



HAL
open science

Cancer-related genes in the transcription signature of facioscapulohumeral dystrophy myoblasts and myotubes

Petr Dmitriev, Ulykbek Kairov, Thomas Robert, Ana Barat, Vladimir Lazar, Gilles Carnac, Dalila D. Laoudj-Chenivresse, Yegor Vassetzky

► To cite this version:

Petr Dmitriev, Ulykbek Kairov, Thomas Robert, Ana Barat, Vladimir Lazar, et al.. Cancer-related genes in the transcription signature of facioscapulohumeral dystrophy myoblasts and myotubes. *Journal of Cellular and Molecular Medicine*, 2014, 18 (2), pp.208-217. 10.1111/jcmm.12182. hal-02540149

HAL Id: hal-02540149

<https://hal.umontpellier.fr/hal-02540149>

Submitted on 10 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cancer-related genes in the transcription signature of facioscapulohumeral dystrophy myoblasts and myotubes

Petr Dmitriev ^{a, b}, Ulykbek Kairov ^{c, d}, Thomas Robert ^a, Ana Barat ^a, Vladimir Lazar ^a, Gilles Carnac ^b, Dalila Laoudj-Chenivresse ^b, Yegor S. Vassetzky ^{a, *}

^a UMR8126, Université Paris-Sud 11, CNRS, Institut de cancérologie Gustave Roussy, Villejuif, France

^b INSERM U1046, Université Montpellier I, Montpellier, France

^c Department of Genomic and Personalized Medicine, Center for Life Sciences, Nazarbayev University, Astana, Kazakhstan

^d National Scientific Shared Lab of Biotechnology, National Center for Biotechnology, Astana, Kazakhstan

Received: February 11, 2013; Accepted: October 8, 2013

Abstract

Muscular dystrophy is a condition potentially predisposing for cancer; however, currently, only Myotonic dystrophy patients are known to have a higher risk of cancer. Here, we have searched for a link between facioscapulohumeral dystrophy (FSHD) and cancer by comparing published transcriptome signatures of FSHD and various malignant tumours and have found a significant enrichment of cancer-related genes among the genes differentially expressed in FSHD. The analysis has shown that gene expression profiles of FSHD myoblasts and myotubes resemble that of Ewing's sarcoma more than that of other cancer types tested. This is the first study demonstrating a similarity between FSHD and cancer cell expression profiles, a finding that might indicate the existence of a common step in the pathogenesis of these two diseases.

Keywords: cancer • rhabdomyosarcoma • Ewing's sarcoma • FSHD • muscular dystrophy

Introduction

Recent studies have shown that murine models of Duchenne muscular dystrophy (DMD) and Limb-girdle muscular dystrophy (LGMD) frequently develop cancer. Mice with mutations in the Dystrophin (DMD), Calpain-3 (LGMD2A) or Dysferlin (LGMD2B) genes are susceptible to malignant tumours originating from the skeletal muscles [1–4] (reviewed in [5]).

Several cases of coincidence of cancer and DMD [6–13] or facioscapulohumeral dystrophy (FSHD) [14–16] have been also reported in humans; however, at the moment, DMD and FSHD patients are not considered to be more susceptible to cancer than the general population. Myotonic dystrophy (MD) patients are known to have higher risk of cancer [17, 18], while no cancer cases are known to be reported among LGMD patients.

Here, we focused on FSHD, an autosomal dominant hereditary neuromuscular disorder, because it is genetically associated with the

same genomic region that is re-arranged or epigenetically modified in various types of cancer. The majority of FSHD patients carry a deletion of the 3.3 kb-long D4Z4 macrosatellite repeats, accompanied by DNA demethylation and chromatin structure alterations within the subtelomeric region of chromosome 4 (4q35) [19, 20] (for review see [21]). The minor form of FSHD (FSHD2) is not associated with D4Z4 repeat contractions, but shares common epigenetic alterations in 4q35 with the major form (FSHD1) of this disease [22].

The rearrangement of the FSHD-associated region in 4q35 has been also found in various tumours, including undifferentiated soft tissue sarcoma [23–25], Ewing's sarcoma [26] and rhabdomyosarcoma [27]. Epigenetic alterations of the same region have been documented in cervical and ovarian cancers [28, 29].

D4Z4 repeats encode a powerful transcription regulator, double homeobox protein 4 (DUX4) [30], playing an important if not the key role in the aetiology of FSHD [31] and a potent enhancer that is capable of regulating a variety of genes [32, 33]. Interestingly, the expression level of DUX4 is altered in cervical cancer [28] and in Ewing's sarcoma where this gene functions as a chimeric oncogene if fused to CIC gene as a result of the t(4;19)(q35;q13.1) translocation [26].

The involvement of 4q35 in FSHD and several types of cancer prompted us to search for similarities in gene expression profiles of FSHD and cancer cells and tissues. We have found that a significant

*Correspondence to: Yegor S. VASSETZKY, UMR8126, Université Paris-Sud 11, CNRS, Institut de cancérologie Gustave Roussy, Villejuif 94805, France. Tel.: +33 142116283 Fax: +33 142115494 E-mail: vassetzky@igr.fr

number of genes differentially expressed in cancer are also differentially expressed in FSHD and, according to statistical criteria, this phenomenon could not be explained by a simple coincidence. We then searched for similarities of expression profiles of FSHD and various types of cancer and found the highest resemblance between FSHD and Ewing's sarcoma expression profiles.

Taken together, the results of this study establish for the first time a link between FSHD and cancer at the level of gene expression.

Materials and methods

Cell culture

Primary human myoblasts from normal individuals and FSHD patients (Table S6) Primary human myoblasts were isolated from skeletal muscles of healthy individuals as described in [34], for details see Table S4), purified with an immuno-magnetic sorting system (Miltenyi-Biotec, Bergisch Gladbach, Germany) by using an anti-CD56/NCAM antibody according to the manufacturer's specifications. CD56-positive myoblasts were seeded in collagen-coated Petri dishes (P1) and cultured in DMEM, 10% FCS, 1% Ultrosor G, at 37°C with 5% CO₂. All experiments were carried out between P1 and P5 to avoid cell senescence. Myoblast purity was determined by staining for Desmin.

Gene expression analysis

Total RNA was isolated from 2×10^6 myoblasts or myotubes by using Trizol (Invitrogen, Carlsbad, CA, USA), 400 ng of total RNA was reverse transcribed by using the High Capacity cDNA Archive kit (Applied Biosystems, AB, Foster city, CA, USA) according to the manufacturer protocol. cDNA was mixed with 2× TaqMan PCR mix (AB) and amplified with TLDA (Taqman Low Density Array; AB) by using Abiprism 7900HT. The expression was analysed by using the $\Delta\Delta C_t$ method [35].

