm wt and DSC3 containing transcripts. Primers used for amplification are listed in Table 3.

Generation of antisera against putative d2 translation product

An ORF of 447 bp within exon d2 was checked for putative translation. A GST-fusion protein was generated by cloning the distal portion of this ORF in vector pGEX2T (GE Healthcare, Muenchen, Germany). In brief, a 279 bp PCR fragment was generated from genomic DNA of a Filipino male control strain Rosetta Blue E. coli care, Muenchen, Germany). In brief, a 279 bp PCR fragment was generated from genomic DNA of a Filipino male control strain Rosetta Blue E. coli (Merck Millipore, Darmstadt, Germany) was transformed with the cloned construct. Positive clones were identified by restriction analysis. Subsequent analytical expression was performed to detect the fusion protein composed of the N-terminal GST domain of 26 kDa and the ~9 kDa corresponding to the distal 87 amino acids of the exon d2 ORF. Since fusion protein of the large-scale expression was insoluble, inclusion bodies were resuspended in CellLytic B II (Sigma-Aldrich, St Louis, USA) and subsequently used for immunization of rabbits according a standard immunization protocol (Biogenes, Berlin, Germany).

Microarray procedure

Cells were harvested from confluent cultures and RNA was extracted according to standard procedures (RNeasy Kit, Qiagen). Biotin-labeled cRNA was prepared using the Ambion® Illumina RNA amplification kit (Ambion Europe, Huntington, Cambridgeshire, UK). 1.5 μg of biotin-labeled cRNA were then hybridized to Sentrix® whole genome bead chips (Human HT-12_v3, Illumina, San Diego, CA, USA) and scanned on the Illumina® BeadStation 500×.

Microarray data analysis and software

Raw data were extracted from the scans of the RNA microar-
rays using Beadstudio 3.1.1.0 software (Beadstudio Gene Expression Analysis Module 3.1.8). All further analyses were performed in R (http://www.r-project.org, version 2.8.0, date last accessed December 7, 2012) using Bioconductor packages (47). Quantile normalization was implemented in the affy package (methods for Affymetrix oligonucleotide arrays). Variable gene expression was defined by a coefficient of vari-
ation (SD/mean) between 0.5 and 10. Determination of calls was based on the detection P-value assessed by the Beadstudio software. A gene was called ‘expressed’ if the detection P-value was <0.05. If P-values were greater, a gene was con-
sidered not expressed. Differentially expressed genes were selec-
ted applying a fold change/P-value filter. Criteria for signif-
ificance were (i) P-values smaller than 0.05, (ii) more than
two-fold expression changes and (iii) difference between mean intensity signals exceeding 100. Genes that were signifi-
cantly differentially expressed were referred to as ‘target genes’ and further analyzed. The Benjamini–Hochberg method was applied to adjustment of raw P-values in order

Table 3. Primers used in qPCR

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 5’-3’</th>
<th>Used to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sybr3/4-F</td>
<td>TGCTGACCTCCAAGAATGAAGCCTGTG</td>
<td>Test overexpression in transfected cells</td>
</tr>
<tr>
<td>Sybr3/4-R</td>
<td>AGCACACCTCAATGGGAGTTCTGC</td>
<td></td>
</tr>
<tr>
<td>Ex2/3_qPCR-F</td>
<td>ATGGTTCTGGAAAATTCTATCC</td>
<td>Quantify d2/d3 containing transcripts</td>
</tr>
<tr>
<td>Ex2/3_qPCR-R</td>
<td>CCAGTGGTGTCGGTCTGCTGAC</td>
<td>Quantify d3/d4 containing transcripts</td>
</tr>
<tr>
<td>Sybr3/4-F2</td>
<td>TCAATCCAAGTCGCGCACC</td>
<td></td>
</tr>
<tr>
<td>Sybr3/4-R2</td>
<td>TCTGGACAACATGGTACCCAGA</td>
<td>Quantify TAF1 transcripts</td>
</tr>
<tr>
<td>TAF_Ex30-F</td>
<td>CCGACTCGATACCTGGTCGTGAG</td>
<td></td>
</tr>
<tr>
<td>TAF_Ex32-R</td>
<td>TGCACCACCAACTGCTTAGC</td>
<td>Quantify GAPDH transcripts</td>
</tr>
<tr>
<td>GAPDH_real-R</td>
<td>GGCATGGACTGTTGTCAATGAG</td>
<td></td>
</tr>
<tr>
<td>GAPDH_real-F</td>
<td>TCTGCCCCATCAGTGCCATTA</td>
<td>Quantify RPS18 transcripts</td>
</tr>
<tr>
<td>RPS18-real-R</td>
<td>ATTCCTCAGTGGTCTCCCG</td>
<td></td>
</tr>
<tr>
<td>RPS18-real-F</td>
<td>CCGAGGTGACGTCG-3</td>
<td></td>
</tr>
</tbody>
</table>

© 2013 Oxford University Press
to control for false discovery. The fold-change was calculated by dividing the mean intensity of the expression of a target gene of one group by the expression intensity of the same gene from the group it was compared with. If this number was <1, the negative reciprocal was used.

Hierarchical cluster analysis was performed using the hcluster method in R. Before clustering, the data were log2 transformed. Distances of the samples were calculated using Pearson correlation and clusters were formed by taking the average of each cluster. Data were deposited in database GEO under accession number GSE41157.

Functional annotation clustering

Functional annotation clustering was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) (48,49). A ‘genelist’ was compiled of all target genes differentially expressed in cells transfected with wt when compared with DSC3 constructs and uploaded to the DAVID web interface. DAVID uses data from HumanHT-12_v3 chip (Illumina) for comparison and calculation of enrichment. Standard annotation categories of the program were used for analysis. These categories have been introduced by the GO Consortium (the gene ontology consortium) and reflect gene function (50). Genes enriched in the uploaded data set (‘functional annotations’) when compared with background (genes on HumanHT-12_v3 chip) were sorted into groups. An ‘enrichment score’ was calculated as the geometric mean of the P-values of all members of a group.

Pathway analysis

An interaction network of the gene products was generated using STRING (http://string-db.org, date last accessed December 7, 2012) (51) at default settings and high level confidence scores (>0.700). Genes were up-loaded that were either up- or down-regulated at least two-fold in cells transfected with d2–d4/wt and d3–d4/wt versus d2–d4/DSC3 and d3–d4/DSC3. The network, i.e. all gene products interacting with at least one other gene product, was transferred to cytoscape (http://cytoscape.org/, date last accessed December 7, 2012) (52) and expression data were overlaid. Up-regulated genes (genes expressed at higher levels in cells transfected with wt constructs) are marked red, down-regulated genes are blue. Intensity of color indicates the degree of dysregulation. Size of nodes indicates the number of interacting gene products. The network was also tested for interaction with known miRNAs using cyTargetLinker Plugin (http://code.google.com/p/cyta rgetlinker/, date last accessed December 7, 2012).

FUNDING

This work was supported by BMBF (BMBF01GI9999).

REFERENCES