An analysis of MyoD-dependent transcription using CRISPR/Cas9 gene targeting in *Xenopus tropicalis* embryos

Caitlin McQueen and Mary Elizabeth Pownall*
Biology Department, University of York, York, United Kingdom
*corresponding author email: betsy.pownall@york.ac.uk

Abstract
Myogenic regulatory factors (MRFs) are known to have essential roles in both the establishment and differentiation of the skeletal muscle cell lineage. MyoD is expressed early in the *Xenopus* mesoderm where it is present and active several hours before the activation of muscle differentiation genes. Previous studies in cultured cells and in *Xenopus* laevis have identified sets of genes that require MyoD prior to differentiation of skeletal muscle. Here we report results from experiments using CRISPR/Cas9 to target the *MyoD* gene in the diploid frog *Xenopus tropicalis*, that are analysed by RNA-seq at gastrula stages. We further investigate our data using cluster analysis to compare developmental expression profiles with that of *MyoD* and *α-cardiac actin*, reference genes for skeletal muscle determination and differentiation. Our findings provide an assessment of using founder (F0) *Xenopus* embryos from CRISPR/Cas9 protocols for transcriptomic analyses and we conclude that although targeted F0 embryos are genetically mosaic for MyoD, there is significant disruption in the expression of a specific set of genes. We discuss candidate target genes in context of their role in the sub-programs of MyoD regulated transcription.

1. Introduction

The MyoD family of bHLH transcription factors are well studied developmental regulators that provide a paradigm for how cell lineages are determined during embryogenesis (Tapscott, 2005). The expression of these key transcriptional regulators is activated in specific cells in the developing embryo by local signals (Emerson, 1993), and effectors downstream of signal transduction pathways act on discrete enhancers present in MRF genes (Carvajal and Rigby, 2010). The activation of any one these master regulators is sufficient to drive skeletal muscle differentiation in many types of cultured cells (Tapscott et al., 1988; Weintraub et al., 1989) and this dominant activity is part due to the ability of these transcription factors to auto- and cross-activate each other's expression (Thayer et al., 1989; Weintraub et al., 1989).

During myogenesis, proliferative myoblasts exit the cell cycle and fuse to form multinucleated myotubes; at the same time the transcription of contractile protein genes is co-ordinately activated. MyoD and Myf5 have overlapping roles that are required for the early commitment of cells to the myogenic lineage, while MRF4 and Myogenin play later roles in myoblast fusion and differentiation (Kaul et al., 2000). Using cell culture based models of myogenesis, together with chromatin immunoprecipitation and transcriptomics, several sub-programmes of gene expression downstream of MyoD have been described and the transcriptional activity of MyoD in myoblasts prior to the activation of contractile protein genes during differentiation has been established (Bergstrom et al., 2002; Cao et al., 2010; Gianakopoulos et al., 2011; Soleimani et al., 2013). The outcome of these studies has been a detailed understanding of how the MyoD transcription factor targets genes for transcription, including its surprising widespread binding throughout the genome, preference for a specific E-box sequence, ability to modify chromatin and interact with other transcriptional regulators (Blum et al., 2012; Conerly et al., 2016; Fong et al., 2015).

In vivo, skeletal muscle cells are derived from the mesoderm and in amniotes the MRFs are first expressed in cells of the early somite (Pownall et al., 2002). However, in frogs and fish the MRFs are expressed in the mesoderm prior to somitogenesis where MyoD and Myf5 genes are transcribed during gastrula stages (Hopwood et al., 1989; Weinberg et al., 1996) and in *Xenopus*, the MyoD protein has been localised to the nascent mesoderm by immunohistochemistry (Hopwood et al., 1992). *Xenopus laevis* MyoD has been shown to be required for the early expression of a set of genes during gastrulation, prior to any activation of contractile gene expression (Maguire et al., 2012). This study took advantage of the time window during *Xenopus* development where the embryo expresses MyoD protein, but myogenic differentiation has not yet begun. Isolating a similar timepoint in amniote embryos would be difficult due to the relatively few myoblasts in the early somite expressing MyoD/Myf5, and the rapid rostral to caudal progression of myogenic differentiation. Therefore, the frog embryo provides a vertebrate model to investigate the requirement for MyoD transcriptional activity prior to myogenic differentiation *in vivo*.