Statistical analysis

qRT-PCR data have been analysed by using one-way ANOVA [36], $P < 0.05$ have been considered significant. To calculate the significance of the lists' intersection, we have used the LOLA tool (www.lola.gwu.edu). For the intersection of two lists, the P -value is determined as the probability of observing an intersection size that occurs by chance and is larger than the given one [37].

Lists of genes

The statistical significance of an intersection of gene lists has been calculated using online service List Of Lists Annotated (LOLA) <http://www.lola.gwu.edu/> [37]. To standardize the lists of genes extracted from different publications, we have used GenBank GeneIDs as reference points. As the GeneID was not usually provided along with the expression data, we have converted available Affymetrix or Ensembl IDs or Gene names to GeneID by using db2db online tool <http://biodbnet.abcc.ncifcrf.gov/>

db/db2db.php [38]. In the cases where Gene name was not recognized by the db2db service, we assumed that the authors of the study used a synonym. In these cases, we provide both Gene names: the major one, recognized by db2db service and the synonym from the original publication.

FSHD-related genes

List A was created by using supplementary data from [39], previously available at www.ucihs.uci.edu/biochem/winokur under 'publications' (Table S2). To our knowledge, at the moment of submission, the data are no longer available for downloading on this server. Genes with fold change >2 and $P < 0.05$ were considered as differentially expressed.

In total, 527 probes from the two types of microarrays used in the study (HuFL and U95) corresponding to differentially expressed genes were retained by the authors. To standardize this list and make it compatible with the lists originating from other studies, we have transformed probe codes to GeneIDs. Some of probe codes corresponded to multiple genes and some probe codes could not be found in available databases, therefore, the resulting list counted 529 unique gene IDs, 301 up- and 238 down-regulated. List B was created by using supplementary data from [40] previously available at www.ucihs.uci.edu/biochem/winokur under 'publications'. To our knowledge, at the moment of submission, the data are no longer available for downloading on this server. Using Affymetrix GCS V3.2 software and undisclosed statistical criteria, the authors attributed an arbitrary value of 1.1 to transcripts exhibiting an increase (I), 1.0 – marginal increase (MI), 0.0 – no change (NC), –1.0 – marginal decrease (MD) and –1.1 – decrease (D). We summed these values across all pairwise comparisons and considered transcripts with a sum of >5.5 as up-regulated and transcripts with a sum of <-5.5 as down-regulated in FSHD. In total, 236 probes codes corresponding to differentially expressed genes were selected according to these criteria. After transforming the list of probe codes to a list of GeneIDs, we obtained a list of 297 unique entries (26 up- and 271 down-regulated GeneIDs). Lists Ca, Cb and Cc were created by using supplementary tables A–G from [41] available for downloading at <http://gefu.cribi.unipd.it/papers/FSHD/>. From these tables, we have retrieved symbols of genes differentially expressed in FSHD biopsies divided into three sublists of genes (a, b and c), corresponding to the groups of patients, containing 194, 164 and 164 gene names respectively. These were converted to three lists of GeneIDs with 207 (112 up- and 95 down-regulated), 177 (118 up- and 59 down-regulated) and 177 (106 up- and 71 down-regulated) unique GeneIDs respectively. List D was created by using data from supplementary table from [42] available at <http://onlinelibrary.wiley.com/doi/10.1634/stemcells.2007-0465/supinfo>. The probe names corresponding to the genes differentially expressed in FSHD mesoangioblasts were converted to 32 unique GeneIDs (20 up- and 12 down-regulated). Lists Ea and Eb were created from Tables S3 and S4 from [43] available at www.neurology.org. The probes corresponding to differentially expressed genes were further subdivided into two sublists a and b, corresponding to the genes differentially expressed in FSHD and other types of muscular dystrophy (318 probes) and genes differentially expressed in FSHD only (134 probes). After the transformation of probe names to GeneIDs, these lists were converted to the Lists Ea and Eb containing 376 and 176 unique GeneIDs respectively. Lists Fa and Fb were created from Table S1 in [44] available for downloading at <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0020966>. Lists of Ensembl IDs of genes differentially expressed in myoblasts or myotubes from FSHD1 patients have been transformed to the two lists of unique GeneIDs containing 393 and 111

unique entries respectively. Lists Ga and Gb were extracted from Table S1 from [45] available online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1209508109/-/DCSupplemental/sd01.xls. The authors have analysed transcriptome profiles of biceps, muscles affected in FSHD and deltoids that are usually spared in FSHD patients and compared them to transcriptome profiles of the same types of muscles from normal subjects. The authors have found 238 GeneIDs differentially expressed (1.2-fold change, $P < 0.01$) in affected FSHD biceps, but not deltoids as compared with normal biceps and 182 GeneIDs differentially expressed according to the same statistical criteria in both biceps and deltoids of FSHD patients as compared with normal subjects. We have extracted these lists from a single Table S1 and named them List Ga and List Gb respectively.

Cancer-related genes

List 1 was constructed by using the data from [46] (Table S1). The authors of the study have conducted a meta-analysis of transcriptomes of 2186 samples representing 20 different cancers from 39 studies and obtained a transcriptome signature comprising 187 genes (list 1 of the present study) of which 117 up-regulated (Table 1 in [46]) and 70 were down-regulated (Table 2 in [46]). Authors claimed that this transcriptome signature could discriminate cancer samples with 92.6% accuracy independently from their tissue of origin. List 2 was constructed by using the data from Table 1 in [47]. The authors of the study used a published census of cancer-related genes [48] to select a representative list of 56 cancer-related genes. List 3 have been created by using data from [49]. The authors of the study have performed transcriptome profiling of 373 samples of 15 different types of benign and malignant tumours by using 17.5K custom-made cDNA arrays. These arrays allowed to cover the expression of 12,947 genes that included 332 cancer biomarkers (Table S1 in [49]) known from literature. The expression analysis of these 332 cancer biomarkers allowed the authors to select from it 56 genes ([49] does not precise which genes these were) that could discriminate between benign and malignant tumours with 88% accuracy. List 4 was created by using data from [50]. The authors have conducted a meta-analysis of transcriptomes of 3209 samples that contained normal tissues, immortalized cell lines, a variety of cancers

pluripotent and partially committed stem cells. As a result, the authors obtained a list of 189 genes (stem cell gene set, Tables S1–S4 in [50]) that could be used as a quantitative measure of stem cell-associated transcriptional activity and could also discriminate histological grades for a variety of human malignancies [50]. List 5 has been created by using the data from [51]. The authors performed transcriptome profiling of five cell lines representing altogether two different systems of cell transformation, one based on an inducible expression of v-Src and another one based on inducible expression of small and large T antigens of SV40. Common transcriptional profile of these cell lines consisted of 343 genes (238 up- and 105 down-regulated, Table S3 in [51]). List 6 is based on the data from [52]. The authors conducted a meta-analysis of 36 transcriptome signatures of tumours originating from 12 representative tissue types. This analysis resulted in identification of a common transcriptome signature of 183 genes from which 67 most significant genes (Fig. 2 in [52]) were differentially expressed in nearly all cancer types available to the authors. List 7 was created from our own large-scale literature search.

