Duchenne muscular dystrophy gene expression is an independent prognostic marker for IDH mutant low-grade glioma

Michael Naidoo[1], Leanne Jones[1], Benjamin Conboy[1], Wael Hamarneh[2], Darwin D'Souza[3], Karen Anthony[1^], Lee R Machado[^1,4]

[1]Centre for Physical Activity and Life Sciences, University of Northampton, University Drive, Northampton, NN1 5PH, United Kingdom. [2]Northampton General Hospital NHS Trust, Northampton, NN1 5BD, United Kingdom. [3]Icahn School of Medicine at Mount Sinai, United States. [4]Department of Genetics and Genome Science, University of Leicester, Leicester LE1 7RH, UK. ^These authors share senior authorship

Corresponding authors: Lee R Machado (lee.machado@northampton.ac.uk) and Karen Anthony (karen.anthony@northampton.ac.uk), University of Northampton, University Drive, Northampton, NN1 5PH, United Kingdom. Tel ++44 1604 89 3476 or 3467

Abstract

Alterations in the expression of the Duchenne muscular dystrophy (DMD) gene have been associated with the development, progression and survival outcomes of numerous cancers including tumours of the central nervous system. We undertook a detailed bioinformatic analysis of low-grade glioma (LGG) bulk RNAseq data to characterise the association between *DMD* expression and LGG survival outcomes. High *DMD* expression was significantly associated with poor survival in LGG with a difference in median overall survival between high and low *DMD* groups of over seven years (P= <0.0001). In a multivariate model, *DMD* expression remained significant (P=0.02) and was an independent prognostic marker for LGG. The effect of *DMD* expression on overall survival was only apparent in isocitrate dehydrogenase (IDH) mutant cases where non-1p/19q co-deleted LGG patients could be further stratified into high/low *DMD* groups. Patients in the high *DMD* group had a median overall survival time almost halve that of the low *DMD* group. The expression of the individual *DMD* gene products Dp71, Dp71ab and Dp427m were also significantly associated with overall survival in LGG which have differential biological effects relevant to the pathogenesis of LGG. Differential gene expression and pathway analysis identifies dysregulated biological processes relating to ribosome biogenesis, synaptic signalling, neurodevelopment, morphogenesis and immune pathways. Genes spanning almost the entirety of chromosome 1p are upregulated in patients with high overall *DMD,* Dp71 and Dp427m expression which worsens survival outcomes for these patients. We confirmed dystrophin protein is variably expressed in LGG tumour tissue by immunohistochemistry and, overall, demonstrate that *DMD* expression has potential utility as an independent prognostic marker which can further stratify IDH mutant LGG to identify those at risk of poor survival. This knowledge may improve risk stratification and management of LGG.

Introduction

The Duchenne muscular dystrophy gene (*DMD*) is located within a common fragile site on the X chromosome and at >2 Mb in size is one of the largest known human genes[1,2]. The major protein product of the gene is the 427 kDa dystrophin protein whose primary function is to maintain skeletal muscle integrity by connecting the actin cytoskeleton to the extracellular matrix[3]. The resulting dystrophin-associated protein complex (DAPC) is essential for muscle function. Smaller dystrophin proteins (Dp) are produced from the *DMD* gene through independent promoter usage and are named according to their size in kDa[3]. Dp71 is a ubiquitous dystrophin protein with prominent neuronal and glial expression throughout the cortex and cerebellum at all stages of development[4]. Dp71 is alternatively spliced to generate four major isoforms which have a large functional diversity including neuronal differentiation, adhesion, cell division, excitatory synapse organisation, nuclear scaffolding and DNA repair[4,5].

Mutations in the *DMD* gene can cause Duchenne or Becker muscular dystrophy depending on whether the reading frame, and therefore dystrophin protein expression, is disrupted[3]. Besides this major role in muscular dystrophy, a growing body of evidence suggests that *DMD* gene mutations and/or changes in *DMD* gene expression are associated with the development, progression and survival outcomes of both myogenic and non-myogenic cancers including tumours of the central nervous system[6–9]. In a study by Luce *et al. DMD* ranked within the top 10% of differentially expressed genes in several tumour versus normal tissue comparisons. In the nervous system, *DMD* was significantly overexpressed in ependymomas and astrocytomas, but was significantly under-expressed in medulloblastoma and non-significantly under-expressed in glioblastoma[8]. In our own preliminary study to determine the effect of high versus low *DMD* gene expression on survival outcomes across all cancers, we found that low-grade glioma (LGG) returned one of the highest hazard ratios. In agreement with Luce *et al.* we posited that these findings may be attributed to Dp71 given it is the predominant *DMD* gene product in the brain and has known roles in proliferation, invasion and migration[5,7,10]. We therefore aimed to undertake the first dedicated study of the involvement of *DMD* and its individual gene products in LGG.

According to the revised 2016 WHO criteria for the classification of central nervous system tumours[11], LGG comprise grade I and II astrocytomas and oligodendrogliomas. A not otherwise specified (NOS) category is also permissible for e.g. a prior diagnosis of oligoastrocytoma which is no longer considered a distinct subtype. Grade III and IV gliomas are considered high-grade, the latter includes the most aggressive type known as glioblastoma, or glioblastoma multiforme (GBM). LGGs account for approximately 11% of all primary brain tumours with a peak incidence between 35 and 44 years of age and an average mean survival time of seven years[11,12]. In most cases LGGs progress to fatal high-grade tumours within 5-10 years[13]. Treatment and management is difficult due to uncertainty over progression and lack of best practice[13]. There is an urgent unmet scientific and clinical need to characterise the mechanisms responsible for LGG development and progression. Genotypic features and prognostic markers such as isocitrate dehydrogenase (IDH) mutation status as well as the presence/absence of the chromosomal 1p/19q co-deletion are used to aid the diagnosis and management of LGG[11]. Both IDH mutation and 1p/19q co-deletion are predictive of better survival but the detailed mechanisms are unknown[14]. Recent changes in surgical practice have improved survival rates with early resections now considered part of

the overall standard of care for LGG[15,16]. However, recurrence remains common and is unpredictable with significant variability across tumour subtypes which presents a continued dilemma for treatment and management[17].

We present, for the first time, an in-depth bioinformatic analysis of the association of *DMD* gene expression with overall survival in LGG. We show that *DMD* is an independent prognostic marker for LGG where high expression is linked to poor overall survival in IDH mutant LGG. Furthermore, we demonstrate that the expression of multiple *DMD* gene products, including Dp71, are also associated with LGG survival outcomes and have differential biological effects relevant to the pathogenesis of LGG.

Results

High *DMD* gene expression is associated with poor survival in low-grade glioma and is an independent prognostic marker

To explore the association of *DMD* gene expression in LGG, we analysed RNAseq data from a LGG (WHO grade II) TCGA dataset using an independent bioinformatic pipeline and cut-point approach (Fig. 1). We confirm in univariate analyses that high *DMD* is associated with poor survival in LGG (HR: 4.15; 95%CI: 1.46, 11.81; P= <0.0001). The median overall survival for the high *DMD* group was 36.79 months compared to 130.7 months for the low *DMD* group, a difference of over seven years. We performed a tumour subtype analysis to determine whether the survival associations are limited to or differ between subtypes (Fig. 1). Univariate subtype analysis revealed that high *DMD* is associated with poor survival across all subtypes which was significant for all except astrocytoma (A) which constituted the smallest number in the cohort (n= 66). Oligodendroglioma had the highest increased risk of poor survival, though with a wide confidence interval (HR: 9.80; 95%CI: 0.56, 170.6; P= <0.0001). Multivariate analysis including tumour subtype as a factor revealed that *DMD* expression remained significant whilst tumour subtype did not (P=0.005, supplementary Table S1). The effect of *DMD* gene expression on glioma survival was specific to only WHO grade II LGG; high-grade (grade III) anaplastic tumours and high-grade (grade IV) glioblastoma TCGA datasets did not result in any significant survival differences when stratified by high/low *DMD* expression (Supplementary Fig. S1). To validate these findings, RNAseq data from the Chinese Glioma Genome Atlas (CGGA) was taken as an independent dataset and replication cohort (Supplementary Fig. S2). Whilst *DMD* expression is associated with survival of low-grade II LGG in the TCGA dataset, the CGGA dataset replicated this finding for only for high-grade III LGG. Overall however, the CGGA data does confirm that *DMD* expression is not significantly associated with survival in the most invasive and aggressive grade IV glioblastoma.

We next performed a multivariate analysis to determine whether *DMD* expression remains significantly associated with survival in a model containing significant clinicopathological variables. In the cohort we used, age at diagnosis and IDH mutation status both had independent Kaplan-Meier statistics of p= <0.01 and were significantly associated with survival in agreement with existing literature[14,18,19] (Supplementary Fig. S3). Note the TCGA dataset splits IDH mutated patients into 1p/19q co-deleted and non-co-deleted groups since this chromosomal co-deletion is also known to have a strong effect on LGG survival outcomes[14,20]. *DMD* expression remained significant in a multivariate analysis with age at diagnosis (<41 years; >41 years) and IDH mutation/co-deletion status (wild-type; IDH mutant co-deleted and IDH mutant non-co-deleted) and is therefore independently prognostic of these for LGG (Supplementary Fig. S3, P= 0.02; HR: 2.12; 95%CI: 1.13, 3.99). IDH mutation status also remained significant in this model but age at diagnosis did not. This provides strong confirmation of a potential role for *DMD* in LGG tumourigenesis and/or disease progression that warrants further investigation.

***DMD* gene expression further stratifies IDH mutant LGG**

Since both IDH mutation status and 1p/19q co-deletion are used clinically to aid prognosis we also investigated, using pairwise comparisons of Kaplan-Meier curves, the survival outcomes of high vs low *DMD* expression in all IDH and co-deletion status groups (Fig. 2). Low *DMD* expression offers a particularly significant survival advantage for IDH mutated patients

regardless of 1p/19q co-deletion status compared with high *DMD* (Fig.2a). In non-1p/19q co-deleted patients, the difference in median survival between high and low *DMD* groups was 67.25 months (>5 years, P=0.0039, Fig. 2b and c). The median survival was undefined for the high *DMD* co-deleted group which contained only four patients, but the same trend is still observed between high and low *DMD* for 1p/19q co-deleted patients (P=0.0089). *DMD* expression was not significantly associated with survival in IDH wild-type patients with both *DMD* high and *DMD* low groups having an overall median survival of approximately 20 months (P=0.8965); this was also replicated in the CGGA dataset (Supplementary Fig. S4). Overall this data demonstrates that *DMD* expression further stratifies IDH mutant patients (p=0.0039 for non-1p/19q co-deleted patients and p =0.0089 for co-deleted patients). Interestingly, high *DMD* expression appears very rare in co-deleted patients (4% [4/107] co-deleted vs 34% [55/161] non-co-deleted), this was also replicated in the CGGA cohort where only 3% (2/58) of co-deleted patients had high *DMD* expression compared to 31% (31/101) of non-co-deleted patients.

