

Analysis of differentially expressed genes in various stages of Duchenne muscular dystrophy by using a network view

D. Wang, Z.J. Wang, X.X. Song, L.H. Pu, X. Li and Y. Wang

Department of Neurology, Xi'an Children's Hospital, Xi'an, Shaanxi, China

Corresponding author: D. Wang E-mail: wangdong6928@yahoo.cn

Genet. Mol. Res. 12 (4): 4480-4488 (2013) Received February 20, 2013 Accepted August 18, 2013 Published October 10, 2013 DOI http://dx.doi.org/10.4238/2013.October.10.13

ABSTRACT. Duchenne muscular dystrophy (DMD), which is caused by mutations in the X-linked dystrophin gene, is a severe and progressive neuromuscular disease with no available cure. By integrating 2 microarray datasets from the Gene Expression Omnibus, we identified differentially expressed genes in 2 stages of DMD and systematically explored their potential disease-related mechanisms using a network view. Twenty differentially expressed genes were detected in various stages of DMD. According to the network with dystrophin as its center, none of the 20 proteins interacts with dystrophin directly. IQ motif-containing GTPaseactivating protein 1 was found in the 2nd-level neighbors with a degree of 21. Microtubule-associated protein tau, membrane metallo-endopeptidase, interleukin 13 receptor alpha 1, and multiple epidermal growth factor-like domains 6 were found in the 3rd-level neighbors. These identifications require further investigation, as this report is the first of possible associations between DMD and these proteins. Analysis of differentially expressed genes through this network view may provide important information for further exploration of underlying mechanisms of DMD.

Key words: Duchenne muscular dystrophy; Differential expression; Protein-protein interaction network

INTRODUCTION

Duchenne muscular dystrophy (DMD), which affects 1 in 3500 newborn boys (Emery, 2002), is a severe and progressive neuromuscular disease caused by mutations in the X-linked dystrophin gene. Absence or defect of the protein dystrophin disrupts its essential function of connecting the subsarcolemmal cytoskeleton to the sarcolemma, resulting in progressive muscle degeneration. Typically, patients without intervention lose independent ambulation by the age of 13 years (Hoffman et al., 1987) and experience cardiac or respiratory failure in their mid- to late 20s, with a mean age of death of approximately 19 years (Bushby et al., 2010a). Cognitive dysfunction might also be present (Anderson et al., 2002). Symptoms of DMD are usually initially observed between the ages of 2 and 5 years (Dubowitz, 1978; Jennekens et al., 1991), although elevated serum creatine kinase level and abnormal muscle histology are always present in patients. Thus, the first 2 years are considered clinically presymptomatic (Pescatori et al., 2007).

Since the genetic etiology of DMD was well analyzed over 20 years ago (Koenig et al., 1987; Kunkel et al., 1987), numerous therapeutic strategies have been proposed, including gene-based therapy, cell-based therapy, and pharmacological agents. Multidisciplinary care has also been recommended (Bushby et al., 2010b). However, these approaches have many hurdles yet to overcome, and no cure for DMD is presently available. Understanding the pathomechanisms for DMD may lead to alternatives to currently inadequate therapeutic strategies. Previous studies have shown that DMD patients exhibit symptoms related to various proteins or molecules, such as growth factors (Gehrig et al., 2012), nitric oxide synthase (Altamirano et al., 2012), phospholipase (Lindahl et al., 1995), and cytokines (De Pasquale et al., 2012), suggesting that "molecule groups" instead of single or a few proteins should be explored in the investigation of the DMD pathomechanism. Protein-protein interaction (PPI) is crucial for all biological processes (Stelzl et al, 2005) because the majority of proteins function with other proteins. Therefore, biological processes should be considered complex PPI networks of interconnected proteins. In other words, analysis from a network view would provide a better understanding of the mechanism underlying DMD progression.

High-throughput experimental strategies enable identification of candidate gene sets associated with the symptoms of DMD patients. However, few of these efforts have focused on the network downstream of dystrophin, which contributes to development and progression from the presymptomatic phase (PP) of the disease to the symptomatic phase (SP) and represents the pathomechanisms of DMD. Knowledge of these networks would facilitate understanding of the molecular mechanisms of DMD.

