Zebrafish dystrophin and utrophin genes: Dissecting transcriptional expression during embryonic development

DANIEL LAI[1], CHUAN-CHING LAN[1], IVONE UN SAN LEONG[1] and DONALD R. LOVE[1,2]

[1]School of Biological Sciences, University of Auckland, Auckland 1142;
[2]Diagnostic Genetics, LabPlus, Auckland City Hospital, Auckland 1148, New Zealand

Received October 21, 2011; Accepted November 23, 2011

DOI: 10.3892/ijmm.2011.865

Abstract. Some genes can encode multiple overlapping transcripts, and this can result in challenges in identifying transcript-specific developmental expression profiles where tools such as RNA *in situ* hybrisations are inapplicable. Given this difficulty, we have undertaken a preliminary analysis of the developmental expression profile of selected transcripts of the dystrophin and utrophin genes of the zebrafish (*Danio rerio*) by targeting unique and common regions of each of these transcripts. The dystrophin and utrophin genes of zebrafish were identified by bioinformatic analysis and the dystrophin gene predictions were confirmed by transcript sequencing. These data enabled primer pairs to be designed in order to determine the expression profiles of unique, but overlapping transcripts, throughout embryonic development using quantitative real time reverse transcription PCR (qRT-PCR). The data indicated the early expression of the short carboxyl-terminal dystrophin transcript, with expression of the full length muscle transcript occurring during myogenesis. Importantly, a composite of these two profiles appeared to comprise the major transcriptional load of the zebrafish dystrophin gene. In contrast, utrophin gene expression was dominated by the full length transcript throughout embryogenesis. The approach described here provided a means by which a gene's transcriptional complexity can be deconvoluted to reveal transcriptional diversity during embryogenesis. This approach, however, required the identification of unique regions for transcript-specific targeting, and an appreciation of alternative splicing events that may compromise the design of primers for qRT-PCR.

Introduction

The human dystrophin gene spans 2.5 million base pairs and is the largest gene in the human genome (1). The dystrophin gene contains at least 7 known promoters that drive the expression of multiple transcripts, together with multiple splice variants (Fig. 1). The largest and most studied isoform of dystrophin, designated Dp427m, is expressed as a 427 kDa protein in human muscle. This protein comprises 3685 amino acids and plays a pivotal role in maintaining the integrity of the dystrophin:glycoprotein complex that connects cytoskeletal actin to the extracellular matrix of muscle cells. The less abundant isoforms of dystrophin are localised to other tissues such as the brain (Dp427b/c and Dp427p), the retina (Dp260), brain and kidney (Dp140), and Schwann cell/peripheral nerve (Dp116); the smallest isoform, Dp71, is ubiquitously expressed (2,3). The dystrophin gene has an equally large paralogue, termed utrophin (4,5), which expresses multiple transcripts (2,3,6-18) (Fig. 1). The full length utrophin isoform is able to fulfil the role of dystrophin when overexpressed in the absence of dystrophin (19).

Human dystrophin orthologues have been identified in many animal and insect species to date, ranging from mammals such as dog (20), cat (21) and mouse (22) to fish such as pufferfish (23) and zebrafish (24-26), together with frog (27), sea urchins (28) and fruit flies (29). Utrophin has been confirmed in many of the species in which dystrophin has been identified (4,30,31).

Importantly, Duchenne and Becker muscular dystrophies (D/BMD) are caused by mutations in the dystrophin gene [reviewed by Manzur *et al* (32)] and several models of D/BMD have been constructed using mice (33) and dogs (20), among others. In addition, there is increasing interest in using the zebrafish to model muscular dystrophies such as DMD (26,34-36).

Given the transcriptional complexity of the dystrophin gene and its role in diseases, mutations generally impact not only on Dp427m but also other dystrophin isoforms. In the absence of dystrophin expression, upregulation of full-length utrophin appears to occur (19). Against this background, little is known of the complexity of the developmental expression profile of dystrophin, as well as the utrophin, gene transcripts. In order to address this deficiency, we turned our attention to the zebrafish (Danio rerio). Zebrafish offers advantages over other model organisms in that they are optically transparent during embryonic development and ex utero development allows for easy manipulations (34,36-38). The main qualities of using zebrafish are their rapid generation cycles, low maintenance requirements and cost effectiveness at high densities.

Correspondence to: Dr Donald R. Love, Diagnostic Genetics, LabPlus, Auckland City Hospital, P.O. Box 110031, Auckland 1148, New Zealand
E-mail: donaldl@adhb.govt.nz

Key words: zebrafish, dystrophin, utrophin, quantitative real-time PCR

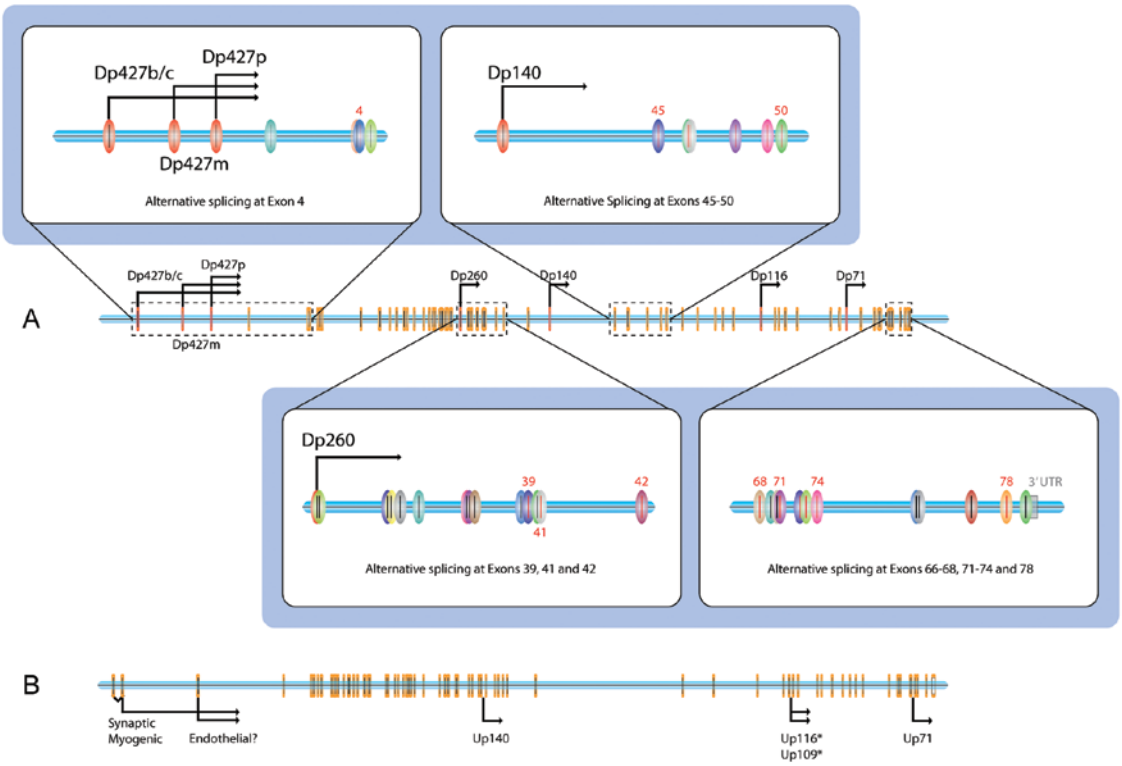

Figure 1. Human dystrophin gene exons and promoters. (A) The middle section shows the human dystrophin gene, the relative location of all exons, and the principal transcripts that are expressed by the gene. The exons involved in alternative splicing are shown in the expanded views. The unique first exons for the various dystrophin isoforms are shown, together with their relative positions in the gene. (B) The start sites of the various isoforms of the human utrophin gene are shown. In the case of Up116 and Up109, they have only been isolated from mouse to date (18).

The aim of our study was to provide for the first time a developmental expression profile of dystrophin and utrophin gene transcripts. Expression profiling posed challenges due to the transcriptional complexity of both genes: nucleotide sequences of each gene's transcripts are largely identical (albeit varying in length), and any differences are largely confined to short unique 5'-end sequences. These challenges highlighted the difficulty in applying a conventional RNA in situ hybridisation approach, which also suffers from sensitivity issues, therefore we targeted multiple exons in a quantitative real-time PCR (qRT-PCR) strategy.

Materials and methods

Primers. Primers were designed using the free online primer design software Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Advanced settings were used where appropriate to minimise the formation of secondary structures while reducing non-specific products being amplified by the primers. All primers designed were subjected to manual quality analysis using NetPrimer (http://www.premierbiosoft.com/netprimer/). qRT-PCR primers (Table I) were designed to produce product sizes between 100-160 base pairs with an approximate T_m of 60˚C. All primers were designed so that the product would encompass at least 2 exons, with preference given to primers that annealed to exon/exon boundaries, providing that primer quality was not compromised.

