Full-length dystrophin is expressed across human tissues and *DMD* downregulation commonly occurring in tumours coincides with Duchenne-like molecular alterations

Nancy Alnassar[1], Malgorzata Borczyk[2], Georgia Tsagkogeorga[3,4], Michal Korostynski[2], Namshik Han[3,5], Dariusz C Górecki[1*]

[1]School of Pharmacy and Biomedical Sciences, University of Portsmouth, Portsmouth, UK

[2]Laboratory of Pharmacogenomics, Maj Institute of Pharmacology PAS, Krakow, Poland

[3]Milner Therapeutics Institute, University of Cambridge, Cambridge, UK

[4]STORM Therapeutics Ltd, Babraham Research Campus, Cambridge, UK

[5]Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge, UK

*to whom correspondence should be addressed

Abstract

Mutations in the *DMD* gene, encoding dystrophins, cause Duchenne muscular dystrophy (DMD). However, loss of *DMD* expression impacts a broader spectrum of cells than those affected in DMD. Tumours were used as a model to investigate whether *DMD* loss across dissimilar tissues evokes related outcomes. Transcriptomic, proteomic, and mutation datasets from forty-nine healthy tissues and corresponding primary tumours (10,542 samples), and 180 related tumour cell lines were analysed. *DMD* expression was widespread across healthy tissues at levels comparable to common housekeeping genes, with corresponding full-length dystrophin identified by mass spectrometry. *DMD* expression was reduced in 80% of analysed tumours due to transcriptional downregulation not somatic mutations. The full-length transcript encoding Dp427m was decreased significantly in 68% of tumours, while Dp71 showed variability of expression. Hierarchical clustering analysis of the most highly expressed *DMD* transcripts revealed six clusters distinguishing malignant from healthy tissues. Transcriptomes of primary tumours and corresponding tumour cell lines with low *DMD* expression showed enrichment of specific pathways in the downregulated genes, compared to tumours and cell lines with higher *DMD* expression. Pathways consistently identified, such as ECM-receptor interaction, calcium signalling and PI3K-Akt, are also altered in DMD. Thus, *DMD* transcription occurs across a spectrum of healthy tissues and the molecular signature associated with its downregulation, which frequently occurs in malignancies, is concordant with changes found in Duchenne muscles, even though these malignancies originate from tissues not commonly associated with dystrophin expression or function. Therefore, the importance of the *DMD* gene extends beyond its roles identified in DMD.

Introduction

Duchenne muscular dystrophy (DMD) is a debilitating and lethal neuromuscular disorder. Diagnosis is made between the age of two and five, but transcriptomes from pre-symptomatic DMD patients reveal typical dystrophic abnormalities (1). In fact, studies of human foetuses (2–4) and in various animal models (5–7) revealed that the pathology starts during prenatal development and continues into adulthood. The first DMD defects are detectable in developing cells even before their differentiation into muscle (8). Given that muscle regeneration replicates processes occurring in muscle development and that some developmental mechanisms are reactivated in tumours, it is intriguing that changes in *DMD* gene expression are increasingly being described in various malignancies (9). Moreover, *DMD* downregulation affects cell proliferation (10,11), adhesion, migration, and invasion (12) – abnormalities that are commonly associated with tumour development. Gaining some understanding of whether these alterations, in a range of diverse cells, have a common origin and lead to a related outcome would expand our knowledge of the role of the *DMD* gene and potentially open up new treatment avenues.

Such analysis must take into consideration the complexity of *DMD*, the largest human gene known, with 79 exons and eight independent tissue-specific promoters driving the expression of distinct dystrophin isoforms. Three promoters control the expression of 14-kb full-length transcripts encoding 427 kDa isoforms (Dp427). The Dp427m is expressed in myofibers and muscle stem (satellite) cells. In myofibers, it interacts with the dystrophin-associated protein complex (DAPC) (13) with structural and scaffolding roles and an involvement in the regulation of various signalling pathways (14). Yet, in satellite cells, Dp427 has a different interactome and it is essential for asymmetric cell divisions (11,15), and in myoblasts the loss of its expression results in abnormalities of cell proliferation and migration (10,16,17). Dp427c and Dp427p full-length dystrophins are expressed in various neurons (18,19), where their loss during development has been linked to the neuropsychological impairment in DMD.

Moreover, intragenic promoters give rise to transcripts encoding truncated isoforms: Dp260 in the retina, Dp140 in the CNS, Dp116 in Schwann glia (20), and Dp71, which is the most ubiquitous *DMD* product. Alternative splicing adds further structural and functional diversity (21,22). *DMD* transcripts are summarised in Supplementary Figure 1.

Importantly, loss of Dp427 expression, which is both necessary and sufficient for the occurrence of Duchenne MD, was also associated with increased metastasis of tumours with myogenic programs (17) and soft tissue sarcomas (23). In contrast to this potential tumour suppressor role of the full-length dystrophin, Dp71 expression was essential for myogenic tumour cells growth (17,23). Interestingly, a putative role for the *DMD* gene was also suggested in various non-myogenic cancers, including carcinomas (24–26), melanoma (27), leukaemia (28), lymphoma (24), and CNS tumours (29–31).

Yet, many of the tumours showing phenotypic changes associated with the decreased *DMD* gene expression originate from healthy tissues that are not known to express the full-length dystrophin protein. Seemingly paradoxical, this observation is not limited to tumours: myoblasts, lymphocytes, and endotheliocytes express 14-kb dystrophin transcripts but are not known to synthesise detectable levels of dystrophin protein. Nevertheless, *DMD* downregulation leads to significant functional abnormalities in these cells (16,32).

We hypothesised that malignancy can be used as a model to investigate changes in *DMD* gene expression across healthy tissues and corresponding tumours and aid our understanding of the overall role of this gene, which clearly extends beyond Duchenne MD. Therefore, we studied *DMD* mRNA and protein expression across various healthy tissues and matching tumours and explored transcriptomic alterations in primary tumours and corresponding tumour cell lines with altered *DMD* expression to identify putative downstream molecular pathways that could be associated with *DMD* dysregulation across human tissues.

Results

1. Significant expression of dystrophin transcript and protein in a range of healthy tissues

Expression data for 24 different types of healthy tissues was examined. *DMD* expression levels across this spectrum of normal tissues averaged 66% of *GAPDH* (range between 27 and 82%) and 108% (52 to 132%) of *HPRT1* housekeeping genes levels, with the lowest expression found in whole blood (Supplementary Table 1). Given this widespread presence of considerable *DMD* transcript levels, we investigated whether it is accompanied by protein expression. Analysis of dystrophin protein across healthy tissue samples in the Human Protein Atlas gave confusing results. The antibody HPA002725, directed against amino acids 186-333 of the full-length dystrophin, detected moderate cytoplasmic and/or membranous staining in a range of tissues in addition to the expected staining in skeletal and cardiac muscle and the CNS. However, the antibody HPA023885, raised against amino acids 2843-2992 and therefore supposed to detect all dystrophin isoforms upstream of Dp71, showed staining in skeletal and cardiac muscle, while other tissues were negative. This staining pattern disagrees with the established expression of Dp260, Dp140 and Dp116, which is broader than that of Dp427 isoforms, and so it could not be accurate. Therefore, we have interrogated mass spectrometry (MS) protein expression datasets available at Proteomics DB. MS identified the full-length dystrophin protein in a variety of normal tissues (Supplementary Figure 2), in agreement with the Human Protein Atlas data obtained with the antibody HPA002725 detecting the full-length isoform. Furthermore, dystrophins expression was ranked in the top 25% of MS-quantified proteins in a range of tissues, such as fallopian tubes, oesophagus, uterus, bladder, colon, prostate, rectum, in addition to heart (PAXdb: Protein Abundance

Database). These data indicate that alterations in *DMD* expression might impact tumours that originate from tissues not commonly associated with dystrophin protein expression or function.

2. Downregulation of *DMD* gene expression across malignancies

We investigated RNA-seq expression data for 25 different types of primary tumours from the TCGA database (carcinomas, melanoma, lymphoma, and leukaemia) and their corresponding healthy tissues. The analysed primary tumours were acute myeloid leukemia (LAML), adrenocortical cancer (ACC), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical and endocervical cancer (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), diffuse large B-cell lymphoma (DLBC), esophageal adenocarcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney chromophobe cell carcinoma (KICH), kidney clear cell carcinoma (KIRC), kidney papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), prostate adenocarcinoma (PRAD), rectal adenocarcinoma (READ), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), uterine carcinosarcoma (UCS) and uterine corpus endometrioid carcinoma (UCEC).

Global *DMD* gene expression was reduced in 20 out of 25 primary tumours in comparison to their corresponding healthy tissues (Figure 1; Supplementary Table 2). The largest difference in expression was found in primary breast invasive carcinoma (LogFC = -3.7), and the smallest in primary kidney papillary cell carcinoma (LogFC = -1). *DMD* gene expression was increased in primary tumours in two out of 25 comparisons, namely in primary thyroid carcinoma and diffuse large B-cell lymphoma, with LogFC values of 0.9 and 4.8, respectively. No significant expression changes were found in three comparisons (acute myeloid leukaemia, kidney clear cell, and chromophobe cell carcinomas).