Expression profile scoring of FSHD samples has been conducted as described in [53] for expression profiles of tumour tissue samples. In total, 96 genes have been used for scoring, for each of them, the authors have assigned a rank R, different for each of the four categories, EWS, RMS, NB and BL (Table S3 in [53] and Table S5 in the present study). The contribution of a given gene to scoring is inversely proportional to its rank. To contribute to a given category, the level of gene expression must correspond to the sign in the scoring table *e.g.* '+1' corresponds to up- and '-1' to down-regulated genes. To determine the contribution of a gene to a given category, the following schema was used. If the level of expression of a gene corresponds to the sign in a given category in the scoring table, the contribution B of the gene to a given tumour category is calculated according to formula $B = 1 - C \times R$ where C is the coefficient and R is the rank. The coefficient was calculated independently for every category as follows: $C = 1/HR$ where HS is the highest rank for a given category resulting in $C_{EWS} = 0.000543$, C_{RMS} and $C_{not RMS} = 0.000499$, C_{NB} and $C_{not NB} = 0.000445$, C_{BL} and $C_{not BL} = 0.000980$. If the level of expression of a given gene did not correspond to the sign in the scoring table, the contribution of a gene to a given category was considered 0. To

Table 1 The lists of cancer-related genes used in the current study. For full lists of genes refer to Table S1

| List | Reference | Samples | Cancers types | Original gene list | Unique GeneIDs | Platform |
|------|------------|---------|--|--|----------------|----------------------|
| 1 | [46] | 2186 | 20 | 187 genes (117 genes up- and 70 down-regulated in cancer) | 190 | Meta-analysis |
| 2 | [47] | N/A | N/A | 56 genes (22 oncogenes and 34 tumour suppressors) causally related to cancer | 56 | Literature search |
| 3 | [49] | N/A | N/A | 332 genes | 277 | Literature search |
| 4 | [50] | 3209 | Unknown | 189 genes | 189 | Meta-analysis |
| 5 | [51] | 5 | 2 model systems of cell transformation | 343 genes (239 up- and 104 down-regulated in cancer) | 348 | Affymetrix U133 2.0A |
| 6 | [52] | 3700 | 12 | 183 genes (67 most significant) | 80 | Meta-analysis |
| 7 | This study | | | 400 genes | | Literature search |

Table 2 Lists of genes differentially expressed in FSHD. For full lists of genes refer to Table S2

| List | Reference | Samples | Patients analysed | Deregulated genes (probes) | Recognized GeneIDs | Statistical criteria | Microarray |
|------|-----------|----------------|--|----------------------------|--------------------------------|--|--|
| A | [39] | Biopsies | Three FSHD patients, three normal participants, pairwise comparisons | 230 | Total: 529; Up: 301; Down: 238 | Twofold change, $P < 0.05$ | Affymetrix HuFL GeneChip |
| B | [40] | Myoblasts | Nine FSHD patients, six normal participants, pairwise comparisons | 297 | Total: 296 Up:26; Down: 270 | Not described | Affymetrix U95A GeneChip |
| Ca | [41] | Biopsies | 11 FSHD patients | 194 | Total: 207; Up: 112; Down: 95 | SD<0.5 for single patient gene replicate | Custom cDNA array |
| Cb | | | Group A (>4 D4Z4 repeats) | 164 | Total: 177; Up: 118; Down: 59 | | |
| Cc | | | Group B (3-4 D4Z4 repeats) | 164 | Total: 177; Up: 106; Down: 71 | | |
| | | | Group C (1 D4Z4 repeat) | 164 | Total: 177; Up: 106; Down: 71 | | |
| D | [42] | Mesangioblasts | Two FSHD patients, two normal participants | 32 | Total: 32; Up: 20; Down: 12 | Not described | HG-Focus GeneChip Affymetrix |
| Ea | [43] | Biopsies | 19 FSHD, 11 MD patients and 30 normal participants | 318 | Total: 326; Up: 186; Down: 140 | 1.5-fold change, $P < 0.001$ | Affymetrix HG133A HG133B |
| Eb | | | FSHD, not MD versus Normal | 134 | Total: 156; Up: 51; Down: 105 | | |
| Fa | [44] | Myoblasts | Seven FSHD1 patients, two FSHD2 patients, eight normal participants | 367 | Total: 395; Up: 133; Down: 262 | Twofold change; $P < 0.01$ | Affymetrix human exon 1.0 ST arrays |
| Fb | | Myotubes | FSHD1 versus control | 129 | Total: 111; Up: 61; Down: 50 | | |
| Ga | [45] | Biopsies | 29 FSHD patients, 21 normal participants | 272 | Total: 272; Up: 141; Down: 131 | $P < 0.01$, 1.2-fold change | Affymetrix GeneChip Gene 1.0 ST arrays |
| Gb | | | FSHD biceps AND deltoid | 162 | Total: 162; Up: 58; Down: 104 | | |

determine the score of an expression profile, scores of every gene are summed. Using these formulas, we have calculated the scores of 'ideal' tumours (where all genes have the direction of expression corresponding to the signs in the scoring table). The resulting scores are: S_{EWS} and $S_{not\ EWS} = 62.644$; S_{RMS} and $S_{not\ RMS} = 64.041$; S_{NB} and $S_{not\ NB} = 70.408$; S_{BL} and $S_{not\ BL} = 70.243$.

Results

Lists of cancer- and FSHD-related genes

We have first compiled seven lists containing 56–343 published cancer-related genes (Table 1). These lists resulted either from a meta-analysis of transcriptomic data (Lists 1, 4 and 6), direct transcriptome analysis of cancer samples (List 5) or were assembled from cancer-related genes known from the literature (Lists 2 and 3). List 7 has been specifically created for this study by using bibliographical search of genes that have been previously used as biomarkers in various types of cancer. In total, Lists 1–7 assimilate transcriptome profiles of 9100 samples representing 35 different cancer types (Table S1).

Twelve lists of genes differentially expressed in FSHD have been extracted from six previously published transcriptome studies and one combined transcriptome–proteome [41] study (Table 2). In total, these lists represent transcriptional profiles of 86 FSHD patients and 73 healthy controls. Of these, eight lists of differentially expressed genes resulted from transcriptome analysis of skeletal muscle biopsies of FSHD patients (Lists A, Ca, Cb, Cc, Ea, Eb, Ga, and Gb), two

lists resulted from the analysis of FSHD primary myoblasts (List B, Fa), one list resulted from the transcriptome analysis of FSHD myotubes (List Fb) and one list resulted from the transcriptome analysis of mesoangioblasts isolated from FSHD patients (List D; Table S2).