Primers were ordered from Invitrogen Corporation (http://www.invitrogen.com) with all primers used for qRT-PCR subjected to quality checks. These checks involved the confirmation of only one PCR amplification product following electrophoresis on a 1.5% agarose 1X Tris/Borate/EDTA (TBE) gel stained with ethidium bromide. All PCR products and excised gel bands were subjected to spin-column purification using a GE Healthcare Illustra GFX PCR DNA and Gel Band Purification kit according to the manufacturer's recommended instructions (GE Healthcare Life Sciences).

Sequencing. Sequencing reactions were carried out by the Genomics Unit, School of Biological Sciences, University of Auckland. The ABI PRISM™ BigDye Terminator sequencing kit version 3.1 was used to incorporate fluorescent dyes as recommended by the manufacturer. Thermal cycling was carried out using an Applied Biosystems GeneAmp® PCR System 9700. Unincorporated fluorescent dyes were removed using Agencourt® CleanSEQ® magnetic beads. Capillary electrophoresis of the sequencing products was carried using the ABI PRISM™ 3130xl Genetic Analyzer.

Zebrafish husbandry. Wild-type zebrafish (Hollywood Fish Farms, Auckland, New Zealand) were grown and maintained at densities of 3-8 fish per 2.75 liters tank, in a water recirculation rack system (Aquatic Habitats) located inside a dedicated zebrafish facility. The zebrafish were exposed to 14 h of light per day with temperatures kept between 26-28˚C. Water quality was monitored on a daily basis. Adult fish were fed a range of dry fish food supplemented with live Artemia, and juveniles were fed a mix of dry fish food and live Rotifer until at least two weeks of age. Adult male and female fish were

Table I. Primers used for qRT-PCR.

Primer designations	Target Gene	Target exons	Primer sequence
ZFdysX01F5	Dystrophin	1-2	GAAGCAGTGAACCCACAGGAC
ZFdysX02R5			TGAGTTTATCCATTTGGTGAAGG
ZFdysX02F1	Dystrophin	2-3	CCTTCACCAAATGGATAAACTCAC
ZFdysX03R1			GCCAACCAGAACCTCCAGAAG
ZFdysX75F5	Dystrophin	75-76	CGGCAACACAAGGGACGA
ZFdysX76R5			GAGGCAGTGGAGGGTGAGGA
ZFdysDp71X01F2	Dystrophin	Dp71-63	GCTTTTTCCACTGTAGAGAGGG
ZFdysX63R2			TGTTGTTTGGGTCTGGTGATTT
ZFutrX01F1	Utrophin	1-2	CGTGGGTGTGATGTCAGAG
ZFutrX02R3			CTTGGAGAACTGAGCGTTGA
ZFutrX60F1	Utrophin	69-70	AGATTCTGCAGGCAGTGGA
ZFutrX70R1			AGGCTGACCGTACTGAAACC
EF1aF	Elongation factor 1α		CTGGAGGCCAGCTCAAACAT
EF1aR			ATCAAGAAGAGTAGTACCGCTAGCATTAC
Rpl13aF	Ribosomal protein L13α		TCTGGAGGACTGTAAGAGGTATGC
Rpl13aR			AGACGCACAATCTTGAGAGCAG

kept mixed until one week prior to spawning at which point they were separated by gender. Embryos were collected after spawning and thoroughly rinsed. The embryos were visually inspected to ensure that they were at the same developmental stage and then kept immersed in 0.5X E3 media (1X E3: 5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$ and 0.33 mM $MgSO_4$) in 15 mm plastic Petri dishes at a density of 20/dish and kept at a constant 28˚C. Embryo E3 media was changed every 24 h.

Developmental time course. The zebrafish embryonic developmental stages described in this paper are reported as hours post fertilisation (hpf) at 28˚C: 3.5 hpf (high stage), 10 hpf (bud stage), 11 hpf (3-somite stage), 14 hpf (10-somite stage), 16 hpf (14-somite stage), 19.5 hpf (21-somite stage), 22 hpf (26-somite stage), 25 hpf (prim-6 stage), 31 hpf (prim-16 stage), 42 hpf (high pec stage), 60 hpf (pec fin stage) and 72 hpf (protruding mouth stage).

RNA extraction and purification. At each appropriate developmental stage, 20 embryos were inspected to ensure the correct developmental stage was reached, then they were pooled and snap frozen. RNA extraction and purification was carried out according to the protocol described earlier (39).

Reverse transcription. One microgram total-RNA from each extraction was reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen Corporation) primed with random hexamers (Invitrogen Corporation) in a 20 μl reaction volume according to the manufacturer's instructions. The final SuperScript III inactivation step (70˚C for 15 min) was omitted and cDNAs were diluted 1:16 before use in qRT-PCR assays.

Quantitative real-time reverse transcription polymerase chain reaction. All qRT-PCR assays were carried out using an Applied Biosystems 7900HT platform running SDS v2.3 software (Applied Biosystems). Reactions were set up using an Eppendorf epMotion 5075 (Eppendorf) automated liquid handling platform to minimise pipetting errors. Ten microliter reaction volumes comprised the following: 5 μl Platinum SYBR-Green qPCR SuperMix-UDG with ROX (Invitrogen Corporation), 1 μl of 2 μM forward primer, 1 μl of 2 μM reverse primer and 2.6 μl of template; the template was either 16X diluted cDNA, or water in the case of negative controls. The expression of two zebrafish reference genes were used as normalisers, *Rpl13a* and *EF1a* (40), with all reactions performed in triplicate. Cycling programs consisted of 40 cycles of 15 sec at 94˚C and 60 sec at 60˚C. Dissociation curve analysis was performed for each qRT-PCR run.

qRT-PCR data analysis. The data obtained from the SDS v2.3 software were inspected manually to omit any well that showed abnormal amplification curves, followed by manual baseline and threshold setting adjustments as recommended by the SDS manual. Amplification efficiencies were determined using LinRegPCR v11.1 (41,42). Baseline and threshold corrected Ct values, together with amplification efficiencies, were processed as described in the geNorm v3.5 manual (version: July 8, 2008) (43). This processing provided relative expression levels (arbitrary quantities) while allowing for normalisation against the two reference genes.

Results

Dystrophin and utrophin sequences. At the start of this study, the available sequences for zebrafish dystrophin were fragmented and incomplete. These limited data were used to design primers in order to amplify overlapping fragments of the full length zebrafish dystrophin muscle transcript, which