CRISPR/Cas9 protocols are particularly effective in *Xenopus* (Liu et al., 2016; Shigeta et al., 2016): synthetic guide RNAs (gRNAs) are transcribed *in vitro* and co-injected with Cas9 protein directly into to the one-cell embryo, circumnavigating any delays from transcription and translation inherent to plasmid based protocols used in other systems. Here we used CRISPR/Cas9 to target the MyoD gene in the diploid frog *Xenopus tropicalis* and analysed gene expression at gastrula stages by RNAseq. We show that founder embryos targeted for MyoD, while genetically mosaic, show specific disruption in the expression of genes previously identified as MyoD targets. We further classify sets of genes identified in our RNAseq analysis according to their expression profile during normal development. This study provides an insight into the earliest genetic targets of MyoD *in vivo*, as well as an assessment of founder embryos derived from genetic targeting using CRISPR/Cas9 protocols and their usefulness in transcriptomic analyses.

2. Results

2.1 Gene targeting X. tropicalis MyoD using CRISPR/Cas9

Detecting genetic disruption of MyoD in embryos

Xenopus tropicalis is a diploid frog and, as such, genetic methods are simplified using this model. We targeted *X. tropicalis MyoD* (Fisher 2003) using CRISPR/Cas9 in order to identify genes that require MyoD for their expression in the early mesoderm, prior to myogenic differentiation. A synthetic guide RNA (gRNA) was designed against a sequence in exon 1 coding for the amino terminal part of the bHLH domain, and injected this together with Cas9 protein directly into 1- to 4-cell embryos. In order to test mutagenesis, single embryos were collected for sequencing analysis. Non-homologous end-joining (NHEJ) that repairs DNA after cleavage by Cas9 results in random insertions or deletions (INDELs), therefore genomic DNA was extracted from each embryo and the targeted region of the *MyoD* gene was amplified by PCR and cloned such that different individual mutations could be identified. A total of 35 embryos were collected and between 3- 15 clones were sequenced from each individual. Of these 35 embryos, 31 contained at least one mutated sequence (targeting of 88.6%; Figure 1A), indicating a high efficiency of gene targeting. As expected, each embryo differed in the proportion of mutant sequences, some returning all mutated sequences, and others showing only 50% mutant sequences.

Characterising alleles

A detailed analysis of the nature of mutations produced by Cas9 targeting, each sequence mutation was characterised to determine whether it was an insertion, a deletion or a point mutation. The average targeting efficiency throughout sequenced embryos shows that 78% of returned sequences were mutated (Figure 1B), of which the majority were deletions (77%) rather than insertions or point mutations (Figure 1C). We found that the level of mosaicism within a single embryo was high, and Figure 1D shows 10 sequences from a single embryo aligned to the predicted wild type *MyoD* sequence. As predicted, all mutation events occur and the near the protospacer-adjacent motif (PAM) where NHEJ results in many different alleles (Shigeta 2016). Figure 1E depicts the proportion of the alleles that code for frameshift mutations as a result of either a deletion or an insertion. We found that only a small proportion of sequences (2.3%) represent in-frame insertions. Frameshift mutations represent the majority of mutated sequences identified, however, this equates to less than half of all sequences returned (43.7%). This highlights a caveat when using F0 embryos for genetic analyses; although CRISPR/Cas9 targeting results in a very high proportion of mutated alleles in an individual embryo, in this case, less than half of these mutations will result in a truncated protein or a genetic null.