  Given the above findings and the established link between 1p/19q co-deletion and LGG survival, we detected the differentially expressed genomic regions in high versus low *DMD* samples using Position RElated Data Analysis (PREDA[21]). Many statistically significant chromosomal regions with expression changes among neighbouring genes were identified (Fig. 3). Notably, genes covering the whole of the p-arm of chromosome 1 are up-regulated as well as many on chromosome 19q. Thus, 1p/19q genes are upregulated in patients with high *DMD* expression which worsens survival outcomes for these patients. High *DMD* expression almost halved the overall median survival of non-1p/19q co-deleted patients we observed in Fig. S3a from 105 to 64 months (>3-year difference) whilst low *DMD* expression maintained the more favourable median survival time (131 months) expected of this group.

Network and pathway analysis of differentially expressed genes (DEGs) implicates development and ribosome biogenesis

To aid future investigation into the functional role(s) of the *DMD* gene in LGG tumourigenesis we undertook a detailed bioinformatic analysis of the DEGs in the TCGA LGG (grade II) cases with high versus low *DMD* expression. We used an integrated web application, iDEP, for data pre-processing and to identify the DEGs (DESeq2 method). The volcano and M-A plots show a subtle transcriptomic response (Fig. 4a). A total of 106 DEGs were identified. These included 32 down-regulated genes and 74 up-regulated genes (Fig. 4a, Supplementary Data S1). The Search Tool for the Retrieval of Interacting Genes (STRING) was used to identify the protein-protein interaction (PPI) network connectivity of the 106 DEGs (Fig. 4b). The resulting PPI network contained 101 nodes forming a core of 47 highly connected proteins with 60 edges, the PPI enrichment p-value was <0.0001 (7.92e-13) using a median confidence score of 0.4. The expected number of edges for a random set of proteins of similar size was 20 which strongly suggests functional intersection of the identified DEGs. STRING additionally detected that the homeobox domain is overrepresented in the proteins encoded by the down-regulated genes hinting at a role for these DEGs in embryonic development. To identify hub genes within the DEGs, we used the Cytohubba plugin on Cytoscape. The top 20 nodes were ranked by the Maximal Clique Centrality (MCC) and Density of Maximum Neighbourhood Component (DMNC) algorithms (Fig. 4c). The list of hub genes from each

algorithm are displayed (Fig. 4d) where 13 genes were common to both algorithms, several of which do indeed have roles in development.

To examine the functional annotation of the DEGs we used the enrichment analysis (gene ontology [GO] biological processes) for DEGs tool in iDEP (Fig. 4e-f). The results are in line with the functions of the identified hub genes in Cytoscape. The up-regulated genes are enriched in processes relating to development and morphogenesis; the down-regulated genes are also enriched in aspects of development including neurodevelopment and processes relating to muscle contraction. As well as analysing the function of the DEGs, we performed a separate pathway analysis using the fold-change values of all the genes in our dataset to identify coherently altered pathways upon high vs low *DMD* expression. We used the generally applicable gene set enrichment (GAGE) method within iDEP and the genes were annotated either according to GO biological processes or with Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Fig. 5). The most significant result returned by the KEGG annotation was a down-regulation of the ribosome pathway which is reflected also in the GO biological processes annotation which returned high significance for terms such as ribosome biogenesis. Thus, in high vs low *DMD* cases, ribosome pathways are downregulated and selected DEGs are enriched in processes relating to development.

We additionally assessed the co-expression behaviour of the top 1000 variable genes across all samples using weighted gene co-expression network analysis (WGCNA, Fig. 6). This analysis identifies groups of genes whose expression levels are similar and tend to co-activate. A consensus LGG network of 882 genes that are correlated in all samples was divided into five modules (Supplementary Data S2) and the network of the top 20 genes displayed for each module (Fig. 6a-b). GO enrichment analyses was performed on the five modules (Fig. 6c). The 402 turquoise module genes are significantly involved in biological processes related to cell signalling, synaptic signalling and transmission and nervous system development. The 364 blue module genes are involved in many aspects of embryonic, tissue and cell morphogenesis and development. The 52 brown module genes are involved in myelination, gliogenesis and glial cell differentiation. The 36 yellow module genes are involved in mitosis, nuclear division and regulation of the cell cycle. Finally, the 28 green module genes are involved in inflammatory response, chemotaxis and response to cytokine stimulus. It is to be noted that the large turquoise and blue modules are less specific in nature with the blue module having a lower degree of co-expression. The DEGs identified above were well represented with 55 DEGs contained within the entire network (supplementary table S2). Venn analysis reveals that 44 of the DEGs are found in the blue module linked to morphogenesis and development; the remaining DEGs are found in the turquoise (10 DEGs) and green (1 DEG) modules (Fig. 6d). Our WGCNA analysis confirms a role for the DEGs identified above in development and in the same biological processes as the commonly co-expressed genes in LGG. WGCNA additionally highlights relevant and specific genes involved in gliogenesis and glial cell differentiation (brown module) and cell growth and division (yellow module). In summary, high *DMD* expression is significantly linked to poor survival outcomes in IDH mutant LGG. In patients with high *DMD* expression, ribosome pathways are downregulated and selected DEGs are enriched in processes relating to morphogenesis and development.

The association of *DMD* gene expression with LGG survival is replicated by multiple *DMD* gene products

The *DMD* gene is complex. As well as the major full-length 427 KDa dystrophin essential for muscle function, several other *DMD* gene variants are produced from independent promoters, many of which are themselves alternatively spliced to produce multiple isoforms[3]. Although literature has previously linked the *DMD* gene to numerous cancers[6], none have considered the many gene products produced by the *DMD* gene and the extent of their contributions to pathogenesis and associated survival outcomes. To begin to identify the relative contribution of individual *DMD* gene products we extracted RNAseq isoform data from the same TCGA cases analysed above. We plotted individual *DMD* gene product expression levels for all patients and determined that the four major Dp71 isoforms (Dp71, Dp71a, Dp71b and Dp71ab) and Dp427m are the most abundant in LGG tissue (Supplementary Fig. S5). There is a weak positive correlation between their expression and total *DMD* expression; this was significant for Dp71a, Dp71ab and Dp427m (Supplementary Fig. S5). We therefore sought to determine the expression of which Dp71 isoform(s) and/or Dp427m gene products were associated with LGG survival. To test this, we repeated our analysis above using individual *DMD* gene product expression data (Fig. 7a). We found that high expression of the Dp71, Dp71ab and Dp427m gene products were significantly associated with poor LGG survival. These findings were replicated across the oligodendroglioma subtype accounting for multiple testing (Supplementary Fig. S6). We moved forward with a detailed analysis of Dp71, Dp71ab and Dp427m since they were significant across the whole LGG cohort.

In a multivariate analysis with the significant clinicopathological variables of age at diagnosis (<41 years; >41 years) and IDH mutation status (wild-type; IDH mutant co-deletion and IDH mutant non-co-deletion), none of the individual gene products remained significant (Supplementary Tables S3-6). However, pairwise comparisons of the Kaplan-Meier curves for all IDH and co-deletion status groups show that in keeping with our findings for *DMD* expression above, low Dp71, Dp71ab and Dp427m expression offers a survival advantage for IDH mutated patients compared with high expression (Fig. 7b). As with *DMD,* the expression of Dp71, Dp71ab and Dp427m were clearly not associated with survival in IDH wild-type patients with each *DMD* low and *DMD* high group having an overall median survival time of approximately 20 months (e.g. P=0.9038 for Dp71). Thus, high Dp71, Dp71ab and Dp427m are significantly associated with poor survival in IDH mutant LGG but do not offer any additional prognostic utility beyond that of total *DMD* expression.

High expression of individual *DMD* gene products have differential biological effects relevant to the pathogenesis of LGG.

We repeated DEG and pathway analysis on the TCGA RNAseq dataset configured for either high vs low Dp71, Dp71ab or Dp427m expression (Fig. 8). The number of DEGs identified for high vs low Dp71 was 167; 30 of which are in common with those identified for high vs low *DMD* expression (Fig. 8a). A more extensive transcriptomic response was observed with high Dp71ab and Dp427m expression where 422 and 617 DEGs were identified with 38 and 55 in common to *DMD*, respectively. The full list of DEGs are provided (Supplementary Data S1) and the PPI networks for each set of DEGs (Fig. 8b) illustrate very large and extensively

connected networks for Dp71ab (380 nodes, 2205 edges [497 expected], PPI enrichment value p=<1.0e-16) and Dp427m (549 nodes, 2856 edges [1058 expected], PPI enrichment value: p= <1.0e-16) in comparison to Dp71 (147 nodes, 78 edges [35 expected], PPI enrichment value p =3.84e-10). The hub genes identified for each network are listed in Supplementary Table S7 and the pathways enriched in DEGs in Fig. 8c, note no significant enrichment was found for the Dp71 DEGs. There is similarity between Dp427m and Dp71ab DEGs (214 in common), with both sets of DEGs significantly enriched in synaptic signalling and nervous system development pathways. These pathways were also significantly altered between high vs low *DMD* expression groups and were also returned in the WGCNA of the whole LGG dataset.

Pathway analysis of all the genes, independent of DEGs, identifies three pathway clusters differentially altered depending on which *DMD* gene product is highly expressed (Fig. 8d). High *DMD* and high Dp71 expression significantly down-regulates ribosome biogenesis and protein targeting pathways whilst high Dp71ab and Dp427m expression significantly down-regulates pathways related to synaptic signalling. High Dp427m expression additionally up-regulates immune and leukocyte cell pathways. Some of these pathways were also featured in the WGCNA modules (Fig. 6).