A

```
   1 GGGGATTCAC AAAAAATTCA CTGCTGGATC TGAGGTTCAG TTTGTTCACA  50
  51 GGGCGTTGGT GAGGGCTCCT TCCTTTTTCT GTTTATAAAA TTGTAGGATT 100
 101 AAAGGACCCA ACGTTGGCTT GGATCTTATG TCATGGCCGA AGCAGTGAAC 150
 151 CCACAGGACC AATGGGAGGA AGGTTTTGAG GACGAATTCG GAGAAATCAT 200
 201 CAAAACTCGA TCAGATGAAA GAGAAGACGT TCAAAAGAAA ACCTTCACCA 250
 251 AATGGATAAA CTCACAGTTT GCTAAGACAA GAAGACCTCC TATTGATGAC 300
 301 CTCTTCACTG ACCTGTGTGA TGGCCGACGT CTTCTTGAAC TTCTGGAGGT 350
 351 TCTGGTTGGC CATGAAATAG TTAAAGAACG TGGCTTCACT CGAGTGCACT 400
 401 CCCTTAACAA TGTTAACAGG GCCCTGCAGA TCCTTCAGAA GAACAATGTT 450
 451 GATCTGGTGA ACATTGGAGG AGCTGACATT GTAGATGGGA ATCATAAACT 500
 501 GACCCTGGGG CTCATCTGGA GCATCATTCT CCACTGGCAG GTTAAGGATG 550
 551 TGATGAAAGA TGTCATGGCG GACTTACAGC AGACCAACAG CGAAAAGATC 600
 601 CTGTTAAGCT GGGTCAGGCA GTCCCTCAAA AACTACCAAG ACGTCAACGT 650
 651 GGTCAACTTC TCCAGCAGCT GGGCCGACGG TTTCGCTTTC AACGCTCTCA 700
 701 TCCACAGCCA CAGGCCGGAG CTGTTCAGCT GGAGTGTAGT GGAGCAACAG 750
 751 GATAATGCCA TTGAGAGACT GGATCATGCC TTCGGTGTTG CGGAGAAGAG 800
 801 TTTAGGAATC GACCGGCTGC TGGACCCCGA GGATGTTGCT ACAGTCCATC 850
 851 CAGATAAGAA ATCCATCATC ATGTACGTTA CCTCCCTCTT CAAAGTGCTG 900
 901 CCCCATGGCG TGAGTTTGGA GGCCATTCAA GAGGTGGAGA CCCTCCCTCG 950
 951 AGCCGCAGTG ACCAAGGAGG AGCACGTCCT CTATCAAACC CAACAGCGGT 1000
1001 ACTCTCAGCA GATCACAGTC AGCGTGGCCC AGAGTCGCGT TCGCAGCCCG 1050
1051 TCGCCCTCCT ATAAGCCGAG ATTTAAAAGC TATGCCTTCA CTCAAGCCGC 1100
1101 CTACGTCAAG ACACCTGAAC AGCAGAGGAA GTTCCTCATC GCACAGTCTC 1150
1151 CAGACAAGGC CGATGAACTG CGACCGTCCC CCAGTCCCCT GCCGCAGGGG 1200
1201 CTAAATGAGC TGGAAAGCTA CCAGAGTGCC CTGGAGGAGG TCTTGACCTG 1250
1251 GCTGCTCTCT GCGGAGGACG GCCTGCAAGC ACAGCCACCC ATCTCTTCTT 1300
1301 TTGTAGAAGA AGTCAAAGAG CAGTTTCACA CTCATGAGGG CTACATGGTG 1350
1351 GAGCTGACGT CTCACCAGGG CAGCGTGGGC CGGGTCCTGA AAGCGGGAAG 1400
1401 TGTGTTGTTG TCTGGAGGGC AGCTGACTGA TGATGAAGAG AGGGAGGTGC 1450
1451 GGGAACAGAT GAATCTGCTC AACTCGCGCT GGGAACATCT GCGTGTGGCC 1500
1501 AGTATGGAGC GACAGAGCAG GCTTCATGAA GTTTTGATGG ACCTTCAGCA 1550
1551 CCAGCAGCTG AAGCAGCTCT CTGATTGGCT GGACACGACG GAGACTCGTA 1600
1601 TTAAAAGGAT GGGAGCGCAG GCTTTGGGGC CTGAACTTGA TGACATAAAG 1650
1651 CGGCAAATCG AAGAGCAGAA ACTCTTACAG GAAGATCTGG AGCTGGAACA 1700
1701 AGTGCGAGTG AATTCACTGA CTCATATGGT GGTGGTTGTA GATGAGAATA 1750
1751 GCGGGGATGG CGCCACCGCA GCCTTGGAGG AAAAGCTGCA GAACCTTGGG 1800
1801 GAACGATGGG CAGCGATCTG TAAATGGACT GAAGAGCGTT GGATCCTCCT 1850
1851 TCAAAAGATC CTGCTGTGCT GGCAGCATTT CTCTGAGGAG CAGCTGCTGT 1900
1901 TTGATTCTTG GCTGACTCAA AAGGAGGAGT TGGTCCAGTC CATCCAAAGC 1950
1951 AGTGGCACAA ATGACCCAAA TGAAGTAGCC GCCAACCTGC GGAAGCTGGC 2000
2001 GATTTTAAAA GCAGACCTGG AGCTGAAGAG GCAGACCATG GACAAGCTTT 2050
2051 GCTCGCTCGT CCAGGATCTG CTGACTAACA TCAAGAGCAA AGAGGCCGCT 2100
2101 GGGAAGCTGG AAGCAAAGCT GGAGAGGTTT GCTCAGCGCT GGGACAAATT 2150
2151 GGTGCAGGCG CTTCAGCTCA CCAGCACCAA GATTTCAACT ATTGTCACCA 2200
2201 CATCCCAGTC GGAGATCACG CACACAACCA TGGCAACTGT CACCAAGGTG 2250
2251 ACCACGAACC AGAAGAAGAT GGTGAAGCAT ACTAAGGAGG GCATGTCCAC 2300
2301 CCCTCCACCT CAGAAAAAGA GACAGATTGT CGTGGATTCA GAACTGCGGA 2350
2351 AAAGATTTGA TGTGGACTTC ACTGAAATTC ACAGTTTTAT GACTCGGTCC 2400
2401 GAGGCTGTTT TGCAAAATCC CGAGTTCTCA ATCTCCCGTA AAGAAGGCAG 2450
2451 TGTGGCAGAT CTCTATGAAA AAGTGCTGGC AATTGACAGA GAAAAACCCG 2500
2501 AAAAGTTCAG GAAGTTGCAA GAAGCCACTC GCTCAGCCCA GGCTTTGGTG 2550
2551 GATCAACTTA CCAGCGACGG TCAGAACTCA GAGGACATCC AGCAAGCGGC 2600
2601 TCAGCAGTTG CGTGCTCGAT GGGTGGACTT CTGCGCTCTC CTGGCCGAGC 2650
2651 GGCTGGCGTG GTTGGCCTAC CAGACCAAAG TGCTGGCTTT CTACAATCTC 2700
2701 TTTCAGCAGC TGGAGCAGGC TGCGGCCACG GCGGAGAACT GGCTGAAGGT 2750
2751 GCAGTCGCCA CCCGCATGTG AACCCGAACC TCTGAGGATC CAGCTGGAGC 2800
2801 GCTGCAGCGA TGAAATCTCT CGCCTCTCGG CTCTTCAACC TCAAGTGGCC 2850
2851 AAACTGCACG AACAACTCGA GGAGCTGCGG CAGAAAGAAG AGACGCCGGT 2900
2901 TCTGTTTGAC GCGGACATTT CTGCCTTCCA AGAGCACTAC CACCAAGTCC 2950
2951 TGGAGGACCT GAGGGCCAGG GAACGACAGC TCGTTTTGGT CCAGTCTAGT 3000
3001 CTTCCTCCAG CACGTTATAA GGATGTGATG GCAGCGCTGT TGGCCTGGCT 3050
3051 ACAGCAGTGT GAGAATAAAC TAGCCATCCC TTCTACAGCA GTGACTGAGT 3100
3101 ATCCCGTCAT GGAGCAGAGA CTCAAAGATA TCAAGGCAAT CCAAGCATCC 3150
3151 CAGAAGGAGC ATCAGGGTGA CGTGGATGAC CTCAATAAGA TGGCTGAGCA 3200
3201 GGTTTTCCAG AAAGCTCCAC CTGAGATCTG TCAGAAGTAC CGGACAGAGC 3250
3251 TGGATAACGT GATGGTGCGC TGGAGACGCA TATCGGAGCA GCTGGAGGAG 3300
3301 AACATCCAGA AACTGCAGGA TCACATGACC AAGCTACAGC AGTTCCAGAA 3350
3351 TGACACAAAG ACTCTGCAGA AGTGGATGGC AGAGGTGGAT GTTTTCCTGA 3400
3401 ATGAGGAATG GCCAGCTCTG GGAGACGCAG AGGCTTTGGA GAAACAGTTG 3450
3451 GAACAGTGCA CGGCTCTGGT GAATGACATC CACACCGTCC AACCCAGTCT 3500
3501 GAACGGCATC AATGAGGTGG GACTGGCCTT GAAGAGGGAA GCTGAGACGC 3550
3551 CGTTTGCCAT CAAAATACAG AAGATGCTTG ATGAGCTCAA TGCGCAGTGG 3600
3601 GAGCTCATTT GCAAACAGGC TTATGCAAAG AAATCAGCCC TGAAAGGAGG 3650
3651 GCTGGACATG ACGGTGAGTT TGAGGAAGGA GATGCAGGAG ATGCAGGAGT 3700
3701 GGATCACTCA GGCCGAGGAA GAGTACCTGG AGAGAGACTT CCAGTACAAA 3750
3751 ACACCTGAGG AGCTTCATAA AGCAGTGGAG GAGCTGAAGC GGGCTCAAGA 3800
3801 GGAAGTCCAT CAAAAGGAAA CAAAGGTGAA GCTGCTGACG GATAAAGTCA 3850
3851 CCAATTTCAT TTCAAAGGCT CCTCCCGCGG CCCATGATGC CCTGAAGGCA 3900
3901 GAACTGGATG TTTTGACTTC TAACTACCAG CGCCTGTGCA GCCGACTCGA 3950
3951 TGGAAAATGC AAGACTTTAG AGCAAGTGTG GGCATGCTGG TGTGAGCTGC 4000
4001 TGTCTTATTT GGAGTTGGAG AATGCCTGGA TGGACCTACT TGAAAAGAAA 4050
4051 CTTGATGAAA CGGAAGGGCT TCAAGGAGGC ATAGAGGAAA TTGAAGAAGC 4100
4101 TTTGACTTCT TTAGACACCA TGATTAGAGA GCATCCTGAA TACAACCGTA 4150
4151 ACCAGATCCG TGAATTAGCG CAGACCCTGA TGGATGGCCG AGTCCTGGAT 4200
```

Figure 2. Nucleotide sequence and predicted amino acid sequence of zebrafish Dp427m.