Our hypothesis is that dysfunctional genes/proteins and their interactions are strongly related to specific states of the disease and can capture the pathogenic characteristics of corresponding disease stages. Differentially expressed genes between SP and PP may reflect central elements in the progression of DMD from PP to SP. Integrating microarray data from both PP and SP, we identified differentially expressed genes/proteins between SP and PP and carried out biological network analysis for their interactions with the DMD protein to get a better understanding of the molecular mechanism underlying DMD progression.

MATERIAL AND METHODS

Microarray data

Two datasets (GSE6011 and GSE300741) were downloaded from the Gene Expres-

sion Omnibus (http://www.ncbi.nlm.nih.gov/geo/) database. Sample information for 19 presymptomatic samples, 8 symptomatic samples, and 14 normal samples is listed in Table 1. Both datasets were based on the GPL96 platform: [HG-U133A] Affymetrix Human Genome U133A Array.

Table 1. Chara	cteristics of	the samples.			
GEO accession	ID	Phase	Age (years)	Gender	Dystrophin expression
GSM139515	PP01	Presymptomatic	0.1	M	Protein absent; mRNA normal
GSM139516	PP02	Presymptomatic	0.2	M	Protein absent; mRNA reduced
GSM139517	PP03	Presymptomatic	0.3	M	Protein absent; mRNA reduced
GSM139519	PP04	Presymptomatic	0.4	M	Protein absent; mRNA reduced
GSM139520	PP05	Presymptomatic	0.4	M	Protein absent; mRNA reduced
GSM139521	PP06	Presymptomatic	0.5	M	Protein absent; mRNA weakly reduced
GSM139522	PP07	Presymptomatic	0.5	M	Protein absent; mRNA reduced
GSM139523	PP08	Presymptomatic	0.6	M	Protein absent; mRNA reduced
GSM139524	PP09	Presymptomatic	0.7	M	Protein absent; mRNA reduced
GSM139525	PP10	Presymptomatic	0.7	M	Protein absent; mRNA reduced
GSM139526	PP11	Presymptomatic	0.9	M	Protein absent; mRNA normal
GSM139527	PP12	Presymptomatic	1.0	M	Protein absent; mRNA reduced
GSM139528	PP13	Presymptomatic	1.0	M	Protein absent; mRNA reduced
GSM139529	PP14	Presymptomatic	1.2	M	Protein absent; mRNA reduced
GSM139530	PP15	Presymptomatic	1.2	M	Protein absent; mRNA reduced
GSM139531	PP16	Presymptomatic	1.3	M	Protein absent; mRNA reduced
GSM139532	PP17	Presymptomatic	1.7	M	Protein absent; mRNA reduced
GSM139533	PP18	Presymptomatic	1.7	M	Protein absent; mRNA reduced
GSM139534	PP19	Presymptomatic	1.8	M	Protein absent; mRNA reduced
GSM121357	SP20	Symptomatic	5.0	M	Protein absent
GSM121361	SP21	Symptomatic	8.0	M	Protein absent
GSM121363	SP22	Symptomatic	7.0	M	Protein absent
GSM121368	SP23	Symptomatic	8.0	M	Protein absent
GSM121369	SP24	Symptomatic	7.0	M	Protein absent
GSM139535	SP25	Symptomatic	2.3	M	Protein absent; mRNA reduced
GSM139536	SP26	Symptomatic	3.9	M	Protein absent; mRNA reduced
GSM139537	SP27	Symptomatic	5.1	M	Protein absent; mRNA reduced
GSM139501	C28	Control	0.4	M	Normal
GSM139502	C29	Control	0.5	F	Normal
GSM139503	C30	Control	0.5	F	Normal
GSM139504	C31	Control	0.5	M	Normal
GSM139505	C32	Control	0.5	M	Normal
GSM139506	C33	Control	0.6	M	Normal
GSM139507	C34	Control	0.7	F	Normal
GSM139508	C35	Control	0.9	M	Normal
GSM139509	C36	Control	1.5	M	Normal
GSM139510	C37	Control	2.8	M	Normal
GSM139511	C38	Control	3.0	M	Normal
GSM139512	C39	Control	4.2	M	Normal
GSM139513	C40	Control	5.0	M	Normal
GSM139513 GSM139514	C41	Control	9.0	M	Normal

Detection of differentially expressed genes

Entire data sets including CEL- and SOFT-formatted family files for the 41 samples from the two datasets (GSE6011 and GSE300741) were downloaded. Raw data from the CEL files, which were generated from satisfactory image files, were normalized via robust multi-array analysis (Irizarry et al., 2003) following 3 steps: first, background noise effects and processing artifacts were neutralized with model-based background correction. Second, quantile normalization was used to align expression values to a common scale. Third, data were summarized and an iterative median polishing procedure was used to generate a single expression

value for each probe set. The resulting log2-transformed robust multi-array analysis expression value was derived through probe set-level analysis from the raw CEL files.