PREDA on the high vs low Dp71 and Dp427m datasets resulted in a similar level of expression changes across different chromosomal regions to that observed with high vs low *DMD*, including a significant up-regulation of genes across a large region of chromosome 1p (Supplementary Fig. S7). However, when we conducted PREDA on the high vs low Dp71ab dataset, no significant regions were found. In summary, our data indicates that the expression of multiple *DMD* gene products are linked to overall survival in LGG which have differential biological effects relevant to the pathogenesis of LGG.

Dystrophin is expressed in the cytoplasm and nucleus of glial cells in LGG
To confirm whether dystrophin protein is expressed in LGG tissue, and in which cell types, we conducted a pilot immunohistochemistry study on a cohort of 24 LGG cases (18 astrocytoma, one oligodendroglioma and five NOS) using a C-terminal dystrophin antibody which detects all dystrophin proteins. Nine out of the 24 cases were IDH mutant and 15 were IDH wild-type. We observed dystrophin expression in both the cytoplasm and nucleus of glial cells within LGG (Fig. 9). Overall dystrophin staining was most predominant in the nucleus. Notably, the number of positively stained cells, as well as the intensity of nuclear, cytoplasmic and neuropil staining, was variable among cases but the cohort is too small to provide a meaningful survival analysis. These results confirm the feasibility of assessing the association of dystrophin protein expression with survival in future studies of large LGG cohorts.

Discussion

Our results identify a new subset of IDH mutant LGG patients who have a significantly poor survival outcome. IDH mutant LGG patients with high *DMD* expression in their tumours survive approximately six years less than those with low expression; this appears to be particularly relevant for non-1p/19q co-deleted patients since high *DMD* expression was exceptionally rare amongst 1p/19q co-deleted cases. In non-co-deleted patients, survival outcomes are already poorer than for co-deleted patients[20] and we demonstrate that if high *DMD* expression occurs then the outcome is even worse. Interestingly, we found that 1p/19q genomic regions are co-ordinately up-regulated in LGG cases with high *DMD* expression which may contribute to pathogenesis. Whilst we have confirmed dystrophin protein is expressed in LGG tumour tissue, further work is needed to experimentally confirm our findings in for e.g. a large immunohistochemical cohort as well as to investigate the potential involvement of *DMD* gene product(s) in LGG tumourigenesis. It is unknown for example whether the *DMD* gene plays a driver role in the cancers it has been linked to, or whether what we (and others[6]) have described are passenger effects. Determining the prognostic utility of *DMD* gene expression and the drug target potential of specific *DMD* gene products may lead to improved risk stratification and the development of new therapeutic strategies for LGG. The Dp71 isoforms have not been studied in low grade glioma before and studies on *DMD* in cancer have not fully considered the complexity of the gene and the roles its individual gene products may play.

Of note, in our CGGA replication cohort, the above findings were replicated in a different grade of tumour (high-grade III versus low-grade II). The decision by neuropathologists as to whether to score a tumour grade II or grade III is difficult and very subjective; practices may vary world-wide. Whilst both cohorts are well matched in terms of age of onset, sex and IDH mutation status, there may also be additional population-specific differences and/or mutations that might be relevant for stratifying the patients before considering *DMD* as a marker for prognosis. We think it of importance however, that in our study both cohorts confirmed that *DMD* expression is not significantly associated with survival in the most invasive and aggressive grade IV glioblastoma.

Our findings extend those of Luce *et al.* who reported in 2017 that *DMD* is significantly overexpressed in astrocytoma and non-significantly underexpressed in glioblastoma[8]. One study has linked Dp71 to high-grade glioblastoma where Dp71 expression appears decreased in glioblastoma cell lines and inversely correlated with the Ki-67 tumour proliferative index in tumour tissue[22]. However, the Ki-67 index is considered a poor predictor of survival in glioblastoma[23] and further work is needed to determine the true effect (if any) of altered Dp71 expression on survival outcomes in high-grade glioblastoma. Our data from two independent cohorts suggests that overall *DMD* expression is significantly associated with survival outcomes in only the less invasive, IDH mutant, gliomas indicating a likely role for *DMD* in the early stages of gliomagenesis.

Both Dp427m and the Dp71 isoforms form DAPCs in the brain[4]. Other components of this complex such as the dystroglycans have also been linked to the progression of primary brain tumours including gliomas[24,25]. Day *et al.* demonstrated that dystroglycan plays a role in maintaining tumour supporting glioma stem cells in the extracellular matrix (ECM)-rich

perivascular niche. Surface markers on these glioma stem cells are strong independent prognostic markers of low-grade glioma progression and survival[26]. Thus, considering our own data implicating ECM-linked dystrophin proteins in low-grade glioma, an area of future investigation could be whether higher levels of these structural protein(s) more effectively support the LGG tumour microenvironment.

We demonstrate that high *DMD* expression is coupled with a significant down-regulation of ribosome pathways including biogenesis and that the DEGs are enriched in biological processes relating to development and morphogenesis. Malignant gliomas recapitulate steps in neurodevelopment to form organ-like structures[27] and high *DMD* expression may support such a strategy. *TP53* mutation and *ATRX* loss are characteristic of IDH mutant, non-1p/19q co-deleted LGG[11]. *TP53* encodes the tumour suppressor p53 which is tightly linked to ribosome activity[28] and *ATRX* encodes a chromatin remodelling protein essential for development. Thus, high *DMD* expression effects biological pathways relevant to the functions of known LGG biomarkers as well as overall cellular functions related to LGG tumourigenesis. Our analysis does not implicate a single *DMD* gene product but rather we observed an overall comparable effect of high expression on survival across both full length and short *DMD* gene products. There were however some differences in the DEGs and pathway analyses between the individual gene products, namely there was a more widespread transcriptomic response when Dp71ab and Dp427m were highly expressed than observed for high Dp71 or overall *DMD* expression. Dp427m expression was also the only gene product associated with an alteration of biological process related to the immune system.

Overall, our data identifies *DMD* expression as an independent prognostic marker for LGG and highlights a potentially important role for *DMD* gene product(s) in the progression of low-grade glioma. This knowledge may help reduce and manage the unpredictable nature of LGG progression and recurrence by improving risk stratification.

Methods

RNAseq and clinical datasets

The Cancer Genome Atlas (TCGA) LGG dataset and corresponding clinical data was downloaded from cBioPortal after querying for the *DMD* gene (Brain Lower Grade Glioma, TCGA, PanCancer Atlas 2018). The data extracted was mRNA expression, RSEM (batch normalized from Illumina HiSeq_RNASeqV2). The replication dataset was downloaded from the Chinese Glioma Genome Atlas (CGGA), dataset ID: mRNAseq_693. The Firebrowse portal was used to extract TCGA RNASeq isoform data from the LGG mRNASeq archives (illuminahiseq_rnaseqv2-RSEM_isoforms_normalized MD5) and the case IDs matched to those obtained from cBioPortal. *DMD* transcript IDs were matched to specific transcripts using the table browser tool from the UCSC genome browser. We note however, that the transcript expression estimates by RSEM may not be 100% accurate. Note the TCGA PanCancer Atlas 2018 dataset downloaded from cBioPortal contains anaplastic grade III cases which are identifiable by cross matching with clinical data from Firebrowse. Unless otherwise stated, only the grade II (non-anaplastic) cases were used in our analyses.

Cut-point selection and survival analysis

X-tile (version 3.6.1, Yale University 2003-2005[29]) was used to dichotomise the datasets into high and low *DMD* (or individual gene product) expression groups using a minimal p-value approach. The optimal cut-point value (defined as the brightest pixel on the X-tile plot of chi-squared log-rank values) generated by X-tile was used for survival analysis. A cut-point was generated for each gene product and used across all tumour subtypes. The cut-point used for the TCGA total *DMD* expression dataset was 1183.5 RSEM. The cut-points for Dp427m, Dp71, Dp71a, Dp71b and Dp71ab were 312.1, 278.6, 465.5, 0.01 and 336.2 RSEM respectfully. The cut-point for the CGGA total *DMD* expression dataset was 8.0 FPKM. Plots showing the variability of low versus high expression are presented in Supplementary Fig. S8. Kaplan-Meier curves were analysed using the log-rank test in GraphPad and multivariate Cox regression analysis was conducted in SPSS. Age at diagnosis data was split into young and old groups using the mean (41 years). Unless otherwise stated, significance was set at 0.05 and asterisks used to indicate the level of significance: *P = ≤ 0.05, **P = ≤ 0.01, ***P = ≤ 0.001 and ****P = ≤ 0.0001. For multiple pairwise comparisons the alpha value was set at 0.017 correcting only for the comparisons that were planned (three high vs low groups).

Identification of differentially expressed genes (DEGs)

We used the integrated web application, iDEP 0.93[30], hosted at http://ge-lab.org/idep/ for data pre-processing and log transformation of normalised expression values. The full genomic TCGA data was downloaded from cBioPortal, case IDs were matched to the *DMD* (or *DMD* gene product) expression data and formatted into low and high groups using the cut-points derived from X-tile. iDEP encompasses many R packages for bioinformatic analysis of RNAseq data. DEGs were identified using the DESeq2 method within iDEP (false discovery rate (FDR) cut-off of 0.05 and a minimum fold-change of 2). Heatmaps, volcano and M-A plots were generated in iDEP. Functional enrichment analysis of DEGs was performed in iDEP using

gene ontology (GO) biological processes, enrichment trees and networks were generated in iDEP.

Protein-protein interaction analysis

The online Search Tool for the Retrieval of Interacting Genes (STRING) was used to identify the protein-protein interaction (PPI) network connectivity of DEGs. The minimum required interaction score was set to median confidence at 0.4. The resulting network files were imported to Cytoscape v3.8.2 and analysed using the Cytohubba plugin. The top 20 nodes were ranked by the Maximal Clique Centrality (MCC) and Density of Maximum Neighbourhood Component (DMNC) algorithms and the extended subnetwork displayed. Venn diagrams were created using an online tool at http://bioinformatics.psb.ugent.be/webtools/Venn/.

Pathway analysis

Pathway analysis for high-low comparisons was performed in iDEP using the generally applicable gene set enrichment (GAGE) method and the genes were annotated according to GO biological processes and Kyoto Encyclopaedia of Genes and Genomes (KEGG). The minimum and maximum gene set sizes were set to 15 and 2000 respectively, the pathway significance cut-off (FDR) was set to 0.2, and the top 30 pathways were retrieved. Pathway trees were created in iDEP.