Cancer-related genes are differentially expressed in FSHD

Based on the fact that the same genomic region is involved in the pathogenesis of both FSHD and cancer, we hypothesized that FSHD and cancer expression profiles might have a significant number of the common differentially expressed genes.

To test our hypothesis, we have systematically searched for the common genes shared by the lists of genes differentially expressed in FSHD and cancer-related, and found a statistically significant overlap between these lists (Table 3). Interestingly, the significance of these intersections depended on the type of samples used for transcriptome profiling and the list source. The lists that were produced from meta-analysis of cancer samples (*e.g.* Lists 1, 4 and 6) have demonstrated the most significant overlap with List Fa and List B, both of which are composed of genes differentially expressed in FSHD myoblasts (Table 3). Lists that were produced from simple literature search of cancer-related genes, *e.g.* the Lists 2, 3 and 7, gave the highest statistically significant intersection with the List A composed of genes differentially expressed in FSHD biopsies.

The number of cancer-related genes among genes differentially expressed in FSHD varied for different studies. For some (*e.g.* Lists A,

Table 3 *P*-values of intersection of gene lists of FSHD- and cancer-related genes. For the expanded version of this Table refer to Table S3

| | Total genes/list | List 1 190 | List 2 56 | List 3 277 | List 4 189 | List 5 348 | List 6 80 | List 7 576 |
|---------|------------------|---------------|--------------|---------------|---------------|---------------|--------------|---------------|
| List A | 539 | 6 (ns) | 7*** | 30*** | 11*** | 7*** | 4* | 49*** |
| List B | 297 | 2** | 3** | 15*** | 3*** | 2*** | 1*** | 20*** |
| List Ca | 207 | 1 (ns) | | 15*** | | 1 (ns) | | 10** |
| List Cb | 117 | | | 14*** | | 2* | | 9** |
| List Cc | 177 | | | 14*** | | 2* | | 9** |
| List D | 32 | 1 (ns) | | | 1 (ns) | | | 3* |
| List Ea | 376 | 7* | 1 (ns) | 15*** | 9* | 9*** | | 15*** |
| List Eb | 176 | 1 (ns) | 1 (ns) | 5** | 3 (ns) | 4** | | 8** |
| List Fa | 395 | 14*** | 3* | 12*** | 24*** | 5*** | 9*** | 23*** |
| List Fb | 111 | 2 (ns) | | 4* | 2 (ns) | 1** | 1 (ns) | 4 (ns) |
| List Ga | 272 | 2 (ns) | | 13*** | 1 (ns) | 1 (ns) | | 11** |
| List Gb | 220 | 2 (ns) | | 8*** | 1 (ns) | 1 (ns) | | 9** |

P* < 0.05; *P* < 0.01; ****P* < 0.001.

B and Fa), a highly significant intersection with all the cancer-related lists has been observed. For others, we could obtain a statistically significant intersection with only a few lists of cancer-related genes (Table 3). The Reasons for this discrepancy are unknown and probably linked to the study design. The description of cancer-related genes that have been found among genes differentially expressed in FSHD can be found in the Table S4.

Finally, we have observed a much higher intersection of the lists of cancer-related genes with the list of genes differentially expressed both in FSHD and MD (List Ea) as compared with the other lists from the same study (e.g. List Eb), that contained genes differentially expressed exclusively in FSHD. The latter result confirms the known link of MD with cancer and indicates that MD and FSHD share the same cancer-related genes (Table 3).

Next, we tested whether cancer-related genes were up- or down-regulated in FSHD as compared with the control. This analysis did not produce a clear result as we could find examples of the presence of cancer-related genes within lists of genes both up- and down-regulated in FSHD (Table S3).

Similarity between FSHD and Ewing's sarcoma gene expression profiles

In addition, we tried to find out whether FSHD transcription signature resembles that of a specific cancer. We have selected 96 genes that

were previously shown to be sufficient for classification of four types of cancers in humans: Ewing's sarcoma, rhabdomyosarcoma, neuroblastoma and Burkitt's lymphoma [53] (for review see [54]) (Table S5) and tested their expression in myoblasts and myotubes from FSHD patients. We have found that 12 of these genes are up-regulated in FSHD myoblasts and 26 are up-regulated in FSHD myotubes as compared with the control. Of these, five genes including ANXA1, GATM, METAP2, PIM2 and PTPN13 are up-regulated both in FSHD myoblasts and in differentiated FSHD myotubes, seven genes including ELF1, FHL1, GAS1, IFI16, IGFBP5, KDSR and KIF3C were specifically up-regulated in FSHD myoblasts, while 20 genes (ALDH7A1, ATN1, BIN1, CCND1, CD99, CKB, CNN, CTNN1, FCGRT, FGFR4, FHL1, GAP43, GATA2, IL4R, KIF3C, MYC, NFIB, PFN2, PTPRF, TP53I3) were specifically up-regulated in FSHD myotubes (Fig. 1).

Interestingly, five of 12 cancer-classifier genes differentially expressed in FSHD myoblasts were specific to Ewing's sarcoma (EWS), while only 2/12, 2/12 and 3/12 of cancer classifiers were specific to rhabdomyosarcoma, neuroblastoma and Burkitt's lymphoma respectively (Fig. 1). In FSHD myotubes, 11 of 26 differentially expressed genes were specific to Ewing's sarcoma, seven of 26 were specific to rhabdomyosarcoma, six of 26 were specific to neuroblastoma and only one of 26 were specific to Burkitt's lymphoma.

To quantitatively evaluate the similarity of the resulting gene expression profile of FSHD cells to cancer cell lines, we have used the