Figure 2. Continued.

```
 8401 CACAAGGCTT TCAGAAGAGA GCTGGGTGCC AAAGAACCTG TAATCAATGG 8450
 8451 AACTCTTGAC AATGCGAAAA CCTTCCTCGC TGAGATGCCT CGTGAGGGCC 8500
 8501 TGAAACAGAG ACCTGGACAA AAAGATGTAA GTCCTGAGGA GCGGGTTCAA 8550
 8551 AACGTGGGCC GCATATTGCG TAAAGAAGTG GAGGATGTAA CGGTGCGATG 8600
 8601 GAAGAATCTG GGTGCAGCTT CTGTGGACTG GCAGCAACAG CTGGAACTGG 8650
 8651 CCTTGGAGAG GCTGATGGAG CTCCAGGACG CCCAGGATCA GCTGGACTAC 8700
 8701 AAGCTACGAC AGACTGAGAG CGTGAAGAAT TCTTGGAAGC CTGTTGGGGA 8750
 8751 GCTGCTTGTA GACGATTTGC AAAACCACAT AGACAGAGTG AAGGCATTTC 8800
 8801 AGGAAGAGAT AGCTCCCATT CAGGATAACG TAAATCATGT CAACCAGCTG 8850
 8851 GCCTCCACAT TTAGACCGTC TGACATTCAG CTTTCTCCAG ACAACCTGAG 8900
 8901 CAGAATCGAT GACCTCAACA TGAGATGGAG GCTTCTGCAG ATCTCCATTG 8950
 8951 AGGAGCATCT GAGTCAGCTG ACCACAGCCT TTAAGGACTT GGGACCCTCT 9000
 9001 CAGAATTTCC TTCATGCGTC TGTTGAAAGC CCTCTCGAAC GCTCTATTTC 9050
 9051 ACCCAACAAT GTCCCCTACT ATATCAATCA CCAGACCCAA ACAACATGTT 9100
 9101 GGGACCACCC AAAGATGGCA GAACTCTACC AGTCATTAGG GGATCTCAAC 9150
 9151 AACGTGCGGT TCTCGGCATA CAGGACGGCA ATGAAGCTCA GACGAATGCA 9200
 9201 GAAAGCCCTC TGTTTGGATC TTCTGAGCAT GCCTGCAGCC TGTGAAGCCT 9250
 9251 TTGAGCAGCA CAATCTCAAA CAGAACGAGC AGTTCATGGA CATCGTGCAG 9300
 9301 GTGATCAACT GTCTGACCAG CATCTACGAC CGTCTGGAGC AGCAGCACAG 9350
 9351 CAGCCTGGTC AACGTGCCTC TCTGTGTGGA CATGTGTCTC AACTGGCTGC 9400
 9401 TCAACGTTTA CGATACAGGA CGAGCTGGGA AGATTCGTAC CCTATCCTTC 9450
 9451 AAAACAGGAA TAATCTCTTT GTGCAAAGCT CACCTTGAAG ATAAGTACAG 9500
 9501 ATTTTTTATTT CGAGAGGTGG CCAGTGCCAC AGGCTTCTGT GACCAGCGGC 9550
 9551 GCCTCGGCCT CCTCCTGCAT GATGCCATTC AGATCCCCAG GCAGCTGGGT 9600
 9601 GAAGTGGCGT CTTTCGGAGG GAGCAATATT GAGCCCAGTG TGCGCAGCTG 9650
 9651 CTTTCAGTTC GCCAATAACA AACCGGAGTT AGAGGCTTCA GTCTTCCTGG 9700
 9701 ACTGGATGCG TTTAGAACCT CAGTCGATGG TTTGGCTTCC TGTCCTTCAC 9750
 9751 CGTGTAGCGG CCGCTGAGAC AGCAAAGCAC CAGGCTAAAT GCAACATTTG 9800
 9801 TAAGGAATGT CCTATTATTG GCTTCAGGTA CCGAAGTTTA AAGCACTTTA 9850
 9851 ACTATGATAT CTGCCAAAGC TGCTTCTTTT CTGGCAGAGT GGCCAAAGGT 9900
 9901 CACAAGATGC AGTACCCTAT GGTTGAATAT TGCACACCGA CGACGTCAGG 9950
 9951 AGAGGATGTG AGAGACTTTG CCAAGGTGTT AAAGAACAAG TTCAGGACAA 10000
10001 AGCGCTATTT TGCCAAGCAC CCTCGCATGG GTTACCTTCC CGTCCAGACC 10050
10051 ATCCTTGAAG GAGACAACAT GGAGATTCCT GTTACACTGA TCAACTTCTG 10100
10101 GCCTGTTGAC CATCGGCCTG CATCTTCCCC TCAACTCTCC CATGACGACA 10150
10151 CCCACTCTCG CATCGAGCAT TACGCTAGCA GGGATGATGA GCATCTGCTG 10200
10201 ATCCAGCACT ACTGCCAGAG TCTGAATCAA GGCTCTCCTC TCAGCCAGCC 10250
10251 GCAGAGCCCC GCTCAGATCC TCATCTCAAT GGAGACTGAA GAGAAGGGAG 10300
10301 AGCTGGAGAG AGTGCTCAAT GATCTGGAGC AGGAAAACAG GAAGCTGCAA 10350
10351 GCGGAGTACG ATCGTCTGAA AAAGGCACAC GATCACAAGG GTCTGTCACC 10400
10401 GTTGCCTTCA CCTCCACAGA TGCTTCCGGT GTCGCCTCAA AGTCCACGCG 10450
10451 ACGCAGAACT AATCGCAGAA GCCAAACTAC TGCGGCAACA CAAGGGACGA 10500
10501 TTGGAGGCAA GAATGCAAAT CCTGGAGGAT CACAACAAGC AGCTGGAATC 10550
10551 GCAGCTTACA CGCCTAAGAC AGTTACTCGA GCAGACTGAG TCCAAGGTGA 10600
10601 ATGGCACTGC CCTGTCCTCA CCCTCCACTG CCTCTCCGAG ATCTGACACC 10650
10651 AGCCTGGCCT CACTGCGTGT GGCCGCAAGC CAAACCACAG AGACCATGGG 10700
10701 TGATGATGAG CTGTCCAGTC CCACCCAGGA TGCAAGCACT GGATTAGAAG 10750
10751 ACGTCATCGA GCAGCTCAAC AACTCCTTCC CTCACAGCCA AGGAGGCGGA 10800
10801 CGGCTAAATC CATGAGAGGT CCCAATGTGG GAAGTCTTTT CCACATGGTT 10850
10851 GATAACATTG GACATGCCAT GGAGTCGCTG GTTCATGTAA TAACAGAGGA 10900
10901 GCAGGATTTA GACTGAAGAC GTCTTCTCCT CGTTGCATGC TTTTGTAGTG 10950
10951 TCAACAACTG GACCGGATAT GTTTACAATG GGGAATATCA ATAAAAATCT 11000
11001 ATTTTTCTGA AGGA                                       11014
```

B

```
    1 MAEAVNPQDQ WEEGFEDEFG EIIKTRSDER EDVQKKTFTK WINSQFAKTR 50
   51 RPPIDDLFTD LCDGRRLLEL LEVLVGHEIV KERGFTRVHS LNNVNRALQI 100
  101 LQKNNVDLVN IGGADIVDGN HKLTLGLIWS IILHWQVKDV MKDVMADLQQ 150
  151 TNSEKILLSW VRQSLKNYQD VNVVNFSSSW ADGFAFNALI HSHRPELFSW 200
  201 SVVEQQDNAI ERLDHAFGVA EKSLGIDRLL DPEDVATVHP DKKSIIMYVT 250
  251 SLFKVLPHGV SLEAIQEVET LPRAAVTKEE HVLYQTQQRY SQQITVSVAQ 300
  301 SRVRSPSPSY KPRFKSYAFT QAAYVKTPEQ QRKFLIAQSP DKADELRPSP 350
  351 SPLPQGLNEL ESYQSALEEV LTWLLSAEDG LQAQPPISSF VEEVKEQFHT 400
  401 HEGYMVELTS HQGSVGRVLK AGSVLLSGGQ LTDDEEREVR EQMNLLNSRW 450
  451 EHLRVASMER QSRLHEVLMD LQHQQLKQLS DWLDTTETRI KRMGAQALGP 500
  501 ELDDIKRQIE EQKLLQEDLE LEQVRVNSLT HMVVVVDENS GDGATAALEE 550
  551 KLQNLGERWA AICKWTEERW ILLQKILLCW QHFSEEQLLF DSWLTQKEEL 600
  601 VQSIQSSGTN DPNEVAANLR KLAILKADLE LKRQTMDKLC SLVQDLLTNI 650
  651 KSKEAAGKLE AKLERFAQRW DKLVQALQLT STKISTIVTT SQSEITHTTM 700
  701 ATVTKVTTNQ KKMVKHTKEG MSTPPPQKKR QIVVDSELRK RFDVDFTEIH 750
  751 SFMTRSEAVL QNPEFSISRK EGSVADLYEK VLAIDREKPE KFRKLQEATR 800
  801 SAQALVDQLT SDGQNSEDIQ QAAQQLRARW VDFCALLAER LAWLAYQTKV 850
  851 LAFYNLFQQL EQAAATAENW LKVQSPPACE PEPLRIQLER CRDEISRLSA 900
  901 LQPQVAKLHE QLEELRQKEE TPVLFDADIS AFQEHYHQVL EDLRARERQL 950
  951 VLVQSSLPPA RYKDVMAALL AWLQQCENKL AIPSTAVTEY PVMEQRLKDI 1000
 1001 KAIQASQKEH QGDVDDLNKM AEQVFQKAPP EICQKYRTEL DNVMVRWRRI 1050
 1051 SEQLEENIQK LQDHMTKLQQ FQNDTKTLQK WMAEVDVFLN EEWPALGDAE 1100
 1101 ALEKQLEQCT ALVNDIHTVQ PSLNGINEVG LALKREAETP FAIKIQKMLD 1150
 1151 ELNAQWELIC KQAYAKKSAL KGGLDMTVSL RKEMQEMQEW ITQAEEEYLE 1200
 1201 RDFQYKTPEE LHKAVEELKR AQEEVHQKET KVKLLTDKVT NFISKAPPAA 1250
 1251 HDALKAELDV LTSNYQRLCS RLDGKCKTLE EVWACWCELL SYLELENAWM 1300
 1301 DLLEKKLDET EGLQGGIEEI EEALTSLDTM IREHPEYNRN QIRELAQTLM 1350
 1351 DGRVLDELIH KKVEDYNTRW DELMQRASQR RQQLEKSLHW AQENDKTLRL 1400
 1401 IQDSLNTTDR HLTAYIADGI DAAQIPQEAQ KIQTELNGHE VTLDDMKKKA 1450
```