Co-expression network construction and functional enrichment

Weighted gene co-expression network analysis (WGCNA) was performed in iDEP. The top 1000 variable genes were included with a soft threshold of five and a minimum module size of 20 genes. Module networks of the top 20 genes in each module were created in the network tab of iDEP using an edge threshold of 0.4. Functional enrichment analysis was performed using GO biological processes in iDEP and the resulting enriched pathway tables for each module were exported. The adjusted P values were converted to -log10 and a heatmap created in GraphPad.

Analysis of regional variations of genomic features

Position RElated Data Analysis (PREDA) was conducted within iDEP using a minimum FDR of 0.01 and minimum statistic of 0.5.

Immunohistochemistry

Haematoxylin and eosin (H&E) and unstained formalin-fixed paraffin-embedded (FFPE) sections from 24 LGG cases were obtained from University Hospital Southampton NHS Foundation Trust as part of BRAIN UK and under the extended ethical approval of the South Central - Hampshire B Research Ethics Committee (REC reference: 19/SC/027, IRAS project ID: 262890). The cohort included 18 cases of astrocytoma, one case of oligodendroglioma and five cases of 'not otherwise specified'. Nine out of the 24 cases were IDH mutant and 15 were IDH wild-type as previously determined by the providing centre. Sections were deparaffinised

with xylene and rehydrated through graded alcohol. Antigen retrieval was achieved by immersing sections in pH 6.0 citrate buffer and heating in a 800W microwave for 10 minutes at high power and 10 minutes at low power. Immunohistochemistry was performed using a C-terminal anti-dystrophin antibody (Abcam 15277) at a 1:100 dilution followed by a Tyramide SuperBoost™ kit (Invitrogen) according to manufactures' instructions. Staining was visualised using 3,3'-diaminobenzidine tetra hydrochloride (DAB), and counterstained with haematoxylin.

Data availability statement

All datasets underpinning this publication are openly available from the University of Northampton Research Explorer at doi: 10.24339/ae20690c-1bdf-48b0-8c5d-c88fbc40ad4a

References

1. McAvoy, S., Ganapathiraju, S., Perez, D. S., James, C. D. & Smith, D. I. DMD and IL1RAPL1: Two large adjacent genes localized within a common fragile site (FRAXC) have reduced expression in cultured brain tumors. *Cytogenet. Genome Res.* **119**, 196–203 (2007).

2. McAvoy, S. *et al.* Non-random inactivation of large common fragile site genes in different cancers. *Cytogenet. Genome Res.* **118**, 260–269 (2007).

3. Muntoni, F., Torelli, S. & Ferlini, A. Dystrophin and mutations: One gene, several proteins, multiple phenotypes. *Lancet Neurology* **2**, 731–740 (2003).

4. Naidoo, M. & Anthony, K. Dystrophin Dp71 and the Neuropathophysiology of Duchenne Muscular Dystrophy. *Molecular Neurobiology* **57**, 1748–1767 (2020).

5. Ash, A., Machado, L., Raleigh, S. M. & Anthony, K. Neuropathophysiology of Duchenne muscular dystrophy: involvement of the dystrophin isoform Dp71 in cell migration and proliferation. *Neuromuscul. Disord.* **28**, S13–S14 (2018).

6. Jones, L., Naidoo, M., Machado, L. R. & Anthony, K. The Duchenne muscular dystrophy gene and cancer. *Cell. Oncol.* **1**, 19–32 (2020).

7. Wang, Y. *et al.* Dystrophin is a tumor suppressor in human cancers with myogenic programs. *Nat. Genet.* **46**, 601–6 (2014).

8. Luce, L. N., Abbate, M., Cotignola, J. & Giliberto, F. Non-myogenic tumors display altered expression of dystrophin (DMD) and a high frequency of genetic alterations. *Oncotarget* **8**, 145–155 (2017).

9. Mauduit, O. *et al.* Recurrent DMD Deletions Highlight Specific Role of Dp71 Isoform in Soft-Tissue Sarcomas. *Cancers (Basel).* **11**, 922 (2019).

10. Tan, S. *et al.* Knocking down Dp71 expression in A549 cells reduces its malignancy in vivo and in vitro. *Cancer Invest.* **34**, 16–25 (2016).

11. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).

12. Ostrom, Q. T. *et al.* CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro. Oncol.* **17**, iv1–iv62 (2015).

13. Dixit, K. & Raizer, J. Newer Strategies for the Management of Low-Grade Gliomas. *Oncology (Williston Park).* **31**, (2017).

14. Youssef, G. & Miller, J. J. Lower Grade Gliomas. *Curr. Neurol. Neurosci. Rep.* **20**, (2020).

15. Solomons, M. *et al.* Survival and Seizure Control have improved for Adult Low-Grade Gliomas over the last eleven years. *Neuro. Oncol.* **21**, iv4–iv4 (2019).

16. Jakola, A. S. *et al.* Surgical resection versus watchful waiting in low-grade gliomas. *Ann. Oncol.* **28**, 1942–1948 (2017).

17. Fukuya, Y. *et al.* Tumor recurrence patterns after surgical resection of intracranial low-grade gliomas. *J. Neurooncol.* **144**, 519–528 (2019).

18. Olar, A. *et al.* IDH mutation status and role of WHO grade and mitotic index in overall survival in grade II–III diffuse gliomas. *Acta Neuropathol. 2015 1294* **129**, 585–596 (2015).

19. Yan, H. *et al.* IDH1 and IDH2 Mutations in Gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).

20. Zhao, J., Ma, W. & Zhao, H. Loss of heterozygosity 1p/19q and survival in glioma: a meta-analysis. *Neuro. Oncol.* **16**, 103 (2014).

21. Ferrari, F., Solari, A., Battaglia, C. & Bicciato, S. PREDA: an R-package to identify regional variations in genomic data. *Bioinformatics* **27**, 2446–2447 (2011).

22. Ruggieri, S. *et al.* Dp71 expression in human glioblastoma. *Int. J. Mol. Sci.* **20**, (2019).

23. Alkhaibary, A., Alassiri, A. H., AlSufiani, F. & Alharbi, M. A. Ki-67 labeling index in glioblastoma; does it really matter? *Hematol. Oncol. Stem Cell Ther.* **12**, 82–88 (2019).

24. Day, B. W. *et al.* The dystroglycan receptor maintains glioma stem cells in the vascular niche. *Acta Neuropathol.* **138**, 1033 (2019).

25. Calogero, A. *et al.* Altered expression of α-dystroglycan subunit in human gliomas. *Cancer Biol. Ther.* **5**, 441–448 (2006).

26. Bourkoula, E. *et al.* Glioma-associated stem cells: A novel class of tumor-supporting cells able to predict prognosis of human low-grade gliomas. *Stem Cells* **32**, 1239–1253 (2014).

27. Jung, E. *et al.* Emerging intersections between neuroscience and glioma biology. *Nat. Neurosci. 2019 2212* **22**, 1951–1960 (2019).

28. Golomb, L., Volarevic, S. & Oren, M. p53 and ribosome biogenesis stress: The essentials. *FEBS Lett.* **588**, 2571–2579 (2014).

29. Camp, R. L., Dolled-Filhart, M. & Rimm, D. L. X-Tile: A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-Point Optimization. *Clin. Cancer Res.* **10**, 7252–7259 (2004).

30. Ge, S. X., Son, E. W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinforma. 2018 191* **19**, 1–24 (2018).

31. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

Acknowledgements

Author contributions

Additional information

The authors declare no competing interests.

Figure legends

Figure 1. High *DMD* expression is significantly associated with poor survival in LGG. (a) TCGA RNAseq data from WHO grade II LGG cases was dichotomised into high (blue) and low (red) *DMD* expressing groups and survival analysis performed in GraphPad using the log-rank test. Tumour subtype analysis was also performed. Numbers in brackets are median overall survival times in months. (b) Forest plot revealing the log-rank hazard ratio with 95% confidence intervals and number of patients for each group. A: astrocytoma; OD: oligodendroglioma; NOS: not otherwise specified.

Figure 2. Pairwise comparison of Kaplan-Meier curves identifies a subgroup of IDH mutated LGG patients with poor survival. The Kaplan-Meier survival curves for high (blue) vs low (red) *DMD* expression for (a) IDH mutation status and (b) 1p/19q co-deletion status were compared. Numbers in brackets are median overall survival times in months. The table provides the log-rank test p values for each planned comparison from (b); the alpha value was adjusted to 0.017 to correct for multiple testing.

Figure 3. Genomic regions significantly dysregulated in WHO grade II LGG cases with high vs low *DMD* expression. Significantly altered regions are highlighted with coloured boxes.

Figure 4. Exploratory analysis of the DEGs in LGG cases with high versus low *DMD* expression. (a) Volcano and M-A plots of the log fold change of all genes. Upregulated and downregulated genes are indicated by red and blue points respectively. (b) STRING PPI network of DEGs with 60 edges (versus 20 expected) and an enrichment p value of 7.92e-13. Edge thickness indicates confidence, disconnected nodes are hidden. (c) Identification of significant hub genes from DEGs using the Cytohubba MCC and DMNC algorithms within Cytoscape. Colour represents ranking based on corrected p values from red to yellow, the expanded subnetwork is displayed by blue nodes. (d) Venn analysis of both sets of hub genes reveals 13 common significant hub genes. (e) Network tree visualisation of the enriched pathways in DEGs using the GO biological processes annotation, dot size corresponds to adjusted p values. (f) Visualisation of the relationship among enriched GO categories. Connected gene sets share more genes, colour of node represents adjusted p values.

Figure 5. Pathway analysis reveals high *DMD* expression downregulates biological processes related to ribosome biogenesis. iDEP was used to conduct pathway analysis using the generally applicable gene set enrichment (GAGE) method and the genes were annotated according to GO biological processes (a) or with Kyoto Encyclopaedia of Genes and Genomes (KEGG, b)[31]. The size of the dot corresponds to the adjusted P values.

Figure 6. WGCNA of the TCGA WHO grade II LGG dataset identified a network of 882 genes divided into five co-expression modules. (a) Gene dendrogram. Colours are randomly assigned except grey which represents areas unassigned to a module. (b) Networks of the top 20 genes for the entire network and each module individually. (c) Visualisation of the GO enrichment analysis for each module; heatmap was produced in GraphPad. (d) Venn analysis to identify common genes returned by both WGCNA and DEG analysis.