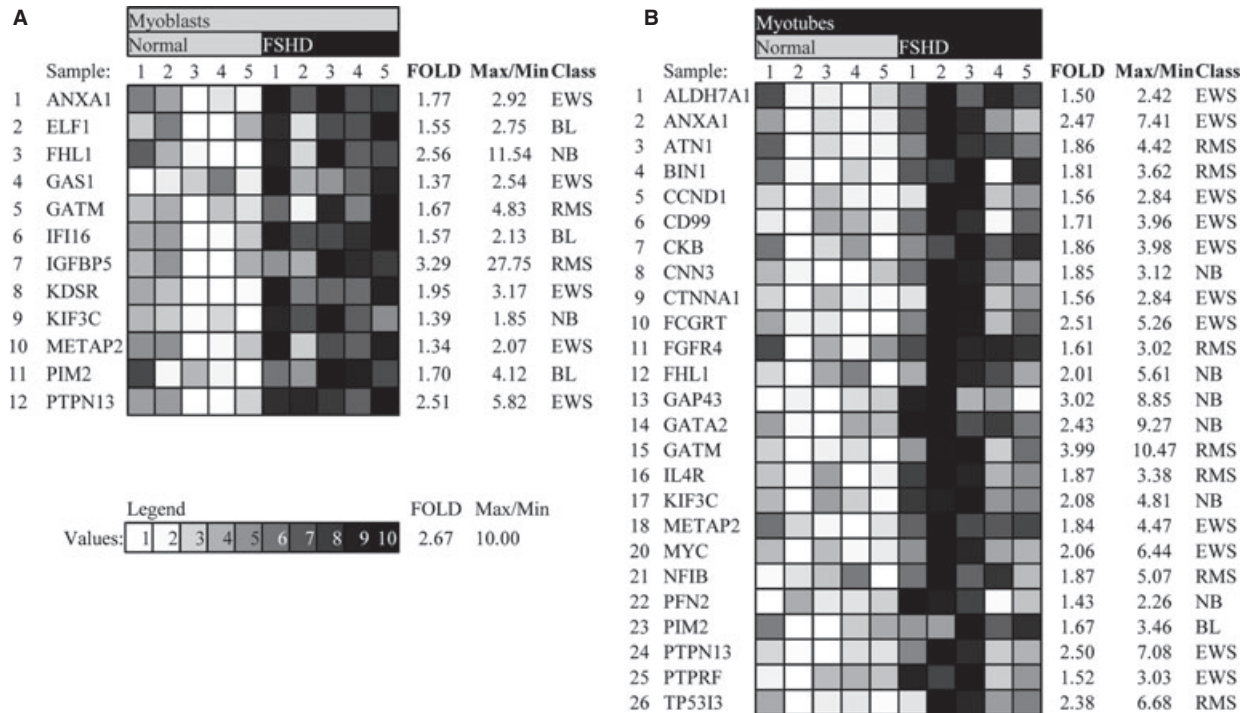


Fig. 1 Cancer-classifier genes differentially expressed in facioscapulohumeral dystrophy (FSHD) myoblasts (A) and myotubes (B). Genes that are differentially expressed both in FSHD myoblasts and myotubes as compared to healthy controls are shown in bold. (C) grey intensity corresponds to the expression level of the genes with darker squares corresponding to higher expression levels. The Sample description can be found in Table S6.

scoring procedure described in [53]. This procedure attributes a maximal score of 1 to each of the four evaluated cancers, and 0 to samples that are unrelated to cancer, scores between 0.4 and 1.0 being sufficient to establish a correct diagnosis [53] (Fig. 2B and C). The procedure attributed to FSHD myoblasts score 0.124 and to FSHD myotubes 0.233 for Ewing's sarcoma followed by rhabdomyosarcoma, neuroblastoma and Burkitt's lymphoma (Fig. 2A). This score was insufficient to result in a diagnosis of cancer (Fig. 2C); however, it clearly indicated that FSHD and cancer samples have similarities in their gene expression profile. Interestingly, transcription signatures of FSHD myoblasts and myotubes resembled those of Ewing's sarcoma more than other types of cancer (Fig. 2A).

Discussion

A link between cancer and muscular dystrophy has been previously demonstrated in the murine models of Duchenne and Limb-girdle muscular dystrophies. Specifically, it was found that the mice with the mutation in Dystrophin, Dysferlin and Calpain-3 were susceptible to spontaneous formation of rhabdomyo-, fibro- and liposarcomas derived from skeletal muscle tissue [1–4] (reviewed in [5]).

In human patients, only MD is currently associated with an increased risk of thyroid cancer and choroidal melanoma [17, 18]. Higher cancer risk in DMD patients was suspected, but not confirmed