Figure 2. Continued.

```
1451 MEVDASEKVI GEIDATLDKL LQVKGKFRLF QKPANFDQRL RECERVLEEV 1500
1501 KVKLGVLSIR SVEQEVVQSQ LEQCMKFYKN LSEVKSEVET VIKTGRQIVQ 1550
1551 KQQTEQPKEL DDRLTALKLA YNDLGSQVTE GKQELEKILK LLRKFRKEVN 1600
1601 SLTEWLATTD GELTRRSSVE GMPSDLDAEL AWAKAAQEDT EHHKPQLKLV 1650
1651 RELAGTLKGL LHSQENLIDD KVSLLNCNWI AVTSRSEQWL KLLLDYQNQM 1700
1701 KTLEQNISQI NTWMDRAEEK LDEMDSQGCI EHVIKGLRME LEEMKGKVEA 1750
1751 VQVLAEDLIK NTGEHCKAQV KPKLEQLNQR FDIVARRILM GQASSQELDE 1800
1801 YHRQANIWLQ VLDEEIKLGE SLKEEDFLED ATIDEEALNE LFLKGENLLK 1850
1851 RTPGGEKREA VREKHNLLHD RYDTLKNLRA LRKKKALALA SQWYQFCKKS 1900
1901 DDMMQWLDRI EKSIRELPDP PEEPRVKAVG YEIDQQRPKL EDLRGLGRVL 1950
1951 SEGGASRMVE PRLLPINKRW TELDVNFTRV RHKSELQFLL QCISENEALL 2000
2001 NSPERWSTAF INLPQQEKCL KEVKVNLDKL EGPVAEAVGR GASQPEEGLL 2050
2051 VQLLRTNWEN LKMLYQDRLK RLEKAKKFNE ELKMLDNWLT DAERTIMKYE 2100
2101 QDPINNRDHL KELQAGLEKQ EAAVKGLNAL GTDLSPQCSK DDRDHIKQQL 2150
2151 ASINSRWAKV SNQLTEIKRR SAGAKMLLAE LNEDMGEFQS WLDDAEAVAA 2200
2201 LPVEAGRKEQ LSATLEKVKA RVVELPSRKQ ALHQINSKGS SLPADKVKPL 2250
2251 EKQLKVINMR WAKVSTDLPE KQRQIEDLLR DLSLYQEQLS KLSIWASNTK 2300
2301 TQLEQSPTAV DPKIEDDVKE KKPAVETLLA KERPPCQPEK GQYDGLSADW 2350
2351 TSIQVLLKDW KDKCQLAAVT LTGSAAGDAA LDKFNKSWTE LDDWLALLDH 2400
2401 MVQTQRVMVG DLDEINEMTV KLKSALQDME QRCPQLNKQI TAAQNLKNKT 2450
2451 NNPETRATIT DRIEKLQAHW EDSHAKLTAR VLTLQNMYKD SSDWLEARKR 2500
2501 VEPLIKKANE KLESWKKVSH SVEDLKGQNA DVKQLSKDLQ QWQTQMNVTN 2550
2551 ELANKLLTLY ADDDTSKVKQ MTESMNLAWA NIKKRAGDKE ADLEAGLRQL 2600
2601 QHYYLDLEKF LNWLTEAETT ANVLQDATFK EGLLENPATV RHLLEQWQDL 2650
2651 QAEIDAHRET YHSLDENGHR IVSSLEGTDN AVVLQKRLDD MGQRWHELCN 2700
2701 KVMSIRPYLD AGVDQWKHLH MSLQELLNWL QLKREELEKQ KPVGGDVPTV 2750
2751 HQQLLTHKAF RRELGAKEPV INGTLDNAKT FLAEMPREGL KQRPGQKDVS 2800
2801 PEERVQNVGR ILRKEVEDVT VRWKNLGAAS VDWQQQLELA LERLMELQDA 2850
2851 QDQLDYKLRQ TESVKNSWKP VGELLVDDLQ NHIDRVKAFQ EEIAPIQDNV 2900
2901 NHVNQLASTF RPSDIQLSPD NLSRIDDLNM RWRLLQISIE EHLSQLTTAF 2950
2951 KDLGPSQNFL HASVESPLER SISPNNVPYY INHQTQTTCW DHPKMAELYQ 3000
3001 SLADLNNVRF SAYRTAMKLR RMQKALCLDL LSMPAACEAF EQHNLKQNEQ 3050
3051 FMDIVQVINC LTSIYDRLEQ QHSSLVNVPL CVDMCLNWLL NVYDTGRAGK 3100
3101 IRTLSFKTGI ISLCKAHLED KYRFLFREVA SATGFCDQRR LGLLLHDAIQ 3150
3151 IPRQLGEVAS FGGSNIEPSV RSCFQFANNK PELEASVFLD WMRLEPQSMV 3200
3201 WLPVLHRVAA AETAKHQAKC NICKECPIIG FRYRSLKHFN YDICQSCFFS 3250
3251 GRVAKGHKMQ YPMVEYCTPT TSGEDVRDFA KVLKNKFRTK RYFAKHPRMG 3300
3301 YLPVQTILEG DNMETPVTLI NFWPVDHPPA SSPQLSHDDT HSRIEHYASR 3350
3351 DDEHLLIQHY CQSLNQGSPL SQPQSPAQIL ISMETEEKGE LERVLNDLEQ 3400
3401 ENRKLQAEYD RLKKAHDHKG LSPLPSPPQM LPVSPQSPRD AELIAEAKLL 3450
3451 RQHKGRLEAR MQILEDHNKQ LESQLTRLRQ LLEQTESKVN GTALSSPSTA 3500
3501 SPRSDTSLAS LRVAASQTTE TMGDDELSSP TQDASTGLED VIEQLNNSFP 3550
3551 HSQGGGRLNP *                                         3561
```

Figure 2. Nucleotide sequence and predicted amino acid sequence of zebrafish Dp427m. (A) Dp427m nucleotide sequence based on double-stranded sequencing of overlapping amplicons derived from reverse-transcribed RNA isolated from zebrafish embryos. Untranslated regions are shown in red, translated regions in blue, bases at exon-exon boundaries are underlined in turquoise and shaded in yellow, and translation start and stop codons are shaded in blue. (B) The translated protein sequence.

Figure 3. Differences between the actual and predicted sequences of zebrafish Dp427m. Five regions of the dystrophin amino acid sequence are shown that differ between the Vega predicted zebrafish dystrophin sequence (Vega Transcript ID: OTTDART00000044458; Vega Zdys) and that predicted from our cDNA sequence (Zdys).

were subsequently sequenced. The entire experimentally determined nucleotide sequence together with the predicted amino acid sequence are shown in Fig. 2. During the course of this study, updates to the zebrafish genome sequence assembly allowed for a more accurate prediction of the full length dystrophin gene; however, while this predicted sequence is largely correct, there are a few regions in the distal half of the full length transcript that do not agree with our sequence data (Fig. 3). This discrepancy may be due to alternative splicing or minor sequencing artefacts. Our sequence also differs from that reported by Guyon *et al* (44), which is a composite of direct cDNA sequence data and EST sequences. This earlier sequence carries an additional exon (number 72, compared to annotations of the human Dp427m transcript, and so probably represents an alternatively spliced transcript), but also 45 amino acids at positions 495-539 (inclusive), which appears to be a repeat of amino acids at positions 585-629 (inclusive).