Statistical *t*-tests with multiple test correction using the Benjamini and Hochberg procedure (Benjamini and Hochberg, 1995) were carried out for the symptomatic-presymptomatic pairs to detect differentially expressed genes with the threshold of significantly expressed genes set at 0.01. The differentially expressed genes were detected as follows: First, data from the symptomatic-presymptomatic pair of patients were used to detect differentially expressed genes. Second, data from controls were divided into 2 subsets according to age (older than 2 years and younger than 2 years). Normal age-related differentially expressed genes were detected in these sets. Finally, differentially expressed genes correlated with DMD progression were finalized as those detected in the symptomatic-presymptomatic patients pair but not identified in the control pair. Up- or down regulation of differentially expressed genes were determined with fold change. All of these procedures were carried out using R statistical software (v2.14.1) with BioConductor, limma packages (3.12.1), and libraries (Smyth et al., 2005).

PPI network analysis

We used Cytoscape (V 2.8.3; http://www.cytoscape.org/) and the Human Protein Reference Database (release date, 2009) (Keshava Prasad et al., 2009) to construct the network related to the DMD protein. Up to 3rd-level neighbors were allowed. First-level neighbors included proteins that interacted with the DMD protein directly. Second-level neighbors included proteins that interacted with 1st-level neighbors directly, whereas 3rd-level neighbors included proteins that interacted with 2nd-level neighbors directly. Differentially expressed proteins correlated with DMD progression identified as described above were checked to confirm their interaction with the DMD protein to characterize the PPI network changes that occurred with DMD progression.

RESULTS

Of the well-characterized human genes in the Affymetrix Human Genome U133A Array, 28 genes were found to be differentially expressed in the symptomatic-presymptomatic patient pairs. Eight of those genes were excluded because they were also differentially expressed in the 2 control pairs. The 20 remaining genes (Table 2) were considered to play important roles in the progression of DMD. An expression heatmap of the 20 differentially expressed genes in all samples is shown in Figure 1. According to the PPI network analysis, none of the proteins encoded by these 20 genes was found in the 1st-level neighbors of the DMD protein. In the 2nd-level network, the IQ motif containing GTPase-activating protein 1 (IQGAP1), which had a high degree of 21, was found. In the 3rd-level network, we found IQGAP1 protein and 4 additional proteins-microtubule-associated protein tau (MAPT), membrane metallo-endopeptidase, interleukin receptor 13 alpha 1, and multiple epidermal growth factor-like domains 6-with degrees of 7, 2, 1, and 1, respectively. Degree is the simplest topological index, corresponding to the number of proteins or nodes directly connected to a given node or protein. Proteins with higher degrees are connected to more proteins and tend to play important roles in body development and metabolism. Illustrations of the 2nd- and 3rd-level PPI networks are shown in Figures 2 and 3, respectively.

Symbol	Full name	P
ARHGAP32	Rho GTPase activating protein 32	1.92E-03
CCDC69	Coiled-coil domain containing 69	5.41E-03
CRTAP	Cartilage associated protein	1.76E-04
FAM149A	Family with sequence similarity 149, member A	1.98E-03
FRZB	Frizzled-related protein	1.02E-03
IL13RA1	Interleukin 13 receptor, alpha 1	7.52E-03
IQGAP1	IQ motif containing GTPase activating protein 1	7.72E-03
KCTD9	Potassium channel tetramerisation domain containing 9	9.94E-03
MAPT	Microtubule-associated protein tau	1.79E-03
MEGF6	Multiple EGF-like-domains 6	4.01E-03
MKKS	McKusick-Kaufman syndrome	6.82E-03
MME	Membrane metallo-endopeptidase	1.09E-03
MOSPD1	Motile sperm domain containing 1	1.79E-03
NPY6R	Neuropeptide Y receptor Y6 (pseudogene)	7.52E-03
OXCT1	3-oxoacid CoA-transferase 1	7.39E-03
PHF20	PHD finger protein 20	6.73E-03
PMS2P1	Postmeiotic segregation increased 2 pseudogene 1	7.53E-03
SLC15A2	Solute carrier family 15 (H+/peptide transporter), member 2	5.14E-03
SSPN	Sarcospan (Kras oncogene-associated gene)	7.39E-03
WISP2	WNT1 inducible signaling pathway protein 2	4.01E-03

