bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Downregulation of dystrophin expression occurs across diverse tumors, correlates with the age of onset, staging and reduced survival of patients.

Nancy Alnassar1, Malgorzata Borczyk2, Georgia Tsagkogeorga3,4, Michal Korostynski2, 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 of the *DMD* gene, encoding dystrophins, cause Duchenne muscular dystrophy (DMD). Some tumors also display altered dystrophin expression and recent studies identified a developmental onset of DMD. Given that embryogenesis and carcinogenesis share many mechanisms, we analyzed a broad spectrum of tumors to establish whether dystrophin loss evokes related outcomes. Transcriptomic, proteomic, and mutation datasets from fifty tumor tissues and matching controls (10,894 samples) and 140 corresponding tumor cell lines were analyzed. Interestingly, *DMD* expression was widespread across healthy tissues at levels comparable to housekeeping genes. In 80% of tumors, *DMD* expression was reduced due to transcriptional downregulation and not somatic mutations. The full-length transcript encoding Dp427 was decreased in 68% of tumors, while Dp71 variants showed variability of expression. Hierarchical clustering analysis of *DMD* transcripts distinguished malignant from control tissues. Transcriptomes of primary tumors and tumor cell lines with low *DMD* expression showed enrichment of specific pathways in the differentially expressed genes. Pathways

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

consistently identified: ECM-receptor interaction, calcium signaling and PI3K-Akt, are also altered in DMD muscle. Notably, low *DMD* expression was associated with a more advanced stage, older age of onset, and reduced survival across different tumors. Thus, *DMD* transcription occurs throughout a spectrum of normal tissues. The molecular signature associated with its frequent downregulation in malignancies is concordant with changes found in Duchenne muscles, even though these malignancies originate from tissues never previously associated with dystrophin expression or function. Therefore, the importance of this largest known gene extends beyond its roles identified in DMD, certainly into oncology.

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 fetuses (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 tumors, 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) – traits that are commonly associated with tumor 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 signaling pathways (14). Yet, in satellite cells, Dp427m has a different interactome and it is essential for asymmetric cell divisions (11,15), and in myoblasts the loss

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 summarized 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 tumors with myogenic programs (17) and soft tissue sarcomas (23). In contrast to this potential tumor suppressor role of the full-length dystrophin, Dp71 expression was essential for myogenic tumor 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), leukemia (28), lymphoma (24), and CNS tumors (29–31).

Yet, many of the tumors 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 tumors: myoblasts, lymphocytes, and endotheliocytes express 14-kb dystrophin transcripts but are not known to synthesize detectable levels of dystrophin protein. Nevertheless, *DMD* downregulation leads to significant functional abnormalities in these cells (16,32,33).

We hypothesized that malignancy can be used as a model to investigate changes in *DMD* gene expression across normal tissues and corresponding tumors 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 normal tissues and matching tumors and explored transcriptomic alterations in primary tumors and corresponding tumor cell lines with altered *DMD* expression to identify putative downstream molecular pathways that could be associated with *DMD* dysregulation across human tissues. We also examined the association of *DMD* gene expression with the onset and survival endpoints in tumor patients.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

1. Gene, protein and clinical datasets

The RNA-seq data for tumor (n= 7,894) and control (n= 2,714) tissue samples was obtained from the UCSC Xena Functional Genomics Browser (http://xena.ucsc.edu). The "RSEM norm_count" and "RSEM expected_count" datasets of the TCGA TARGET GTEx cohort were used to examine expression data at the gene and transcript level, respectively, for a variety of cancers in The Cancer Genome Atlas (TCGA) database (https://www.cancer.gov/tcga), different non-diseased tissues in the Genotype-Tissue Expression (GTEx) project (https://gtexportal.org/), and two pediatric hematological malignancies in the TARGET (Therapeutically Applicable Research To Generate Effective Treatments) database (https://ocg.cancer.gov/programs/target). Samples from the TCGA, TARGET and GTEx have been re-analyzed and processed by the same bioinformatic pipeline to eliminate batch effects resulting from different computational processing (34). The expression level of *DMD* gene/ transcripts was compared between tumor tissues and histologically normal tissues adjacent to the tumors (NATs) from the TCGA (16 comparisons). When expression data for NATs was not available in the TCGA, a comparison was made between *DMD* gene/ transcripts expression in tumor samples from the TCGA and healthy tissue samples in the GTEx database (nine comparisons). Database sources in each comparison are stated in Supplementary Table 1. The level of *DMD* transcripts was also compared between two TARGET pediatric malignancies and GTEx healthy samples.

The Proteomics DB data resource (https://www.proteomicsdb.org/vue/) was used to examine the expression of the full-length dystrophin protein 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 (35).

Tissue specificity (TS) scores (i.e. the enrichment) of dystrophin protein across a variety of normal tissues was visualized at (https://tsomics.shinyapps.io/RNA_vs_protein/), as well as the correlation between *DMD* gene and protein expression in 32 different tissue types covering all major organs including skeletal muscle (36).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (37).

Dystrophin protein expression in normal tissues was also 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 (38).

Mutation and somatic copy number alteration (SCNA) data for primary tumor samples was derived from the datasets "somatic mutation (SNP and INDEL) – MC3 public version" and "gene-level copy number (gistic2_thresholded)" of the TCGA Pan-cancer (PANCAN) cohort available at the UCSC Xena Browser. The SCNA data was generated by the GISTIC2 algorithm, which generates putative gene copy number specific calls (39). It should be noted that deep deletions and amplifications are considered as biologically relevant for individual genes by default. Due to purity and ploidy differences between samples, and because these calls are usually not manually reviewed, there may be false positives and false negatives (40,41).

DMD gene expression and mutation data for tumor cell lines from the Cancer Cell Line Encyclopedia was obtained from cBioPortal website (https://www.cbioportal.org).

The $\log_2$ (norm_value + 1) miRNA mature strand expression data was obtained from the dataset "miRNA mature strand expression – Batch effects normalized miRNA data" of the TCGA PANCAN cohort at the UCSC Xena browser.

Data for the American Joint Committee on Cancer (AJCC) pathologic tumor stage and the patient's age at the initial diagnosis was derived from the dataset "Phenotype – curated clinical data" of the TCGA PANCAN cohort. Tumor samples were allocated into the following age groups: 14-29 years, 30-39, 40-49, 50-59, 60-69, 70-79 and 80 and above years. These groups corresponded with the age groups for GTEx donors obtained from the GTEx portal with two exceptions. The GTEx data did not have any samples from patients older than 79 years-old or younger than 20 years-old.

The normalized log-transformed RNA-seq data used to compare gene expression in primary tumor samples with low vs. high *DMD* expression in 15 TCGA primary tumors was obtained

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

from the UCSC Xena Browser. Links to database sources can be found in Supplementary File 2.

Expression data for the *DMD* gene in tumor cell lines was obtained from the "Expression 21Q3 Public" dataset using the Cancer Dependency Map (DepMap) portal (https://depmap.org/portal), which includes RNA-seq expression data for 1,379 tumor cell lines from 37 lineages (42).

Gene expression microarray data for skeletal muscle samples from DMD patients (n= 12) and healthy donors (n= 11) was obtained from the dataset GSE1004 available at the Gene Expression Omnibus (GEO) from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/geo/).

Survival data was derived from the dataset "Phenotype – curated clinical data" of the TCGA PANCAN cohort and the dataset "Phenotype – TARGET donor phenotype" of the TARGET PANCAN cohort available at the UCSC Xena Browser.

The STRING database was used to construct dystrophin's protein-protein interaction (PPI) network (43).

2. *DMD* transcript identification and classification

The "RSEM expected_count" dataset of the TCGA TARGET GTEx cohort in the UCSC 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. However, we note that estimates of transcript expression by RSEM may not be 100% accurate. 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 analysis of differential *DMD* gene and transcript expression between tumor tissues and corresponding controls

Statistical analysis of the expression level of the *DMD* gene and transcripts was performed using the UCSC Xena Browser, using the two-tailed Welch's t-test. The *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, and when this adjustment gave a value above 1, the corrected *p*-values were set at 1. The Bonferroni-corrected *p*-values were then compared

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with the overall level of α (α= 0.05). Log-transformed expression data was downloaded from the UCSC Xena Browser and the mean of the *DMD* gene and transcripts expression level for tumor (A) and control (B) samples was calculated. The following formula was then used to calculate Log Fold Change (LogFC): log (A/B) = log (A) – log (B). The UCSC 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 tumors and corresponding control 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 standardized to a sum of 1 in order to compare the relative expression levels of these transcripts in different tumor and control 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 2). 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 analyzed tissues using Kruskal Wallis test and Dunn's multiple comparisons test.

5. Association between *DMD* gene expression and mutation status

Tumor samples from 23 tumor types with available mutation and SCNAs data (n= 6,751) were used to perform a univariate general linear model (GLM) analysis to determine whether there is a significant difference in *DMD* gene expression between samples with different mutation/SCNA types and samples with no mutations/SCNAs in the *DMD* gene locus using SPSS statistical software (v.28.0.0.0). Gender was selected as a random factor to account for gender differences between groups. Estimated Marginal (EM) means were calculated, and *p*-values were adjusted using the Bonferroni correction.

Next, primary tumor samples and tumor cell line samples (n= 921) were ranked according to their level of *DMD* expression and those at the bottom 30% and top 30% were selected for

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

further analysis. A Chi Square test was used to determine whether there is an association between *DMD* expression level and exonic mutations, non-exonic mutations, and SCNAs in tumor samples and between *DMD* mutations and expression level in tumor cell lines.

6. **Association between *DMD* gene expression and cancer stage and patient's age at the initial diagnosis**

Tumor samples from 18 tumor types with available data for cancer stage, gender, and patient's age at the initial diagnosis (n= 6,118) were classified into four groups based on the cancer stage: II, II, III and IV. A univariate GLM analysis was performed to determine whether there is a significant difference in *DMD* gene expression between these four stages. Gender and age were selected as random factors, and EM means were calculated for the four groups. The *p*-values were adjusted using the Bonferroni correction.

7. **Identification of differentially expressed genes (DEGs) and miRNAs between primary tumor samples with low and high levels of *DMD* gene expression**

Gene expression in primary tumors samples with low vs. high levels of *DMD* gene expression from 15 different primary tumors was compared. For each tumor, 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 DEGs 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 DEGs between Groups A and B, and *p*-values were adjusted using the Bonferroni correction. LogFC was calculated using the formula (LogFC): log (A/B) = log (A) – log (B). Genes with an FDR-corrected *p*-value < 0.1, a Bonferroni-corrected *p*-value < 0.05, and a |LogFC| value >= 0.7 were considered differentially expressed in Group A compared to Group B. 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%20tumor%20samples). Similarly, the differentially expressed miRNAs were identified.