The utrophin gene proved difficult to predict using bioinformatic approaches, and the sequences appeared to be fragmentary using the currently available zebrafish genome builds. The utrophin sequence presented here is based on an incomplete utrophin gene prediction, which primarily covers the amino and carboxyl-terminal regions (Fig. 4). A comparison of the predicted exons for zebrafish utrophin and those from human utrophin via ClustalX DNA alignment showed 64.3 and 55.9% of the base pairs to be identical for the amino and carboxyl-terminal regions, respectively.

Developmental expression profiling. Expression profiling of the zebrafish dystrophin gene involved targeting four distinct regions of this gene in order to allow for the simultaneous analysis of its transcriptional diversity; our own sequence data was used for designing most of the primers. The muscle-specific first exon was targeted to follow the expression of the muscle isoform, the second exon of Dp427m allowed the profiling of all full-length transcripts, the expression of Dp71 was followed by targeting this transcript's unique first exon (determined bioinformatically), and the full transcriptional load of the dystrophin gene was determined by targeting exons 75-76 (Fig. 5).

The expression profile of Dp427m approximates that of total full length dystrophin transcripts (Fig. 5A and B) with expression starting at 16 hpf, coinciding with myogenesis, and a reduced rate of expression after 30 hpf. These data suggest that for the most part, the Dp427m isoform is dominant among full length isoforms expressed during zebrafish embryo development. The expression profile of the ubiquitous Dp71 isoform (Fig. 5C) suggests maternal inheritance as post-zygotic transcription begins at 3-3.5 hpf (45). Dp71 transcripts decrease to their lowest level at approximately 16 hpf, and are then largely maintained at a low steady state level. Interestingly, the expression profile of total dystrophin gene transcripts (Fig. 5D) appears to be a composite of Dp71 and Dp427 expression profiles.

Total utrophin gene transcript expression and full length Utr427 expression exhibit similar profiles (Fig. 5E and F), suggesting that the full length Utr427 is the dominant isoform that is expressed by the utrophin gene during embryonic development. The data suggest maternal inheritance of Utr427 transcripts with low level *de novo* expression starting at approximately 16 hpf.

```
>Zebrafish_Utrophin_Exon_01
ATGGCCACCGTGGGTGTGATGTCAGAGGCGTCTGAGGGCGGAGACAACAACGAATTCACT
GACATCATCAAGTGGAGATCAG

>Zebrafish_Utrophin_Exon_02
ATGAACATGACGCGGTGCAGAAGAAGACCTTCACCAAGTGGATCAACGCTCAGTTCTCCA
AG

>Zebrafish_Utrophin_Exon_03
ACGGGGAAAGCGGCGATCAAAGACATGTTCACTGACCTGAGAGACGGCAGAAAACTGCTG
GACCTGCTGGAGGGACTGACCGCAAACACACTG

>Zebrafish_Utrophin_Exon_05
GTGGAGCTGGTGAATATTGGAGGGACAGATATAGTTGATGGCAATCATAAGCTGACTCTG
GGCCTCATCTGGAGCATCATCCTGCACTGGCAG

>Zebrafish_Utrophin_Exon_07
GCCTCATGGCTTCAGCTGGGAGAAGGTGGTCAGTCTGACGCCTGTGGAGAGACTGGAGCA
CGCTTTCACCTTCGCCAAAGACCAGCTCAACATCGAGAAGCTGCTGGATCCAGAAG

>Zebrafish_Utrophin_Exon_08
ATGTGGCGGTGCAGCTGCCGGATAAGAAGTCCATCATCATGTACGTGACCTCACTGTTCG
CCGTCTTACCCAATGACATCACCATGGATGACATCAGAGAGGTGGAGACGCTGCCCCGAC
GCTACAAGGTGGAGACCGACGAGGGACACCCGGTGAGT

>Zebrafish_Utrophin_Exon_09
GTGGAGGTGGACAGCAGCAGCAGCAGCAGCACACGGGCCGAGACCCTGGCGGAGCTGCAG
ATGGAGGTGAATCTGGACTCGTATCAGATCACTCTGGAGGAGGTGCTGACCTGGCTGCTG
TCGGCCGAAGACACGCTGCACATGCAGGAGGACGTTTCTGAAGACGTGGAGGAGGTCAAG
GACCAATTCCACACGCATGAG

>Zebrafish_Utrophin_Exon_11
GCTTCATGAGGTCCTGATGGACCTGCAGCAGCAGCAGCTTCAGCAGCTCTCTGATTGGCT
GACACAGACAGAGGGGCGGATCCGGAAGATGGAGATTGAGCCAATAGCAGGAGACATACA
GGGATACCTGACCCAGATAGAGCAGCACAAG

>Zebrafish_Utrophin_Exon_12
GCATTGCAGAATGACCTGGAGGCGGAGCAGGTGAAGGTGAACTCTCTCACACACATGGTT
GTGGTTGTGGATGAGAACAGCGGTGAGAGCGCCACTGCTGCCCTGGAGGAGCAACTGCAG

>Zebrafish_Utrophin_Exon_13
TCACTGGGTGAGCGCTGGGCGGCAGTGTGTCGCTGGACGGAGGAGCGCTGGAACAAACTG
GATGAAACTCAGCTGGTGTGGCAGAGGCTGCTTGATGACCAG

>Zebrafish_Utrophin_Exon_14
AGTTTATTTGGGTCATGGCTAACCGAAAAGGAGAAGGCCTTGGGTGAAGTGCAGACCTGT
GATTTTAAAGACCCCAGTGACATTAATGCCAGCGTCCGCAGATTGGCT

>Zebrafish_Utrophin_Exon_15
AGTCTTAAGGAGGATATGGCCCAGAAGCGCAGGACTTTGAGCGCTCTGTCAGACGCTGGT
CAAGATGTGATGGTGCTGTTGAACAGTCCAGCCGCAGCTCAAAGGATCGAATCCGACACA
GAACTTTTGACCCTACGCTGGGACAACCTGGTCCAGAAACTAGAGGACTGCTCCAGCCAG


>Zebrafish_Utrophin_Exon_63
GTGACCAATAAGGTTGAGCTGGAGCCGAGGCAGTTTATTGATTGGATGAGGCTGGAGCCG
CAGTCTATGGTTTGGCTGCCGGTTCTACACCGAGTTGCTGCGGCAGAGACGGCGAAACAT
CAGGCCAAGTGCAACATCTGCAAAGAGTTCCCCATCGTCGGCTTCAG

>Zebrafish_Utrophin_Exon_64
ATATCGCAGTTTAAAGCACTTCAACTATGACGTCTGCCAGAGCTGTTTCTTCTCTGGTCG
CACAGCGAAGGGTCATAAACTGAACTACCCGATGGTGGAGTACTGCACACCG

>Zebrafish_Utrophin_Exon_66
TCCGGTCACTCTCATCAGCATGTGTCCGGAGCACTATGA

>Zebrafish_Utrophin_Exon_67
GCTCTGTCCATCACCGCAGCTCCCTCATGATGACACACACTCACGGATAGGGGAGTACGC
CAACAG

>Zebrafish_Utrophin_Exon_68
GCTGGCACAGATGGAGCGATCCAATGGCTCTCTGCCCACCGACAGCAGCTCGGCCACTGG
AAGCAT

>Zebrafish_Utrophin_Exon_69
GGACGAGGAGCACGCTTTGATCCTGCAGTACTGCCAGACATTGGGAGGGGAGGGCTCATC
TTTCAGCCAGCCACAAAGCCCCGCCCATATCAACAGGACCCCCGCCCACAACACCCACAG
CCCCTCCTACCTGCTCCAGAGCCCCGCCCAGATTCTGCAGGCAGTGGAGAGGGAGGAGCG
AGGAGAGCTGGAGCGAATCATAGCACGGCTGGAGGACGATCAGCG

>Zebrafish_Utrophin_Exon_70
GGTTCTGCAGAGGGAGTATGAAGAACTGCGGTTTCAGTACGGTCAGCCTGGCGCCGCAGC
TGCTAGCGGAGCCCCAGATGAAGCTGATCTGCTAGCTGAAGCTAAACTCCTGCGACAGCA
CAAGGGACGACTGGAGGCCCGCATGCAGATTCTGGAGGACCACAACAAACAGCTGGAGTC
TCAGCTTTACCGACTGCGCCAGCTGCTGCATCAG

>Zebrafish_Utrophin_Exon_72
ACGAGTTTCTCCATCAGTGTGCTAACAGTAACTCTGCTCTGGCGGAGGTTATGGATCAGA
TTAACCACAGTTTTCCTGCATGCTCTC

>Zebrafish_Utrophin_Exon_73
CGTCGAGTGTGTCCTCAAGACCGAAG
```

Figure 4. Predicted zebrafish utrophin gene exon sequences.