Figure 7. The expression of multiple *DMD* gene products are significantly associated with LGG survival outcomes. (a) LGG TCGA RNAseq data for each *DMD* isoform was dichotomised into high (blue) and low (red) expression groups and survival analysis performed in GraphPad using the log-rank test. Numbers in brackets are median overall survival times in months. (b) Kaplan-Meier survival curves for high (blue) vs low (red) Dp71, Dp71ab or Dp427m expression for each IDH mutation status group were compared. Numbers in brackets are median overall survival times in months. The tables provide the log-rank test p values for each planned comparison; the alpha value was adjusted to 0.017 to correct for multiple testing.

Figure 8. Exploratory analysis of the DEGs in LGG cases with high versus low Dp71, Dp71ab and Dp427m expression. (a) Venn analysis of the DEGs identified from high vs low overall *DMD*, Dp71, Dp71ab and Dp427m expression. The five DEGs common to all are listed. (b) STRING PPI networks of DEGs. Edge thickness indicates confidence, disconnected nodes are hidden. (c) Network tree visualisation of the enriched pathways in DEGs using the GO biological processes gene annotation, dot size corresponds to adjusted p values. Note no significant enrichment was found for Dp71 DEGs. (d) Pathway analysis using GO biological processes gene annotation. The heatmap was produced in GraphPad and displays the top 10 GO terms for high vs low *DMD*, Dp71, Dp71ab and Dp427m. Upregulated (red) and downregulated (blue) pathways are indicated.

Figure 9. Immunohistochemistry for dystrophin in LGG. Representative images of grade II astrocytoma (A), gemistocytic astrocytoma (B), oligodendroglioma (C) and NOS (D). Images were taken at 20x magnification except for image B which was taken at 40x magnification to better visualise the gemistocytic staining.

Figure 1.

a

b

TCGA cancer type	Gene expression		No. of patients	Hazard Ratio (95% CI)	P value
LGG (all)	DMD		321	4.15 (1.46, 11.81)	**<0.0001**
A	DMD		66	3.00 (0.30, 30.27)	0.1521
OD	DMD		117	9.80 (0.56, 170.6)	**<0.0001**
NOS	DMD		134	3.19 (0.91, 11.29)	**0.0042**

Figure 2

a

IDH status

P=<0.0001****

- Low *DMD* IDH mut (139.03)
- Low *DMD* IDH wt (19.96)
- High *DMD* IDH mut (63.55)
- High *DMD* IDH wt (21.01)

b

1p/19q co-deletion status

- Low *DMD* IDH mutant co-del (169.8)
- Low *DMD* IDH mutant non-co-del (130.8)
- Low *DMD* IDH wt (19.96)
- High *DMD* IDH mutant co-del (undefined)
- High *DMD* IDH mutant non-co-del (63.55)
- High *DMD* IDH wt (21.01)

c

	High DMD mutant co-del	High DMD mutant non-co-del	High DMD wt
Low DMD mutant co-del	**0.0089**		
Low DMD mutant non-co-del		**0.0039**	
Low DMD wt			0.8965

Figure 3

Figure 4

Figure 5

a

- 6e-06 RRNA metabolic process
- 3e-06 RRNA processing
- 3e-08 Ribosome biogenesis
- 2e-09 Ribonucleoprotein complex biogenesis
- 4e-06 NcRNA metabolic process
- 3e-06 NcRNA processing
- 1e-03 Ribonucleoprotein complex subunit organization
- 1e-03 Ribonucleoprotein complex assembly
- 4e-05 Nuclear-transcribed mRNA catabolic process
- 3e-06 Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
- 2e-08 SRP-dependent cotranslational protein targeting to membrane
- 2e-08 Cotranslational protein targeting to membrane
- 9e-07 Establishment of protein localization to endoplasmic reticulum
- 2e-07 Protein targeting to ER
- 2e-04 Protein localization to endoplasmic reticulum
- 3e-04 Establishment of protein localization to membrane
- 2e-05 Protein targeting to membrane
- 1e-04 Viral transcription
- 6e-05 Viral gene expression
- 2e-05 Translational initiation
- 1e-03 MRNA catabolic process
- 7e-05 Mitochondrial translational termination
- 7e-05 Mitochondrial translational elongation
- 7e-05 Translational termination
- 4e-05 Translational elongation
- 1e-04 Mitochondrial gene expression
- 6e-05 Mitochondrial translation
- 1e-03 Mitochondrial ATP synthesis coupled electron transport
- 1e-03 ATP synthesis coupled electron transport
- 1e-03 Respiratory electron transport chain

b

- 1e-01 Oxytocin signaling pathway
- 9e-02 Circadian entrainment
- 9e-02 Serotonergic synapse
- 1e-01 Long-term depression
- 9e-02 Gap junction
- 5e-02 Inflammatory mediator regulation of TRP channels
- 3e-02 Platelet activation
- 1e-01 Pancreatic secretion
- 9e-02 Salivary secretion
- 2e-02 Calcium signaling pathway
- 1e-01 Taste transduction
- 9e-02 Rap1 signaling pathway
- 2e-02 Regulation of actin cytoskeleton
- 9e-02 Proteoglycans in cancer
- 1e-01 Chemokine signaling pathway
- 9e-02 Axon guidance
- 1e-01 Osteoclast differentiation
- 9e-02 C-type lectin receptor signaling pathway
- 1e-01 Toll-like receptor signaling pathway
- 9e-02 Yersinia infection
- 9e-02 AGE-RAGE signaling pathway in diabetic complications
- 9e-02 Neutrophil extracellular trap formation
- 9e-02 Leishmaniasis
- 9e-02 Inositol phosphate metabolism
- 9e-02 Phosphatidylinositol signaling system
- 8e-04 Herpes simplex virus 1 infection
- 1e-02 Malaria
- 9e-02 Spliceosome
- 8e-13 Ribosome
- 1e-04 Oxidative phosphorylation

Figure 6

Figure 7

a

Dp71
Dp71 Low (130.8)
Dp71 High (62.96)
P= 0.0025**

Dp71a
Dp71a Low (117.4)
Dp71a High (146.1)
P= 0.0794

Dp71b
Dp71b Low (117.4)
Dp71b High (154.4)
P= 0.8228

Dp71ab
Dp71ab Low (130.8)
Dp71ab High (87.45)
P= 0.0002***

Dp427m
Dp427m Low (130.8)
Dp427m High (65.75)
P= 0.0216*

b

IDH status Dp71
- Low Dp71 mutant co-del (154.35)
- Low Dp71 mutant non-co-del (130.78)
- Low Dp71 wt (19.96)
- High Dp71 mutant co-del (Undefined)
- High Dp71 mutant non-co-del (62.96)
- High Dp71 wt (21.01)

IDH status Dp71ab
- Low Dp71ab mutant co-del (169.8)
- Low Dp71ab mutant non-co-del (130.8)
- Low Dp71ab wt (19.96)
- High Dp71ab mutant co-del (Undefined)
- High Dp71ab mutant non-co-del (87.45)
- High Dp71ab wt (21.01)

IDH status Dp427m
- Low Dp427m mutant co-del (169.8)
- Low Dp427m mutant non-co-del (130.8)
- Low Dp427m wt (21.01)
- High Dp427m mutant co-del (Undefined)
- High Dp427m mutant non-co-del (65.75)
- High Dp427m wt (16.83)

	High Dp71 mutant co-del	High Dp71 mutant non-co-del	High Dp71 wt
Low Dp71 mutant co-del	0.4530		
Low Dp71 mutant non-co-del		0.0617	
Low Dp71 wt			0.9038

	High Dp71ab mutant co-del	High Dp71ab mutant non-co-del	High Dp71ab wt
Low Dp71ab mutant co-del	0.7487		
Low Dp71ab mutant non-co-del		0.0201	
Low Dp71ab wt			0.7200

	High Dp427m mutant co-del	High Dp427m mutant non-co-del	High Dp427m wt
Low Dp427m mutant co-del	0.8549		
Low Dp427m mutant non-co-del		0.0590	
Low Dp427m wt			0.5195

Figure 8

a

Venn diagram:
- DMD
- Dp71: 96
- Dp71ab: 154
- Dp427m: 316
- 15, 14, 186
- 15 (DMD only)
- 7
- 0
- 3, 5, 13
- 23, 3
- 24

Common to all	
Gene	**Direction**
TRIM67	Down
TLX1	Down
LHX5	Down
PHEX	Up
RASSF9	Up

b

Dp71

Dp71ab

Dp427m

c

Dp71ab

● Down ● Up

- 4e-08 Animal organ development
- 6e-09 Circulatory system development
- 3e-08 Skeletal system morphogenesis
- 9e-12 Embryonic skeletal system development
- 1e-10 Embryonic skeletal system morphogenesis
- 8e-09 Embryonic organ morphogenesis
- 2e-08 Chordate embryonic development
- 3e-10 Skeletal system development
- 5e-09 Embryo development
- 1e-10 Embryonic morphogenesis
- 2e-09 Animal organ morphogenesis
- 5e-08 Extracellular matrix organization
- 2e-08 Mitotic sister chromatid segregation
- 5e-39 Synaptic signaling
- 2e-39 Chemical synaptic transmission
- 2e-28 Cell-cell signaling
- 2e-18 Modulation of chemical synaptic transmission
- 4e-17 Ion transport
- 2e-17 Regulation of transport
- 1e-17 Regulation of postsynaptic membrane potential
- 2e-19 Regulation of membrane potential
- 5e-17 Neurotransmitter secretion
- 9e-19 Neurotransmitter transport
- 2e-17 Vesicle-mediated transport in synapse
- 1e-24 Nervous system development