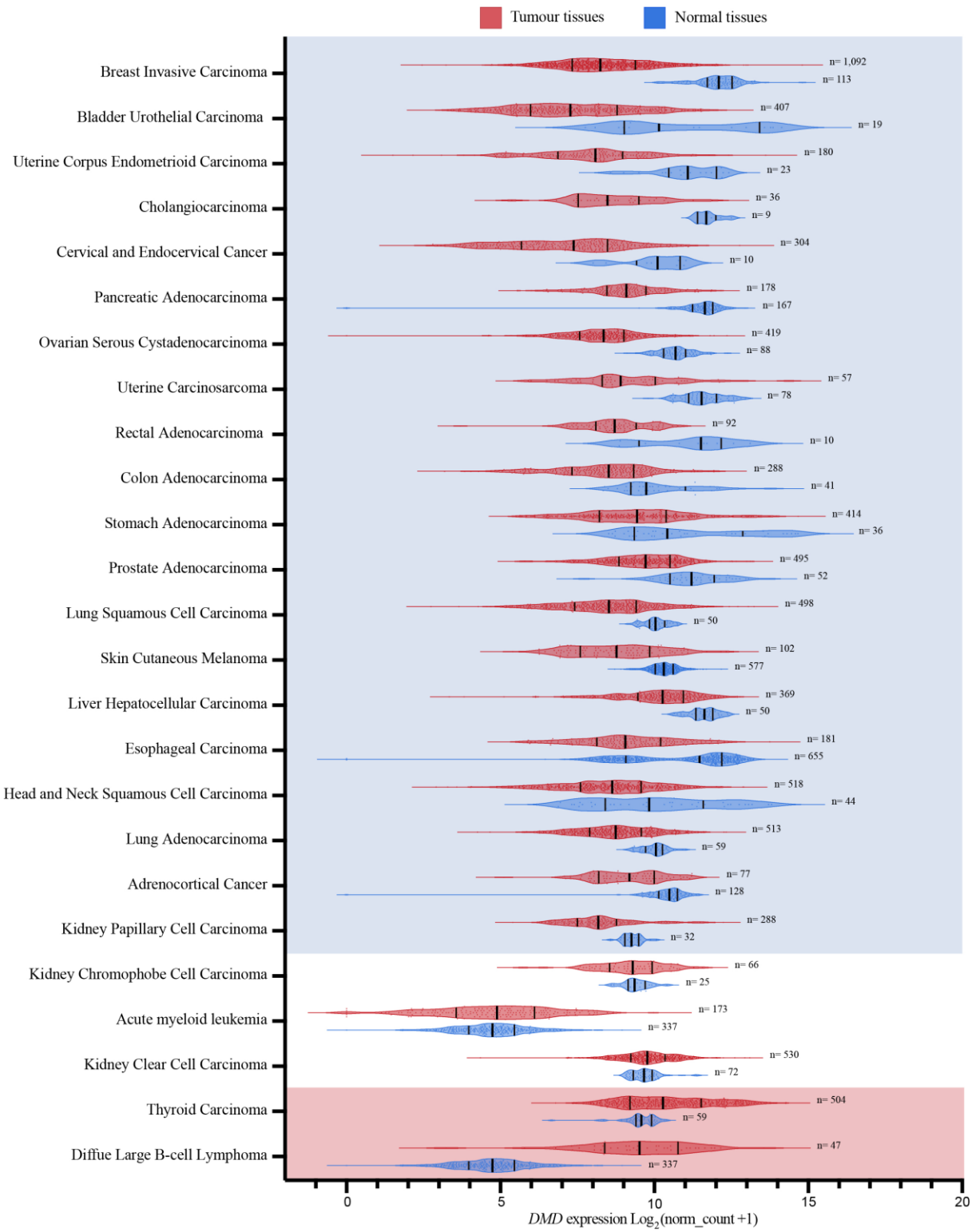

Figure 1. *DMD* gene expression level in primary tumours and corresponding healthy tissues. Red and blue violin blots represent tumour and healthy tissues, respectively. Vertical black lines represent the median and quartiles. Comparisons (n= 20) with a statistically significant downregulation of *DMD* expression in tumour samples compared to healthy tissues are highlighted in blue, and comparisons (n= 2) with a statistically significant upregulation of *DMD* expression in tumour samples compared to healthy tissues are highlighted in red (corrected

$p < 0.05$, Supplementary Table 2). P values were calculated using the Xena Browser using a two-tailed Welch's t test and adjusted for multiple testing using the Bonferroni correction.

Given the well-known multiplicity of transcripts originating from the *DMD* gene and the variability in their expression patterns, a hierarchical clustering analysis was performed to identify changes in *DMD* expression profile at the transcript level across various healthy tissues and corresponding tumours.

3. Hierarchical clustering analysis of the relative expression of *DMD* transcripts distinguishes tumour and healthy tissues

Of the top ten highly expressed transcripts (Table 1), nine were predicted to be protein-coding (Supplementary Figure 3). Three of these mRNAs were not canonical *DMD* gene transcripts. Of those, ENST00000358062.6 encoding H0Y304 protein is a poorly annotated transcript (Ensemble). The start of its CDS is unknown due to a 5' truncation of the available sequence. The 5' sequence upstream of the first exon (equivalent to exon 48 of the full-length transcript) is composed of the last 50 base pairs of the intron located between exons 47-48. Therefore, it is not clear whether this is a pre-mRNA sequence undergoing co-transcriptional splicing or a mature mRNA with this part of the intron spliced in. If the latter, the predicted protein encoded by such a transcript would have an N-terminus longer than Dp140 with valine as the putative initiation codon (UniProt). The transcript encoding H0Y864 appears to be a partial 4-exon transcript, whose predicted protein sequence does not encode any functional domains (Supplementary Figure 3). ENST00000475732.1 is a two-exon sequence not predicted to encode a protein.

Alternative splicing is a discernible feature of *DMD* transcripts found in the analysed samples, with the splice variant Dp71b lacking exon 78 being the predominant one. The mRNA encoding Dp140c differs from the canonical Dp140 transcript as it lacks exons 28 to 31 (equivalent to exons 71 to 74 of the full-length transcript) (Supplementary Figure 3).

Table 1. The top ten highly expressed *DMD* transcripts in the analysed tumour and healthy tissues and their predicted protein products. The identity percentage of unknown vs. known dystrophin isoforms is shown. Source: Ensemble and UniProt.

Transcript ID	Number of exons	Protein size (aa)	UniProt match
ENST00000357033.8	79	3,685	A0A075B6G3 (99.973 % sequence identity to **Dp427m**)
ENST00000619831.4	51	2,341	A0A087WV90 (99.957 % sequence identity to **Dp260-2**)

ENST00000358062.6	30	1,386	H0Y304
ENST00000541735.5	32	1,115	A0A5H1ZRR9 (99.91 % sequence identity to **Dp140c**)
ENST00000343523.6	25	970	A0A5H1ZRP7 (98.454 % sequence identity to **Dp116**)
ENST00000378702.8	18	617	P11532-7 (**Dp71**)
ENST00000378723.7	17	635	P11532-6 (**Dp71b**)
ENST00000361471.8	16	622	P11532-5 (**Dp71ab**)
ENST00000493412.1	4	134	H0Y864
ENST00000475732.1	2	None	

Hierarchical clustering analysis of these ten highly expressed transcripts in healthy and tumour tissues yielded six main clusters (Figure 2, Supplementary Table 3 reports other dendrogram cut-offs). The relative expression of the transcript encoding Dp427m was significantly lower in the clusters composed mainly of tumours compared to clusters with control tissues, while the relative expression of Dp71 variants was higher in clusters with tumours (Supplementary Figure 4). Transcripts ENST00000493412.1 and ENST00000475732.1 did not show significantly different levels of expression between any of the clusters (Supplementary Figure 4) and the differential expression of ENST00000358062.6 was not followed due to its uncertain annotation.

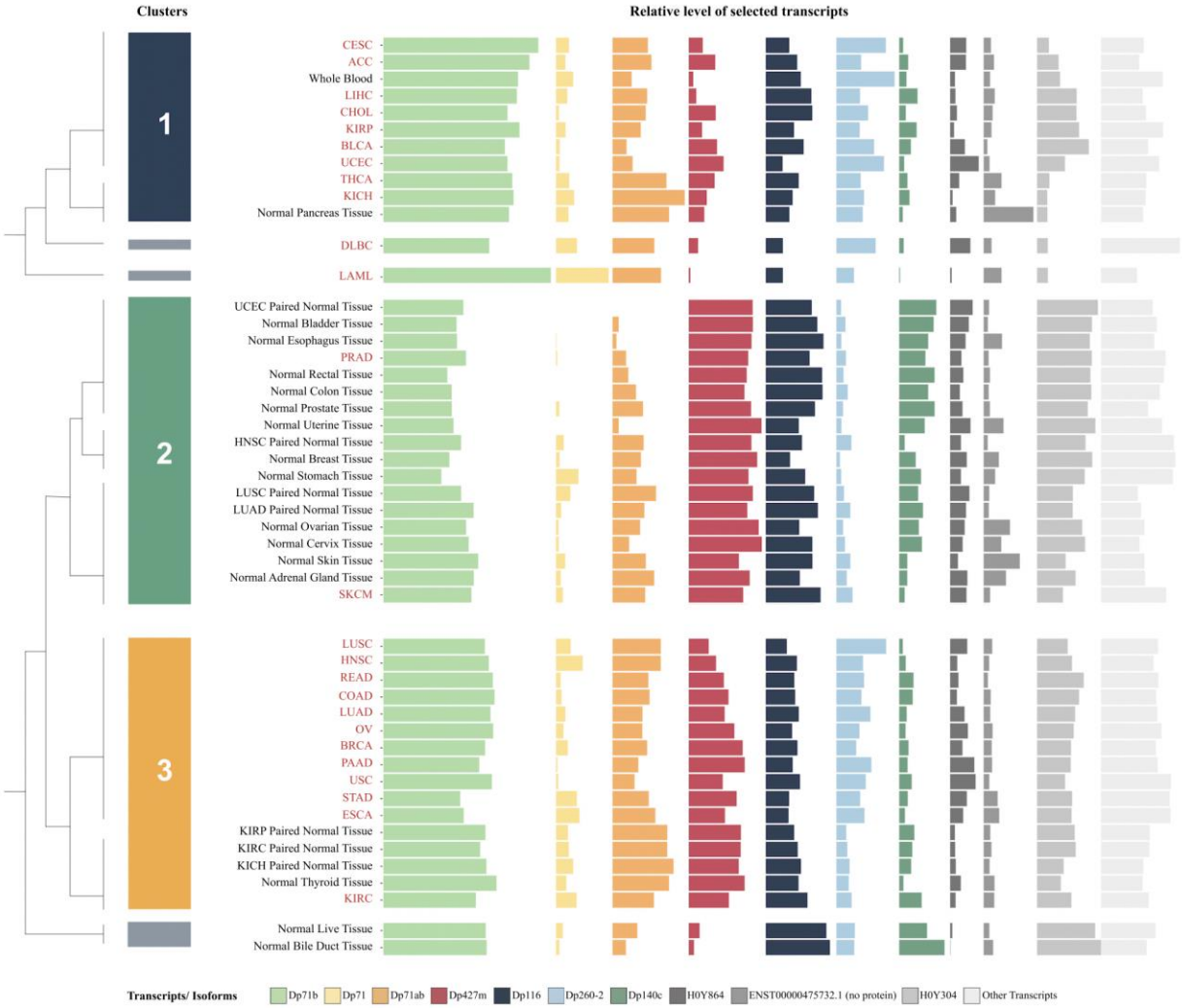

Figure 2. Bar plots representing the relative expression levels of the top 10 highly expressed *DMD* transcripts in tumour and healthy tissues and a dendrogram displaying the clusters identified by the hierarchical clustering analysis. Abbreviations: CESC (Cervical And Endocervical Cancer), ACC (adrenocortical cancer), LIHC (Liver Hepatocellular Carcinoma), CHOL (cholangiocarcinoma), KIRP (Kidney Papillary Cell Carcinoma), BLCA (Bladder Urothelial Carcinoma), UCEC (Uterine Corpus Endometrioid Carcinoma), THCA (Thyroid Carcinoma), KICH (kidney chromophobe cell carcinoma), DLBC (diffuse large B-cell lymphoma), LAML (Acute Myeloid Leukemia), PRAD (Prostate Adenocarcinoma), HNSC (Head And Neck Squamous Cell Carcinoma), LUSC (lung squamous cell carcinoma), LUAD (lung adenocarcinoma), SKCM (skin cutaneous melanoma), READ (rectal adenocarcinoma), COAD (colon adenocarcinoma), OV (Ovarian Serous Cystadenocarcinoma), BRCA (Breast Invasive Carcinoma), PAAD (Pancreatic Adenocarcinoma), USC (Uterine Carcinosarcoma), STAD (Stomach Adenocarcinoma). ESCA (esophagus adenocarcinoma), KIRC (Kidney Clear Cell Carcinoma).