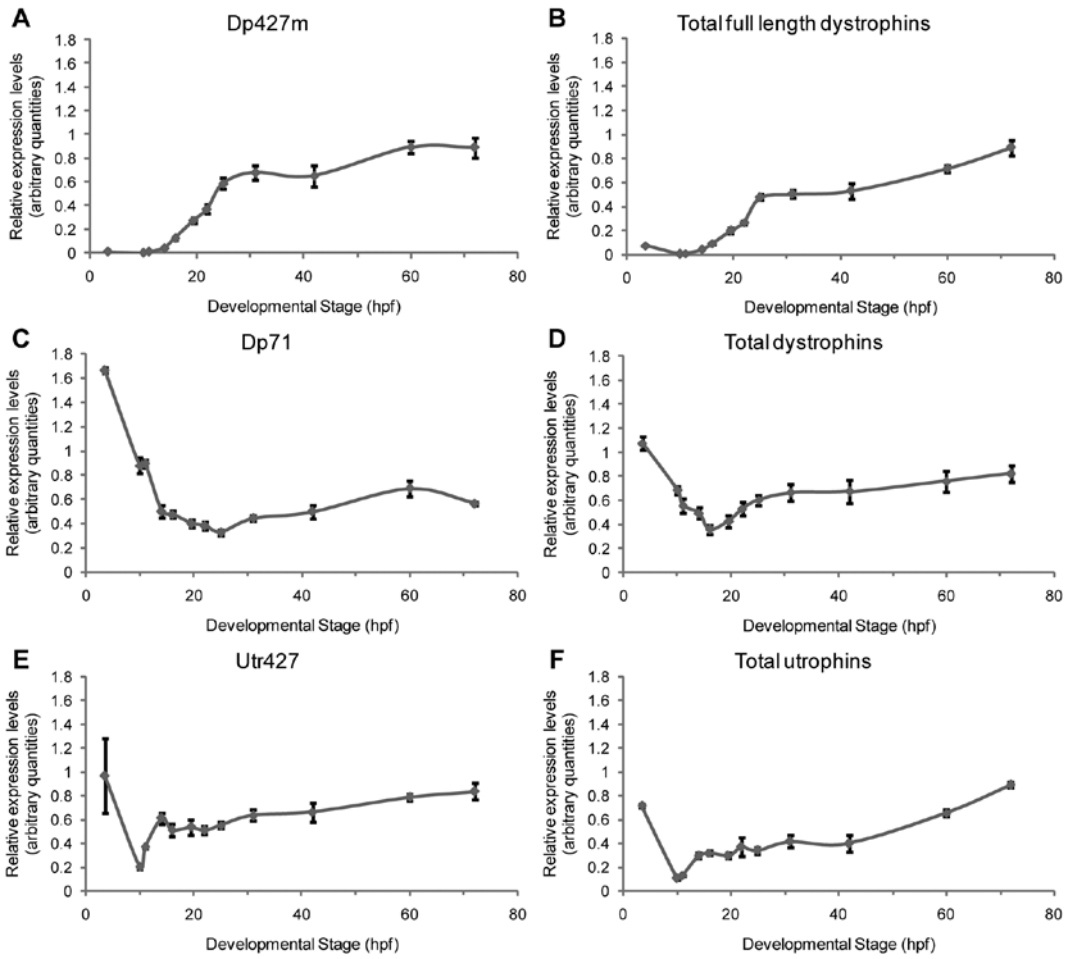

Figure 5. Developmental expression profiles of dystrophin and utrophin gene transcripts. The developmental stage in hpf is shown on the X-axis, and the relative expression levels (arbitrary units) are shown on the Y-axis. The latter have been normalised against the two reference genes *Rpl13a* and *EF1a*.

Discussion

The data presented here used an approach that allowed for the simultaneous expression profiling of the Dp427m transcript, all full-length dystrophin transcripts, the Dp71 transcript, and all transcripts expressed by the dystrophin gene. This approach involved targeting common exons that are expressed among isoforms, and targeting unique exons to allow for the profiling of specific isoforms. In addition, the profiles of all full length utrophin gene transcripts, as well as all transcripts expressed by the utrophin gene, were examined throughout zebrafish embryonic development. In terms of the former, it was considered critical to avoid exons involved in alternative splicing, which comprised exons 4 (46), 39, 41-42, 45-50 (47), 66-68 and 71-74 (24,25,48) and 78 (49). As such, the applicability of the approach described here requires knowledge of alternative splicing of target genes. The results reflect expression across the whole embryo as development progresses, and represent an average across all tissues and cell types.

The total expression profile of the dystrophin gene comprises that of Dp71 and the full length expression profiles, with Dp71 being the most abundant isoform present during the early stages of embryonic development, up until 16 hpf. From 16 hpf onwards, full length dystrophin expression appears to be present in higher abundance relative to other dystrophin isoforms. It should be noted that while other potential dystrophin transcripts were not investigated, they do not appear to be major players based on the expression profiles reported here. Experimental validation of this conclusion would be difficult in the context of qRT-PCR analysis as the sequences of the unique first exons of other dystrophin isoforms thought to exist are currently unknown or are too short to design transcript-specific primers.

In contrast to dystrophin, in which the predominantly expressed isoforms during embryonic development change from Dp71 to the full length dystrophin, the dominant isoform of utrophin that is expressed throughout development appears to be the full length Utr427. This conclusion does not rule out the presence of other as yet uncharacterised isoforms, but their expression levels would not be expected to be significant given our analysis herein.

The caveat to looking at gene expression of a whole embryo, as opposed to individual tissues, is that an absence of noticeable expression changes between two developmental stages does not imply the absence of differences between individual tissues. Conversely a small change in expression across the whole embryo may represent significant changes in a single tissue type. For example, a moderate increase in the levels of total full length dystrophins could reflect a significant increase in the expression of non-muscle isoforms.

Localised expression, not undertaken here, might be evident by RNA *in situ* analysis (24,25), but as suggested earlier, the design of appropriate probes suffers from sensitivity and specificity issues that are not easily resolved in the case of the two genes studied. With these effects in mind, the total expression profile should be interpreted with caution as it does not reflect expression in different tissues, which is driven by tissue-specific promoters. Ideally for genes with tissue-specific expression such as dystrophin, tissue-specific investigations would be preferred over whole embryo studies. However, attempting to isolate individual tissue types from a developing embryo would be technically challenging. In conclusion, the expression data are the first reported attempt to follow multiple dystrophin and utrophin gene transcripts simultaneously throughout vertebrate development. The approach offers opportunities to determine the impact of targeted dystrophin gene mutagenesis in the zebrafish on the expression profile of the dystrophin gene, and the possible upregulation of utrophin gene transcripts.