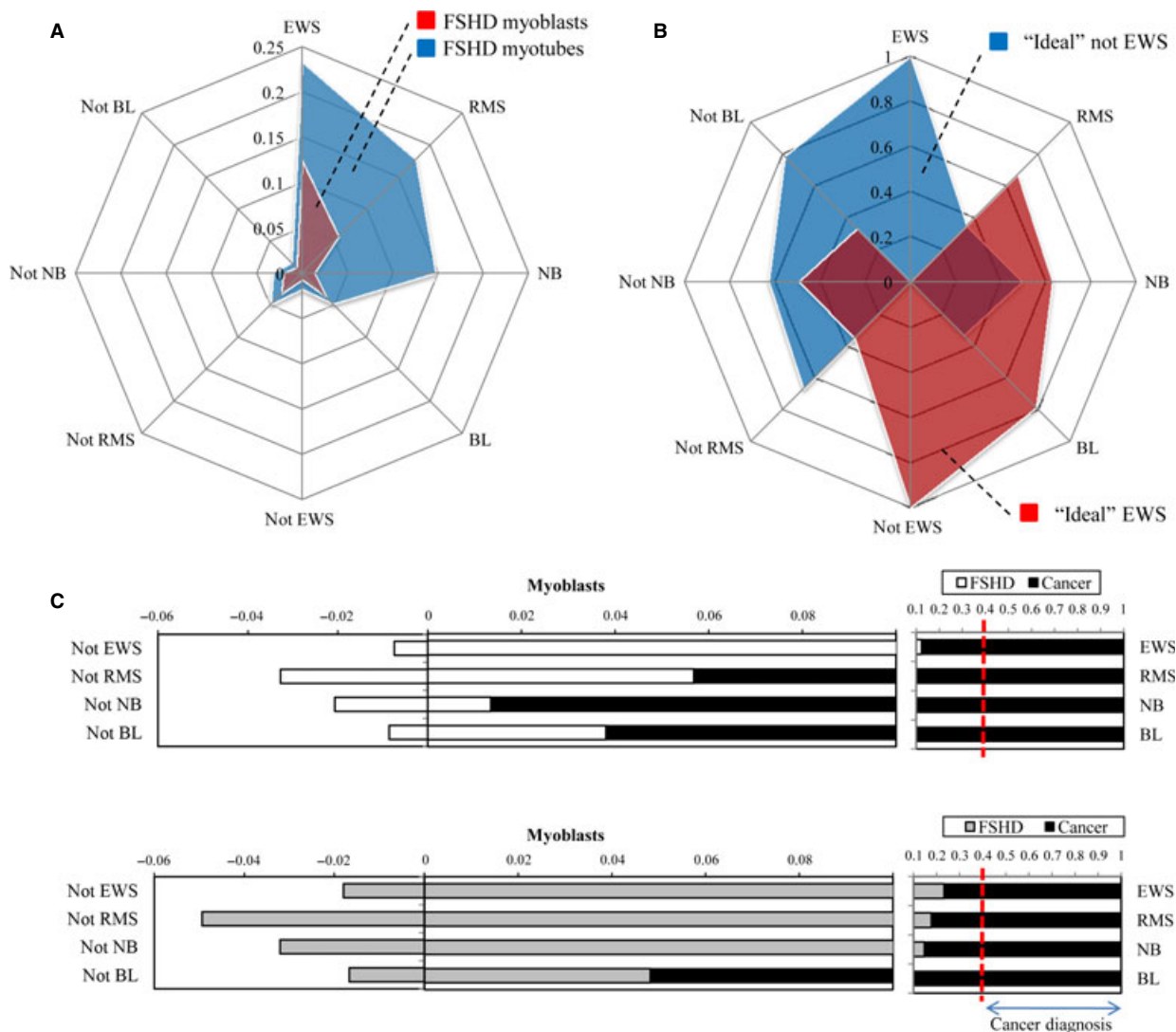


Fig. 2 A radar chart indicating the similarity of the facioscapulohumeral dystrophy (FSHD) samples to the one of four frequent types of cancer (A). The numbers correspond to relative scores of FSHD gene expression profiles normalized to maximal scores of the 'ideal' expression signatures of tumour samples. (B) The same scoring system was applied to the 'ideal EWS' and the 'ideal not EWS' expression profiles. (C) Linear representation of the same data as in A. The limit, below which the algorithm cannot recognize the sample as a 'true' cancer is indicated to the right.

[2, 13] To our knowledge, no such study has been conducted in the case of FSHD, the third most common muscular dystrophy in the world.

In the present work, we used meta-analysis to establish a link between FSHD and cancer at the level of gene expression. We have demonstrated that genes differentially expressed in FSHD contain more cancer-related genes than could be expected by chance. To exclude the possibility that these genes merely represent the tissue-specific markers unrelated to oncogenic process, we have selected studies that have analysed transcriptome signatures of a large number of cancer samples representing multiple cancer types of different origin.

We have also measured the expression level of 96 tumour markers and used a previously published scoring algorithm [53] to find out whether FSHD gene expression signature profile resembles any specific type of cancer. We expected, that the expression profiles of myogenic cells from FSHD patients would be more similar to the expression profile of rhabdomyosarcoma, a tumour originating from skeletal muscle tissue. Instead, we have found FSHD cells are more similar to Ewing's sarcoma. The resulting score, however, was insufficient to classify FSHD samples as true cancer samples, which is in agreement with the fact that none of the patients who participated in the study was diagnosed with cancer.

The reasons for the similarity of FSHD and tumour cell expression profiles remain unknown. One possibility is that this similarity originates from inflammation, fibrosis or oxidative stress that frequently accompany tumour development and are also observed in FSHD patients.

Alternatively, a certain similarity of FSHD and cancer gene expression profiles might be caused by the nature of the genetic defect in FSHD patients that affects the genomic region that is also altered in several types of cancer. The involvement of the same genomic region in both dystrophy and cancer is not unique to FSHD; this is also the case for the rearrangements in Dystrophin gene [55]. Intriguingly, DUX4, a powerful transcription factor encoded in 4q35, is involved in the pathological mechanism of both FSHD [56] and Ewing's sarcoma [26].

Another plausible explanation of the similarity of FSHD and cancer gene expression profiles is based on pre-mRNA splicing. Pre-mRNA splicing is perturbed in FSHD [57, 58] [59], probably, because of elevated expression of FRG1, a novel splicing regulator [57, 60]. As alteration in pre-mRNA splicing is a common phenomenon in cancer [61, 62], one could hypothesize that an abnormal pre-mRNA splicing might be a mechanism linking FSHD to cancer. In support of this hypothesis, we have observed that many cancer-related genes differentially expressed in FSHD are also differentially expressed in MD, a disease that has been shown to increase the risk of cancer in human patients [17, 18]. Myotonic dystrophy is caused by CTG and CCTG microsatellite repeat expansion affecting the function of splicing factors in the cells of the MD patients [63]. Therefore, an altered pre-mRNA splicing could potentially initiate the process of tumour formation in MD; however, currently, this hypothesis is missing a solid experiment support [64, 65].

The link at the level of gene expression that we have established between FSHD and cancer does not imply that FSHD patients might

have a higher incidence of cancer, as compared to general population. Although several cases of cancer in FSHD patients have been reported, FSHD is not considered as a cancer-predisposing condition. The latter is supported by the mouse models of FSHD [57, 66, 67] that have not been reported to have higher than average incidence of tumours. However, the conclusions about cancer incidence in muscular dystrophy patients made using mouse models should be handled with caution. Indeed, high incidence of cancer observed in *mdx* mouse model is not supported by observations in human patients, who do not demonstrate higher incidence of cancer [2]. Conversely, while the patients with MD suffer from an increased incidence of cancer, mouse models of this disease are not susceptible for cancer [68]. To our knowledge, no cases of concomitant LGMD and cancer have been described, while LGMD mouse models are susceptible to cancer.

This discrepancy may be explained by certain limitations of mouse models to represent human diseases. For example, *mdx* mice do not suffer from muscle wasting and satellite cell pool depletion as human patients do. Therefore, high incidence of tumours in *mdx* mice might be explained by the regenerative environment of a permanently damaged muscle, favouring oncogenic transformation of muscle satellite cells in *mdx* mice, but not in DMD patients, where satellite cells are rare [2].

To conclude, we consider that to determine whether FSHD patients are more prone to cancer than general population, one should not rely on mouse models of this disease, but rather carry out a retrospective examination of medical histories of FSHD patients.

Our study may also have an impact on the development of FSHD therapy strategies. Recently, several anti-cancer drugs have been proven efficient in the mouse model of DMD [69, 70]. The similarity of gene expression profile linking FSHD to cancer, discovered in our study, may provide the basis for examination of the usability of anti-cancer agents in FSHD.

Acknowledgements

The research has been supported by grants from the Association Française contre les Myopathies (AFM) to YSV and DL. AB was a recipient of the IRCSET-Marie Curie International Mobility Fellowships in Science Engineering and Technology, Ireland. P.D. was supported by the Association "Amis FSH" and the University Montpellier I. We thank Cécile Cassan for critical reading of the manuscript.

Conflicts of interest

The authors confirm that there are no conflicts of interest.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1. Lists of cancer-associated genes.

Table S2. Lists of genes differentially expressed in FSHD.

Table S3. Intersection of lists of cancer-related genes and lists of genes differentially expressed in FSHD.

Table S4. Description of cancer-related genes found among genes differentially expressed in FSHD.

Table S5. Scoring table for tumour classification.