8. **Identification of DEGs in tumor cell lines with low and high levels of *DMD* gene expression**

We performed two comparisons in tumor cell lines from two distinct categories, carcinoma, and sarcoma. A two-class comparison analysis was conducted using the DepMap custom

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analysis tool (https://depmap.org/portal/interactive/). The "Expression 21Q2 Public" was selected as the input dataset, and the cell lines with low *DMD* expression were used as the "in" group cell lines, while cell lines with high *DMD* expression were used as the "out" group cell lines. The analysis compared gene expression between the selected groups and generated estimates of effect size and corresponding q-values. The DEGs in each of these two comparisons (q-value < 0.05) were identified.

9. Identification of DEGs in skeletal muscle samples from DMD patients in comparison to healthy skeletal muscle

The GEO2R tool at the NCBI website (http://www.ncbi.nlm.nih.gov/geo/geo2r/) was used to identify the DEGs between 12 skeletal muscle samples collected from DMD patients and 11 samples collected from healthy donors. The default parameters of the GEO2R tool were used, and the NCBI-generated annotations were used to display the list of DEGs. The *p*-values were adjusted using the FDR correction. Genes with an FDR-adjusted *p*-value < 0.1 and a |LogFC| value >= 0.7 were considered differentially expressed between DMD and normal skeletal muscle samples. The R script generated by the GEO2R tool is available at (https://github.com/nancyalnassar/DMD-gene-in-cancer/blob/main/GEO2R%20Script).

10. Identification of DEGs between hematological malignancies samples with low and high levels of Dp71 expression

We compared gene expression in hematological malignancies samples with low vs. high levels of Dp71 expression from two TCGA and two TARGET studies. Samples (n= 286) were ranked based on the sum expression level of Dp71, Dp71b and Dp71ab. Samples were then divided into three groups: Group A (samples at the bottom 33.3% of Dp71 expression), Group B (samples at the top 33.3% of Dp71 expression) and Group C (the remaining samples with medium level of Dp71 expression). Using the R v4.0.4, an ANOVA was preformed to identify the DEGs between the three groups, and *p*-values were adjusted using the FDR correction. Next, Post-Hoc Pairwise T tests were performed to identify the DEGs between Groups A and B, and *p*-values were adjusted using the Bonferroni correction. Genes with an FDR-corrected *p*-value < 0.1, a Bonferroni-corrected *p*-value < 0.05, and a |LogFC| value >= 0.7 were considered differentially expressed in Group A compared to Group B.

11. Functional enrichment analysis

The EnrichR tool (44) was used to identify the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) Biological Process terms enriched in the DEGs

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in the previously mentioned comparisons. In comparisons where the number of DEGs 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.

12. Survival analysis

Samples (n= 6,931) from 15 different TCGA primary tumors were ranked based on their level of *DMD* gene expression. Next, the following survival endpoints were compared between patients at the bottom 25% of *DMD* expression and those at the top 25% of *DMD* expression: overall survival (OS), progression-free interval (PFI), disease-specific survival (DSS) and disease-free interval (DFI).

Samples (n= 271) from two TCGA and two TARGET studies of hematological malignancies were ranked based on the sum expression level of Dp71, Dp71b and Dp71ab. Next, OS was compared between patients at the bottom 25% of Dp71 expression and those at the top 25% of Dp71 expression. Kaplan-Meier curves were generated in GraphPad and analyzed using the log-rank test.

Results

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

Expression data for 17 healthy tissues as well as skeletal muscle tissue from the GTEx database was examined. These heathy tissues were adrenal glands, bladder, breast, cervix, colon, esophagus, kidney, liver, lung, ovary, pancreas, prostate, skin, stomach, thyroid, uterus, and whole blood. *DMD* expression levels across this spectrum of healthy tissues were compared with the expression levels of two housekeeping genes (HKGs) *PKG1* and *HMBS*, which were identified as HKGs across 32 tissues in the GTEx database (36).

DMD expression averaged 79.3% of *PKG1* (range between 30.8 and 94.6%) and 114.7% of *HMBS* (43.7 to 137.4%) expression levels with the lowest expression found in whole blood (Supplementary Table 3). In skeletal muscle, *DMD* expression was 94.6% of *PKG1* and 138.2% of *HMBS*.

DMD expression relative to *PKG1* and *HMBS* was compared between skeletal muscle tissue and the previously mentioned 17 healthy tissues using a Kruskal-Wallis and Dunn's multiple comparisons test. The expression of *DMD* relative to *PKG1* in skeletal muscle was comparable to that in the pancreas, uterus, and bladder tissue (adjusted *p*-value > 0.9999), and

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DMD expression relative to *HMBS* was comparable to that in the uterus, ovary, and bladder tissue (adjusted *p*-value > 0.9999).

Given this widespread presence of considerable *DMD* transcript levels, we investigated whether it is accompanied by protein expression. We 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). Moreover, quantitative profiling of the proteome of 201 samples from 32 tissues in the GTEx database identified dystrophin as a housekeeping protein, as it was present in all the 32 tissues analyzed (36). Dystrophin protein expression was statistically significantly and positively correlated with *DMD* gene expression in those samples (Spearman correlation = 0.67, BH-adjusted *p*-value < 0.1).

Finally, in the Protein Abundance Database (PAXdb), dystrophin expression was ranked in the top 25% of MS-quantified proteins in a range of tissues, such as fallopian tubes, esophagus, uterus, bladder, colon, prostate, rectum, in addition to the heart. Thus, data from three databases demonstrated significant expression of dystrophin protein in a range of healthy tissues. Therefore, we investigated whether alterations in *DMD* expression might occur in tumors that originate from tissues not commonly associated with dystrophin protein expression or function.

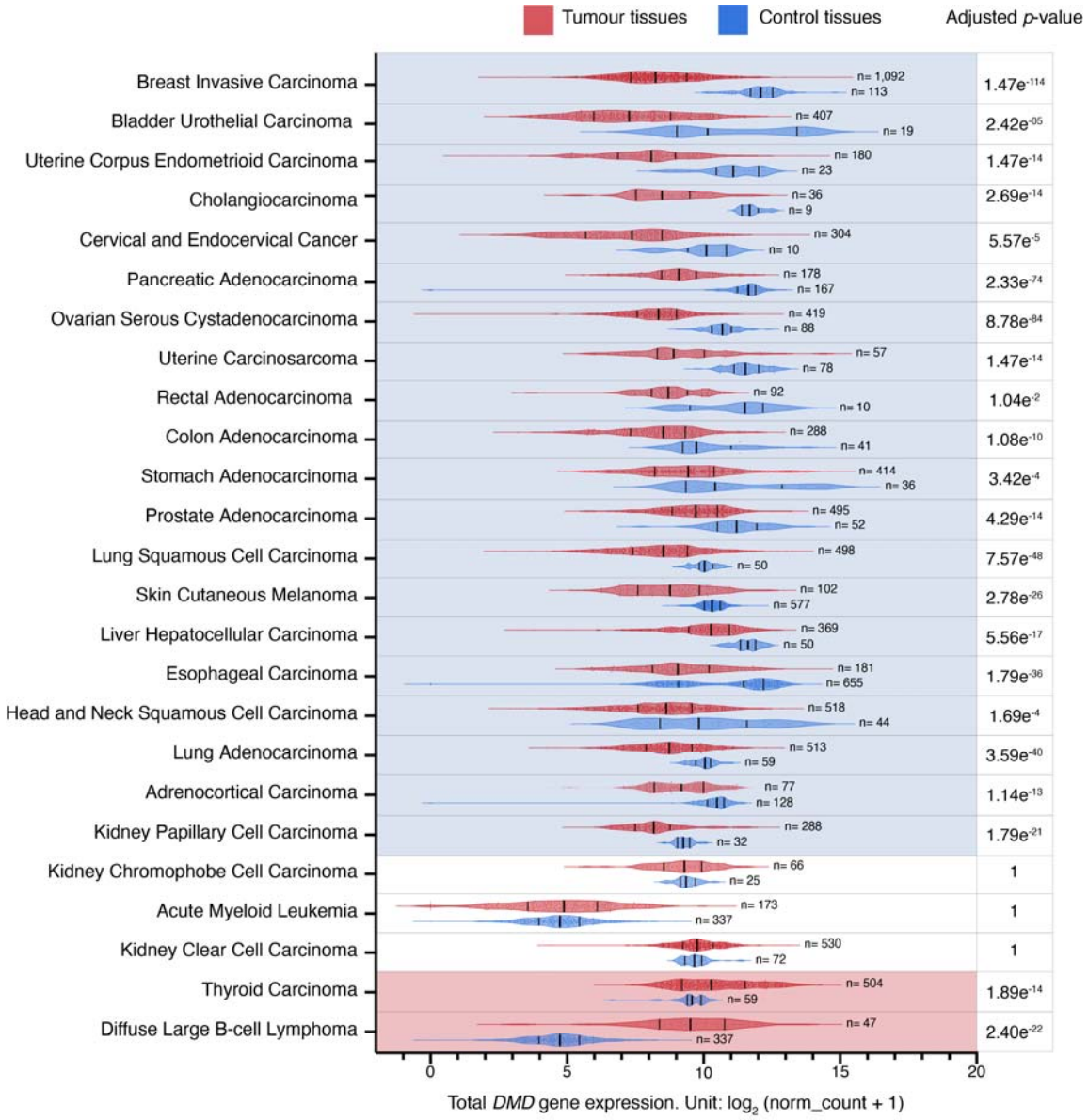
2. Downregulation of *DMD* gene expression across malignancies

We investigated RNA-seq expression data for 25 different types of primary tumors from the TCGA database (carcinomas, melanoma, lymphoma, and leukemia) and their corresponding NATs from the TCGA or healthy GTEx tissues. The analyzed primary tumors were acute myeloid leukemia (LAML), adrenocortical carcinoma (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 carcinoma (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

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 tumors in comparison to their corresponding control tissues (Figure 1; Supplementary Table 1). 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 tumors 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 leukemia, kidney clear cell, and chromophobe cell carcinomas).



bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1. *DMD* gene expression level in primary tumors and corresponding control tissues. Red and blue violin plots represent tumor and control tissues, respectively. Vertical black lines represent the median and quartiles. Comparisons (n= 20) with a statistically significant downregulation of *DMD* expression in tumor samples compared to controls are highlighted in blue, and comparisons (n= 2) with a statistically significant upregulation of *DMD* expression in tumor samples compared to controls are highlighted in red ($p < 0.05$, Supplementary Table 1). The *p*-values were calculated using the UCSC Xena Browser using a two-tailed Welch's t test and adjusted for multiple testing using the Bonferroni correction.