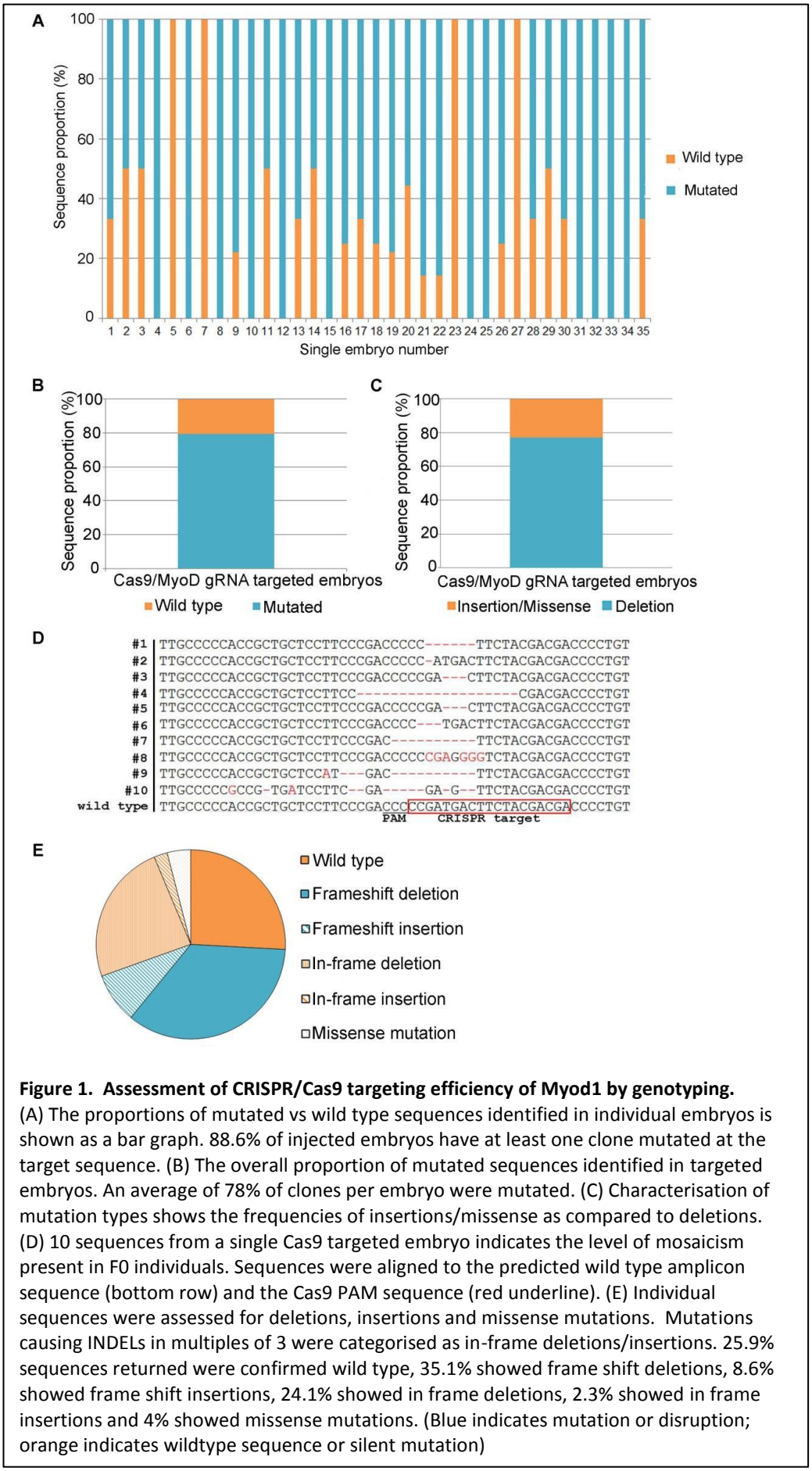

Figure 1. Assessment of CRISPR/Cas9 targeting efficiency of Myod1 by genotyping.
(A) The proportions of mutated vs wild type sequences identified in individual embryos is shown as a bar graph. 88.6% of injected embryos have at least one clone mutated at the target sequence. (B) The overall proportion of mutated sequences identified in targeted embryos. An average of 78% of clones per embryo were mutated. (C) Characterisation of mutation types shows the frequencies of insertions/missense as compared to deletions. (D) 10 sequences from a single Cas9 targeted embryo indicates the level of mosaicism present in F0 individuals. Sequences were aligned to the predicted wild type amplicon sequence (bottom row) and the Cas9 PAM sequence (red underline). (E) Individual sequences were assessed for deletions, insertions and missense mutations. Mutations causing INDELs in multiples of 3 were categorised as in-frame deletions/insertions. 25.9% sequences returned were confirmed wild type, 35.1% showed frame shift deletions, 8.6% showed frame shift insertions, 24.1% showed in frame deletions, 2.3% showed in frame insertions and 4% showed missense mutations. (Blue indicates mutation or disruption; orange indicates wildtype sequence or silent mutation)