Dp427m

● Down ● Up

- 1e-05 Positive regulation of cell population proliferation
- 7e-07 Cell population proliferation
- 1e-05 Regulation of multicellular organismal process
- 1e-05 Positive regulation of multicellular organismal process
- 1e-05 Positive regulation of phosphorus metabolic process
- 2e-05 Cellular response to oxygen-containing compound
- 2e-05 Cellular response to endogenous stimulus
- 2e-05 Cell adhesion
- 2e-05 Positive regulation of interleukin-6 production
- 1e-05 Regulation of interleukin-8 production
- 2e-05 Extracellular matrix organization
- 4e-40 Synaptic signaling
- 8e-41 Chemical synaptic transmission
- 2e-32 Cell-cell signaling
- 3e-18 Modulation of chemical synaptic transmission
- 1e-15 Regulation of transport
- 5e-16 Regulation of ion transport
- 2e-14 Cation transport
- 4e-19 Ion transport
- 2e-14 Vesicle-mediated transport in synapse
- 9e-15 Neurotransmitter transport
- 7e-16 Signal release
- 1e-15 Nervous system development

d

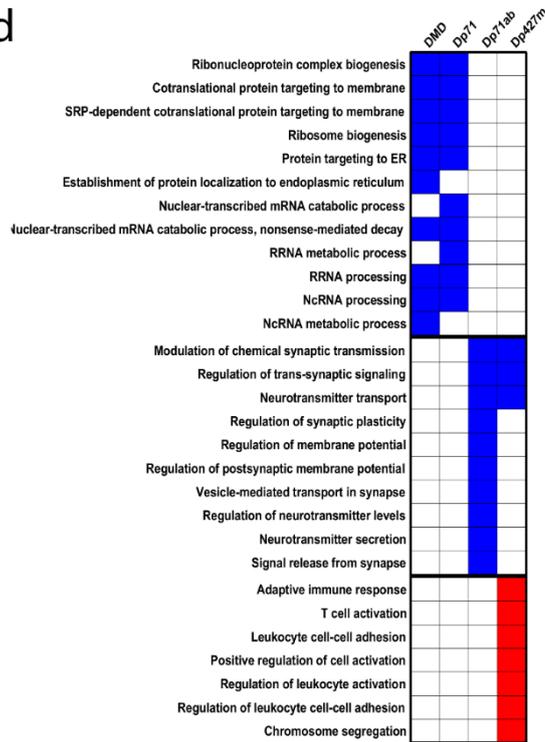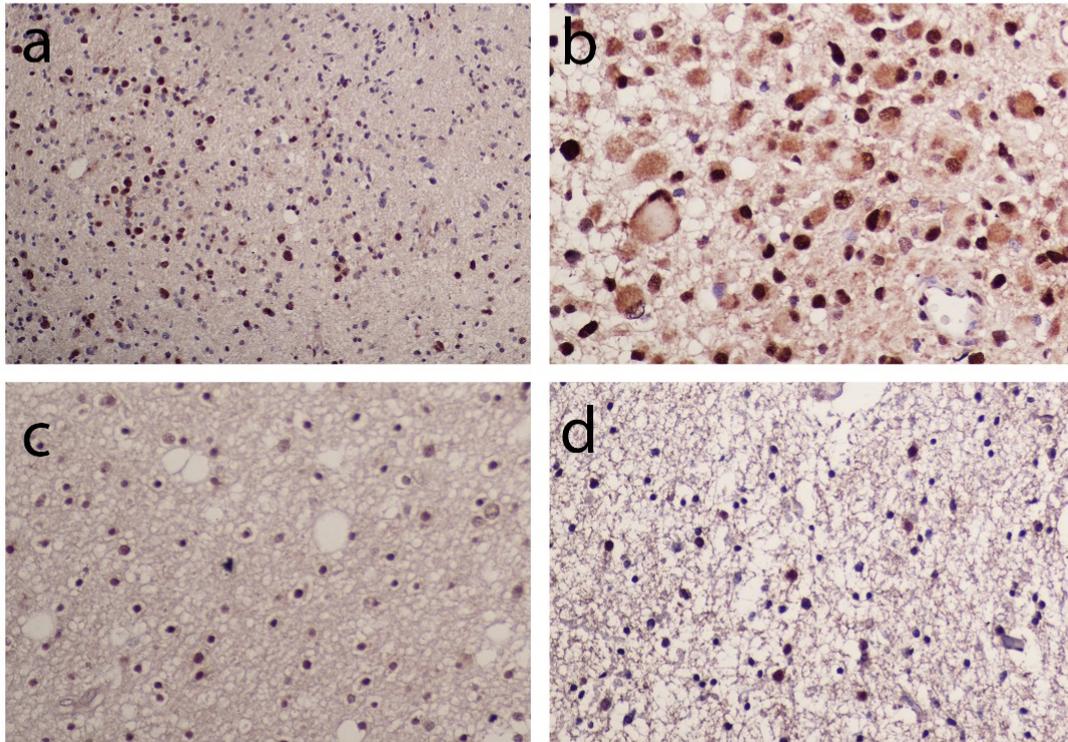|	DMD	Dp71	Dp71ab	Dp427m
Ribonucleoprotein complex biogenesis	■	■		
Cotranslational protein targeting to membrane	■	■		
SRP-dependent cotranslational protein targeting to membrane	■	■		
Ribosome biogenesis	■	■		
Protein targeting to ER	■	■		
Establishment of protein localization to endoplasmic reticulum		■		
Nuclear-transcribed mRNA catabolic process		■		
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay		■		
RRNA metabolic process		■		
RRNA processing	■	■		
NcRNA processing	■	■		
NcRNA metabolic process	■	■		
Modulation of chemical synaptic transmission			■	■
Regulation of trans-synaptic signaling			■	■
Neurotransmitter transport			■	■
Regulation of synaptic plasticity			■	
Regulation of membrane potential			■	■
Regulation of postsynaptic membrane potential			■	■
Vesicle-mediated transport in synapse			■	■
Regulation of neurotransmitter levels			■	■
Neurotransmitter secretion			■	■
Signal release from synapse			■	■
Adaptive immune response				■
T cell activation				■
Leukocyte cell-cell adhesion				■
Positive regulation of cell activation				■
Regulation of leukocyte activation				■
Regulation of leukocyte cell-cell adhesion				■
Chromosome segregation				■

27

Figure 9

Supplementary table S1. SPSS output for multivariate Cox regression analysis of *DMD* expression with tumour subtype

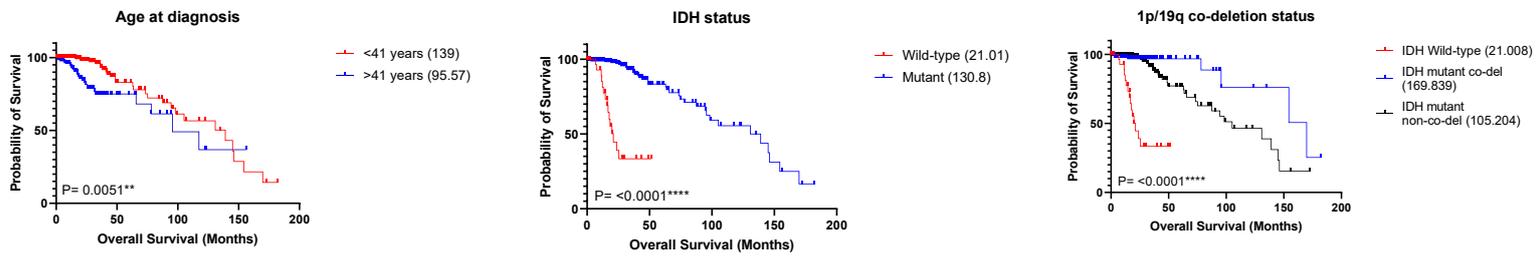|	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
DMD low=0; high=1	0.892	0.316	7.955	1	**0.005**	2.441	1.313	4.539
Tumour subtype			4.426	2	0.109			
Oligodendroglioma vs. astrocytoma	0.027	0.439	0.004	1	0.951	1.028	0.435	2.427
NOS vs. astrocytoma	0.638	0.416	2.353	1	0.125	1.893	0.838	4.279
Age at diagnosis	0.562	0.322	3.045	1	0.081	1.755	0.933	3.299
IDH mutation status	-2.163	0.381	32.300	1	**<0.001**	0.115	0.055	0.242
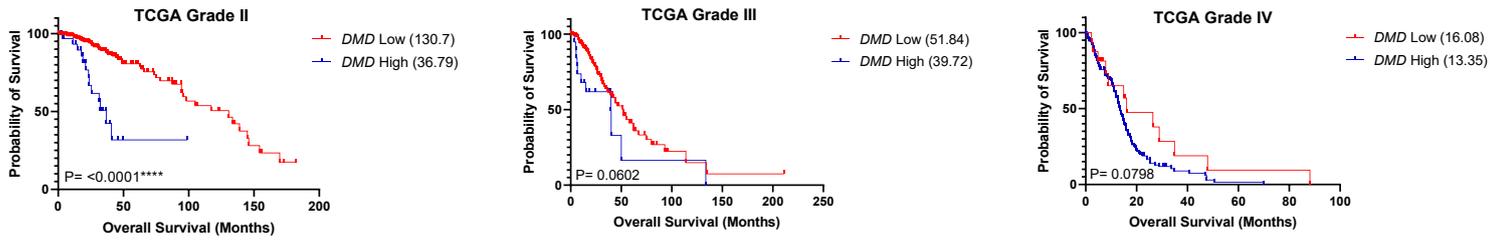
Supplementary figure S1. ***DMD* expression is only significantly associated with survival in lower-grade gliomas.** TCGA RNAseq data from WHO LGG grade II, III or IV was dichotomised into high (blue) and low (red) *DMD* expressing groups and survival analysis performed in GraphPad using the log-rank test. Numbers in brackets are median overall survival times in months.
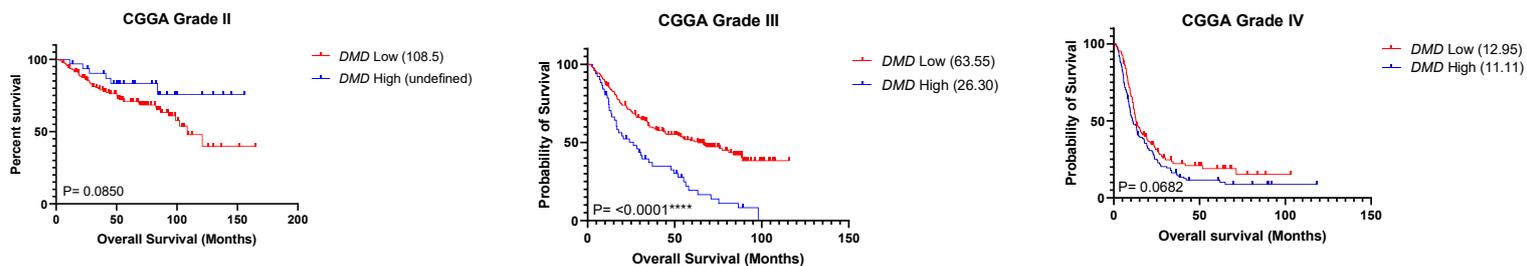
Cohort	Grade II	Grade III	Grade IV
TCGA	✔	×	×
CGGA	×	✔	×