To examine the impact of the batch effect resulting from differences in processing between TCGA and GTEx samples (45) in 13 analyses, where *DMD* expression was compared between TCGA tumors and their corresponding NATs, we also compared *DMD* expression between these tumors and corresponding healthy tissues from the GTEx database, We found consistent results in 11 out of 13 comparisons, including BRCA, BLCA, UCEC, COAD, STAD, PRAD, LUSC, LIHC, LUAD, KICH and THCA (Supplementary Table 4). In KIRP, *DMD* expression was reduced compared to kidney NAT in TCGA datasets (adjusted *p*-value= $9.30e^{-22}$), however, no statistically significant difference was found between *DMD* expression in KIRP and GTEx healthy kidney tissue samples (adjusted *p*-value= 1). Moreover, there was no statistically significant difference in *DMD* expression between KIRC and paired kidney NAT samples (adjusted *p*-value= 1), while *DMD* expression was found upregulated in KIRC samples compared to GTEx healthy kidney tissue (adjusted *p*-value= $2.77e^{-08}$). These inconsistencies might result from small sample sizes leading to a reduction in statistical power.

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 profiles at the transcript level across various control tissues and corresponding tumors.

3. Hierarchical clustering analysis of the relative expression of *DMD* transcripts distinguishes tumor 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 analyzed samples, with the splice variant Dp71b lacking exon 78 having the highest mean abundance level of all *DMD* transcripts across the analyzed tumor and control tissues. 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 analyzed tumor and control 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**)

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ENST00000493412.1	4	134	H0Y864
ENST00000475732.1	2	None	

Hierarchical clustering analysis of these ten highly expressed transcripts in tumor and control tissues yielded six clusters. Supplementary Table 2 reports other dendrogram cut-offs. We focused mainly on three of these clusters (Figure 2) that contained the majority of tumor and control tissues. The first cluster was composed of nine tumors and healthy whole blood and pancreas tissue. The second cluster was composed of 16 control tissues (including six healthy GTEx tissues and 10 NATs from the TCGA) as well as SKCM and PRAD tumor samples. Finally, the third cluster contained 12 tumors in addition to kidney and thyroid NATs from the TCGA.

The relative expression of the transcript encoding Dp427m was significantly lower in the first and third clusters, that were composed mainly of tumors, compared to the second cluster with a majority of control tissues, while the relative expression of Dp71 variants was higher in the first and third clusters (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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

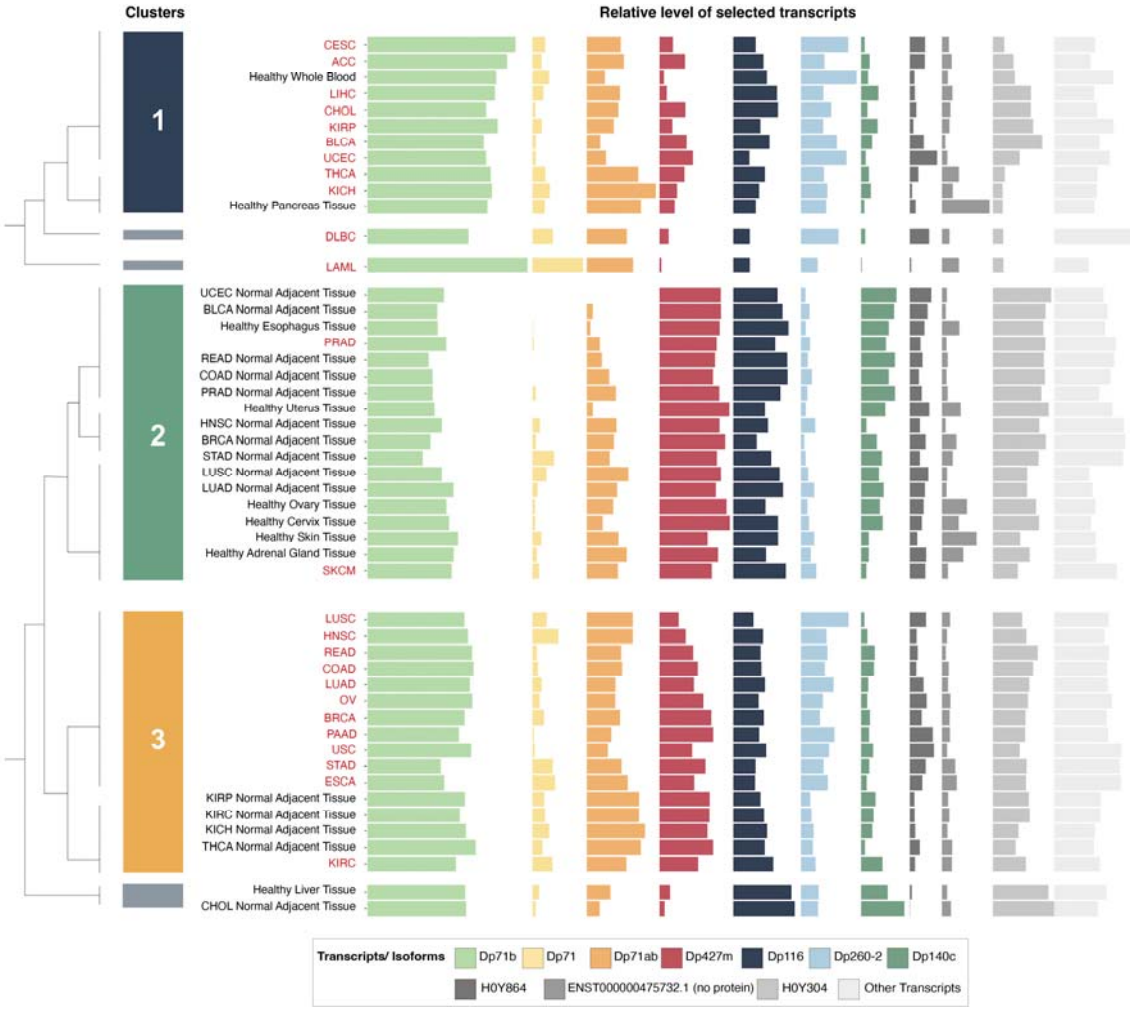


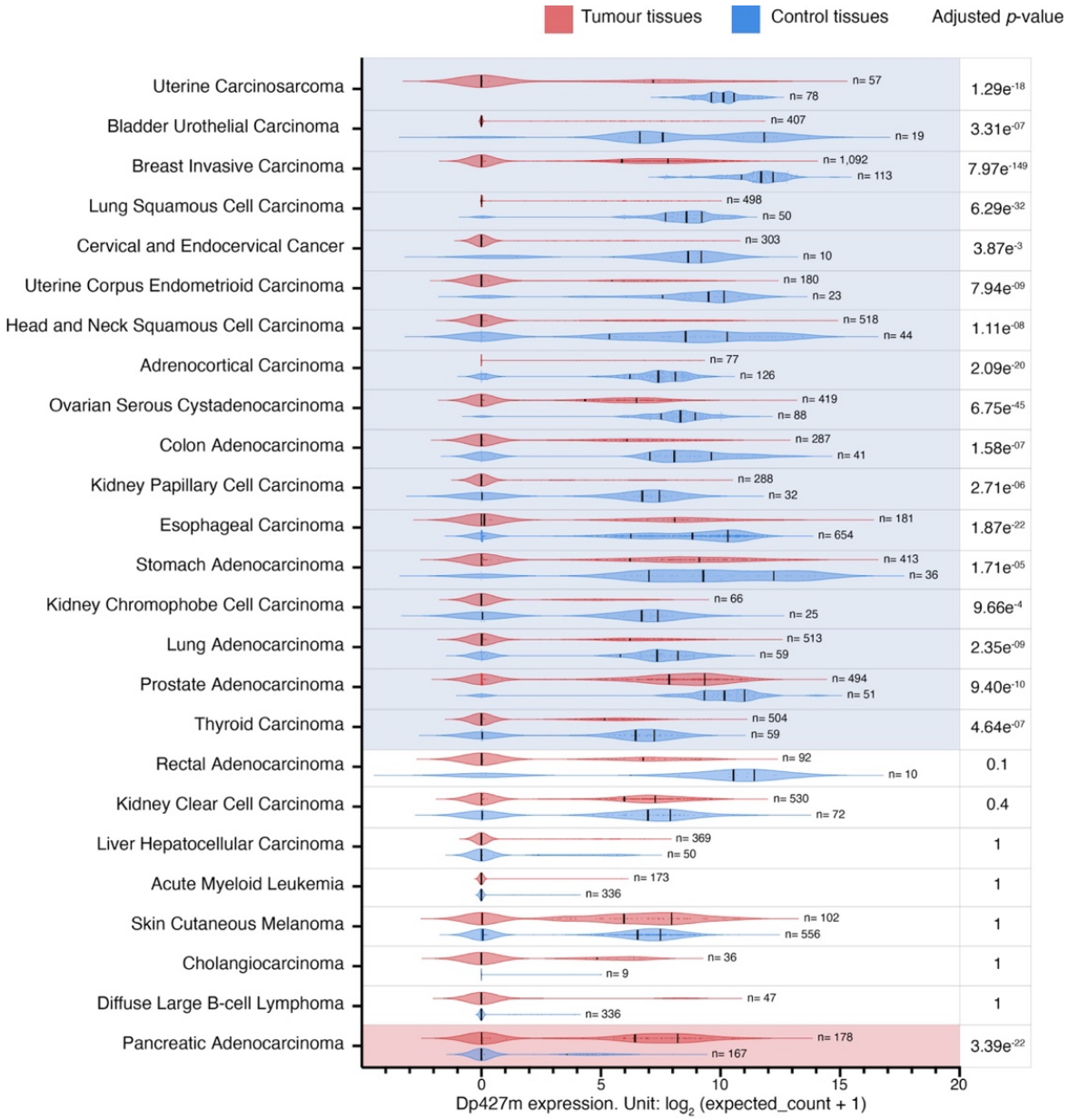
Figure 2. Bar plots representing the relative expression levels of the top 10 highly expressed *DMD* transcripts in tumor and control tissues and a dendrogram displaying the clusters identified by the hierarchical clustering analysis. Abbreviations: CESC (cervical and endocervical cancer), ACC (adrenocortical carcinoma), 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 (esophageal carcinoma), 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, whose loss is responsible for Duchenne muscular dystrophy, and the predominance of the transcript encoding Dp427m (Table 1), its expression patterns were analyzed in more detail. As Dp71 is the isoform most widely

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

expressed across the body and its splice variants were among the top highly expressed *DMD* transcripts in tumor and control tissues (Table 1), we also compared expression patterns of this transcript (Supplementary Table 5). Dp427m expression was decreased in primary tumors 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 20 out of 25 comparisons (a decrease in 17 and an increase in three) (Supplementary Table 5).



bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3. Dp427m transcript expression level in primary tumors and corresponding control tissues. Red and blue violin plots represent tumor and control tissues, respectively. Vertical black lines represent the median and quartiles. Comparisons (n= 17) with a statistically significant downregulation of Dp427m expression in tumor samples compared to controls are highlighted in blue, and one comparison with a statistically significant upregulation of Dp427m expression in tumor samples compared to controls is highlighted in red ($p < 0.05$, Supplementary Table 5). The *p*-values were calculated using the UCSC 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 tumor categories were confounded by the opposing dysregulation of Dp427m and Dp71 variants. Specifically, while *DMD* expression was higher in primary thyroid carcinoma compared to thyroid NAT, Dp427m expression there was lower compared to thyroid NAT 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 healthy pancreas tissue, but this decrease resulted from the lower expression of transcripts encoding Dp71 variants, while Dp427m expression was higher in this tumor type (LogFC = 3.7 compared to healthy pancreas tissue) (Figure 4.B).