Acknowledgements

We thank the Genomics Unit, DNA Sequencing Facility at the School of Biological Sciences, University of Auckland for their sequencing services. We acknowledge the assistance of Dr Andrew Dodd and Dr Rongying Tang who provided the foundation for the research presented herein. We acknowledge the financial assistance of the Maurice and Phyllis Paykel Trust (http://www.paykeltrust.co.nz/) and the School of Biological Sciences, The University of Auckland (http://www.sbs.auckland.ac.nz/), for partial funding of the study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Tinsley JM, Blake DJ, Zuellig RA and Davies KE: Increasing complexity of the dystrophin-associated protein complex. Proc Natl Acad Sci USA 91: 8307-8313, 1994.
2. Lederfein D, Levy Z, Augier N, Mornet D, Morris G, Fuchs O, Yaffe D and Nudel U: A 71 kd protein is a major product of the Duchenne muscular dystrophy gene in brain and other non-muscle tissues. Proc Natl Acad Sci USA 89: 5346-5350, 1992.
3. Austin RC, Howard PL, D'Souza VN, Klamut HJ and Ray PN: Cloning and characterization of alternatively spliced isoforms of Dp71. Hum Mol Genet 4: 1475-1483, 1995.
4. Love DR, Hill DF, Dickson G, Spurr NK, Byth BC, Marsden RF, Walsh FS, Edwards YH and Davies KE: An autosomal transcript in skeletal muscle with homology to dystrophin. Nature 339: 55-58, 1989.
5. Tinsley JM, Blake DJ, Roche A, Fairbrother U, Riss J, Byth BC, Knight AE, Kendrick-Jones J, Suthers GK, Love DR, *et al*: Primary structure of dystrophin-related protein. Nature 360: 591-593, 1992.
6. Nudel U, Zuk D, Einat P, Zeelon E, Levy Z, Neuman S and Yaffe D: Duchenne muscular dystrophy gene product is not identical in muscle and brain. Nature 337: 76-78, 1989.
7. Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM and Kunkel LM: Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature 323: 646-650, 1986.
8. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener CA and Kunkel LM: Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50: 509-517, 1987.
9. Gorecki DC, Monaco AP, Derry JM, Walker AP, Barnard EA and Barnard PJ: Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. Hum Mol Genet 1: 505-510, 1992.
10. Holder E, Maeda M and Bies RD: Expression and regulation of the dystrophin Purkinje promoter in human skeletal muscle, heart, and brain. Hum Genet 97: 232-239, 1996.
11. D'Souza VN, Nguyen TM, Morris GE, Karges W, Pillers DA and Ray PN: A novel dystrophin isoform is required for normal retinal electrophysiology. Hum Mol Genet 4: 837-842, 1995.
12. Lidov HGW, Selig S and Kunkel LM: Dp140: a novel 140 kDA CNS transcript from the dystrophin locus. Hum Mol Genet 4: 329-335, 1995
13. Byers TJ, Lidov HGW and Kunkel LM: An alternative dystrophin transcript specific to peripheral nerve. Nat Genet 4: 77-81, 1993.
14. Burton EA, Tinsley JM, Holzfeind PJ, Rodrigues NR and Davies KE: A second promoter provides an alternative target for therapeutic upregulation of utrophin in Duchenne muscular dystrophy. Proc Natl Acad Sci USA 96: 14025-14030, 1999.
15. Zuellig RA, Bornhauser BC, Knuesel I, Heller F, Fritschy JM and Schaub MC: Identification and characterisation of transcript and protein of a new short N-terminal utrophin isoform. J Cell Biochem 77: 418-431, 2000.
16. Blake DJ, Schofield JN, Zuellig RA, Górecki DC, Phelps SR, Barnard EA, Edwards YH and Davies KE: G-utrophin, the autosomal homologue of dystrophin Dp116, is expressed in sensory ganglia and brain. Proc Natl Acad Sci USA 9: 3697-3701, 1995.
17. Wilson J, Putt W, Jimenez C and Edwards YH: Up71 and up140, two novel transcripts of utrophin that are homologues of short forms of dystrophin. Hum Mol Genet 8: 1271-1278, 1999.
18. Jimenez-Mallebrera C, Davies K, Putt W and Edwards YH: A study of short utrophin isoforms in mice deficient for full-length utrophin. Mamm Genome 14: 47-60, 2003.
19. Matsumura K, Ervasti JM, Ohlendieck K, Kahl SD and Campbell KP: Association of dystrophin-related protein with dystrophin-associated proteins in mdx mouse muscle. Nature 360: 588-591, 1992.
20. Cooper BJ, Winand NJ, Stedman H, Valentine BA, Hoffman EP, Kunkel LM, Scott MO, Fischbeck KH, Kornegay JN, Avery RJ, *et al*: The homologue of the Duchenne locus is defective in X-linked muscular dystrophy of dogs. Nature 334: 154-156, 1988.
21. Carpenter JL, Hoffman EP, Romanul FC, Kunkel LM, Rosales RK, Ma NS, Dasbach JJ, Rae JF, Moore FM, McAfee MB, *et al*: Feline muscular dystrophy with dystrophin deficiency. Am J Pathol 135: 909-919, 1989.
22. Hoffman EP, Brown RH Jr and Kunkel LM: Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 51: 919-928, 1987.
23. Pozzoli U, Elgar G, Cagliani R, Riva L, Comi GP, Bresolin N, Bardoni A and Sironi M: Comparative analysis of vertebrate dystrophin loci Indicate intron gigantism as a common feature. Genome Res 13: 764-772, 2003.
24. Bolaños-Jiménez F, Bordais A, Behra M, Strähle U, Mornet D, Sahel J and Rendón A: Molecular cloning and characterization of dystrophin and Dp71, two products of the Duchenne Muscular Dystrophy gene, in zebrafish. Gene 274: 217-226, 2001.
25. Bolaños-Jiménez F, Bordais A, Behra M, Strähle U, Sahel J and Rendón A: Dystrophin and Dp71, two products of the DMD gene, show a different pattern of expression during embryonic development in zebrafish. Mech Dev 102: 239-241, 2001.
26. Chambers SP, Dodd A, Overall R, Sirey T, Lam LT, Morris GE and Love DR: Dystrophin in adult zebrafish muscle. Biochem Biophys Res Commun 286: 478-483, 2001.
27. Kramarcy NR and Sealock R: Dystrophin as a focal adhesion protein. Collocalization with talin and the Mr 48,000 sarcolemmal protein in cultured Xenopus muscle. FEBS Lett 274: 171-174, 1990.
28. Wang J, Pansky A, Venuti JM, Yaffe D and Nudel U: A sea urchin gene encoding dystrophin-related proteins. Hum Mol Genet 7: 581-588, 1998.
29. Neuman S, Kaban A, Volk T, Yaffe D and Nudel U: The dystrophin/utrophin homologues in Drosophila and sea urchin. Gene 263: 17-29, 2001.
30. Vater R, Young C, Anderson LV, Lindsay S, Blake DJ, Davies KE, Zuellig R and Slater CR: Utrophin mRNA expression in muscle is not restricted to the neuromuscular junction. Mol Cell Neurosci 10: 229-242, 1998.
31. Urasawa N, Wada MR, Machida N, Yuasa K, Shimatsu Y, Wakao Y, Yuasa S, Sano T, Nonaka I, Nakamura A, *et al*: Selective vacuolar degeneration in dystrophin-deficient canine Purkinje fibers despite preservation of dystrophin-associated proteins with overexpression of Dp71. Circulation 117: 2437-2448, 2008.
32. Manzur AY, Kinali M and Muntoni F: Update on the management of Duchenne muscular dystrophy. Arch Dis Child 93: 986-990, 2008.

33. Sicinski P, Geng Y, Ryder-Cook AS, Barnard EA, Darlison MG and Barnard PJ: The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science 244: 1578-1580, 1989.
34. Bassett DI and Currie PD: The zebrafish as a model for muscular dystrophy and congenital myopathy. Hum Mol Genet 12: 265-270, 2003.
35. Bassett DI, Bryson-Richardson RJ, Daggett DF, Gautier P, Keenan DG and Currie PD: Dystrophin is required for the formation of stable muscle attachments in the zebrafish embryo. Development 130: 5851-5860, 2003.
36. Dodd A, Chambers SP and Love DR: Short interfering RNA-mediated gene targeting in the zebrafish. FEBS Lett 561: 89-93, 2004.
37. Dodd A, Curtis PM, Williams LC and Love DR: Zebrafish: bridging the gap between development and disease. Hum Mol Genet 9: 2443-2449, 2000.
38. Guyon JR, Mosley AN, Zhou Y, O'Brien KF, Sheng X, Chiang K, Davidson AJ, Volinski JM, Zon LI and Kunkel LM: The dystrophin associated protein complex in zebrafish. Hum Mol Genet 12: 601-615, 2003.
39. Lan CC, Tang R, Un San Leong I and Love DR: Quantitative real-time RT-PCR (qRT-PCR) of zebrafish transcripts: optimization of RNA extraction, quality control considerations, and data analysis. Cold Spring Harb Protoc 2009: pdb.prot5314, 2009.
40. Tang R, Dodd A, Lai D, McNabb WC and Love DR: Validation of zebrafish (*Danio rerio*) reference genes for quantitative real-time RT-PCR normalization. Acta Biochim Biophys Sin (Shanghai) 39: 384-390, 2007.
41. Ramakers C, Ruijter JM, Deprez RH and Moorman AF: Assumption-free analysis of quantitative real-time PCR data. Neurosci Lett 339: 62-66, 2003.
42. Ruijter JM, Ramakers C, Hoogaars WM, Karlen Y, Bakker O, Van den Hoff MJ and Moorman AF: Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 37: e45, 2009.
43. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A and Speleman F: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: Research0034, 2002.
44. Guyon JR, Goswami J, Jun SJ, Thorne M, Howell M, Pusack T, Kawahara G, Steffen LS, Galdzicki M and Kunkel LM: Genetic isolation and characterization of a splicing mutant of zebrafish dystrophin. Hum Mol Genet 18: 202-211, 2009.
45. Kane DA and Kimmel CB: The zebrafish midblastula transition. Development 119: 447-456, 1993.
46. Torelli S and Muntoni F: Alternative splicing of dystrophin exon 4 in normal human muscle. Hum Genet 97: 521-523, 1995.
47. Sironi M, Cagliani R, Pozzoli U, Bardoni A, Comi GP, Giorda R and Bresolin N: The dystrophin gene is alternatively spliced throughout its coding sequence. FEBS Lett 517: 163-166, 2002.
48. Lidov HG and Kunkel LM: Dp140: alternatively spliced isoforms in brain and kidney. Genomics 45: 132-139, 1997.
49. Ceccarini M, Rizzo G, Rosa G, Chelucci C, Macioce P and Petrucci TC: A splice variant of Dp71 lacking the syntrophin binding site is expressed in early stages of human neural development. Brain Res Dev Brain Res 103: 77-82, 1997.