2.2 Analysis of transcripts in MyoD-targeted embryos

Disruption of MyoD gene transcription and MyoD activity

To further characterise embryos targeted by CRISPR/Cas9, we extracted mRNA from groups of ten embryos at NF stage 11.5 and used qPCR to analyse the expression of *MyoD* and its known target gene *Rbm24* (*Seb4*) (Li et al., 2010). We found that there is a significant decrease in *Myod* expression (P< .01) and *Rbm24* (P<.05) when embryos injected with MyoD gRNA + Cas9 protein are compared to those injected with embryos injected with the same amount of Cas9 protein alone. The results were calculated as relative proportions of expression and repeated for three biological replicates; we found that the relative expression of *Myod* in targeted embryos compared with controls is reduced to 0.57 and *Rbm24* is reduced to 0.77 (Figure 2A).

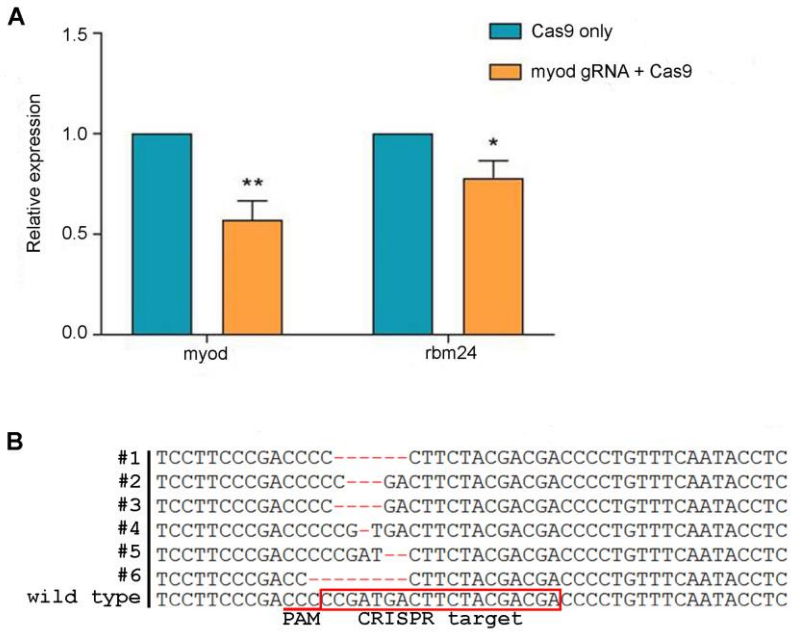

Figure 2. Analysis of transcripts in embryos target for MyoD using CRISPR/Cas9
Experimental and control embryos were collected at NF stage 11.5. (A) qPCR analysis shows the expression of *MyoD* and the known MyoD target gene *Rbm24.* Pair-wise t-tests for the mean relative expression for Cas9 only and Cas9 plus gRNA injected sets for each gene. Error bars represent SEM, * = p<0.05, ** = p<0.01. (B) MyoD sequences returned from mapping raw RNA-Seq reads to the *Xenopus tropicalis myod1* gene. A total of 1200 reads were extrapolated across the three biological replicates and the proportion of reads showing each mutation type was calculated. The mutated sequences are shown aligned to the sequence for wild type MyoD. The total number of reads for each of the mutations shown is as follows: #1= 175 (15%), #2= 13.5 (1%), #3= 219 (18%), #4= 70 (6%), #5= 69.5 (6%), #6= 40 (3%).

Identifying genes that require MyoD using RNA-seq analysis

Three biological repeats for *X.tropicalis* experimental embryos (targeted for *MyoD*) and sibling controls (Cas9-only injected) were collected at NF stage 11.5 and mRNA was extracted for RNA-seq. cDNA libraries were prepared and Illumina deep sequencing resulted in 440 million reads across the 6 samples. RNA-Seq reads were mapped using the *Xenopus tropicalis* genome version 9.0 (Xenbase.org) and FPKM values were established for all genes. Transcripts that align to *MyoD* were analysed for INDELs, and Figure 2B shows that a significant proportion of the reads have deletions in the expected target site adjacent to the PAM. Insertions are less likely to be detected as they would fail

to align with reference genome.