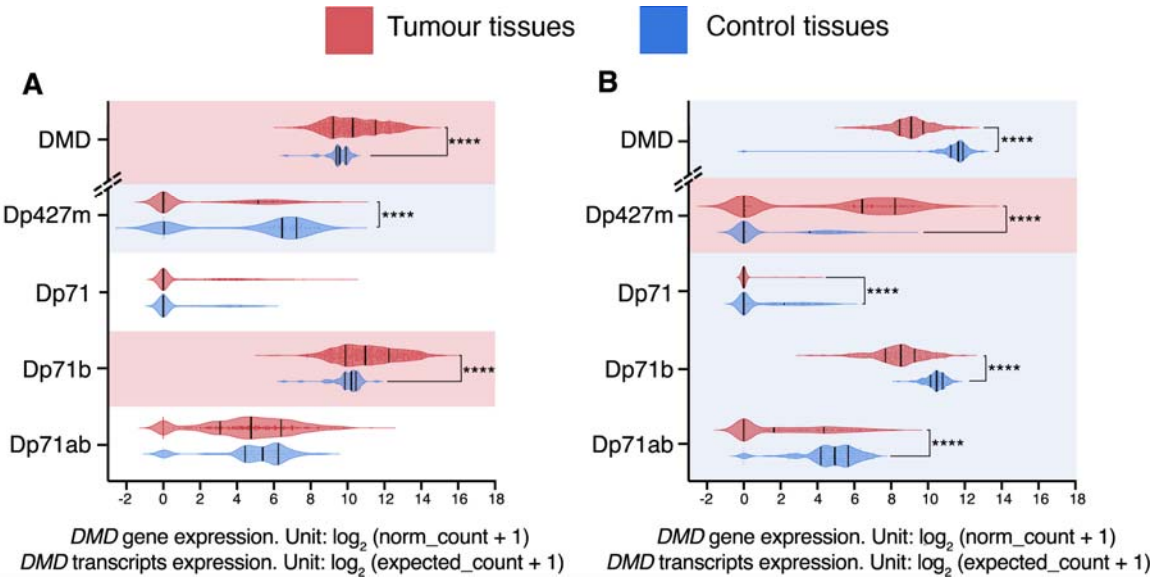


Figure 4. Total *DMD* gene expression and expression of transcripts encoding Dp427m, Dp71, Dp71b and Dp71ab in (A) primary thyroid carcinoma samples (n= 504) compared to thyroid NAT (n= 59) and (B) primary pancreatic adenocarcinoma samples (n= 178) and healthy pancreas tissue (n= 167). Red and blue violin plots represent tumor and control samples, respectively. Vertical black lines represent the median and quartiles. Comparisons where there is a statistically significant decrease in *DMD* gene/ transcript expression in tumor tissue compared to controls are highlighted in blue, while comparisons where there is a statistically significant

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

increase in *DMD* gene/ transcript expression in tumor tissue compared to controls are highlighted in red (****$p < 0.0001$). 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 UCSC Xena Browser. The *p*-values were calculated using the UCSC Xena Browser using a two-tailed Welch's t test and adjusted for multiple testing using the Bonferroni correction.

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

Next, using datasets for samples from 23 of the previously mentioned TCGA tumors (LAML and SKCM samples were excluded as they did not have mutation and SCNA details available) we investigated the association between *DMD* expression and exonic and non-exonic mutations as well as SCNAs in the *DMD* gene.

The majority of tumor samples had no identified exonic mutations in the *DMD* gene locus (6,201 out of 6,751). A univariate GLM analysis was carried out to assess the effect of exonic mutations and gender (to account for the X-chromosome localization of the *DMD* gene) as well as their interaction on *DMD* expression. The GLM indicated a significant effect for exonic mutations on *DMD* expression ($p = 0.007$). There was no effect for gender ($p = 0.703$) or the interaction between exonic mutations and gender ($p = 0.648$). Samples with missense and multiple mutations had significantly lower levels of *DMD* expression compared to samples with no exonic mutations in the *DMD* locus ($p < 0.001$ and $p = 0.043$, respectively) (Figure 5.A). A Chi Square test revealed that there is an overrepresentation of *DMD* mutations in tumor samples with low *DMD* expression ($X^2(1) = 45.44$, $p < 0.0001$). Specifically, 11.7% of samples in the low *DMD* group (237 out of 2,025) had *DMD* exonic mutations compared to 5.73% (116 out of 2,025) in the high *DMD* group. However, in the low *DMD* group, 88.3% of samples (1,788 out of 2,025) did not have any detectable mutations in the exonic regions of the *DMD* locus.

As for non-exonic mutations, 98.2% of tumor samples (6,632 out 6,751) had no mutations in the non-exonic regions of the *DMD* gene. Non-exonic mutations were found to have a significant effect on *DMD* expression ($p = 0.031$), but no effect was found for gender ($p = 0.104$) or the interaction between gender and non-exonic mutations ($p = 0.698$). Samples with intronic mutations had significantly lower *DMD* expression compared to samples without any

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DMD non-exonic mutations (p= 0.02) (Figure 5.B). A Chi Square test revealed that there is an overrepresentation of *DMD* non-exonic mutations in tumor samples with low *DMD* expression ($X^2(1) = 18.71$, $p < 0.0001$). In the low *DMD* group, 3.01% of tumor samples (61 out of 2,025) had *DMD* non-exonic mutations compared to 1.09% of samples (22 out of 2,025) in the high *DMD* group. However, 96.99% of samples in the low *DMD* group (1,964 out of 2,025) did not have any non-exonic mutations in the *DMD* locus.

Regarding SCNAs, 66.7% of samples had a normal copy number of the *DMD* locus (4,503 out of 6,751). Both SCNAs (p< 0.001) and gender (p= 0.006) had a significant effect on *DMD* expression. However, the interaction between the two factors did not have a significant effect. Samples with a normal *DMD* copy number had higher expression compared to samples with deep and shallow deletions and those with gains (p< 0.001). Additionally, samples with deep deletions had lower expression compared to those with amplifications (p= 0.002) (Figure 5.C).

In samples from female patients, we found that there is an overrepresentation of *DMD* SCNAs in the low *DMD* group ($X^2(1) = 36.40$, $p < 0.0001$), where 39.08% of samples (399 out of 1,021) had SCNAs in the *DMD* locus compared to 26.54% in the high *DMD* group (271 out of 1,021). In the low *DMD* group, 60.92% of samples (622 out of 1,021) did not have any SCNAs.

In samples from male patients, we also found that there is an overrepresentation of *DMD* SCNAs in the low *DMD* group ($X^2(1) = 69.82$, $p < 0.0001$), where 44.18% of samples (444 out of 1,005) had SCNAs in the *DMD* locus compared to 26.37% in the high *DMD* group (265 out of 1,005). In the low *DMD* group, 55.82% of samples had no SCNAs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

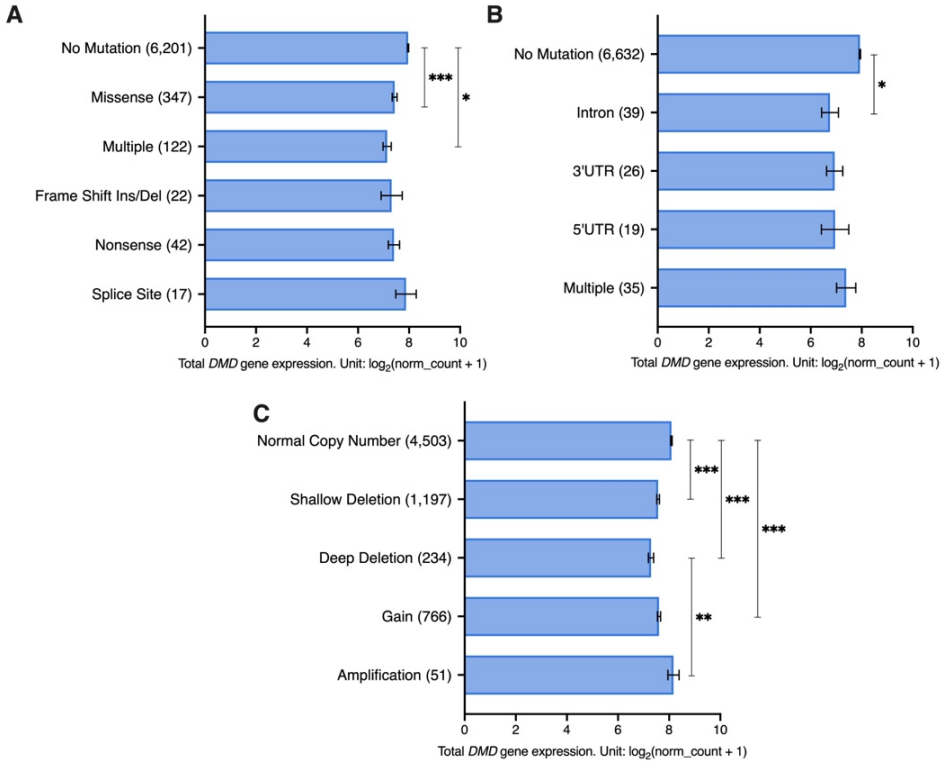


Figure 5. Total *DMD* gene expression in tumor samples from 23 TCGA studies with/ without (A) exonic, (B) non-exonic mutations and (C) SCNAs in the *DMD* gene. The number of samples in each group is indicated in brackets. A deep deletion indicates a deep loss (possibly a homozygous deletion). A shallow deletion indicates a shallow loss (possibly a heterozygous deletion). A gain indicates a low-level gain (a few additional copies, often broad). An amplification indicates a high-level amplification (more copies, often local). Expression data is represented as EM mean ± SEM. The *p*-values were adjusted using the Bonferroni correction. Asterisks indicate statistical significance (**p* <0.05, ***p* <0.01, ****p* < 0.001).

We further investigated the association between *DMD* expression and mutations using data from 921 tumor cell lines (cBioPortal). The majority of tumor 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). We found that there is an overrepresentation of *DMD* mutations in tumor cell lines with low levels of *DMD* expression ($X^2(1) = 9.727$, $p = 0.0018$). However, only 23% of tumor cell lines with low *DMD* gene expression (63 out of 276) had *DMD* mutations but the majority of cell lines (213 out of 276) had low levels of *DMD* expression without any detectable mutations.