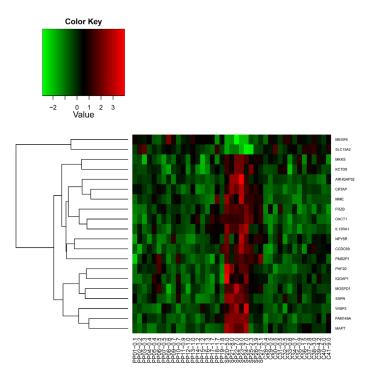


Figure 1. Heatmap for the expression of the 20 genes selected. Information for each sample corresponding to Table 1 are shown in the X axis format as: sample id-age. The 20 differentially expressed genes were combined and then hierarchically clustered to represent the expression patterns with average linkage and Euclidean distance as a measurement of similarity. The expression values were log ratios normalized according to the procedure described in the Methods section. Red and green represent upregulation and downregulation, respectively. Precise color scheme is illustrated in the color key.

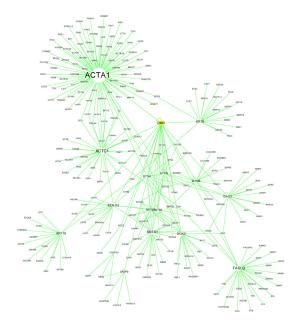


Figure 2. Second-level PPI network for dystrophin. First-level proteins refer to proteins interact with dystrophin directly. Second-level proteins refer to those interact with first-level proteins. This figure illustrate the interaction network for dystrophin, including proteins up to the second level.

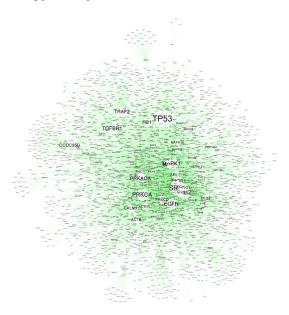


Figure 3. Third-level PPI network for dystrophin. First-level proteins refer to proteins interact with dystrophin directly. Second-level proteins refer to those interact with first-level proteins and third-level proteins refer to those interact with second-level proteins. This figure illustrates the interaction network for dystrophin, including proteins up to the third level.

DISCUSSION

The pathophysiology of DMD involves many secondary changes (Gorospe et al., 1994; Chen et al., 2000; Head, 2010; Altamirano et al., 2012; De Pasquale et al., 2012; Gehrig et al., 2012) owing to mutation in the dystrophin gene. This high complex progression of DMD presents challenges for curing the disease. Understanding the pathomechanisms for DMD from a network view may improve understanding of DMD and contribute potential targets for new therapeutic strategies because most proteins function with other proteins. Herein we detected differentially expressed genes that contribute to DMD progression using 2 data sets from the Gene Expression Omnibus database and characterized their interaction with the DMD protein using a network view.

As shown in Figure 1, 20 genes with differential expression only in SP and PP patient pairs were identified as genes contributing to the progression of DMD. As expected, the DMD gene did not show differential expression between the 2 phases because the DMD mutation exists even when patients are phenotypically indistinguishable from normal individuals. According to the network analysis, none of the proteins encoded by these 20 genes interacts with dystrophin directly, indicating that the biological function of the proteins that interact directly with dystrophin do not change substantially in the progression of DMD.