To produce an overview of the significance of fold changes observed in CRISPR targeted samples, a volcano plot for (log_2) fold change vs ($–log_{10}$) paired t-test P-value was constructed using Python script (Figure 3). Each individual point represents a gene and the dotted line represents a p-value of <0.05. Points in red indicate genes with a fold change of less than 1; that is, where the average FPKM value of experimental samples have not doubled or halved compared to that of the control. Blue points represent genes with a fold change greater than 1 but a P-value of >0.05, so not statistically significant. Yellow points represent genes with a fold change greater than 1 and a P-value of <0.05. The majority of points show a fold change of less than 1 and P-values of >0.05, indicating no significant change in gene expression at NF stage 11.5 in response to CRISPR/Cas9 targeting of MyoD. However, 1165 genes mapped to the *X.tropicalis* genome display significant change and are further analysed in Sections 2.3.

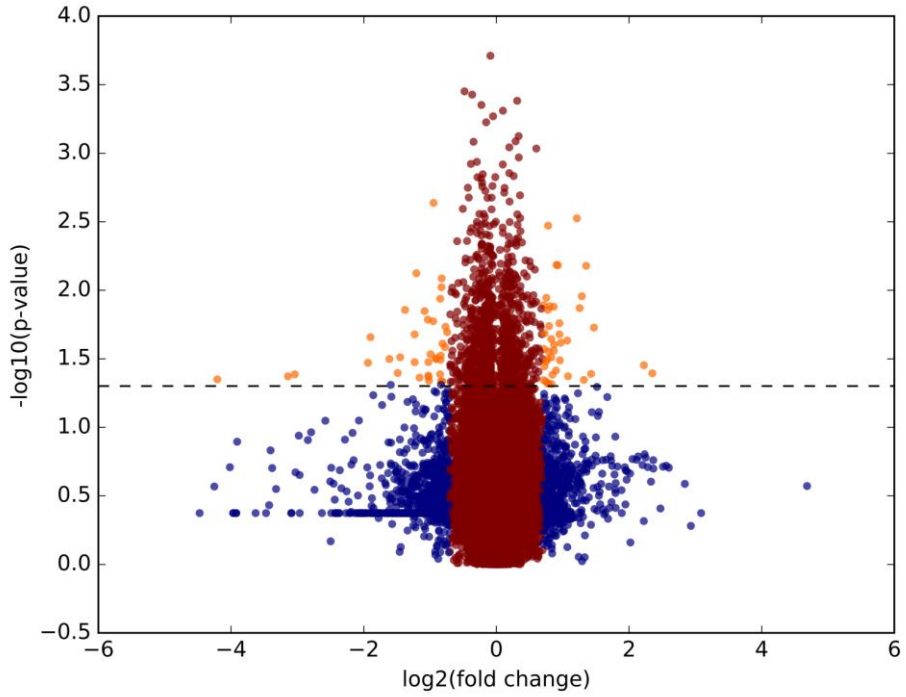

Figure 3. Overview of results from RNA-Seq
RNA-Seq reads were mapped using the *Xenopus tropicalis* genome version 9.0 (Xenbase). FPKM (fragments per kilobase of transcript per million mapped reads) values were calculated to avoid bias towards longer genes by normalising the number of reads per fragment to the length. FPKM values for three biological replicates were analysed by pairwise t-tests comparing expression in Cas9 only control embryos and MyoD CRISPR-targeted samples. A volcano plot showing t test significance value (-log10 p-value) vs fold change (log2) was constructed in Python. Genes in blue indicate a fold change of greater than 1, genes in yellow indicate a fold change greater than 1 and a p value of <0.05.

2.3 Computational analysis of early genetic targets of MyoD

Of the 1165 genes found to be significantly altered in the absence of MyoD, some showed very low expression levels. Therefore, a minimum expression threshold of 5.0 FPKM average for the control samples was applied. In addition, as MyoD expression in targeted samples showed a fold change of 0.71, therefore genes with fold changes in this same range (between 0.68 and 0.91) were selected for further analysis. During our manual curation, we included two genes that fell just outside the criteria cutoff: *FoxC1* (0.84; P<.058) and *Pbx2* (0.93; P<.02). This resulted in a short list of 100 potential target genes (Supplemental Data, Table S1).