4. Changes in the expression of Dp427m and Dp71 transcripts across malignancies

Given the importance of the full-length dystrophin Dp427m, whose loss is responsible for Duchenne muscular dystrophy, and also the predominance of the transcript encoding Dp427m (Table 1), its expression patterns were analysed in more detail. As Dp71 is the isoform most widely expressed across

the body and its splice variants were among the top highly expressed *DMD* transcripts in tumour and healthy tissues (Table 1), we also compared expression patterns of this transcript. (Supplementary Table 4). Dp427m expression was decreased in primary tumours compared to control tissues in 17 out of 25 comparisons with LogFC values ranging from -7.2 to -2.7 in primary uterine carcinosarcoma and primary thyroid carcinoma, respectively (Figure 3). There was a statistically significant change in the expression of transcripts encoding Dp71 in 10 out of 25, Dp71ab in 13 out of 25, while Dp71b was altered in 19 out of 25 comparisons (a decrease in 16 and an increase in three) (Supplementary Table 4).

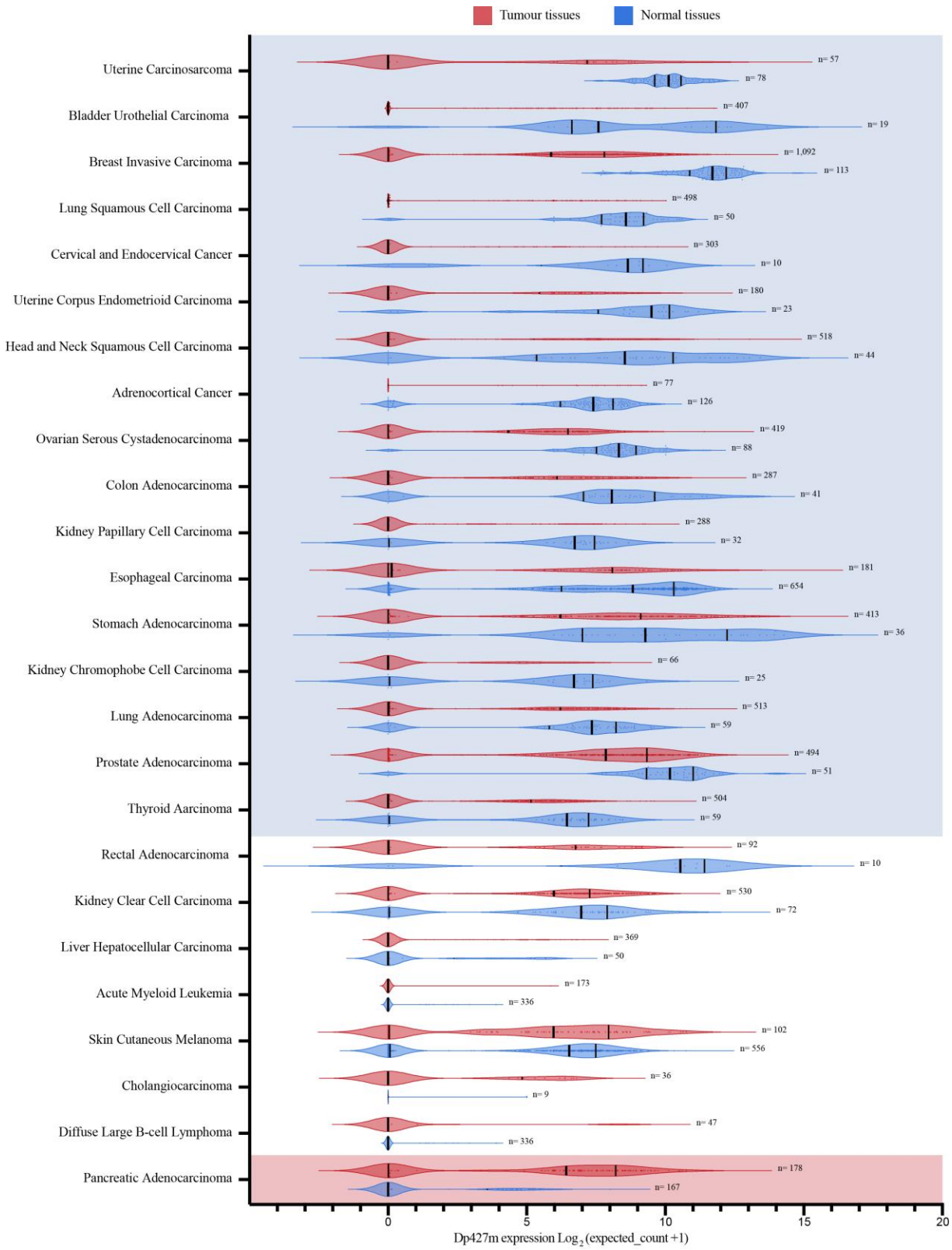

Figure 3. Dp427m transcript expression level in primary tumours and corresponding healthy tissues. Red and blue violin blots represent tumour and healthy tissues, respectively. Vertical black lines represent the median and quartiles. Comparisons (n= 17) with a statistically significant downregulation of Dp427m expression in tumour samples compared to controls are highlighted in blue, and one comparison with a statistically significant upregulation of Dp427m expression in tumour samples compared to controls is highlighted in red (corrected p <

0.05, Supplementary Table 4). P values were calculated using the Xena Browser using a two-tailed Welch's t test and adjusted for multiple testing using the Bonferroni correction.

Further analysis of these *DMD* transcripts showed that changes in overall *DMD* expression levels in two tumours were confounded by the opposing dysregulation of Dp427m and Dp71 variants. Specifically, while *DMD* expression was higher in primary thyroid carcinoma compared to normal thyroid tissue, Dp427m expression there was lower compared to healthy tissue and the increase in the overall *DMD* gene expression resulted from the elevated expression of Dp71b (Figure 4.A). In contrast, in primary pancreatic adenocarcinoma, total *DMD* expression was lower compared to the normal pancreas, but this decrease resulted from the lower expression of transcripts encoding Dp71 variants, while Dp427m expression was higher in this tumour type (LogFC = 3.7 compared to healthy tissue) (Figure 4.B).

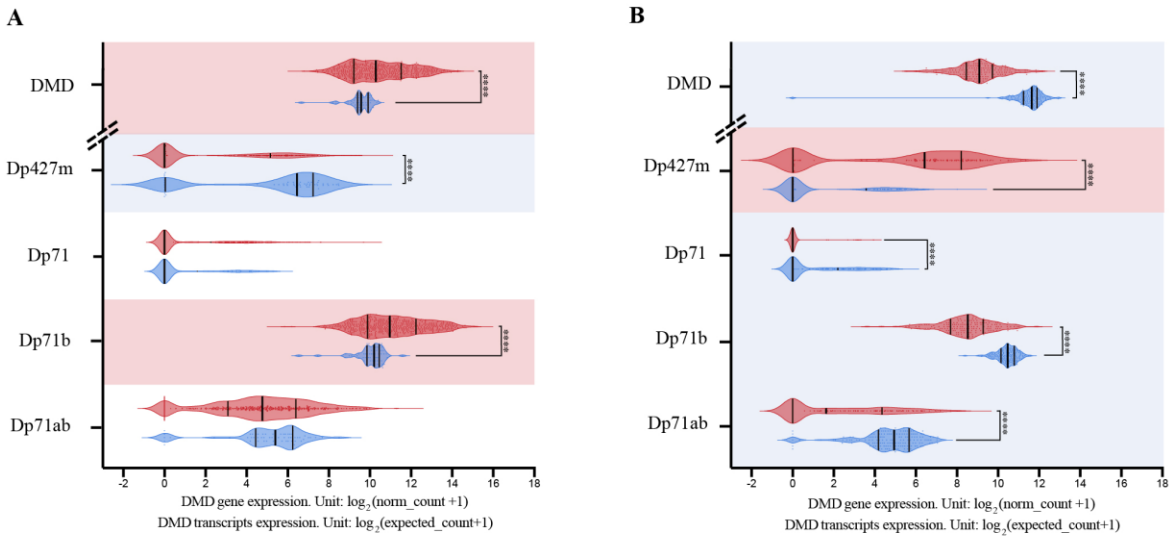

Figure 4. Total *DMD* gene expression and expression of transcripts encoding Dp427m, Dp71, Dp71b and Dp71ab in (A) primary thyroid carcinoma samples (n= 504) compared to adjacent normal thyroid (n= 59) and (B) primary pancreatic adenocarcinoma samples (n= 178) and the normal pancreas (n= 167). Red and blue violin blots represent tumour and healthy samples, respectively. Vertical black lines represent the median and quartiles. Comparisons where there is a statistically significant decrease in *DMD* gene/ transcript expression in tumour tissue compared to healthy controls are highlighted in blue, while comparisons where there is a statistically significant increase in *DMD* gene/ transcript expression in tumour tissue compared to healthy controls are highlighted in red (****corrected $p < 0.0001$). Expression values for the *DMD* gene were derived from the "RSEM norm_count" dataset, while expression values for *DMD* transcripts were derived from the "RSEM expected_count" dataset of the TCGA TARGET GTEx cohort at the Xena browser. P values were calculated using the Xena Browser using a two-tailed Welch's t test and adjusted for multiple testing using the Bonferroni correction.