One of the identified differentially expressed genes encodes IQGAP1, which belongs to the 2nd-level neighbors of DMD and has a high degree of 21. No report about the relationship between IQGAP1 and DMD has been published until now. As a scaffolding protein, IQGAP1 binds directly to an impressive collection of other proteins (Mateer et al., 2003), including F-actin (or alpha actins encoded by *ACTA1* in skeletal muscle), which interacts directly with dystrophin. F-actin is polymerized from G-actin. A previous study (Prins et al., 2008) has shown that skeletal muscle-specific ablation of G-actin does not exacerbate the dystrophin-deficient phenotype in mice. Thus, the relationship with IQGAP1 and deterioration of DMD may not involve the dysfunction of actins. IQGAP1 may affect the progression of DMD through its interaction with Ca²⁺/calmodulin (Joyal et al., 1997; Ho et al., 1999) predominantly via its 4 IQ motifs, because Ca²⁺ and calmodulin modulate numerous cellular functions, including muscle contraction (Berridge et al., 2000). A previous study (Weissbach et al., 1998) has shown that the interaction of IQGAP1 and myosin essential light chain have significant consequences for actomyosin contractility, suggesting another possible mechanism for IQGAP1 in the progression of DMD.

Protein encoded by a gene called *MAPT* was found in the 3rd-level neighbors of dystrophin with a degree of 7. *MAPT* mutations have been associated with several neurodegenerative disorders, such as frontotemporal lobar degeneration (Sieben et al., 2012) and Parkinson disease (Farrer, 2006). Differential expression of MAPT may be related to the cognitive dysfunction of DMD patients; however, further research is needed for confirmation because previous reports have indicated that cognitive dysfunction in DMD patients is non-progressive (Bushby et al., 2010a), whereas the differential expression pattern existed between SP and PP in our study. Membrane metallo-endopeptidase was also found in the 3rd-level neighbors with a degree of 2. Its differential expression may contribute to the psychological state changes of the patients that accompany the progression of DMD because the function of this protein is related to the inactivation of enkephalins, which have been implicated in the regulation of mood, anxiety, reward, euphoria, and pain (Comings et al., 2000).

Among the other proteins that do not interact with dystrophin in up to three levels, sarcospan (SSPN) is especially notable. As a unique tetraspanin-like core component of the dystrophin- and utrophin-glycoprotein complexes in skeletal muscle, SSPN and its modulation of integrin signaling are necessary for extracellular matrix attachment and muscle force development (Marshall et al., 2012). Peter et al. (2008) have also reported that to compensate for the loss of dystrophin, SSPN may increase utrophin-glycoprotein complex levels at the extrasynaptic membrane to stabilize the sarcolemma. Thus, SSPN is worthy of further investigation and may be a potential target for DMD treatment.

In summary, we analyzed differentially expressed genes in the progression of DMD using a network view. The genes identified herein may help to understand the symptoms and especially the progression of DMD and at the same time spur new directions for therapeutic interventions, which are at present inadequate.

REFERENCES

Altamirano F, Lopez JR, Henriquez C, Molinski T, et al. (2012). Increased resting intracellular calcium modulates NF-kappaB-dependent inducible nitric-oxide synthase gene expression in dystrophic mdx skeletal myotubes. *J. Biol. Chem.* 287: 20876-20887.

Anderson JL, Head SI, Rae C and Morley JW (2002). Brain function in Duchenne muscular dystrophy. *Brain* 125: 4-13. Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. S B Met.* 57: 289-300.

Berridge MJ, Lipp P and Bootman MD (2000). The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol.* 1: 11-21.

Bushby K, Finkel R, Birnkrant DJ, Case LE, et al. (2010a). Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. *Lancet Neurol.* 9: 77-93.

Bushby K, Finkel R, Birnkrant DJ, Case LE, et al. (2010b). Diagnosis and management of Duchenne muscular dystrophy, part 2: implementation of multidisciplinary care. *Lancet Neurol.* 9: 177-189.

Chen YW, Zhao P, Borup R and Hoffman EP (2000). Expression profiling in the muscular dystrophies: identification of novel aspects of molecular pathophysiology. *J. Cell Biol.* 151: 1321-1336.

Comings DE, Dietz G, Gade-Andavolu R, Blake H, et al. (2000). Association of the neutral endopeptidase (MME) gene with anxiety. *Psychiatr. Genet.* 10: 91-94.

De Pasquale L, D'Amico A, Verardo M, Petrini S, et al. (2012). Increased muscle expression of interleukin-17 in Duchenne muscular dystrophy. *Neurology* 78: 1309-1314.