References

1. Chamberlain JS, Metzger J, Reyes M, *et al.* Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma. *FASEB J.* 2007; 21: 2195–204.
2. Fernandez K, Serinagaoglu Y, Hammond S, *et al.* Mice lacking dystrophin or alpha sargoglycan spontaneously develop embryonal rhabdomyosarcoma with cancer-associated p53 mutations and alternatively spliced or mutant Mdm2 transcripts. *Am J Pathol.* 2010; 176: 416–34.
3. Hosur V, Kavirayani A, Riefler J, *et al.* Dystrophin and dysferlin double mutant mice: a novel model for rhabdomyosarcoma. *Cancer Genet.* 2012; 205: 232–41.
4. Schmidt WM, Uddin MH, Dysek S, *et al.* DNA damage, somatic aneuploidy, and malignant sarcoma susceptibility in muscular dystrophies. *PLoS Genet.* 2011; doi:10.1371/journal.pgen.1002042.
5. Fanzani A, Monti E, Donato R, *et al.* Muscular dystrophies share pathogenetic mechanisms with muscle sarcomas. *Trends Mol Med.* 2013; 19: 546–54.
6. Roszbach HC, Lacson A, Grana NH, *et al.* Duchenne muscular dystrophy and concomitant metastatic alveolar rhabdomyosarcoma. *J Pediatr Hematol Oncol.* 1999; 21: 528–30.
7. Jakab Z, Szegedi I, Balogh E, *et al.* Duchenne muscular dystrophy-rhabdomyosarcoma, ichthyosis vulgaris/acute monoblastic leukemia: association of rare genetic disorders and childhood malignant diseases. *Med Pediatr Oncol.* 2002; 39: 66–8.
8. Doddihal H, Jalali R. Medulloblastoma in a child with Duchenne muscular dystrophy. *Childs Nerv Syst.* 2007; 23: 595–7.
9. Johnston KM, Zoger S, Golabi M, *et al.* Neuroblastoma in Duchenne muscular dystrophy. *Pediatrics.* 1986; 78: 1170–1.
10. Korones DN, Brown MR, Palis J. “Liver function tests” are not always tests of liver function. *Am J Hematol.* 2001; 66: 46–8.
11. Saldanha RM, Gasparini JR, Silva LS, *et al.* Anesthesia for Duchenne muscular dystrophy patients: case reports. *Rev Bras Anesthesiol.* 2005; 55: 445–9.
12. Svarch E, Menendez A, Gonzalez A. Duchenne muscular dystrophy and acute lymphoblastic leukaemia. *Haematologia (Budap).* 1988; 21: 123–4.
13. van den Akker M, Northcott P, Taylor MD, *et al.* Anaplastic medulloblastoma in a child with Duchenne muscular dystrophy. *J Neurosurg Pediatr.* 2012; 10: 21–4.
14. Blake D, Gilliam C, Warburton D, *et al.* Possible clue for chromosomal assignment of the gene for facioscapulohumeral muscular dystrophy: a family with polyposis. *Ann Neurol.* 1988; 24: 178.
15. Blake D, Brown R, Gilliam T, *et al.* The second family with facioscapulo-humeral muscular dystrophy and familial polyposis coli. *Neurology.* 1989; 39: 404.
16. Kazakov V, Rudenko D, Schulev J, *et al.* Unusual association of FSHD and extramedullary thoracic tumour in the same patient: a case report. *Acta Myol.* 2009; 28: 76–9.
17. Gadalla SM, Lund M, Pfeiffer RM, *et al.* Cancer risk among patients with myotonic muscular dystrophy. *JAMA.* 2011; 306: 2480–6.
18. Win AK, Perattur PG, Pulido JS, *et al.* Increased cancer risks in myotonic dystrophy. *Mayo Clin Proc.* 2012; 87: 130–5.
19. de Greef JC, Frants RR, van der Maarel SM. Epigenetic mechanisms of facioscapulohumeral muscular dystrophy. *Mutat Res.* 2008; 647: 94–102.
20. Cbianca DS, Gabellini D. FSHD: copy number variations on the theme of muscular dystrophy. *JCB.* 2010; 191: 1049–60.
21. Tawil R. Facioscapulohumeral muscular dystrophy. *Neurotherapeutics.* 2008; 5: 601–6.
22. de Greef JC, Lemmers RJ, van Engelen BG, *et al.* Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum Mutat.* 2009; doi: 10.1002/humu.21091.
23. Yoshimoto M, Graham C, Chilton-MacNeill S, *et al.* Detailed cytogenetic and array analysis of pediatric primitive sarcomas reveals a recurrent CIC-DUX4 fusion gene event. *Cancer Genet Cytogenet.* 2010; 195: 1–11.
24. Italiano A, Sung YS, Zhang L, *et al.* High prevalence of CIC fusion with double-homeobox (DUX4) transcription factors in EWSR1-negative undifferentiated small blue round cell sarcomas. *Genes Chromosom Cancer.* 2012; 51: 207–18.
25. Graham C, Chilton-MacNeill S, Zielenska M, *et al.* The CIC-DUX4 fusion transcript is present in a subgroup of pediatric primitive round cell sarcomas. *Hum Pathol.* 2012; 43: 180–9.
26. Kawamura-Saito M, Yamazaki Y, Kaneko K, *et al.* Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. *Hum Mol Genet.* 2006; 15: 2125–37.
27. Sirvent N, Trassard M, Ebran N, *et al.* Fusion of EWSR1 with the DUX4 facioscapulohumeral muscular dystrophy region resulting from t(4;22)(q35;q12) in a case of embryonal rhabdomyosarcoma. *Cancer Genet Cytogenet.* 2009; 195: 12–8.
28. Katargin AN, Pavlova LS, Kissel'ov FL, *et al.* Hypermethylation of genomic 3.3-kb repeats is frequent event in HPV-positive cervical cancer. *BMC Med Genomics.* 2009; doi: 10.1186/1755-8794-2-30.
29. Tsumagari K, Qi L, Jackson K, *et al.* Epigenetics of a tandem DNA repeat: chromatin DNaseI sensitivity and opposite methylation changes in cancers. *Nucleic Acids Res.* 2008; doi: 10.1093/nar/gkn055.
30. Richards M, Coppee F, Thomas N, *et al.* Facioscapulohumeral muscular dystrophy (FSHD): an enigma unravelled? *Hum Genet.* 2011; doi: 10.1007/s00439-011-1100-z.
31. van der Maarel SM, Tawil R, Tapscott SJ. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. *Trends Mol Med.* 2011; 17: 252–8.
32. Petrov A, Allinne J, Pirozhkova I, *et al.* A nuclear matrix attachment site in the 4q35 locus has an enhancer-blocking activity *in vivo*: implications for the facio-scapulo-humeral dystrophy. *Genome Res.* 2008; 18: 39–45.
33. Dmitriev P, Petrov A, Anseau E, *et al.* The Kruppel-like factor 15 as a molecular link between myogenic factors and a chromosome 4q transcriptional enhancer implicated in facioscapulohumeral dystrophy. *J Biol Chem.* 2011; 286: 44620–31.
34. Barro M, Carnac G, Flavier S, *et al.* Myoblasts from affected and non-affected FSHD muscles exhibit morphological