Therefore, downregulation of *DMD* expression in tumors cannot be simply attributed to somatic mutations or copy number alterations within the *DMD* locus but rather involves a regulatory mechanism.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6. Association between *DMD* expression and cancer stage and patients' age

We investigated the association between *DMD* expression and cancer stage in 18 different types of primary tumors with available data for cancer stage, gender, and patient's age at the initial diagnosis. These tumors included ACC, BLCA, BRCA, CHOL, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PAAD, READ, SKCM, STAD, and THCA. The ages of patients ranged from 14 to 90 years old.

A GLM analysis was performed to assess the effect of cancer stage on *DMD* expression while accounting for gender and age differences. Cancer stage was found to have an effect on *DMD* expression ($p = 0.035$). While no effect was identified for gender differences ($p = 0.735$), the interaction between cancer stage and gender was found to have an impact on *DMD* expression ($p = 0.010$). Age of patients was also found to have an impact on *DMD* gene expression ($p < 0.001$) as well as the interaction between age and cancer stage ($p = 0.003$).

Samples from patients with stage I cancer had significantly higher levels of *DMD* gene expression compared to those with stage II (LogFC= 0.72, $p < 0.001$), stage III (LogFC= 0.50, $p < 0.001$), and stage IV cancer (LogFC= 0.69, $p < 0.001$) (Figure 6.A). Moreover, samples from younger patients had significantly higher *DMD* expression compared to samples from older patients (Figure 6.B).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

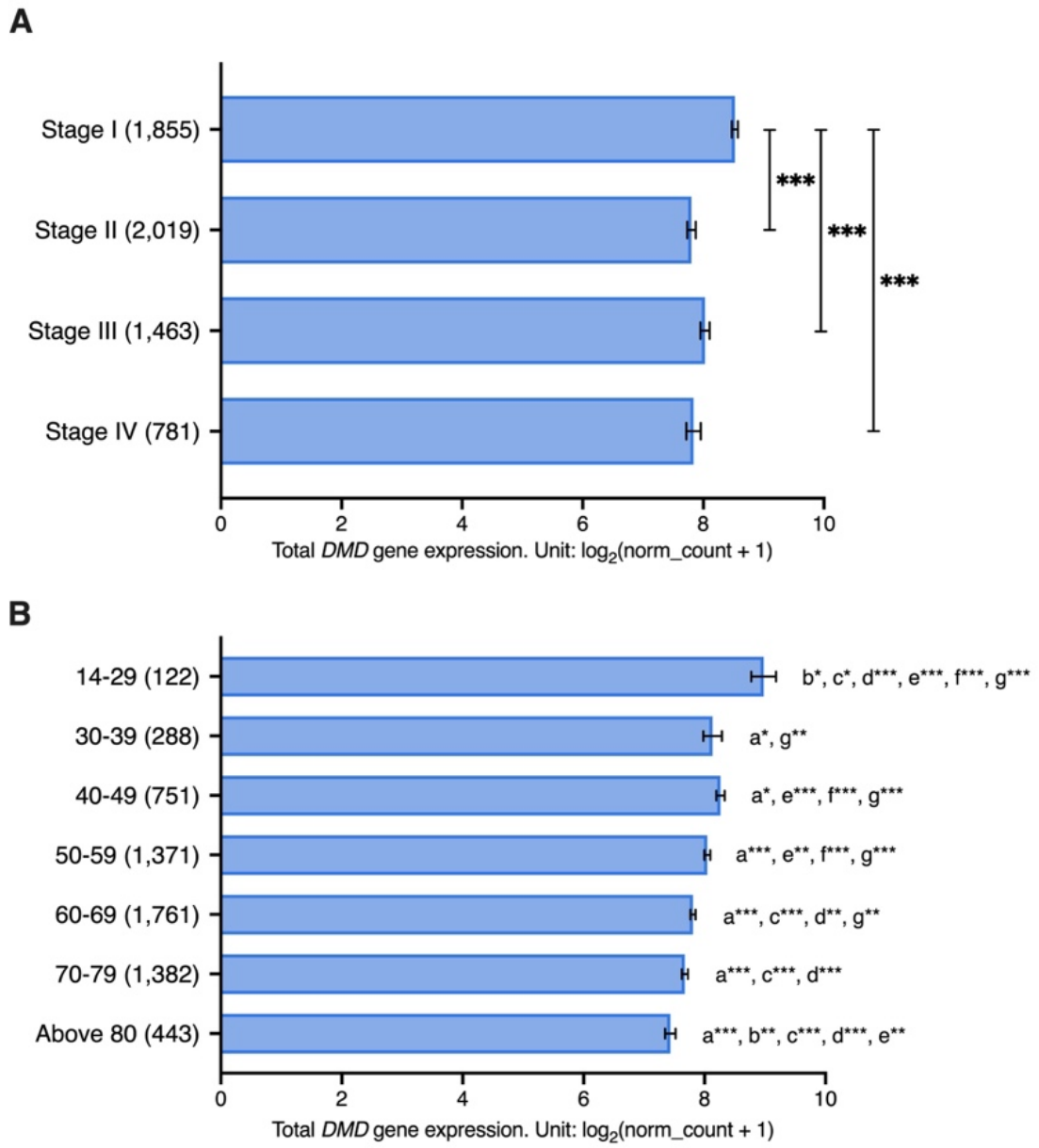


Figure 6. *DMD* gene expression in tumor samples from 18 TCGA tumors with different (A) cancer stages and (B) age groups. Expression data is represented as EM mean ± SEM. The number of samples in each group is indicated in brackets. The *p*-values were adjusted using the Bonferroni correction (a: significant difference from 14-29 age group, b: significant difference from 30-39 age group, c: significant difference from 40-49 age group, d: significant difference from 50-59 age group, e: significant difference from 60-69 age group, f: significant difference from 70-79 age group, g: significant difference from above 80 age group, p^* <0.05, p^{**} <0.01, p^{***} < 0.001). Patients aged between 14 and 29 years old had significantly higher *DMD* expression compared to patients in the 30-39 (LogFC= 0.84, p= 0.02), 40-49 (LogFC= 0.71, p= 0.02), 50-59 (LogFC= 0.93, p< 0.001), 60-69 (LogFC= 1.17, p< 0.001), 70-79 (LogFC= 1.30, p< 0.001), and above 80 age groups (LogFC= 1.54, p< 0.001). Patients aged between 30 and 39 years old had higher *DMD* expression compared to patients in the above 80 age group (LogFC= 0.70, p= 0.001). Patients in the 40-49 age groups had significantly higher *DMD*

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

levels than patients in the 60-69 (LogFC= 0.45, $p < 0.001$), 70-79 (LogFC= 0.59, $p < 0.001$), and above 80 age groups (LogFC= 0.83, $p < 0.001$). Patients in the 50-59 age groups had significantly higher *DMD* levels than patients in the 60-69 group (LogFC= 0.24, $p = 0.005$), 70-79 (LogFC= 0.37, $p < 0.001$), and above 80 age groups (LogFC= 0.61, $p < 0.001$). Patients aged between 60-69 had significantly higher *DMD* levels than patients in the above 80 group (LogFC= 0.37, $p = 0.002$).

Twelve of the tumor tissues used in the previous analysis have corresponding healthy tissues in the GTEx database: adrenal glands, bladder, breast, colon, esophagus, kidney, liver, lung, pancreas, skin, stomach, and thyroid. We examined whether there is an association between *DMD* gene expression and age in these healthy tissue samples (n= 2,863). The ages of donors ranged from 20 to 79 years old, and samples we grouped into six age groups: 20-29, 30-39, 40-49, 50-59, 60-69, and 70-79 years. There was no statistically significant effect for age ($p = 0.484$), gender ($p = 0.647$), or their interaction ($p = 0.591$) on *DMD* expression in these healthy samples.

7. Decreased *DMD* gene expression in primary tumors is associated with specific transcriptomic alterations

Expression of the *DMD* gene is significantly altered in tumors, with the majority having lower expression levels compared to their respective control tissues. This downregulation predominantly affects the transcript encoding the full-length dystrophin and it is due to a regulatory alteration. To identify downstream molecular pathways that could be associated with such *DMD* downregulation (analogous to the impact of full-length dystrophin loss in muscle), we compared transcriptomes of primary tumor samples with low vs. high *DMD* gene expression from 15 different tumor types: BRCA, BLCA, UCEC, CESC, OV, COAD, STAD, PRAD, LUSC, ESCA, HNSC, LUAD, KIRP and THCA, where Dp427m was found to be downregulated compared to control tissues. 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 control tissues for sarcomas was not available in the TCGA TARGET GTEx cohort, and therefore *DMD* gene expression in sarcomas vs. control tissues could not be compared. In each tumor type, the DEGs between samples at the bottom 33.3% and top 33.3% of *DMD* expression were identified and used to perform a pathway enrichment analysis. Figure 7 shows the combined score values (-Log (*p*-value) * odds ratio) for KEGG pathways that were found to be enriched in DEGs in more than 50% of the analyzed primary tumors (adjusted *p*-value < 0.05).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

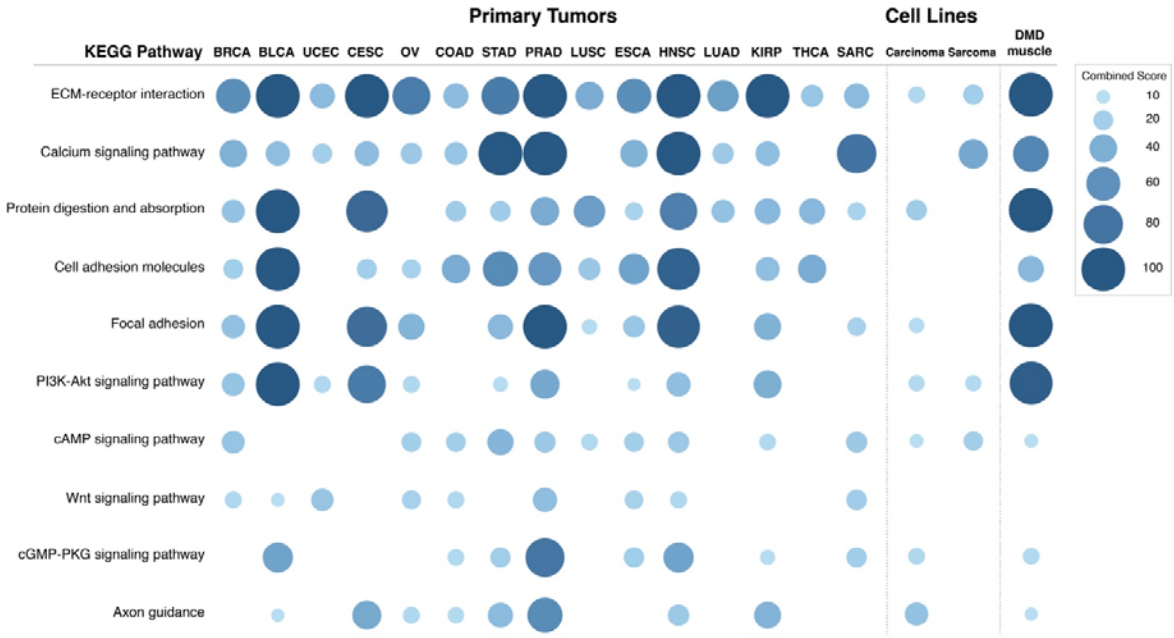


Figure 7. KEGG pathways enriched in the DEGs in primary tumor samples and tumor cell lines with low vs. high *DMD* gene expression as well as in DMD skeletal muscle. Only pathways that were identified in more than 50% of primary tumor samples with low *DMD* expression are displayed (adjusted *p*-value < 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 differentially expressed transcripts in tumors with low *DMD* expression showed enrichment in the ECM-receptor interaction pathway in all 15 primary tumors analyzed. Calcium signaling and protein digestion and absorption were enriched in 13 tumors, cell adhesion molecules in 12, focal adhesion in 11, PI3K and cAMP signaling in 10, Wnt signaling in 9, and cGMP-PKG signaling, and axon guidance were enriched in 8 comparisons.