Dubowitz V (1978). Muscle disorders in childhood. Major Problem. Clin. Pediatr. 16: iii-xiii, 1-282.

Emery AE (2002). The muscular dystrophies. Lancet 359: 687-695.

Farrer MJ (2006). Genetics of Parkinson disease: paradigm shifts and future prospects. Nat. Rev. Genet. 7: 306-318.

Gehrig SM, van der Poel C, Hoeflich A, Naim T, et al. (2012). Therapeutic potential of PEGylated insulin-like growth factor I for skeletal muscle disease evaluated in two murine models of muscular dystrophy. *Growth Horm. IGF Res.* 22: 69-75.

Gorospe JR, Tharp MD, Hinckley J, Kornegay JN, et al. (1994). A role for mast cells in the progression of Duchenne muscular dystrophy? Correlations in dystrophin-deficient humans, dogs, and mice. *J. Neurol. Sci.* 122: 44-56.

Head SI (2010). Branched fibres in old dystrophic mdx muscle are associated with mechanical weakening of the sarcolemma, abnormal Ca²⁺ transients and a breakdown of Ca²⁺ homeostasis during fatigue. *Exp. Physiol.* 95: 641-656.

Ho YD, Joyal JL, Li Z and Sacks DB (1999). IQGAP1 integrates Ca²⁺/calmodulin and Cdc42 signaling. *J. Biol. Chem.* 274: 464-470.

Hoffman EP, Brown RH Jr and Kunkel LM (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 51: 919-928.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249-264.

Jennekens FG, ten Kate LP, de Visser M and Wintzen AR (1991). Diagnostic criteria for Duchenne and Becker muscular dystrophy and myotonic dystrophy. Neuromuscul. Disord. 1: 389-391.

- Joyal JL, Annan RS, Ho YD, Huddleston ME, et al. (1997). Calmodulin modulates the interaction between IQGAP1 and Cdc42. Identification of IQGAP1 by nanoelectrospray tandem mass spectrometry. *J. Biol. Chem.* 272: 15419-15425.
- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, et al. (2009). Human Protein Reference Database 2009 update. *Nucleic Acids Res.* 37: D767-D772.
- Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, et al. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* 50: 509-517
- Kunkel LM, Monaco AP, Hoffman E, Koenig M, et al. (1987). Molecular studies of progressive muscular dystrophy (Duchenne). *Enzyme* 38: 72-75.
- Lindahl M, Backman E, Henriksson KG, Gorospe JR, et al. (1995). Phospholipase A2 activity in dystrophinopathies. Neuromuscul. Disord. 5: 193-199.
- Marshall JL, Chou E, Oh J, Kwok A, et al. (2012). Dystrophin and utrophin expression require sarcospan: loss of alpha7 integrin exacerbates a newly discovered muscle phenotype in sarcospan-null mice. *Hum. Mol. Genet.* 21: 4378-4393.
- Mateer SC, Wang N and Bloom GS (2003). IQGAPs: integrators of the cytoskeleton, cell adhesion machinery, and signaling networks. *Cell Motil. Cytoskeleton* 55: 147-155.
- Pescatori M, Broccolini A, Minetti C, Bertini E, et al. (2007). Gene expression profiling in the early phases of DMD: a constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression. *FASEB J.* 21: 1210-1226.
- Peter AK, Marshall JL and Crosbie RH (2008). Sarcospan reduces dystrophic pathology: stabilization of the utrophinglycoprotein complex. *J. Cell Biol.* 183: 419-427.
- Prins KW, Lowe DA and Ervasti JM (2008). Skeletal muscle-specific ablation of gamma(cyto)-actin does not exacerbate the mdx phenotype. *PLoS One* 3: e2419.
- Sieben A, Van Langenhove T, Engelborghs S, Martin JJ, et al. (2012). The genetics and neuropathology of frontotemporal lobar degeneration. *Acta Neuropathol.* 124: 353-372.
- Smyth GK, Michaud J and Scott HS (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21: 2067-2075.
- Stelzl U, Worm U, Lalowski M, Haenig C, et al. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122: 957-968.
- Weissbach L, Bernards A and Herion DW (1998). Binding of myosin essential light chain to the cytoskeleton-associated protein IQGAP1. *Biochem. Biophys. Res. Commun.* 251: 269-276.