5. Specific transcriptomic alterations are associated with decreased *DMD* gene expression in primary tumours

Expression of the *DMD* gene is significantly altered in tumours, with the majority having lower expression levels compared to their respective healthy tissues. This downregulation predominantly affects the transcript encoding the full-length dystrophin. In order to identify downstream molecular pathways that could be associated with *DMD* downregulation, we compared transcriptomes of primary tumour samples with low vs. high *DMD* gene expression from 15 different tumour types, where Dp427m was found to be downregulated compared to healthy tissues: BRCA, BLCA, UCEC, CESC, OV, COAD, STAD, PRAD, LUSC, ESCA, HNSC, LUAD, KIRP, THCA. In addition, we included sarcoma (SARC) in this comparison because sarcomas originate from tissues (muscle and bone) known to express the *DMD* gene. For clarification, RNA-seq data for sarcoma control tissues was not available, and therefore *DMD* gene expression in sarcomas vs. healthy tissues could not be compared, we could investigate transcriptome changes in sarcoma samples with low vs. high *DMD* gene expression. In each tumour type, the genes differentially expressed between groups of samples with low, medium, and high levels of *DMD* expression were identified (ANOVA FDR-corrected p-value < 0.1). Next, Post-Hoc T tests were performed to identify the differentially expressed genes between samples with low vs. high *DMD* expression (Bonferroni-corrected p-value < 0.05). The downregulated genes (LogFC < -0.5) were used to perform a pathway enrichment analysis. In comparisons where the number of downregulated genes was higher than 1,000, only the top 1,000 genes with the largest |LogFC| were used in the pathway enrichment analysis. Figure 5 shows the combined score values (-log (p value) * odds ratio) for the KEGG pathways that were found to be enriched in the downregulated genes in more than 50% of the analysed primary tumours (adjusted p-value < 0.05). Details of this analysis and the complete set of results can be found in Supplementary File 2. Importantly, an analogous pathway enrichment analysis conducted for the upregulated genes revealed no consistent pathways across the analysed tumours.

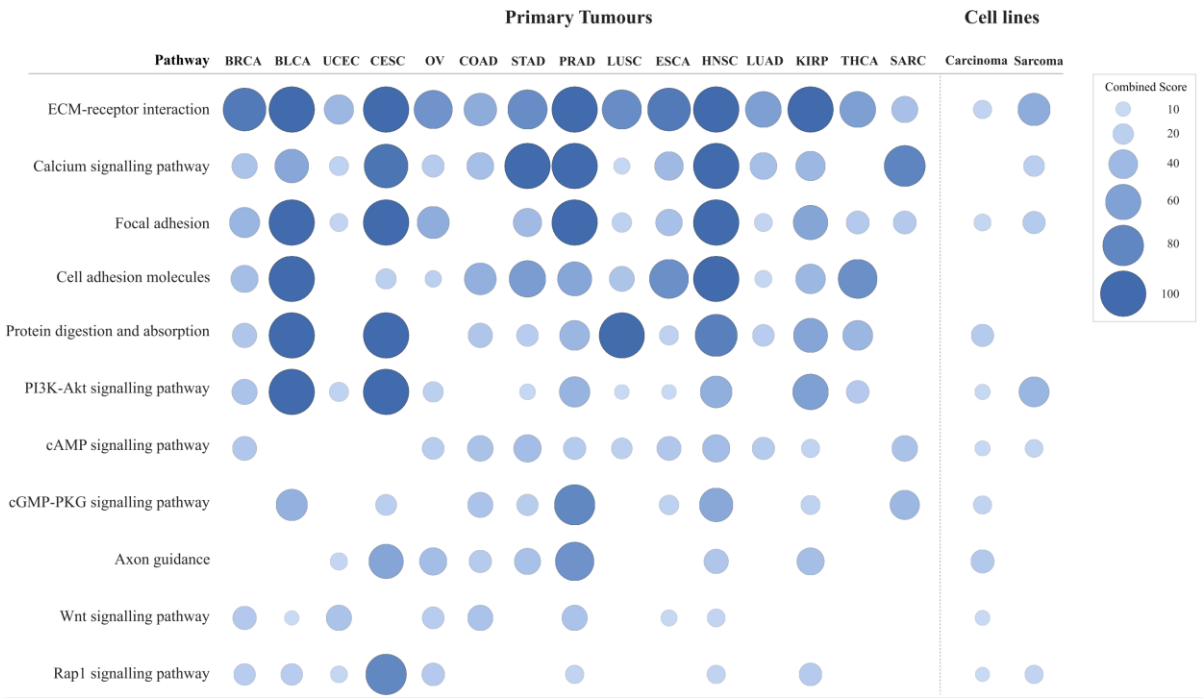

Figure 5. KEGG pathways enriched in the downregulated genes in primary tumour samples and tumour cell lines with low vs. high *DMD* gene expression. Only pathways that were identified in more than 50% of primary tumour samples with low *DMD* expression are displayed (adjusted p value < 0.05). Data is represented as the combined score for the enrichment (-log (p value) * odds ratio). Combined score values greater than 100 were given a value of 100.

The transcripts downregulated in tumours with low *DMD* expression showed enrichment in the ECM-receptor interaction pathway in all 15 primary tumours analysed. Calcium signalling and focal adhesion pathways were enriched in 14 tumours, cell adhesion molecules in 13, protein digestion and absorption and PI3K signalling in 12, cAMP in 11, cGMP in 9, while axon guidance, Wnt and Rap1 signalling pathways were enriched in 8 comparisons.

The specific GO Biological Process terms enriched in the downregulated genes in primary tumours with low vs. high *DMD* expression were also identified (Figure 6). The GO terms enriched in more than 50% of tumours were: extracellular matrix organisation (14 out of 15 tumours), synapse organisation (12), nervous system development, skeletal system development and cell junction assembly (11), regulation of ERK1 and ERK2 cascade, regulation of cell migration, chemical synaptic transmission, and synapse assembly (10), and regulation of cytosolic calcium ion concentration (9).

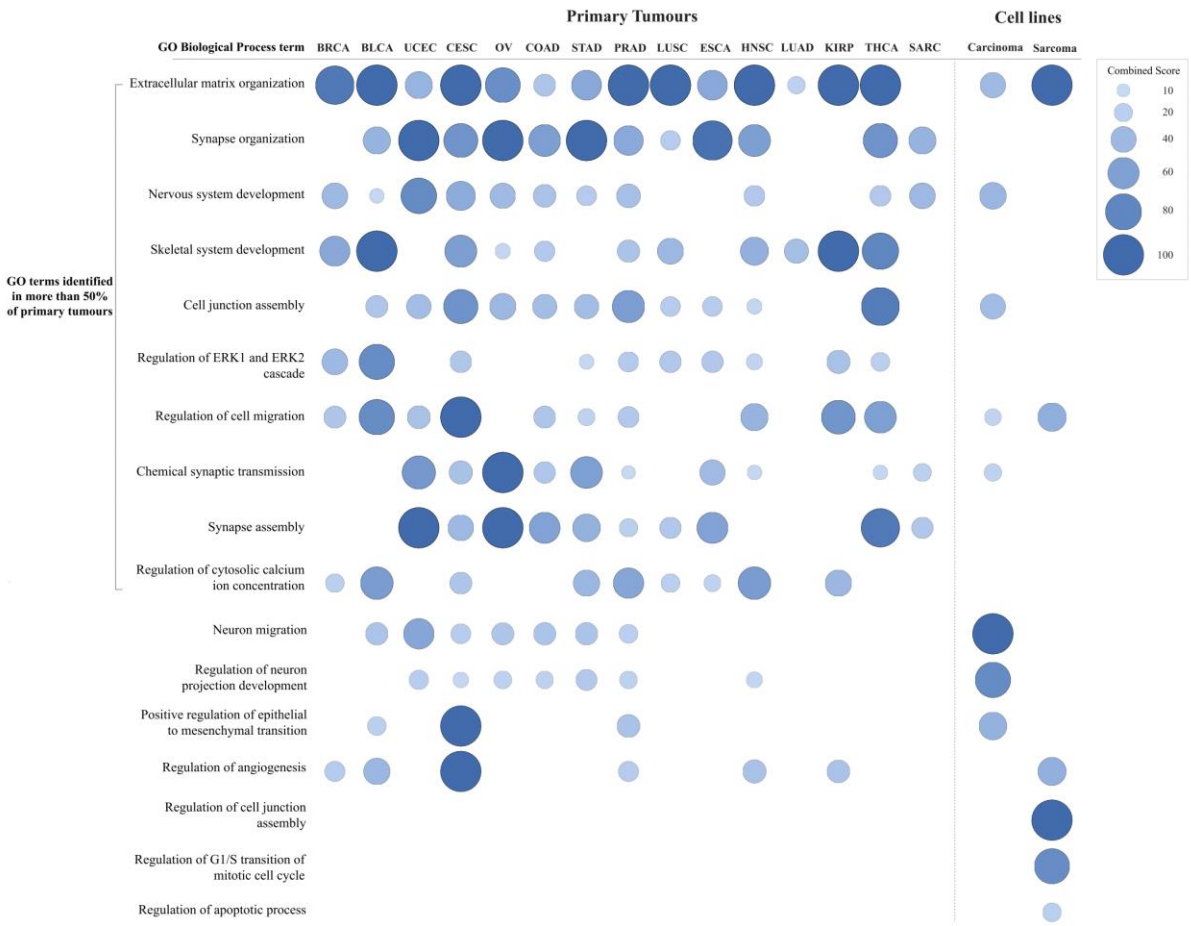

Figure 6. GO Biological Process terms enriched in the downregulated genes in primary tumour samples and tumour cell lines with low vs. high *DMD* gene expression (adjusted p value < 0.05). Data is represented as the combined score for the enrichment (-log (p value) * odds ratio). Combined score values greater than 100 were given a value of 100.

In order to confirm that these transcriptomic changes between primary tumour samples with low vs. high *DMD* expression are associated with *DMD* downregulation and not a result of other factors, we repeated the analysis for each tumour type using three groups of tumour samples identical in size to the groups used in the previous analysis, but that were chosen at random. *DMD* gene expression was not statistically significantly different between the three random groups in all the 15 tumours, and no statistically significant differences were found in the transcriptomes between these groups, confirming the specific gene expression alterations to be associated with *DMD* downregulation (Supplementary File 2).

6. Transcriptomic alterations associated with decreased *DMD* gene expression in tumour cell lines

To further confirm that these transcriptomic changes are evoked by *DMD* dysregulation specifically in cancer cells rather than originate from *DMD* expression in the stromal or infiltrating immune cells present in tumour tissue samples, we conducted a two-class comparison analysis using the DepMap

portal to identify the differentially expressed genes between cell lines of the same tumour origin (carcinoma and sarcoma) but with low vs. high level of *DMD* gene expression. While the *DMD* gene was not found to be essential for tumour cell lines survival (33,34), its altered expression may have an important common effect across malignancies.