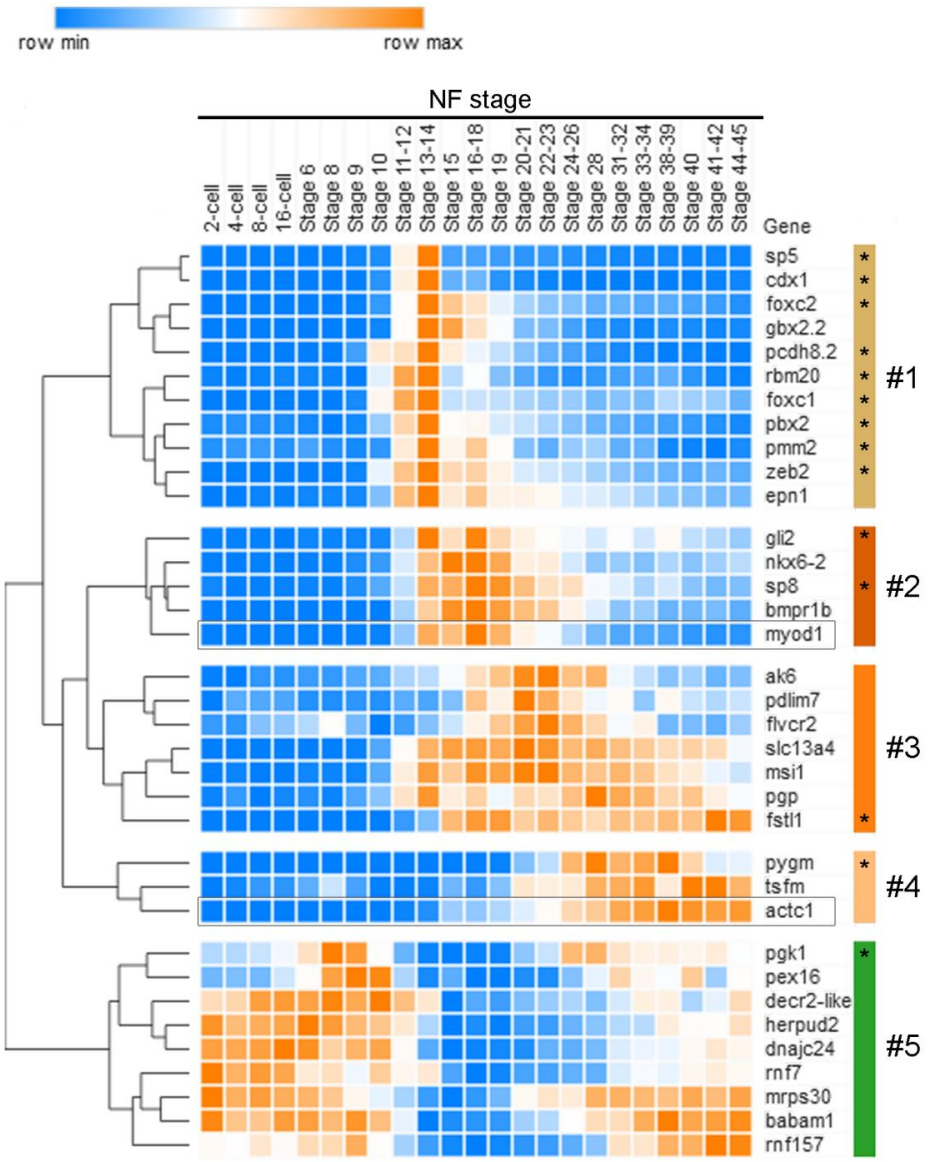

Figure 4. Hierarchical clustering of Myod target genes.
Genes identified as significantly downregulated in targeted samples were further analysed using expression data from (Tan et al., 2013) to create a heatmap. Relative expression is shown as a scale of low (blue) to high (orange). Euclidean distance was used as the metric for hierarchical clustering of complete samples, which resulted in 5 clusters showing distinct expression profiles. (*) indicates genes which have known or predicted roles in myogenic or pre-myogenic cells. The expression profiles for *Myod* and *α-actin* (*actc1*) are boxed and shown to highlight clusters relevant to determination and differentiation respectively.