The specific GO Biological Process terms enriched in DEGs in primary tumors with low vs. high *DMD* expression were also identified (Figure 8). The GO terms enriched in more than 50% of tumors were: extracellular matrix organization (14 out of 15 tumors), axonogenesis (12), regulation of cell migration, synapse organization and nervous system development (11), skeletal system development, regulation of ERK1 and ERK2 cascade (10) and calcium ion transmembrane import into cytosol (8). Lists of DEGs in each comparison and results of the pathway and GO term analysis can be found in Supplementary File 2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

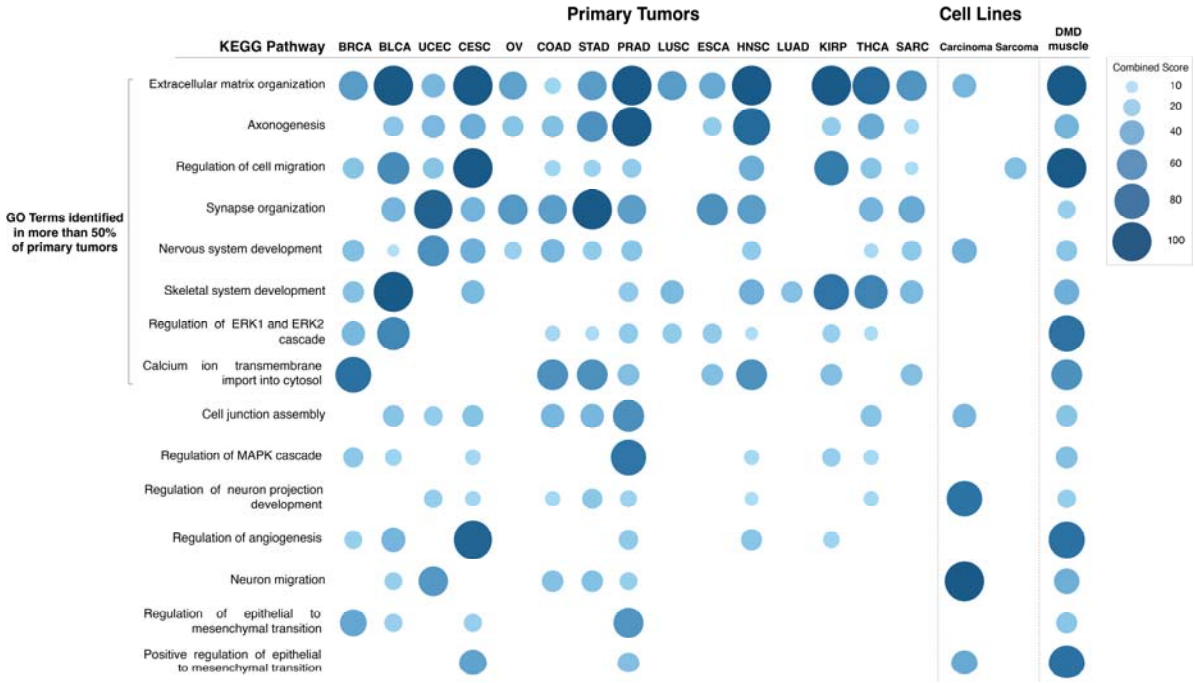


Figure 8. GO Biological Process terms enriched in the DEGs in primary tumor samples and tumor cell lines with low vs. high *DMD* gene expression as well as in DMD skeletal muscle (adjusted *p*-value < 0.05). Data is represented as the combined score for the enrichment (-Log (*p*-value) * odds ratio). Combined score values greater than 100 were given a value of 100.

In order to confirm that these transcriptomic changes between primary tumor 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 tumor type using three groups of tumor 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 tumors, 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).

8. Transcriptomic alterations associated with decreased *DMD* gene expression in tumor cell lines

To further confirm that these transcriptomic changes are evoked by *DMD* downregulation specifically in cancer cells rather than originate from *DMD* expression in the stromal or infiltrating immune cells present in tumor tissue samples, we conducted a two-class comparison analysis using the DepMap portal to identify DEGs between cell lines of the

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

same tumor 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 tumor cell lines survival (46,47), its altered expression may have an important common effect across malignancies.

One-hundred and forty tumor 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 2). 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 (4.07 – 7.6) and low (0 – 0.06) *DMD* expression. The carcinoma cell lines used in this analysis originated from tissues where primary tumors were found to have a lower level of Dp427m transcript compared to control tissues. The two-class comparison analysis between cell lines with low vs. high *DMD* expression identified 998 DEGs for carcinoma, and 543 DEGs for sarcoma cell lines. Interestingly, in carcinoma cell lines with low *DMD* expression, the majority of DEGs were downregulated (n= 976 out of 998). Figure 7 shows the combined score values (-Log (*p*-value) * odds ratio) for KEGG pathways that were found to be enriched in DEGs in these two comparisons (adjusted *p*-value < 0.05).

Pathway enrichment for the DEGs in carcinoma cell lines with low *DMD* expression suggested that *DMD* downregulation may affect the following KEGG pathways: ECM-receptor interaction (p= 0.037), protein digestion and absorption (p= 0.015), focal adhesion (p=0.029), PI3K-Akt signaling (p= 0.021), cAMP signaling (p= 0.044), cGMP-PKG signaling (p= 0.024) and axon guidance (p= 0.002). Notable, these pathways were also enriched in the DEGs in more than 50% of primary tumors with low *DMD* expression.
Similar to carcinoma cell lines, the following KEGG pathways were enriched in the DEGs in sarcoma cell lines with low *DMD* expression: ECM-receptor interaction (p= 0.048), PI3K-Akt signaling (p= 0.041) and cAMP signaling (p= 0.034). The calcium signaling pathway, which was enriched in the DEGs in more than 50% of primary tumors with low *DMD* expression was also enriched in the DEGs in the comparison of sarcoma cell lines (p= $7.24e^{-04}$).

Moreover, GO Biological Process term analysis for DEGs in carcinoma cell lines with low vs. high *DMD* expression (Figure 8 and Supplementary File 2) identified the following terms: extracellular matrix organization (p= $5.22e^{-04}$), cell junction assembly (p= 0.012) and positive regulation of epithelial to mesenchymal transition (p= 0.038). Terms related to the development of the CNS and synaptic transmission: nervous system development (p= $8.39e^{-}$

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

^{05}), neuron migration (p= 8.39e^{-05}) and regulation of neuron projection development (p= 1.04e^{-05}) were also found.

The GO Biological Process term regulation of cell migration was enriched (p= 0.02) in DEGs in sarcoma cell lines with low vs. high *DMD* expression (Figure 8 and Supplementary File 2). Lists of the DEGs in each comparison and results of the pathway and GO term analysis can be found in Supplementary File 2.

9. Transcriptomic alterations in Duchenne skeletal muscle compared to healthy muscle samples

Next, we investigated whether the transcriptomic changes resulting from *DMD* downregulation in primary tumors and tumor cell lines are similar to those observed in skeletal muscles of DMD patients. We compared gene expression data from 12 DMD skeletal muscle and 11 healthy muscle samples. The GEO2R tool identified 1,160 gene to be differentially expressed between DMD and healthy skeletal muscle. Figure 7 shows the combined score values (-Log (p-value) * odds ratio) for the KEGG pathways that were found to be enriched in the top 1,000 DEGs with the highest |LogFC| values in this comparison (adjusted p-value < 0.05).

Pathway enrichment analysis indicated *DMD* downregulation to be associated with changes in the following KEGG pathways: ECM receptor interaction (p= 7.82e^{-11}), calcium signaling pathway (p= 8.80e^{-08}), protein digestion and absorption (p= 3.67e^{-08}), cell adhesion molecules (p= 1.55e^{-04}), focal adhesion (p= 2.87e^{-12}), PI3K-Akt signaling (p= 7.82e^{-11}), cAMP signaling (p= 0.020), cGMP-PKG signaling (p= 0.007) and axon guidance pathways (p= 0.030). These very pathways were enriched in the DEGs in more than 50% of primary tumors with low *DMD* expression, and also enriched in DEGs in carcinoma and sarcoma cell lines with low vs. high *DMD* expression.

GO Biological Process term analysis for DEGs in in DMD skeletal muscle compared to healthy muscle (Figure 8) identified the following terms: extracellular matrix organization (p= 6.25e^{-17}), axonogenesis (p= 9.35e^{-5}), regulation of cell migration (p= 3.39e^{-12}), synapse organization (p= 0.009), nervous system development (p= 1.87e^{-4}), skeletal system development (p= 2.22e^{-4}), regulation of ERK1 and ERK2 cascade (p= 5.15e^{-8}), calcium ion transmembrane import into cytosol (p= 0.001). These GO terms were enriched in DEGs in more than 50% of primary tumors with low *DMD* expression. GO terms that were found to be in common with carcinoma cell lines with low *DMD* expression were: extracellular matrix

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

organization, nervous system development, cell junction assembly (p= 0.004), regulation of neuron projection development (p= 0.005), neuron migration (p= 0.004) and positive regulation of epithelial to mesenchymal transition (p= 2.52e^{-4}). Finally, the GO term, regulation of cell migration, was found to be in common with sarcoma cell lines with low *DMD* expression.

Furthermore, 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 tumors, tumor cell lines and DMD skeletal muscle: extracellular matrix organization (p= 1.10e^{-18}), axon guidance (p= 9.45e^{-10}), focal adhesion (p= 3.27e^{-14}) and PI3K-Akt signaling pathways (p= 9.27e^{-09}) (Supplementary Figure 5).