One-hundred and eighty tumour cell lines were grouped into four groups based on their level of *DMD* expression (low or high) and their origin (carcinoma or sarcoma) (Supplementary File 3). The ranges for *DMD* expression for these cell lines were as follows (Unit: $\log_2$ (TPM+1)): carcinoma cell lines with high (3.3 – 7.4) and low (0 – 0.01) *DMD* expression, sarcoma cell lines with high (4.2 – 7.6) and low (0 – 0.06) *DMD* expression. The carcinoma cell lines used in this analysis originated from tissues where primary tumours were found to have a lower level of Dp427m transcript compared to healthy tissues. The two-class comparison analysis between cell lines with low vs. high *DMD* expression identified 998 differentially expressed genes for carcinoma, and 631 genes for sarcoma cell lines. In carcinoma cell lines with low *DMD* expression, the majority of the differentially expressed genes were downregulated (n= 976 out of 998), and in sarcoma cell lines, over half of the differentially expressed genes were also downregulated (n= 349 out of 631). Figure 5 shows the combined score values (-log (p value) * odds ratio) for the KEGG pathways that were found to be enriched in the downregulated genes in these two comparisons (adjusted p-value < 0.05). Details of this analysis and the complete set of results can be found in Supplementary File 3.

Pathway enrichment for the downregulated genes in carcinoma cell lines with low *DMD* expression suggested that *DMD* downregulation may be associated with changes in the following KEGG pathways: ECM-receptor interaction (p= 0.033), focal adhesion (p=0.025), protein digestion and absorption (p= 0.012), PI3K-Akt signalling (p= 0.027), cAMP signalling (p= 0.037), cGMP-PKG signalling (p= 0.022), axon guidance (p= 0.004), Wnt signalling (p= 0.048) and Rap1 signalling pathways (p= 0.048). These pathways were also enriched in the downregulated genes in more than 50% of primary tumours with low *DMD* expression.

Similar to carcinoma cell lines, the following KEGG pathways were enriched in the downregulated genes in sarcoma cell lines with low *DMD* expression: ECM-receptor interaction (p= 0.006), focal adhesion (p= 0.013), PI3K-Akt signalling (p= 0.001), cAMP signalling (p= 0.045) and Rap1 signalling pathways (p= 0.04). The calcium signalling pathway, which was enriched in the downregulated genes in more than 50% of primary tumours with low *DMD* expression was also enriched in the downregulated genes in the comparison of sarcoma cell lines (p= 0.016).

Moreover, GO Biological Process term analysis for the downregulated genes in carcinoma cell lines with low vs. high *DMD* expression (Figure 6 and Supplementary File 3) identified the following terms: extracellular matrix organisation (p= 0.000) and cell junction assembly (p= 0.01) as well as terms related to the regulation of the developmental processes: positive regulation of epithelial to mesenchymal transition (p= 0.033) and regulation of cell migration (p= 0.039). Finally, terms related

to the development of the CNS and synaptic transmission: nervous system development (p= 0.000), neuron migration (p= 0.000) and chemical synaptic transmission (p= 0.039) were found.

The identified GO Biological Process terms for the downregulated genes in sarcoma cell lines with low vs. high *DMD* expression (Figure 6 and Supplementary File 3) included: extracellular matrix organisation (p= 0.000), regulation of cell migration (p= 0.001) and regulation of cell junction assembly (p= 0.008), which were analogous to those identified for the downregulated genes in carcinoma cell lines with low *DMD* expression. Some of the identified GO terms were unique to sarcoma cell lines: regulation of angiogenesis (p= 0.002), regulation of G1/S transition of mitotic cell cycle (p= 0.004) and regulation of the apoptotic process (p= 0.009).

Importantly, functional enrichment tests of the dystrophin protein-protein interaction network constructed using high confidence PPI information demonstrated that the enriched functional terms were consistent with the pathways found to be significantly enriched in the aforementioned comparisons of primary tumours and tumour cell lines: extracellular matrix organisation (p= 1.10e^{-18}), axon guidance (p= 9.45e^{-10}), focal adhesion (p= 3.27e^{-14}), PI3K-Akt signalling pathways (p= 9.27e^{-09}) (Supplementary Figure 5).

7. Limited correlation between dystrophin and dystrophin-associated proteins expression in tumour cell lines

Given the previous results, we investigated whether the loss of dystrophin in tumour cell lines is associated with dysregulation of the dystrophin-associated proteins. First, we analysed the correlation between *DMD* transcript and protein expression in 360 tumour cell lines for which both the mRNA and protein data were available at the DepMap portal. Pearson's correlation coefficient for *DMD* gene and protein expression indicated a moderate positive correlation, with a value of 0.468 (p = 5.09e^{-21}).
Among the known protein members of the DAP complex, only β-2 syntrophin (*SNTB2*) expression had a moderate positive correlation with dystrophin expression (Pearson's coefficient: 0.527, q-value: 33.05e^{-24}), and dystroglycan (*DAG1*) had a weak positive correlation (Pearson's coefficient: 0.397, q-value: 4.14e^{-12}), while dysbindin (*DTNBP1*) showed a weak negative correlation with dystrophin (Pearson's coefficient: -0.321, q-value: 1.56e^{-08}) across all 366 tumour cell lines for which protein data was available. A few other DAP members showed a weak positive correlation with dystrophin: α-1 syntrophin (*SNTA1*) (Pearson's coefficient: 0.380, q-value: 7.57e^{-09} in 216 cell lines), α-dystrobrevin (*DTNA*) (Pearson's coefficient: 0.368, q-value: 2.60e^{-12} in 348 cell lines). Therefore, dystrophin in tumour cells does not appear to correlate with the presence of its canonical DAPC nor does its loss affect the expression of its typical interacting partners.

8. *DMD* expression downregulation occurs irrespective of somatic mutations within the *DMD* locus

We next investigated the association between *DMD* expression and *DMD* gene single nucleotide mutations within the *DMD* gene region and CNVs using data from 9,801 tumour samples (cBioPortal). The majority of tumour samples had no identified *DMD* mutations (n= 9,073), while the remaining had missense (n= 542), truncating (n= 81), splice (n= 31), inframe (n= 1), and multiple mutations (n= 73). The majority of samples were diploid for the *DMD* gene (n= 6,270), 302 had a deep deletion (i.e., a deep loss, possibly a homozygous deletion), 1854 had a shallow deletion (i.e., a shallow loss, possibly a heterozygous deletion), 1,127 had gains (i.e., a low-level gain of a few additional copies, often broad), 67 had amplifications (i.e., a high-level amplification of more copies, often focal), and 181 samples were not profiled for CNVs.

As expected, tumour samples with *DMD* mutations and CNVs had a significantly lower level of *DMD* expression compared to those with no alterations (Supplementary Figure 6.A). Tumour samples diploid for *DMD* had a higher level of expression compared to those with shallow and deep deletions and those with *DMD* gains of function (Supplementary Figure 6.B.). Tumour samples were then ranked based on their level of *DMD* expression and samples at the top 30% (n= 2,886) and bottom 30% (n= 2,886) were then used to determine whether there is an association between *DMD* mutations/ CNVs and expression level in tumour samples using a Chi Square test. The Chi Square test revealed that there is an overrepresentation of *DMD* mutations and CNVs in tumour samples with low level of *DMD* expression ($X^2(1) = 120.5$, $p < 0.0001$). 47% of tumour samples with low *DMD* gene expression (1,366 out of 2,886) had *DMD* mutations or CNVs. However, 53% of tumour samples (1,520 out of 2,886) had low levels of *DMD* expression and no detectable mutations or CNVs.

We also investigated the association between *DMD* expression and mutations using data from 921 tumour cell lines (cBioPortal). The majority of tumour cell lines had no identified mutations within the *DMD* gene region (n= 773), while the remaining had missense (n= 120), truncating (n= 11), splice (n= 4), and multiple mutations (n= 13). Cell lines were then ranked based on their level of *DMD* expression and those at the top 30% (n= 276) and bottom 30% (n= 276) were selected to determine whether there is an association between *DMD* mutations and expression level in tumour cell lines using a Chi Square test. The Chi Square test revealed that there is an overrepresentation of *DMD* mutations in tumour cell lines with low levels of *DMD* expression ($X^2(1) = 9.727$, $p = 0.0018$). 23% of tumour cell lines with low *DMD* gene expression (63 out of 276) had *DMD* mutations. However, 77 % of cell lines (213 out of 276) had low levels of *DMD* expression without any detectable mutations.

Therefore, DMD expression downregulation cannot be simply attributed solely to somatic mutations within the *DMD* locus but rather involves a regulatory mechanism.

Discussion

The *DMD* gene encodes a spectrum of dystrophin isoforms, but significant expression of the majority of these is restricted to specific tissues. Loss of the full-length dystrophin results in Duchenne muscular dystrophy, while mutations also disrupting other isoforms result in exacerbated phenotypes (35,36). In skeletal myofibers, cardiomyocytes and neurons, which express the highest levels of the full-length dystrophin, this protein has been found to serve as a structural scaffold for proteins engaged in ECM and cell-cell interactions, and in intracellular signalling. However, more recent data demonstrate that the loss of *DMD* expression impacts a broader spectrum of cells than those affected in DMD. In myoblasts (10,16,36,37), lymphocytes (32), endotheliocytes (38,39,40), mesodermal (8), and myogenic cells (11,15), loss of *DMD* expression leads to significant abnormalities. Moreover, the same abnormalities can occur in very distinct cells, e.g., calcium dys-homeostasis was found across all cells (41), and the damaging purinergic phenotype affects myoblasts and lymphocytes (32,36). Some old findings, such as platelets abnormalities (42), have recently been vindicated (43). Yet, these defects cannot be clearly attributed to the loss of interaction between dystrophin and the known dystrophin-associated proteins. Indeed, these cell-autonomous defects appear to affect dystrophic cells, which, when healthy, express the 14-kb *DMD* transcript, but were not shown to produce detectable levels of full-length dystrophins. This phenomenon, where expression of the 14-kb *DMD* transcript in cells such as myoblasts and lymphocytes does not correlate with detectable dystrophin protein, was known for decades but just disregarded as an “illegitimate transcription” (44).

Moreover, changes in *DMD* gene expression appear to affect progression in various malignancies (9), many of which originate from tissues not commonly associated with dystrophin expression or function, in which *DMD* transcription would thus be described as “illegitimate”. Yet, we found *DMD* transcript levels in these tissues to be comparable to the levels of common housekeeping genes. Therefore, we investigated a spectrum of tumours of different origins with the expectation that *DMD* alterations in a range of diverse tissues and cell lines may have a common origin and lead to a related outcome. Indeed, data presented here indicate that the current view on the role of the *DMD* gene in health and disease requires significant re-evaluation.