Temporal expression analysis of potential target genes

To further investigate whether the identified genes are expressed at a time consistent with activation by MyoD, temporal expression profiles were analysed. As MyoD protein is first detected in the mesoderm at NF stage 11, candidate genes with expression prior to these stages are less likely to be bonafide target genes of MyoD. Furthermore, as our analysis was carried out at Stage 11.5, genes coding for contractile proteins or other differentiation specific genes are not expected to be identified by our study, however, we include the expression profile of *actc1* as a reference for this class of genes. RNA-Seq data from a development time course of *Xenopus tropicalis* is available (Tan et al., 2013) and these expression profiles were used for hierarchical cluster analysis of target genes. To do this, expression data was extracted for the 100 short-listed genes (Supplemental data Table S1) and used to create a heatmap of expression levels over a developmental time course (Supplemental data Figure S1). Genes aligning with reference profiles for determination (*myoD*) and differentiation (*actc1*) were selected for further heat mapping and cluster analyses (shown in Figures 4 and 5). Euchlidean distance was used as the metric of linkage for complete samples and a selected cutting point for the clustered dendrogram resulted in the formation of 5 clusters of distinct expression profiles; this is shown in the heat map where orange boxes represent highest expression levels (Figure 4). To determine the expression patterns observed within the 5 clusters, profiles of relative expression for each stage were constructed from the expression data used in the heatmap and clustering.

Clusters 1 and 2 show similar overall expression profiles: genes within these clusters have very low or no maternal expression with earliest notable expression at stage 10. Expression in both clusters increases during gastrula and neurula stages, however, in Cluster 1 expression increase is more rapid, as highest expression is observed at stages 13-14, whilst genes in Cluster 2 show peak expression at stages 16-18. Both clusters then show decreases in gene expression in later stages. Notably, *MyoD* itself is allocated to Cluster 2. Genes located in Cluster 3 also show low or no expression prior to mid-blastula transition (MBT) and the overall expression trend shows increasing expression until early tailbud stages 20-22. Expression then decreases in later stages. Individual gene expression within this cluster however, is more varied than in other clusters. Cluster 4 is a much smaller cluster containing the known MyoD target *alpha-cardiac actin* (*actc1*). Gene expression for this cluster shows delayed gene activation with increases occurring from stage 14 onwards, this increasing expression is maintained through tailbud stages and only decreases slightly in the later tadpole stages. Cluster 4 contains the muscle glycogen phosphorylase *pygm.* Cluster 5 is distinct from all other clusters in that the genes located within this cluster show maternal expression. Expression decreases rapidly after MBT and is at lowest levels during neurula stages, then increases again during tailbud stages (20-28) through to later tadpole stages (31-45).

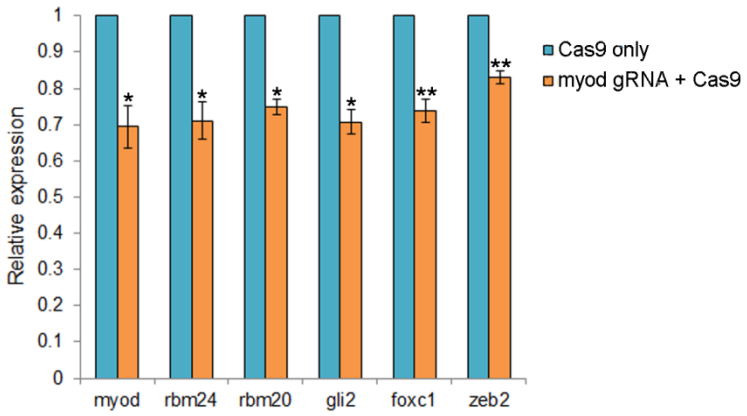

Figure 5

qRT-PCR analysis of identified early targets of MyoD at NF St11.5. Analysis shows the expression of MyoD and the known MyoD target gene Rbm24 alongside predicted target genes Rbm20, Gli2, Foxc1 and Zeb2. Pair-wise t-tests were carried out for the mean relative expression of three biological replicates for Cas9 only and Cas9 plus gRNA injected sets for each gene. Error bars represent SEM, * = $p<0.05$, ** = $p<0.01$.