10. Low *DMD* gene expression is associated with poor survival in patients with fifteen different primary tumor types

Given these similarities between pathways altered in tumors and in the lethal muscle disease, we examined the association between *DMD* gene expression and patients' survival in the following tumor types: BRCA, BLCA, UCEC, CESC, OV, COAD, STAD, PRAD, LUSC, ESCA, HNSC, LUAD, KIRP, THCA and SARC. We compared overall survival (OS), progression-free interval (PFI), disease-specific survival (DSS) and disease-free interval (DFI) endpoints between patients at the bottom 25% of *DMD* expression and those at the top 25% across all the previously mentioned tumor types. OS was lower in the low *DMD* group (HR 1.33; 95% 1.17, 1.51; p< 0.0001) with 2,417 days compared to 3,253 for the high *DMD* group. PFI was also decreased in the low *DMD* group (HR 1.28; 95% 1.14, 1.45; p< 0.0001) with 2,472 days compared to 3,669 days, respectively. Finally, the low *DMD* group had lower DSS (HR 1.46; 95% 1.24, 1.72; p< 0.0001) and DFI (HR1.30; 95% 1.06, 1.59; p= 0.012) compared to the high *DMD* group (Figure 9).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

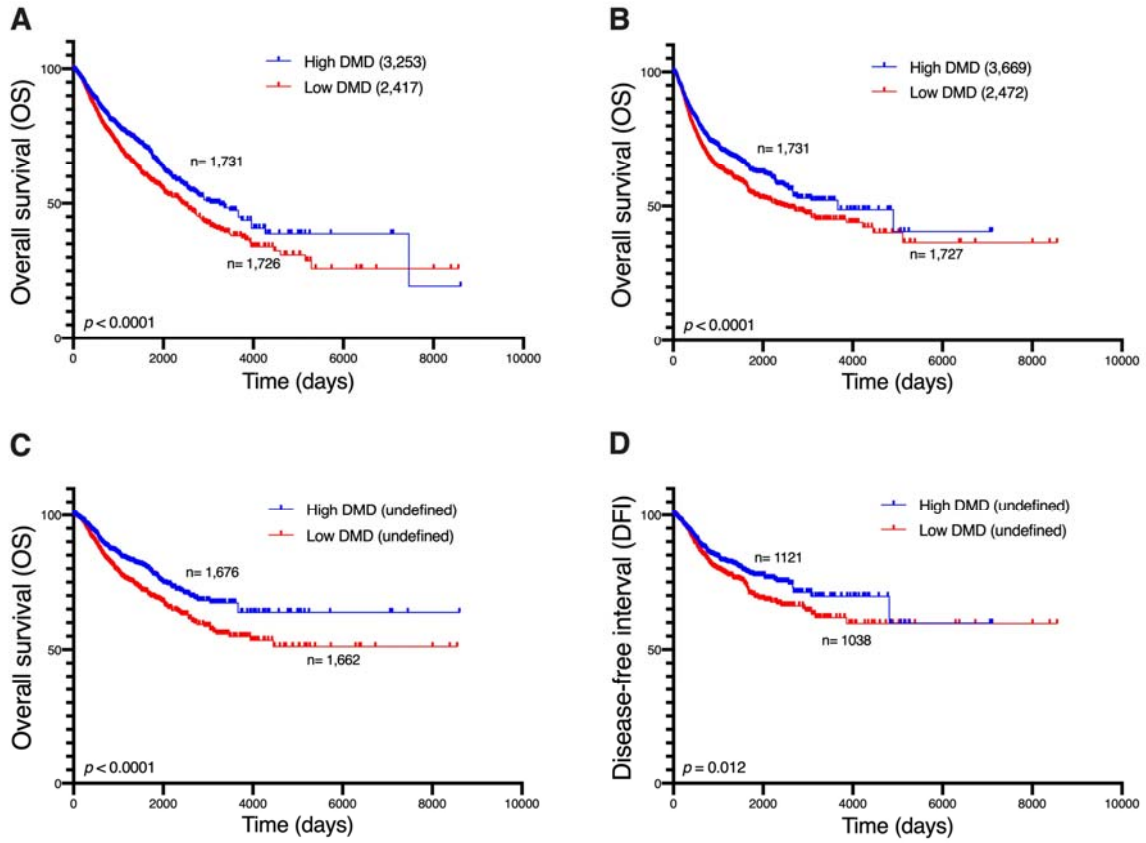


Figure 9. Kaplan-Meier curves of (A) overall survival (B) progression-free interval (C) disease-specific survival and (D) disease-free interval for patients with low or high levels of *DMD* gene expression. Numbers in brackets are overall survival times in days. Numbers of patients in each group and *p*-values for the log-rank test are displayed in the figure.

In order to confirm that these changes in survival outcomes between patients with low vs. high *DMD* expression in their tumors are associated with *DMD* downregulation and not a result of other factors, we repeated the survival analysis using two groups of patients chosen at random, and no statistically significant differences in survival endpoints were found between the two.

11. Transcriptomic alterations in hematological malignancies with low vs. high Dp71 expression

The hierarchical clustering analysis revealed that blood malignancies acute myeloid leukemia (LAML) and diffuse large B-cell lymphoma (DLBC) had a unique pattern of *DMD* transcripts to the point that each of these two malignancies was classified as a separate cluster (Figure 2). While no changes in Dp427m expression were observed in TCGA datasets for LAML and DLBC compared to healthy blood, Dp71 levels including its splice variants Dp71b and Dp71ab were higher in these tumors (Figure 3, Supplementary Table 5). Levels of

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dp71 and its splice variants were also found higher in two TARGET datasets (acute myeloid leukemia and acute lymphoblastic leukemia) compared to GTEx healthy whole blood (Supplementary Table 6). However, we note that age differences between TARGET and GTEx donors might be a confounding factor when interpreting these results.

Next, we compared gene expression between samples from the previously mentioned TCGA and TARGET datasets at the bottom 33.3% and top 33.3% of Dp71 expression across all tumor types. The pathways enriched in the top 1,000 DEGs with the largest |LogFC| were: protein digestion and absorption (p= 8.59e^{-05}), ECM-receptor interaction (p= 3.87e^{-04}), focal adhesion (p=0.003) and PI3K-Akt signaling pathway (p= 0.02). The identified GO Biological Process terms for these genes were extracellular matrix organization (p= 1.57e^{-10}), regulation of angiogenesis (p= 7.94e^{-04}), skeletal muscle development (p= 8.38e^{-04}), positive regulation of MAPK cascade (p= 0.005), regulation of ERK1 and ERK2 cascade (p= 0.006), regulation of cell migration (p= 0.01) and positive regulation of calcium ion import (p= 0.02) (Supplementary File 2). Thus, these pathways and processes were identical with those found in comparisons with low vs. high expression of Dp427 in completely unrelated malignancies. A comparison between random groups of samples from these hematological tumors did not reveal any statistically significant changes in gene expression.

12. Low expression of Dp71 is associated with poor survival in patients with hematological malignancies

We compared OS between hematological malignancies patients with low and high expression of Dp71 and its splice variants. OS was lower in the low Dp71 group (HR 2.39; 95% 1.44, 3.99; p= 0.0003) with 792 days compared to 1,992 for the high Dp71 group (Figure 10). For specificity testing, the survival analysis was conducted using two random groups of patients with these hematological malignancies and no statistically significant difference in overall survival was found.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

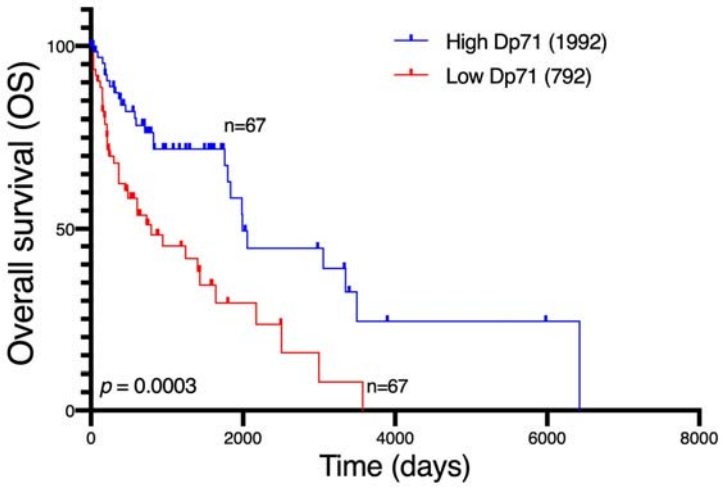


Figure 10. Kaplan-Meier curve of overall survival for hematological malignancies patients with low or high levels of Dp71 expression. Numbers in brackets are overall survival times in days. Numbers of patients in each group and *p*-values for the log-rank test are displayed in the figure.

Thus, while specific dystrophins are differentially regulated across various tumors, the low expression of both full-length dystrophin and the Dp71 variants in the analyzed tumors is associated with analogous molecular alterations and significantly decreased patients' survival.

Discussion

The *DMD* gene encodes a spectrum of dystrophin isoforms, but significant expression of the majority of these appeared to be restricted to specific tissues. Loss of the full-length dystrophin results in Duchenne muscular dystrophy, while mutations additionally disrupting other isoforms result in exacerbated phenotypes (36,37). In skeletal myofibers, cardiomyocytes and neurons, which express the highest levels of the full-length dystrophin, this protein has been described to serve as a structural scaffold for proteins engaged in ECM and cell-cell interactions, and in intracellular signaling. However, more recent data demonstrate that the loss of *DMD* expression impacts a broader spectrum of cells than those affected in DMD. In myoblasts (10,16,49,50), lymphocytes (33), endotheliocytes (32,51,52), 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 multiple cells (53), and the damaging purinergic phenotype affects myoblasts and lymphocytes (33,49). Some old findings, such as platelets

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

abnormalities (54), have recently been vindicated (55). 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" (56).

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 housekeeping genes in some tissues. Therefore, we investigated a spectrum of tumors 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 tumors 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 distribution profile for the full-length dystrophin isoform (Supplementary Figure 2). Interestingly, according to the Human Protein Atlas (38), the antibody HPA002725, directed against amino acids 186-333 of the full-length dystrophin, detected moderate cytoplasmic and/or membrane staining in a range of normal 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. Given the high sensitivity and specificity of the mass-spec (57,58), 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (59) to nuclear localization (22). Therefore, we compared expression of specific *DMD* transcripts that showed differential expression in tumors and respective control tissues.