Firstly, although many tumours with downregulated *DMD* expression originated from healthy tissues previously not known for the expression of the full-length dystrophin, interrogation of proteomics datasets demonstrated a much wider dystrophin distribution profile than indicated by immunodetection (Supplementary Figure 2). Given the ambiguous immunodetection data and the high sensitivity and specificity of the mass-spec (45,46), identification of the full-length dystrophin in a wide spectrum of normal tissues using this latter method is likely to represent the true expression status. Thus, dystrophin may be present in many tissues at low levels and/or in a tightly controlled spatiotemporal manner, which might be missed using standard detection methods.

Moreover, dystrophin deficiency leads to a variety of cell-specific phenotypes, resulting from the loss of function of specific isoforms in a particular cell. This variability is reflected in the unique subcellular localization of dystrophins in different cell types, from subsarcolemmal distribution in muscle through postsynaptic puncta in neurons (47) to nuclear localisation (22). Therefore, we compared expression of specific *DMD* transcripts that showed differential expression in tumours and respective healthy tissues.

We found a significant downregulation of *DMD* gene expression across diverse primary tumours. Both full-length and truncated dystrophin variants were differentially expressed, and hierarchical clustering of the top highly expressed transcripts distinguished tumours from matched healthy tissue samples. A similar trend for *DMD* downregulation across carcinomas has been described previously (24) and our analysis discriminating specific *DMD* transcripts showed that levels of Dp427m mRNA were statistically significantly decreased in the clusters composed mainly of tumour tissues (Supplementary Figure 4), suggesting a specific impact of the loss of the full-length transcript. Indeed, pancreatic adenocarcinoma was the only cancer with higher levels of Dp427m transcript compared to healthy tissue. In contrast, the relative expression of Dp71 splice variants was increased in tumour clusters.

In contrast to primary carcinoma samples, in haematological malignancies, total *DMD* expression and the expression of the full-length and Dp71 transcripts was higher or unchanged when compared to normal blood, which also showed the lowest *DMD* expression relative to housekeeping genes of all tissues analysed (Supplementary Table 1).

Given some evidence of a causative link between *DMD* downregulation and phenotypic changes in tumour cells (17,23) and that alterations in Duchenne, such as increased cell proliferation, abnormalities in adhesion, migration, and invasion (11,12) are commonly associated with tumours, transcriptomes of primary tumour samples as well as tumour cell lines low vs. high level of *DMD* gene expression were compared.

While no causation can be confirmed at this stage, it is important to note that *DMD* dysregulation was associated with specific transcriptomic changes across 15 primary tumours and 180 various tumour cell lines. Functional enrichment analysis showed that the pathways significantly enriched in these downregulated genes in primary tumours and cell lines with low vs. high *DMD* expression were consistent with the functional dystrophin PPI network (Supplementary Figure 5). Pathways altered, including cell adhesion, ECM interactions and PI3K-Akt, correspond with alterations found in Duchenne patients' cells (48,49). The calcium signalling pathway enriched in the downregulated genes in tumours samples with low *DMD* expression from 14 out of the 15 analysed primary tumours, and well as in sarcoma cell lines with low *DMD* expression agrees with the dysregulation of calcium signalling across a whole spectrum of dystrophic cells (reviewed in 41), as do GO terms related to the regulation of the developmental mechanisms (8). Thus, the *DMD* gene may play similar roles in cancer and development, two processes showing biological and molecular similarities (50). The GO term "regulation of G1/S transition of mitotic cell cycle" enriched in the downregulated genes in sarcoma

cell lines with low *DMD* expression agrees with the previous finding that *DMD* gene targeting caused a G0/G1 arrest (51) and could also be connected to the abnormalities in satellite cell division (11,15).

A further indirect indication of the functional significance of the *DMD* gene in this paradigm is that *DMD* downregulation across various malignancies involves regulatory changes, not just results from somatic gene mutations. Although, as expected, the presence of somatic mutations or CNVs was associated with significantly lower levels of expression, in over 50% of tumours and 77% of cell lines, *DMD* downregulation could not be linked to such events. While deletions were described as a causative factor for the downregulation of *DMD* gene expression in some tumours (17,23,30), significantly reduced expression was also found in the absence of deletions or nonsense mutations (27), in agreement with our findings across different malignancies. Moreover, in primary pancreatic adenocarcinoma, Dp427m transcript was increased while Dp71 expression was reduced (Figure 4.B). Given the *DMD* gene structure, such an expression pattern can only be explained by differential regulation. Targeted degradation of dystrophin transcripts was suggested before (52) and recently, an epigenetic mechanism responsible for reduced *DMD* transcript levels has been described (53). These data indicate that, rather than being an effect of random mutations, quite likely to occur in this large gene, *DMD* alterations in tumour cells may have a complex regulatory nature involving mechanisms such as transcriptional regulation, chromatin remodelling or transcript degradation.

Dystrophin in tumour cell lines originating from tissues previously not associated with its expression or function did not correlate strongly with the presence of its established DAPC partners, suggesting that it may have a different role(s) than those in muscle cells. This is unsurprising, given that dystrophin associated proteins are known to differ in different tissues, with muscle and brain being the most notable examples. But even within muscle, dystrophin interactome changes with differentiation, with functionally distinct DAPs in satellite cells and myofibers. In conclusion, specific downregulation of *DMD* gene expression is a common feature across a spectrum of malignancies including those originating from tissues with very low levels or previously not known to express the full-length dystrophin protein. The molecular signature associated with decreased *DMD* expression in tumours and corresponding tumour cell lines is concordant with that found in Duchenne muscular dystrophy. Therefore, more attention should be given to the subtler *DMD* gene functions, beyond those causing the main symptoms of DMD. Such studies have a potential to identify new therapeutic targets for the treatment of this debilitating and still incurable disease.

Methods

1. *DMD* gene and protein dataset sources

The RNA-seq data for tumour (n= 7,828) and healthy (n= 2,714) tissue samples was obtained from the UCSC Xena Functional Genomics Browser (http://xena.ucsc.edu). The "RSEM norm_count" dataset of the TCGA TARGET GTEx cohort was used to examine RNA-seq gene expression data for a variety of cancers in The Cancer Genome Atlas (TCGA) and different non-diseased tissues in the Genotype-Tissue Expression (GTEx) project. Samples from both the TCGA and GTEx have been re-analysed and processed by the same bioinformatic pipeline to eliminate batch effects resulting from different computational processing (54). The level of *DMD* expression was compared between tumour samples and adjacent paired normal tissues from the TCGA (16 comparisons). When expression data for normal samples was not available in the TCGA, a comparison was made between *DMD* gene expression in tumour samples from the TCGA and normal tissue samples in the GTEx database (10 comparisons). Database sources in each comparison are stated in Supplementary Table 2.

Expression data for thirty *DMD* gene transcripts in tumour samples and their corresponding normal tissues was obtained from the dataset "RSEM expected_count" of the TCGA TARGET GTEx cohort using the Xena Browser.

The normalised log-transformed RNA-seq data used to compare gene expression in primary tumour samples with low vs. high *DMD* expression in 15 TCGA primary tumours was obtained from Xena Browser. Links to database sources can be found in Supplementary File 2.

Expression data for the *DMD* gene and dystrophin protein in tumour cell lines was obtained from the "Expression 21Q3 Public" and "Proteomics" datasets, respectively using the Cancer Dependency Map (DepMap) portal (https://depmap.org/portal). This portal includes RNA-seq expression data of 1,379 tumour cell lines from 37 lineages (55) and quantitative proteome profiling by mass spectrometry of 375 tumour cell lines from 22 lineages, dominated by solid organs. In the proteomics dataset, all protein expression values are not absolute, but relative values to the other values for the same protein (56).

The STRING database was used to construct dystrophin's protein-protein interaction (PPI) network (57).

Dystrophin protein expression in normal tissues was examined using the Human Protein Atlas (https://www.proteinatlas.org/), which contains results from immunohistochemistry assays conducted in a variety of tissues and cell types (58). The Proteomics DB data resource

(https://www.proteomicsdb.org/vue/) was also used to examine dystrophin protein expression in normal tissues. This database contains quantitative proteomics data assembled from liquid chromatography tandem-mass-spectrometry (LC-MS/MS) experiments in a variety of normal tissues (59).

The database "PaxDb" (Protein Abundances Across Organisms) was used to examine dystrophin protein abundance data across different normal tissues (https://pax-db.org). This database covers a large number of protein abundance datasets in a variety of tissues. For a specific protein of interest, it provides a table of abundances in all available datasets, along with the rank of this protein in the entire detectable proteome (60).

The cBioPortal website (https://www.cbioportal.org) was used to examine mutations within the *DMD* locus, copy number variations (CNVs) and *DMD* gene expression in 9,801 tumour samples from 32 TCGA PanCancer Atlas studies as well as *DMD* mutations and expression in 921 tumour cell lines from the Cancer Cell Line Encyclopaedia.

2. *DMD* transcript identification and classification

The TCGA TARGET GTEx cohort in the Xena browser contains expression data for thirty *DMD* gene transcripts. The Ensembl database was used for transcript annotation and the identification of coding and non-coding transcripts, and protein products in the UniProt database. Protein products were then aligned to known dystrophin isoforms in the UniProt database to identify the isoforms with the highest percentage of sequence identity to these protein products. The HMMER web server (https://www.ebi.ac.uk/Tools/hmmer/) was used for the graphical representations of Pfam domains of different dystrophin isoforms.

3. Statistical analyses of differentially expressed genes across tumour tissues and corresponding normal controls

Statistical analysis of the expression level of the *DMD* gene and transcripts was performed using the Xena browser, using the two-tailed Welch's t-test. P values were then adjusted for multiple testing using the Bonferroni correction; each of the obtained p values was multiplied by the total numbers of tests (n= 25). When this adjustment gave a value above 1, the corrected p values were set at 1. The Bonferroni-corrected p-values were then compared with the overall level of α (α= 0.05). Log-transformed expression data was downloaded from the Xena Browser and the mean of the *DMD* gene and transcripts expression level for tumour (A) and normal (B) samples was calculated. The following formula was then used to calculate Log Fold Change (LogFC): log (A/B) = log (A) – log (B). The Xena browser's instructions can be found at (https://ucsc-xena.gitbook.io/project/faq/advanced-data-and-datasets#how-do-i-calculate-fold-change-fc).