- differentiation defects. *J Cell Mol Med.* 2010; 14: 275–89.
35. **Livak KJ, Schmittgen TD.** Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) Method. *Methods.* 2001; 25: 402–8.
 36. **Georgin P, Gouet M.** *Statistiques avec Excel 2000.* Paris: Eyrolles, 2000.
 37. **Cahan P, Ahmad AM, Burke H, et al.** List of lists-annotated (LOLA): a database for annotation and comparison of published microarray gene lists. *Gene.* 2005; 360: 78–82.
 38. **Mudunuri U, Che A, Yi M, et al.** bioDBnet: the biological database network. *Bioinformatics.* 2009; 25: 555–6.
 39. **Winokur ST, Chen YW, Masny PS, et al.** Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol Genet.* 2003; 12: 2895–907.
 40. **Winokur ST, Barrett K, Martin JH, et al.** Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromuscul Disord.* 2003; 13: 322–33.
 41. **Celegato B, Capitanio D, Pescatori M, et al.** Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. *Proteomics.* 2006; 6: 5303–21.
 42. **Morosetti R, Mirabella M, Gliubizzi C, et al.** Isolation and characterization of mesoangioblasts from facioscapulohumeral muscular dystrophy muscle biopsies. *Stem Cells.* 2007; 25: 3173–82.
 43. **Osborne RJ, Welle S, Venance SL, et al.** Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology.* 2007; 68: 569–77.
 44. **Cheli S, Francois S, Bodega B, et al.** Expression profiling of FSHD-1 and FSHD-2 cells during myogenic differentiation evidences common and distinctive gene dysregulation patterns. *PLoS ONE.* 2011; doi: 10.1371/journal.pone.0020966.
 45. **Rahimov F, King OD, Leung DG, et al.** Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. *Proc Natl Acad Sci USA.* 2012; 109: 16234–9.
 46. **Lu Y, Yi Y, Liu P, et al.** Common human cancer genes discovered by integrated gene-expression analysis. *PLoS ONE.* 2007; 2: e1149.
 47. **Fankhauser N, Cima I, Wild P, et al.** Identification of a gene expression signature common to distinct cancer pathways. *Cancer Inform.* 2012; 11: 139–46.
 48. **Futreal PA, Coin L, Marshall M, et al.** A census of human cancer genes. *Nat Rev Cancer.* 2004; 4: 177–83.
 49. **Basil CF, Zhao Y, Zavaglia K, et al.** Common cancer biomarkers. *Cancer Res.* 2006; 66: 2953–61.
 50. **Palmer NP, Schmid PR, Berger B, et al.** A gene expression profile of stem cell pluripotentiality and differentiation is conserved across diverse solid and hematopoietic cancers. *Genome Biol.* 2012; 13: R71.
 51. **Hirsch HA, Iliopoulos D, Joshi A, et al.** A transcriptional signature and common gene networks link cancer with lipid metabolism and diverse human diseases. *Cancer Cell.* 2010; 17: 348–61.
 52. **Rhodes DR, Yu J, Shanker K, et al.** Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci USA.* 2004; 101: 9309–14.
 53. **Khan J, Wei JS, Ringner M, et al.** Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med.* 2001; 7: 673–9.
 54. **Greer BT, Khan J.** Diagnostic classification of cancer using DNA microarrays and artificial intelligence. *Ann N Y Acad Sci.* 2004; 1020: 49–66.
 55. **Mitsui J, Takahashi Y, Goto J, et al.** Mechanisms of genomic instabilities underlying two common fragile-site-associated loci, PARK2 and DMD, in germ cell and cancer cell lines. *Am J Hum Genet.* 2010; 87: 75–89.
 56. **Lemmers RJ, van der Vliet PJ, Klooster R, et al.** A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science.* 2010; 329: 1650–3.
 57. **Gabellini D, D'Antona G, Moggio M, et al.** Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature.* 2006; 439: 973–7.
 58. **Davidovic L, Sacconi S, Bechara EG, et al.** Alteration of expression of muscle specific isoforms of the fragile X related protein 1 (FXR1P) in facioscapulohumeral muscular dystrophy patients. *J Med Genet.* 2008; 45: 679–85.
 59. **Pistono M, Shiue L, Cline MS, et al.** Rbfox1 downregulation and altered calpain 3 splicing by FRG1 in a mouse model of Facioscapulohumeral muscular dystrophy (FSHD). *PLoS Genet.* 2013; doi: 10.1371/journal.pgen.1003186.
 60. **van Koningsbruggen S, Straasheijm KR, Sterrenburg E, et al.** FRG1P-mediated aggregation of proteins involved in pre-mRNA processing. *Chromosoma.* 2007; 116: 53–64.
 61. **Venables JP.** Aberrant and alternative splicing in cancer. *Cancer Res.* 2004; 64: 7647–54.
 62. **Faustino NA, Cooper TA.** Pre-mRNA splicing and human disease. *Genes Dev.* 2003; 17: 419–37.
 63. **Hino S, Kondo S, Sekiya H, et al.** Molecular mechanisms responsible for aberrant splicing of SERCA1 in myotonic dystrophy type 1. *Hum Mol Genet.* 2007; 16: 2834–43.
 64. **Mueller CM, Hilbert JE, Martens W, et al.** Hypothesis: neoplasms in myotonic dystrophy. *Cancer Causes Control.* 2009; 20: 2009–20.
 65. **Zemtsov A.** Association between basal, squamous cell carcinomas, dysplastic nevi and myotonic muscular dystrophy indicates an important role of RNA-binding proteins in development of human skin cancer. *Arch Dermatol Res.* 2010; 302: 169–70.
 66. **Krom YD, Thijssen PE, Young JM, et al.** Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. *PLoS Genet.* 2013; doi: 10.1371/journal.pgen.1003415.
 67. **Pandey SN, Cabotage J, Shi R, et al.** Conditional over-expression of PITX1 causes skeletal muscle dystrophy in mice. *Biol Open.* 2013; 1: 629–39.
 68. **Gomes-Pereira M, Cooper TA, Gourdon G.** Myotonic dystrophy mouse models: towards rational therapy development. *Trends Mol Med.* 2011; 17: 506–17.
 69. **Ito T, Ogawa R, Uezumi A, et al.** Imatinib attenuates severe mouse dystrophy and inhibits proliferation and fibrosis-marker expression in muscle mesenchymal progenitors. *Neuromuscul Disord.* 2013; 23: 349–56.
 70. **Dorchies OM, Reutenauer-Patte J, Dahmane E, et al.** The anticancer drug tamoxifen counteracts the pathology in a mouse model of duchenne muscular dystrophy. *Am J Pathol.* 2013; 182: 485–504.