Despite the advantage that using NATs as control samples in cancer studies reduces individual-and anatomical site-specific confounding factors and eliminates technical interlaboratory differences, it was found that these tissues are distinct from healthy and tumor tissues and represent a unique intermediate state between them (60). Although we showed that the results of comparing *DMD* expression between TCGA tumors and their corresponding NATs were consistent with those of comparing *DMD* expression between TCGA tumors and healthy GTEx tissues in 11 out of 13 comparisons, the unique transcriptomic profile of NATs might explain why thyroid NAT and kidney NATs from three different kidney TCGA tumors clustered with tumor tissues in the third cluster. The opposite was observed in the second cluster where PRAD and SKCM samples clustered with control tissues. This might be the result of a high proportion of non-tumor cells in these tumor samples.

We found a significant downregulation of *DMD* gene expression across diverse primary tumors. Both full-length and truncated dystrophin variants were differentially expressed, and hierarchical clustering of the top highly expressed transcripts distinguished tumors from corresponding control 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 tumor tissues (Supplementary Figure 4), suggesting a specific impact of the loss of the full-length transcript. Indeed, pancreatic adenocarcinoma was the only primary tumor with higher levels of Dp427m transcript compared to healthy pancreas tissue. In contrast, the relative expression of Dp71 splice variants was increased in tumor clusters.

In contrast to primary carcinoma samples, in hematological 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 analyzed (Supplementary Table 3). In contrast,

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

human and mouse hematopoietic stem cells were found to express Dp71, but its expression was decreasing with cell differentiation (manuscript submitted), which agrees with whole blood showing the low *DMD* expression. Interestingly, while Dp71 expression is found upregulated in hematological malignancies, its low levels were found associated with the very same dystrophic molecular alterations in cancer cells.

Given some evidence of a causative link between *DMD* downregulation and phenotypic changes in tumor 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 malignancy, transcriptomes of primary tumor samples as well as tumor cell lines with 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 tumors and 140 various tumor cell lines. Functional enrichment analysis showed that the pathways and GO Biological Process terms significantly enriched in DEGs in primary tumors and cell lines with low vs. high *DMD* expression were consistent with the pathways and GO terms enriched in DEGs in DMD skeletal muscle as well as the functional dystrophin PPI network (Supplementary Figure 5). Key pathways altered, including cell adhesion, ECM interactions and PI3K-Akt signaling, correspond to alterations found in Duchenne patients' cells (61,62). The calcium signaling pathway enriched in DEGs in tumors samples with low *DMD* expression from 13 out of the 15 analyzed primary tumors, and well as in sarcoma cell lines with low *DMD* expression agrees with the dysregulation of calcium signaling across a whole spectrum of dystrophic cells (reviewed in (53)), 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 (63).

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 (Supplementary Table 7). This is unsurprising, given that PAPs 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.

The importance of the *DMD* gene in tumorigenesis is supported by the finding that low *DMD* expression was associated with poor survival outcomes in patients with 15 different types of

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tumors (14 carcinomas and sarcoma). The overall survival of cancer patients with decreased *DMD* expression in tumors was 27 months lower than that of patients with high *DMD* expression. However, since the low and high *DMD* groups used in this analysis were composed of 15 tumor types, the number of samples for each tumor type is highly variable between the two groups and this could possibly be a confounding factor when interpreting the results. In other studies, mutations in the *DMD* gene were associated with poor overall survival of patients of two out of 11 analyzed tumors, namely uterine corpus endometrioid carcinoma and breast invasive carcinoma (24). Dystrophin protein was also identified as a survival biomarker in upper gastrointestinal cancer patients as poorer survival was observed in patients with low compared to high levels of dystrophin protein (64).

In our analyses, the overall survival of patients suffering from hematological malignancies with decreased Dp71 expression was about 39 months lower than that of patients with high Dp71 expression. However, in low-grade glioma (65) and B-cell chronic lymphocytic leukemia (28), high Dp71 expression was associated with poor patients' survival.

We found *DMD* expression to be associated with tumor stage. Samples from patients with the first stage had significantly higher levels of *DMD* expression compared to higher stages after controlling for age and gender differences.

DMD expression was also found to decrease with the age of onset, as samples from younger patients had higher *DMD* expression compared to samples from older patients. This association between *DMD* expression and age was unique to tumor tissues, as no such association was found by us in the corresponding healthy tissues from the GTEx database and also in a meta-analysis that identified genes with age-associated expression in human peripheral blood samples (66).

A further indirect indication of the functional significance of the *DMD* gene in tumors is that *DMD* downregulation across various malignancies involves regulatory changes, not just results from somatic gene mutations. Although, as expected, the presence of some types of somatic mutations and SCNAs was associated with lower levels of *DMD* expression, in about 88% of tumor samples and 77% of cell lines, *DMD* downregulation could not be linked to exonic mutations in the *DMD* gene. Moreover, *DMD* downregulation was not a result of SCNAs in about 61% of samples from female patients and about 56% of samples from male patients. While deletions were described as a causative factor for the downregulation of *DMD*

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

gene expression in some tumors (17,23,30), significantly reduced expression was also found in the absence of deletions or nonsense mutations (27), in agreement with our comprehensive analysis 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 (67) and recently, an epigenetic mechanism responsible for reduced *DMD* transcript levels has been described (68). These data indicate that, rather than simply being an effect of random mutations, quite likely to occur in this large gene, *DMD* alterations in tumor cells may have a complex regulatory nature involving mechanisms such as transcriptional regulation, chromatin remodeling or transcript degradation. Given that miRNAs might be responsible for the differential regulation of *DMD* expression in tumor samples, we investigated but did not find any differentially expressed miRNAs in tumor samples with low vs. high *DMD* expression that were common to all the 15 primary tumor types analyzed (Supplementary Table 8).

In conclusion, specific alterations in *DMD* gene expression are 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 tumors and corresponding tumor 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. Moreover, given the poor survival rate of patients with tumors downregulating dystrophin, *DMD* gene may be important in oncology.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

helpful comments on the manuscript and many anonymous reviewers whose critique helped improving this work.

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.H. (1977) Muscle nuclear changes in fetuses at risk for Duchenne muscular dystrophy. *J Med Genet*, **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.H. (1974) Muscle histology in fetuses at risk for Duchenne muscular dystrophy. *Clin Genet*, **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. *J Comp Pathol*, **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)

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Myogenesis modelled by human pluripotent stem cells: a multi-omic study of Duchenne myopathy early onset. *J Cachexia Sarcopenia 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. (2022) Loss of full-length dystrophin expression results in major cell- autonomous abnormalities in proliferating myoblasts. *Elife*, **11**, e75521.
11. Chang, N.C., Chevalier, F.P. and Rudnicki, M.A. (2016) Satellite Cells in Muscular Dystrophy - Lost in Polarity. *Trends Mol Med*, **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. *J Mol Cell Biol*, **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. *Physiol Rev*, **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. *Nat Med*, **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. *Nat Genet*, **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. *Hum Mol Genet*, **1**, 505–510.
20. Duan, D., Goemans, N., Takeda, S., Mercuri, E. and Aartsma-Rus, A. (2021) Duchenne muscular dystrophy. *Nat Rev Dis 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. *Hum Mol Genet*, **4**, 1475–1483.
22. Naidoo, M. and Anthony, K. (2020) Dystrophin Dp71 and the Neuropathophysiology of Duchenne Muscular Dystrophy. *Mol Neurobiol*, **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 (Basel)*, **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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 Invest*, **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. *Leuk 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. *Nat Commun*, **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 Neuropathol*, **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. *Int J Mol Sci*, **20**, 7–10.
32. 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. *Arterioscler Thromb Vasc Biol*, **33**, 2867–2876.
33. 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. *Am J Physiol Cell Physiol*, **267**, C886–C892.
34. 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. *Nat Biotechnol*, **38**, 669–673.
35. 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 Res*, **48**, D1153–D1163.
36. Jiang, L., Wang, M., Lin, S., Jian, R., Li, X., Chan, J., Dong, G., Fang, H., Robinson, A.E., Aguet, F., *et al.* (2020) A Quantitative Proteome Map of the Human Body. *Cell*, **183**, 269–283.
37. 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.
38. 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 (1979)*, **347**, 1260419.
39. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R. and Getz, G. (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*, **12**, 1–14.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

40. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, **2**, 401–404.
41. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., *et al.* (2013) Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal Complementary Data Sources and Analysis Options. *Sci Signal*, **6**, pl1–pl1.
42. DepMap, B. (2021) DepMap 21Q4 Public. 10.6084/m9.figshare.16924132.v1.
43. 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 Res*, **47**, D607–D613.
44. 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 Res*, **44**, W90–W97.
45. Schmidt, F., List, M., Cukuroglu, E., Köhler, S., Göke, J. and Schulz, M.H. (2018) An ontology-based method for assessing batch effect adjustment approaches in heterogeneous datasets. *Bioinformatics*, **34**, i908–i916.
46. 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. *Nat Commun*, **9**, 1–13.
47. 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.
48. 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.
49. 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.
50. 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. *J Neuromuscul Dis*, **8**, S205–S222.
51. 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 Res Cardiol*, **107**, 1–9.
52. 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.
53. Zabłocka, B., Górecki, D.C. and Zabłocki, K. (2021) Disrupted calcium homeostasis in duchenne muscular dystrophy: A common mechanism behind diverse consequences. *Int J Mol Sci*, **22**, 11040.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

54. Forst, J., Forst, R., Leithe, H. and Maurin, N. (1998) Platelet function deficiency in Duchenne muscular dystrophy. *Neuromuscular Disorders*, **8**, 46–49.
55. 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.
56. 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.
57. Hale, J.E. (2013) Advantageous Uses of Mass Spectrometry for the Quantification of Proteins. *Int J Proteomics*.
58. 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 Mol Biol*, **673**, 211–222.
59. 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.
60. Aran, D., Camarda, R., Odegaard, J., Paik, H., Oskotsky, B., Krings, G., Goga, A., Sirota, M. and Butte, A.J. (2017) Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nat Commun*, **8**, 1–14.
61. 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.
62. 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.
63. Ma, Y., Zhang, P., Wang, F., Yang, J., Yang, Z. and Qin, H. (2010) The relationship between early embryo development and tumourigenesis. *J Cell Mol Med*, **14**, 2697–2701.
64. Stephens, N.A., Skipworth, R.J.E., Gallagher, I.J., Greig, C.A., Guttridge, D.C., Ross, J.A. and Fearon, K.C.H. (2015) Evaluating potential biomarkers of cachexia and survival in skeletal muscle of upper gastrointestinal cancer patients. *J Cachexia Sarcopenia Muscle*, **6**, 53–61.
65. Naidoo, M., Jones, L., Conboy, B., Hamarneh, W., D'Souza, D., Anthony, K. and Machado, L.R. (2022) Duchenne muscular dystrophy gene expression is an independent prognostic marker for IDH mutant low-grade glioma. *Sci Rep*, **12**, 1–14.
66. Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., *et al.* (2015) The transcriptional landscape of age in human peripheral blood. *Nat Commun*, **6**, 1–14.
67. 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.
68. 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. *Proc Natl Acad Sci U S A*, **117**, 15664–16464.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.04.486990; this version posted November 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.