Supplementary figure S2. A replication cohort analysis confirms the detrimental effect of high *DMD* expression on glioma survival is limited to the less invasive gliomas. CGGA RNAseq data from grade II, III or IV LGG cases was dichotomised into high (blue) and low (red) *DMD* expressing groups and survival analysis performed in GraphPad using the log-rank test. Numbers in brackets are median overall survival times in months. The table highlights for which tumour grade *DMD* expression was significant for both TCGA and CGGA cohorts.

a

b

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Age at diagnosis 0=<41; 1=>41	0.494	0.309	2.552	1	0.11	1.638	0.894	3.001
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2			35.888	2	**<.001 (1.6105E-8)**			
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(1)	-2.785	0.523	28.37	1	**<.001 (1.002E-7)**	0.062	0.022	0.172
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(2)	-1.985	0.377	27.778	1	**<.001 (1.3609E-7)**	0.137	0.066	0.287
DMD low=0; high=1	0.752	0.323	5.412	1	**0.02**	2.12	1.126	3.994

Supplementary figure S3. *DMD* expression remains significantly associated with survival in a multivariate model containing important clinicopathological variables. (a) Age at diagnosis and IDH mutation status are both significantly associated with survival in the TCGA LGG grade II cohort using the log-rank test. Numbers in brackets are median overall survival times in months. (b) *DMD* expression remained significant in a multivariate Cox regression analysis with age (<41 years; >41 years) and IDH mutation/co-deletion status (wild-type; IDH mutant co-deleted and IDH mutant non-co-deleted) and is therefore independently prognostic of these for LGG. Multivariate analysis was conducted using SPSS.

CGGA grade III IDH status

CGGA grade III IDH wt

CGGA grade III IDH mut
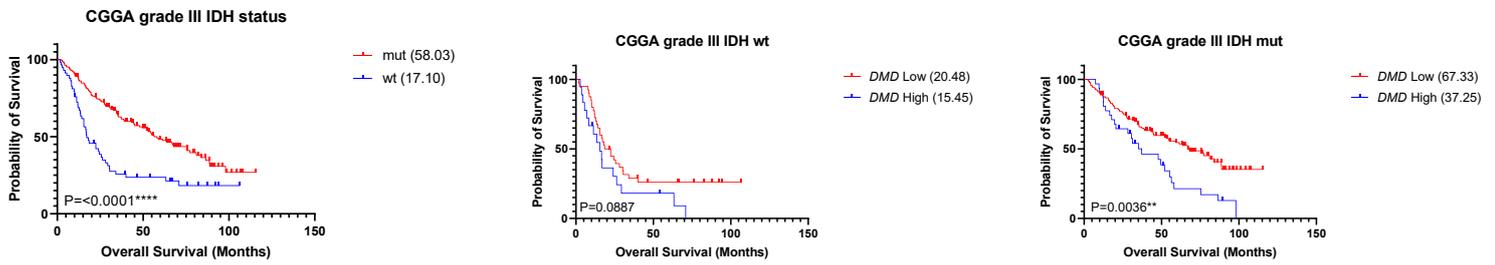
Supplementary figure S4. A replication cohort analysis confirms the detrimental effect of high *DMD* expression on glioma survival is limited to the less invasive, IDH mutant, gliomas. CGGA RNAseq data from grade III LGG cases was dichotomised into high (blue) and low (red) *DMD* expressing groups and analysed according to IDH mutation status. Survival analysis was performed in GraphPad using the log-rank test. This data replicates our finding that high *DMD* expression is significantly linked to poor survival in only IDH mutant cases. Numbers in brackets are median overall survival times in months.

**Supplementary table S2. DEGs present in WGCNA modules.**

Names	Total	Genes
DEGs present in turquoise module	10	NDST3
		TMPRSS3
		EPHA6
		SLC39A12
		SP8
		CDKL5
		VGF
		WIF1
		GABRA2
		SLC7A10
DEGs present in blue module	44	CREB3L1
		ZNF676
		MOXD1
		LEFTY2
		TRIM67
		HOXA13
		HCG22
		COL28A1
		DUSP4
		SEMA3A
		AMH
		SHISA9
		BARHL1
		POU4F1
		GALR1
		LHX5
		MSX2
		TFPI
		TRPM3
		MSTN
		RPE65
		CDH7
		SULF1
		PDGFD
		MADCAM1
		ETV4
		C10orf105
		STC2
		OR4N2
		LRRC14B
		COL11A1
		CLEC4F

|		DDIT4L
		ACAN
		RELN
		EMILIN3
		AR
		MMEL1
		PCDHGA3
		ZFR2
		GDF8
		ACY3
		TNNT1
		LHX1
DEGs present in green module	1	IL1B

a
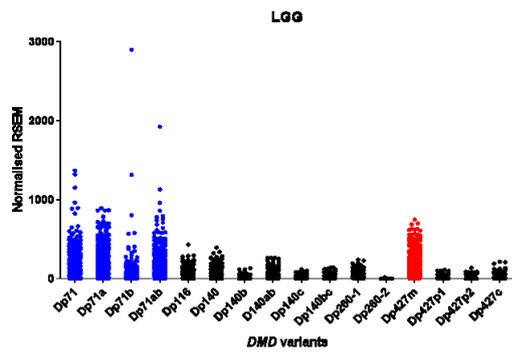
b
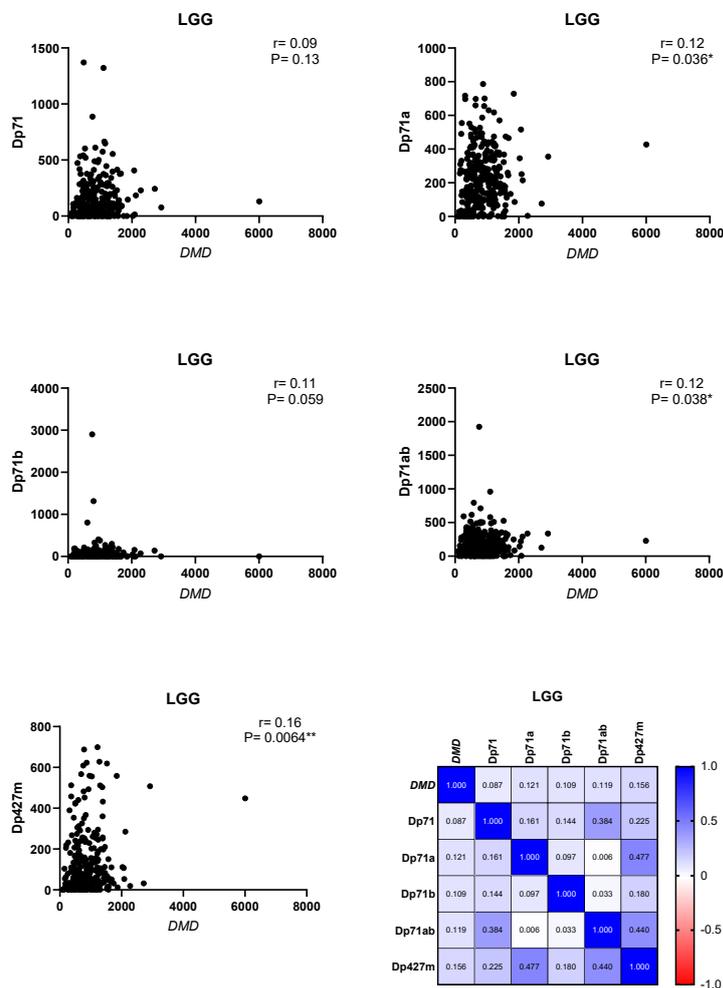
Supplementary figure S5. The four Dp71 isoforms and Dp427m are the most predominant *DMD* gene products expressed in LGG tissue. (a) RNAseq isoform data was extracted from the TCGA cohort and the expression level (normalised RSEM) for each *DMD* gene product per sample was plotted. (b) Scatter graphs for Spearman r correlation analysis of each gene product against total *DMD* gene expression. The correlation matrix represents r values for all correlation combinations.
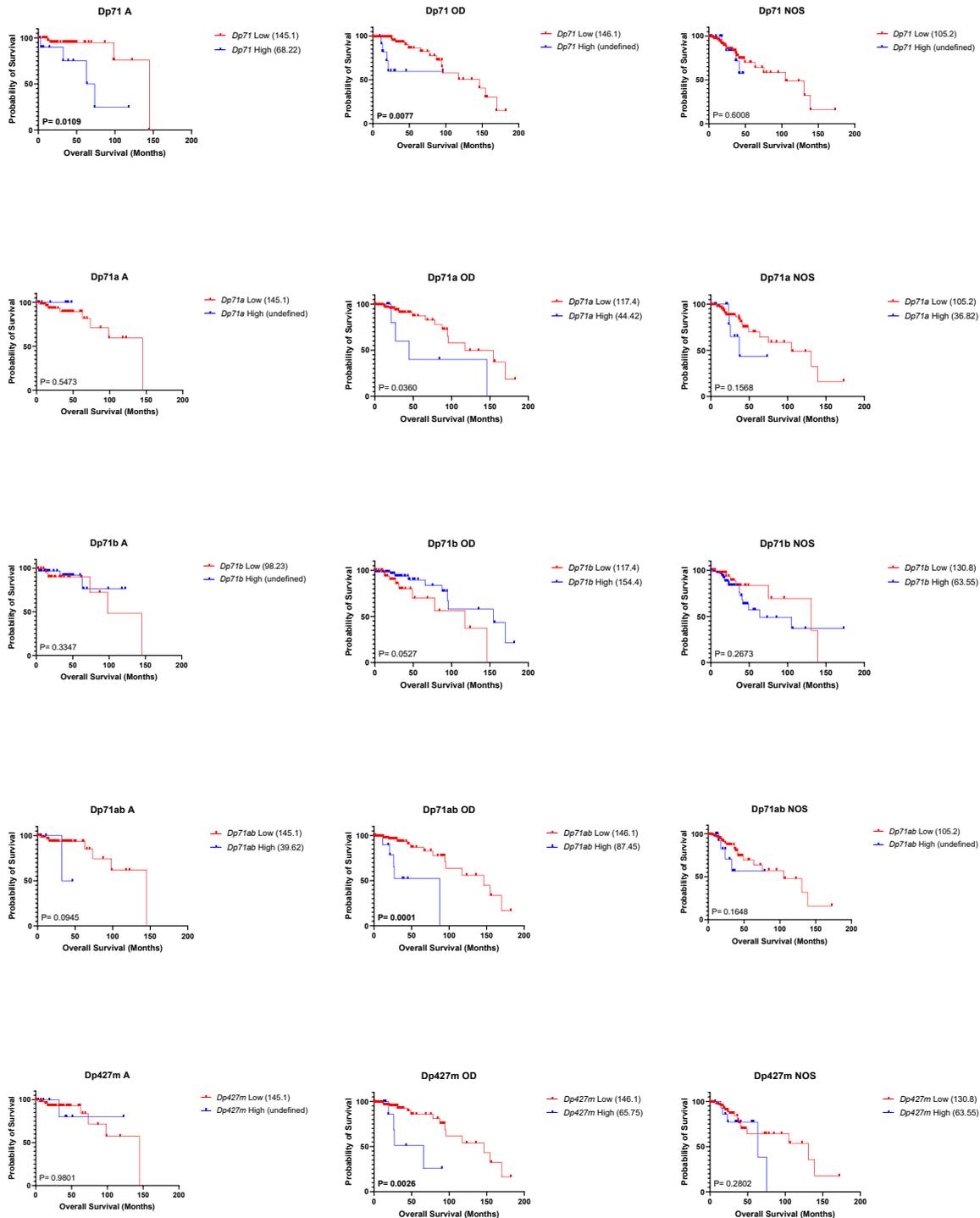
Supplementary figure S6. Tumour subtype analysis for each *DMD* gene product confirms that the expression of multiple *DMD* gene products are associated with LGG survival across all subtypes. The TCGA LGG cohort was divided into tumour subtype groups, each group was further dichotomised into high (blue) or low (red) *DMD* expressing groups. Survival analysis was performed in GraphPad using the log-rank test. The alpha value was adjusted to 0.017 to correct for multiple testing (five separate analyses; three tests per gene product), bold P values indicate significance. A: astrocytoma; OD: oligodendroglioma; NOS: not otherwise specified. Numbers in brackets are median overall survival times in months.