4. Hierarchical clustering analysis of *DMD* transcripts

Thirty *DMD* gene transcripts were ranked based on their mean abundance levels in primary tumours and corresponding normal tissues, and the top ten highly expressed transcripts were identified. The expression values of the remaining twenty transcripts were collapsed into one value called "other transcripts". Expression values of the top ten highly expressed *DMD* transcripts as well as the "other transcripts" value were then standardised to a sum of 1 in order to compare the relative expression levels of these transcripts in different tumour and normal tissues. An UPGMA hierarchical clustering analysis was then performed using the R v4.0.4 on an Euclidean distance matrix with the hclust average method. The dendrogram was cut at different levels to yield 3-7 clusters (Supplementary Table 3). Six was chosen as the optimal number of clusters as it was the smallest number that provided more than two clusters with around 10 tissues, and higher numbers of clusters yielded additional smaller clusters with less than three tissues. The R script used to perform the analysis is available at (https://github.com/nancyalnassar/DMD-gene-in-cancer/commit/ca99cba54c769c5be760e1f04a09010685647d42). Next, the expression level of these ten transcripts were compared across three clusters that contained the majority of the analysed tissues using Kruskal Wallis test and Dunn's multiple comparisons test.

5. **Identification of differentially expressed genes between primary tumour samples with low and high levels of *DMD* gene expression and functional enrichment analysis**

Gene expression in primary tumours samples with low vs. high levels of *DMD* gene expression from 15 different primary tumours was compared. For each tumour, samples were divided into three groups: Group A (samples at the bottom 33.3% of *DMD* expression), Group B (samples at the top 33.3% of *DMD* expression) and Group C (the remaining samples with medium level of *DMD* expression). Using the R v4.0.4, an Analysis of Variance (ANOVA) was preformed to identify the differentially expressed genes between the three groups, and p-values were adjusted using the False Discovery Rate (FDR) correction. Next, Post-Hoc Pairwise T tests were performed to identify the differentially expressed genes between Groups A and B, and p-values were adjusted using the Bonferroni correction. Genes with an FDR-corrected p-value lower than 0.1 and a Bonferroni-corrected p-value lower than 0.05 were considered to be statistically significantly different between the two groups. LogFC was calculated using the formula (LogFC): log (A/B) = log (A) – log (B). Genes with a LogFC value lower than -0.5 were considered to be downregulated in Group A compared to Group B, and genes with a LogFC value higher than 0.5 were considered to be upregulated. The R script used to perform the analysis is available at (https://github.com/nancyalnassar/DMD-gene-in-cancer/blob/main/Differential%20gene%20expression%20in%20primary%20tumour%20samples). The EnrichR tool (61) was then used to identify the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) Biological Process terms enriched in the downregulated genes. In comparisons where the number of downregulated genes was higher than 1,000, only the top 1,000 genes with the largest |LogFC| were used in the pathway and GO term enrichment analysis.

6. **Identification of differentially expressed genes in** tumour **cell lines with low and high levels of *DMD* gene expression and functional enrichment analysis**

We performed two comparisons in tumour cell lines from two distinct categories, carcinoma, and sarcoma. A two-class comparison analysis was conducted using the DepMap portal. The analysis computed a moderated estimate of the differences across the "Expression 21Q2 Public" dataset between cell lines with low and high levels of *DMD* expression. The differentially expressed genes in each of these two comparisons (q-value < 0.05) were identified. The EnrichR tool was then used to identify the KEGG pathways and GO Biological Process terms enriched in those genes.

7. **Correlation of dystrophin expression with other members of the DAP complex**

The Data Explorer tool at the DepMap portal was used to calculate Pearson's correlation between dystrophin and other members of the DAP complex (those encoded by *SGCA*, *SGCB*, *SGCD*, *SGCE*, *SGCG*, *SGCZ*, *DAG1*, *SSPN*, *LAMA2*, *DTNA*, *DTNB*, *SNTA1*, *SNTB1*, *SNTB2*, *SNTG1*, *SNTG2*, *CAV3*, *SYNC*, *DTNBP1* and *SYNM* genes) across all tumour cell lines using data from the "Proteomics" database.

8. **Association between *DMD* gene expression and mutation/ CNV status**

Data for 9,801 tumour samples and 921 cell lines from 32 TCGA PanCancer Atlas studies was downloaded from cBioPortal. Tumour samples with mutation and CNVs data (n= 9,620), and tumour cell lines with mutation data (n=921) were ranked according to their level of *DMD* expression and those at the top 30% and bottom 30% were selected for further analysis (n= 2,886 in each group of tumour samples and n= 276 in each group of tumour cell lines). A Chi Square test was used to determine whether there is an association between *DMD* mutations/ CNVs and expression level in tumour samples and between *DMD* mutations and expression level in tumour cell lines.

Acknowledgements

Funding from the University of Portsmouth (under the Strategic PhD studentships scheme) is gratefully acknowledged. The authors thank Thomas Krag and Javier. F Novo for very helpful comments on the manuscript.

Author contributions

NA carried out all the analyses with support from MB, GT, MK and NH and wrote the initial manuscript. DCG designed the project and wrote the final version of the manuscript with assistance from all co-authors.

Competing interest declaration

Authors declare no competing interests.

Additional information

All data are presented in the manuscript and the Supplementary information file.

References

1. Pescatori, M., Broccolini, A., Minetti, C., Bertini, E., Bruno, C., D'amico, A., Bernardini, C., Mirabella, M., Silvestri, G., Giglio, V., *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. *The FASEB Journal*, **21**, 1210–1226.
2. Vassilopoulos, D. and Emery, A.E. (1977) Muscle nuclear changes in fetuses at risk for Duchenne muscular dystrophy. *Journal of Medical Genetics*, **14**, 13–15.
3. Emery, A.E. (1977) Muscle histology and creatine kinase levels in the foetus in Duchenne muscular dystrophy. *Nature*, **266**, 472–473.
4. Toop, J. and Emery, A.E. (1974) Muscle histology in fetuses at risk for Duchenne muscular dystrophy. *Clinical Genetics*, **5**, 230–233.
5. Bassett, D.I., Bryson-Richardson, R.J., Daggett, D.F., Gautier, P., Keenan, D.G. and Currie, P.D. (2003) Dystrophin is required for the formation of stable muscle attachments in the zebrafish embryo. *Development*, **130**, 5851–5860.
6. Nguyen, F., Cherel, Y., Guigand, L., Goubault-Leroux, I. and Wyers, M. (2002) Muscle lesions associated with dystrophin deficiency in neonatal golden retriever puppies. *Journal of Comparative Pathology*, **126**, 100–108.
7. Merrick, D., Stadler, L.K.J., Larner, D. and Smith, J. (2009) Muscular dystrophy begins early in embryonic development deriving from stem cell loss and disrupted skeletal muscle formation. *DMM Disease Models and Mechanisms*, **2**, 374–388.
8. Mournetas, V., Massouridès, E., Dupont, J.B., Kornobis, E., Polvèche, H., Jarrige, M., Dorval, A.R.L., Gosselin, M.R.F., Manousopoulou, A., Garbis, S.D., *et al.* (2021) Myogenesis modelled by human pluripotent stem cells: a multi-omic study of Duchenne myopathy early onset. *Journal of Cachexia, Sarcopenia and Muscle*, **12**, 209–232.
9. Jones, L., Naidoo, M., Machado, L.R. and Anthony, K. (2021) The Duchenne muscular dystrophy gene and cancer. *Cellular Oncology*, **44**, 19–32.
10. Gosselin, M.R., Mournetas, V., Borczyk, M., Bozycki, L., Korostynski, M., Robson, S., Pinset, C. and Gorecki, D.C. (2021) Loss of full-length dystrophin expression results in major cell- autonomous abnormalities in proliferating myoblasts. *bioRxiv*.
11. Chang, N.C., Chevalier, F.P. and Rudnicki, M.A. (2016) Satellite Cells in Muscular Dystrophy - Lost in Polarity. *Trends in Molecular Medicine*, **22**, 479–496.
12. Young, C.N.J., Chira, N., Róg, J., Al-Khalidi, R., Benard, M., Galas, L., Chan, P., Vaudry, D., Zabłocki, K. and Górecki, D.C. (2018) Sustained activation of P2X7 induces MMP-2-evoked cleavage and functional purinoceptor inhibition. *Journal of Molecular Cell Biology*, **10**, 229–242.

13. Muntoni, F., Torelli, S. and Ferlini, A. (2003) Dystrophin and mutations: One gene, several proteins, multiple phenotypes. *Lancet Neurology*, **2**, 731–740.
14. Allen, D.G., Whitehead, N.P. and Froehner, S.C. (2016) Absence of dystrophin disrupts skeletal muscle signaling: roles of Ca2+, reactive oxygen species, and nitric oxide in the development of muscular dystrophy. *Physiological reviews*, **96**, 253–305.
15. Dumont, N.A., Wang, Y.X., von Maltzahn, J., Pasut, A., Bentzinger, C.F., Brun, C.E. and Rudnicki, M.A. (2015) Dystrophin expression in muscle stem cells regulates their polarity and asymmetric division. *Nature Medicine*, **21**, 1455–1463.
16. Yablonka-Reuveni, Z. and Anderson, J.E. (2006) Satellite cells from dystrophic (Mdx) mice display accelerated differentiation in primary cultures and in isolated myofibers. *Developmental Dynamics*, **235**, 203–212.
17. Wang, Y., Marino-enriquez, A., Bennett, R.R., Zhu, M., Eilers, G., Lee, J., Henze, J., Fletcher, B.S., Gu, Z., Fox, E.A., *et al.* (2014) Dystrophin Is a Tumor Suppressor in Human Cancers with Myogenic Programs. *Nature genetics*, **46**, 601–606.
18. Nudel, U., Zuk, D., Einat, P., Zeelon, E., Levy, Z., Neuman, S. and Yaffe, D. (1989) Duchenne muscular dystrophy gene product is not identical in muscle and brain. Duchenne muscular dystrophy gene product is not identical in muscle and brain. *Nature*, ***337***, 76–78.
19. Górecki, D.C., Monaco, A.P., Derry, J.M.J., Walker, A.P., Barnard, E.A. and Barnard, P.J. (1992) Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. *Human Molecular Genetics*, **1**, 505–510.
20. Duan, D., Goemans, N., Takeda, S., Mercuri, E. and Aartsma-Rus, A. (2021) Duchenne muscular dystrophy. *Nature Reviews Disease Primers*, **7**, 1–19.
21. Austin, R.C., Howard, P.L., D'souza, V.N., Klamut, H.J. and Ray, P.N. (1995) Cloning and characterization of alternatively spliced isoforms of Dp71. *Human Molecular Genetics*, **4**, 1475–1483.
22. Naidoo, M. and Anthony, K. (2020) Dystrophin Dp71 and the Neuropathophysiology of Duchenne Muscular Dystrophy. *Molecular Neurobiology*, **57**, 1748–1767.
23. Mauduit, O., Delcroix, V., Lesluyes, T., Pérot, G., Lagarde, P., Lartigue, L., Blay, J.-Y. and Chibon, F. (2019) Recurrent DMD Deletions Highlight Specific Role of Dp71 Isoform in Soft-Tissue Sarcomas. *Cancers*, **11**, 922.
24. Luce, L.N., Abbate, M., Cotignola, J. and Giliberto, F. (2017) Non-myogenic tumors display altered expression of dystrophin (DMD) and a high frequency of genetic alterations. *Oncotarget*, **8**, 145–155.
25. Tan, S., Tan, J., Tan, S., Zhao, S., Cao, X., Chen, Z., Weng, Q., Zhang, H., Wang, K.K., Zhou, J., *et al.* (2016) Decreased Dp71 expression is associated with gastric adenocarcinoma prognosis. *Oncotarget*, **7**, 53702–53711.
26. Tan, S., Tan, S., Chen, Z., Cheng, K., Chen, Z., Wang, W., Wen, Q. and Zhang, W. (2016) Knocking down Dp71 expression in A549 cells reduces its malignancy in vivo and in vitro. *Cancer Investigation*, **34**, 16–25.
27. Körner, H., Epanchintsev, A., Berking, C., Schuler-Thurner, B., Speicher, M.R., Menssen, A. and Hermeking, H. (2007) Digital karyotyping reveals frequent inactivation of the Dystrophin/DMD gene in malignant melanoma. *Cell Cycle*, **6**, 189–198.
28. Nikitin, E.A., Malakho, S.G., Biderman, B. v., Baranova, A. v., Lorie, Y.Y., Shevelev, A.Y., Peklo, M.M., Vlasik, T.N., Moskalev, E.A., Zingerman, B. v., *et al.* (2007) Expression level of lipoprotein lipase and dystrophin genes predict survival in B-cell chronic lymphocytic leukemia. *Leukemia and Lymphoma*, **48**, 912–922.
29. Gallia, G.L., Zhang, M., Ning, Y., Haffner, M.C., Batista, D., Binder, Z.A., Bishop, J.A., Hann, C.L., Hruban, R.H., Ishii, M., *et al.* (2018) Genomic analysis identifies frequent deletions of Dystrophin in olfactory neuroblastoma. *Nature Communications*, **9**, 1–6.

30. Juratli, T.A., McCabe, D., Nayyar, N., Williams, E.A., Silverman, I.M., Tummala, S.S., Fink, A.L., Baig, A., Martinez-Lage, M., Selig, M.K., *et al.* (2018) DMD genomic deletions characterize a subset of progressive/higher-grade meningiomas with poor outcome. *Acta Neuropathologica*, **136**, 779–792.
31. Ruggieri, S., de Giorgis, M., Annese, T., Tamma, R., Notarangelo, A., Marzullo, A., Senetta, R., Cassoni, P., Notarangelo, M., Ribatti, D., *et al.* (2019) Dp71 expression in human glioblastoma. *International Journal of Molecular Sciences*, **20**, 7–10.
32. Ferrari, D., Munerati, M., Melchiorri, L., Hanau, S., di Virgilio, F. and Baricordi, O. (1994) Responses to extracellular ATP of lymphoblastoid cell lines from Duchenne muscular dystrophy patients. American Journal of Physiology - Cell Physiology, 267, C886–C892.
33. McFarland, J.M., Ho, Z. v., Kugener, G., Dempster, J.M., Montgomery, P.G., Bryan, J.G., Krill-Burger, J.M., Green, T.M., Vazquez, F., Boehm, J.S., *et al.* (2018) Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nature Communications*, **9**, 1–13.
34. Dempster, J.M., Rossen, J., Kazachkova, M., Pan, J., Kugener, G., Root, D.E. and Tsherniak, A. (2019) Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. *bioRxiv*, p.720243.
35. Young, C.N.J., Gosselin, M.R.F., Rumney, R., Oksiejuk, A., Chira, N., Bozycki, L., Matryba, P., Łukasiewicz, K., Kao, A.P., Dunlop, J., *et al.* (2020) Total Absence of Dystrophin Expression Exacerbates Ectopic Myofiber Calcification and Fibrosis and Alters Macrophage Infiltration Patterns. *American Journal of Pathology*, **190**, 190–205.
36. Yeung, D., Zabłocki, K., Lien, C., Jiang, T., Arkle, S., Brutkowski, W., Brown, J., Lochmuller, H., Simon, J., Barnard, E.A., *et al.* (2006) Increased susceptibility to ATP via alteration of P2X receptor function in dystrophic mdx mouse muscle cells. *The FASEB Journal*, **20**, 610–620.
37. Goswami, M. v., Tawalbeh, S.M., Canessa, E.H. and Hathout, Y. (2021) Temporal Proteomic Profiling During Differentiation of Normal and Dystrophin-Deficient Human Muscle Cells. *Journal of Neuromuscular Diseases*, **8**, S205–S222.
38. Palladino, M., Gatto, I., Neri, V., Straino, S., Smith, R.C., Silver, M., Gaetani, E., Marcantoni, M., Giarretta, I., Stigliano, E., *et al.* (2013) Angiogenic impairment of the vascular endothelium: A novel mechanism and potential therapeutic target in muscular dystrophy. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **33**, 2867–2876.
39. Dabiré, H., Barthélémy, I., Blanchard-Gutton, N., Sambin, L., Sampedrano, C.C., Gouni, V., Unterfinger, Y., Aguilar, P., Thibaud, J.L., Ghaleh, B., *et al.* (2012) Vascular endothelial dysfunction in Duchenne muscular dystrophy is restored by bradykinin through upregulation of eNOS and nNOS. *Basic Research in Cardiology*, **107**, 1–9.
40. Kodippili, K., Thorne, P.K., Laughlin, M.H. and Duan, D. (2021) Dystrophin deficiency impairs vascular structure and function in the canine model of Duchenne muscular dystrophy. *Journal of Pathology*, **254**, 589–605.
41. Zabłocka, B., Górecki, D.C. and Zabłocki, K. (2021) Disrupted calcium homeostasis in duchenne muscular dystrophy: A common mechanism behind diverse consequences. *International Journal of Molecular Sciences*, **22**, 11040.
42. Forst, J., Forst, R., Leithe, H. and Maurin, N. (1998) Platelet function deficiency in Duchenne muscular dystrophy. *Neuromuscular Disorders*, **8**, 46–49.
43. Schorling, D.C., Müller, C.K., Pechmann, A., Borell, S., Rosenfelder, S., Kölbel, H., Schara, U., Zieger, B. and Kirschner, J. (2021) Impaired secretion of platelet granules in patients with Duchenne muscular dystrophy – results of a prospective diagnostic study. *Neuromuscular Disorders*, **31**, 35–43.
44. Chelly, J., Gilgenkrantz, H., Hugnot, J.P., Hamard, G., Lambert, M., Recan, D., Akli, S., Cometto, M., Kahn, A. and Kaplan, J.C. (1991) Illegitimate transcription. Application to the analysis of

truncated transcripts of the dystrophin gene in nonmuscle cultured cells from duchenne and becker patients. *Journal of Clinical Investigation*, **88**, 1161–1166.

45. Hale, J.E. (2013) Advantageous Uses of Mass Spectrometry for the Quantification of Proteins. *International Journal of Proteomics*.
46. Zhang, G., Ueberheide, B.M., Waldemarson, S., Myung, S., Molloy, K., Eriksson, J., Chait, B.T., Neubert, T.A. and Fenyö, D. (2010) Protein quantitation using mass spectrometry. *Methods in molecular biology (Clifton, N.J.)*, **673**, 211–222.
47. Lidov, H.G., Byers, T.J., Watkins, S.C. and Kunkel, L.M. (1990) Localization of dystrophin to postsynaptic regions of central nervous system cortical neurons. Localization of dystrophin to postsynaptic regions of central nervous system cortical neurons. *Nature*, ***348***, 725–728.
48. Dogra, C., Changotra, H., Wergedal, J.E. and Kumar, A. (2006) Regulation of phosphatidylinositol 3-kinase (PI3K)/Akt and nuclear factor-kappa B signaling pathways in dystrophin-deficient skeletal muscle in response to mechanical stretch. *Journal Cellular Physiology*, **208**, 575–585.
49. Haslett, J.N., Sanoudou, D., Kho, A.T., Han, M., Bennett, R.R., Kohane, I.S., Beggs, A.H. and Kunkel, L.M. (2003) Gene expression profiling of Duchenne muscular dystrophy skeletal muscle. *Neurogenetics*, **4**, 163–171.
50. Ma, Y., Zhang, P., Wang, F., Yang, J., Yang, Z. and Qin, H. (2010) The relationship between early embryo development and tumourigenesis. *Journal of Cellular and Molecular Medicine*, **14**, 2697–2701.
51. Kittler, R., Pelletier, L., Heninger, A.K., Slabicki, M., Theis, M., Miroslaw, L., Poser, I., Lawo, S., Grabner, H., Kozak, K., *et al.* (2007) Genome-scale RNAi profiling of cell division in human tissue culture cells. *Nature Cell Biology*, **9**, 1401–1412.
52. Górecki, D.C., Łukasiuk, K., Szklarczyk, A. and Kaczmarek, L. (1998) Kainate-evoked changes in dystrophin messenger RNA levels in the rat hippocampus. *Neuroscience*, **84**, 467–477.
53. García-Rodríguez, R., Hiller, M., Jiménez-Gracia, L., van der Pal, Z., Balog, J., Adamzek, K., Aartsma-Rus, A. and Spitali, P. (2020) Premature termination codons in the DMD gene cause reduced local mRNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 15664–16464.
54. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., *et al.* (2020) Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*, **38**, 669–673.
55. DepMap, B. (2021) DepMap 21Q4 Public. 10.6084/m9.figshare.16924132.v1.
56. Nusinow, D.P., Szpyt, J., Ghandi, M., Rose, C.M., McDonald, E.R., Kalocsay, M., Jané-Valbuena, J., Gelfand, E., Schweppe, D.K., Jedrychowski, M., *et al.* (2020) Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell*, **180**, 387-402.e16.
57. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., *et al.* (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, **47**, D607–D613.
58. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
59. Samaras, P., Schmidt, T., Frejno, M., Gessulat, S., Reinecke, M., Jarzab, A., Zecha, J., Mergner, J., Giansanti, P., Ehrlich, H.C., *et al.* (2020) ProteomicsDB: A multi-omics and multi-organism resource for life science research. *Nucleic Acids Research*, **48**, D1153–D1163.
60. Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D. and von Mering, C. (2015) Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics*, **15**, 3163–3168.

61. Kuleshov, M. v., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, **44**, W90–W97.