Supplementary table S3. SPSS output for multivariate Cox regression analysis of Dp71

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Age at diagnosis 0=<41; 1=>41	0.547	0.314	3.029	1	0.082	1.728	0.933	3.2
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2			37.025	2	**<.001 (9.1225E-9)**			
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(1)	-2.924	0.511	32.691	1	**<.001 (1.0803E-8)**	0.054	0.02	0.146
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(2)	-1.957	0.389	25.263	1	**<.001 (5.002E-7)**	0.141	0.066	0.303
Dp71 low=0; high=1	0.523	0.362	2.081	1	0.149	1.687	0.829	3.432

Supplementary table S4. SPSS output for multivariate Cox regression analysis of Dp71a

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Age at diagnosis 0=<41; 1=>41	0.587	0.308	3.638	1	0.056	1.799	0.984	3.289
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2			40.335	2	**<.001 (1.7435E-9)**			
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(1)	-3.011	0.51	34.842	1	**<.001 (3.5762E-9)**	0.049	0.018	0.134
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(2)	-2.035	0.383	28.201	1	**<.001 (1.0932E-7)**	0.131	0.062	0.277
Dp71a low=0; high=1	0.133	0.405	0.108	1	0.742	1.142	0.517	2.525

Supplementary table S5. SPSS output for multivariate Cox regression analysis of Dp71ab

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Age at diagnosis 0=<41; 1=>41	0.51	0.316	2.595	1	0.107	1.665	0.895	3.095
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2			37.613	2	**<.001 (6.8004E-9)**			
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(1)	-2.919	0.513	32.446	1	**<.001 (1.2256E-8)**	0.054	0.02	0.147
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(2)	-1.987	0.386	26.569	1	**<.001 (2.5435E-7)**	0.137	0.064	0.292
Dp71ab low=0; high=1	0.66	0.394	2.807	1	0.094	1.936	0.894	4.191

Supplementary table S6. SPSS output for multivariate Cox regression analysis of Dp427m

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Age at diagnosis 0=<41; 1=>41	0.477	0.314	2.302	1	0.129	1.611	0.87	2.981

	B	S.E.	Wald	df	Sig.	Exp(B)		
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2			44.151	2	**<.001 (2.586E-10)**			
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(1)	-3.054	0.505	36.552	1	**<.001 (1.4868E-9)**	0.047	0.018	0.127
IDH mutation status wt=0; mutant co-del=1; mutant non-co-del=2(2)	-2.209	0.393	31.671	1	**<.001 (1.8259E-8)**	0.11	0.051	0.237
Dp427m low=0; high=1	0.67	0.386	3.008	1	0.083	1.954	0.917	4.165

Supplementary table S7. Identified hub genes amongst the DEG PPI networks for high vs low Dp71, Dp71ab or Dp427m expression

|	Dp71	Dp71ab	Dp427m
MCC	PGF, TLX1, AQP4	SYNPR, GRIN1, SLC6A7, VSTM2B, SYN1, HTR5A, FAM163B, SNCB, MAG, SCRT1, LY6H, CCK, GABRG2, SNAP25, GABRD, SLC32A1, CA10	NGB, SYNPR, GRIN1, SYN1, HTR5A, SNCB, HPCA, SCRT1, CPNE6, HRH3, CCK, OPALIN, GABRG2, SNAP25, GABRD, SLC32A1, ATP2B3
DMNC	TGFB2, MFAP5, SLC2A10	KIF20A, ESPL1, E2F7, CENPA, CDCA2, KIF14, ASPM, BUB1B, DLGAP5, KIF4A, KIF18A, KIF23, DEPDC1, SLC8A2, TMEM151B, CDCA8, KIF2C	KIF20A, CENPF, CDCA2, P2RY12, KIF14, ASPM, GRM7, CENPI, SOHLH1, PRR11, S1PR3, CLSPN, DLGAP5, KIF18A, TTK, DEPDC1, TMEM151B
Common	GPR143, NPB, GJB6, RYR3, TMPRSS3, NDST3, NDST4, SDC4, SLC26A4, TMC1, P2RY2, GNG12, GPR17, GALR1, HS6ST3, GJA1, GPC5	CENPF, CALY, PHYHIP	SLC6A, CALY, PHYHIP

MCC: Maximal Clique Centrality; DMNC: Density of Maximum Neighbourhood Component.
Common: hub genes returned with both MCC and DMNC algorithms.

a
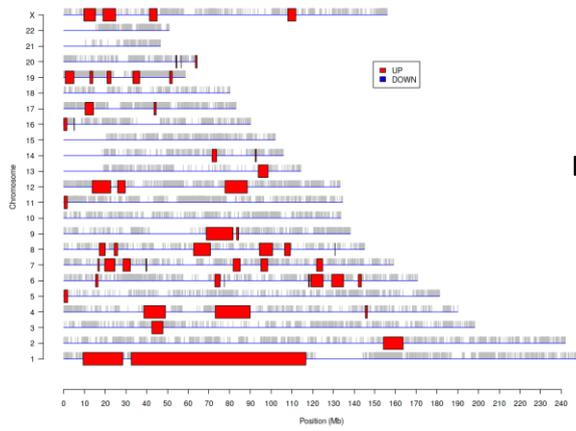
Dp71

b
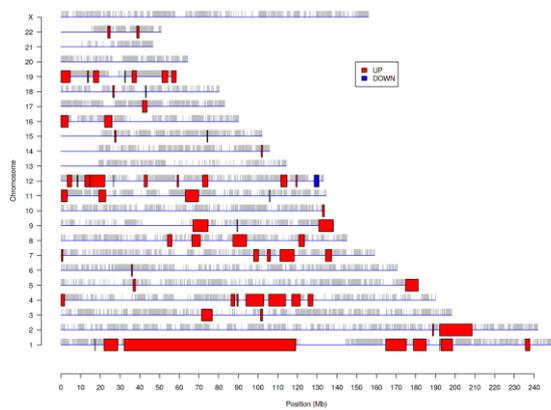
Dp427m

Supplementary figure S7. PREDA reveals genomic regions significantly altered upon high vs low *DMD* gene product expression. PREDA was performed in iDEP for (a) high vs low Dp71 and (b) Dp427m. No significant regions were found using PREDA with the high vs low Dp71ab dataset.

TCGA *DMD*   **CGGA *DMD***   **TCGA transcripts**
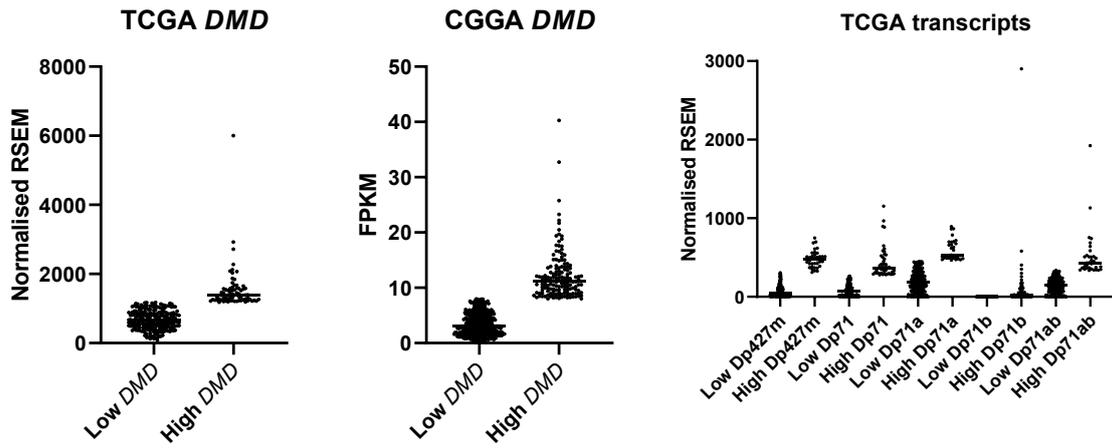
Supplementary figure S8. Plots to illustrate the variability of dystrophin expression across both low and high expression groups. Data is shown for total *DMD* expression in both TCGA and CGGA cohorts. *DMD* transcript data is available for the TCGA dataset only.

Supplementary data S1. Complete list of identified DEGs for high versus low *DMD*, Dp71, Dp71ab and Dp427m expression in LGG. Provided as a separate xlsx. file.

Supplementary data S2. Complete list of genes identified by WGCNA that are correlated in all LGG samples. Module number and colour for each gene are indicated. Provided as a separate xlsx. file.