Hierarchical cluster analysis was used to highlight genes that show developmentally relevant expression profiles, and strengthen their status as candidates for early genetic targets of MyoD; clusters 1 and 2 include genes identified in other studies as myogenic or pre-myogenic genes (highlighted by an asterix and discussed in Section 3). In order to validate whether any of these genes require MyoD for their expression during gastrula stages, we injected gRNA targeting MyoD together with Cas9 protein and directly assayed the expression of several candidate genes in these embryos at stage 11.5 as compared to Cas9 only injected embryos using qPCR (Figure 5). Figure 5 shows that *Rbm20*, *Rbm24*, *gli2*, *FoxC1*, and *Zeb2* (aka *XSip1*), as well as *myoD* itself, are all significantly down-regulated in targeted embryos. This validation supports the notion that our gene targeting and transcriptomic analysis of founder embryos has provided a robust list of candidates genes regulated by MyoD prior to the onset of skeletal muscle differentiation.

Gene symbol	ENSEMBL ID	Average FPKM Control	Average FPKM Experimental	Experimental Relative expression	p-value
bmpr1b	ENSXETG00000019220	8.01	5.45	0.68	0.02
pgp	ENSXETG00000016097	6.12	4.52	0.74	0.03
gbx2.2	ENSXETG00000003293	42.10	31.57	0.75	0.01
sp8	ENSXETG00000030115	12.27	9.28	0.76	0.05
nkx6-2	ENSXETG00000023614	20.42	15.51	0.76	0.02
tsfm	ENSXETG00000009653	5.08	3.96	0.78	0.05
decr2-like	ENSXETG00000010329	12.93	10.08	0.78	0.04
rbm20	ENSXETG00000025245	7.04	5.51	0.78	0.04
zeb2	ENSXETG00000000237	18.67	14.88	0.80	0.01
foxc2	ENSXETG00000016387	80.68	65.42	0.81	0.03
babam1	ENSXETG00000025571	9.14	7.47	0.82	0.02
gli2	ENSXETG00000011189	12.23	10.04	0.82	0.01
sp5	ENSXETG00000025407	64.04	53.11	0.83	0.03
pmm2	ENSXETG00000004549	45.85	38.21	0.83	0.04
pygm	ENSXETG00000034136	123.57	103.71	0.84	0.04
foxc1	ENSXETG00000000594	73.17	61.58	0.84	0.06
fstl1	ENSXETG00000018009	25.29	21.31	0.84	0.02
pex16	ENSXETG00000001027	8.44	7.15	0.85	0.05
slc13a4	ENSXETG00000008163	18.80	15.99	0.85	0.03
ak6	ENSXETG00000018174	14.39	12.24	0.85	0.03
pgk1	ENSXETG00000007447	19.96	17.15	0.86	0.02
mrps30	ENSXETG00000017716	11.89	10.23	0.86	0.04
flvcr2	ENSXETG00000027282	8.63	7.47	0.87	0.04
cdx1	ENSXETG00000010282	77.18	66.87	0.87	0.03
pdlim7	ENSXETG00000007240	13.77	11.97	0.87	0.01
msi1	ENSXETG00000012216	55.14	47.91	0.87	0.00
rnf7	ENSXETG00000014753	99.16	86.80	0.88	0.04
pcdh8.2	ENSXETG00000008792	73.31	64.29	0.88	0.01
herpud2	ENSXETG00000013111	5.58	4.91	0.88	0.04
rnf157	ENSXETG00000019548	5.75	5.16	0.90	0.05
epn1	ENSXETG00000022662	42.43	38.19	0.90	0.04
dnajc24	ENSXETG00000008179	14.14	12.76	0.90	0.03
pbx2	ENSXETG00000005223	169.56	158.17	0.93	0.02

Table 1: A shortlist of genes identified as early targets of MyoD. After heat-mapping and cluster analysis, only genes located within clusters showing developmentally relevant expression profiles were shortlisted as early Myod targets. The list was manually curated using existing available expression profiles and published literature to curate the 33 genes shortlisted as early targets of Myod.

3. Discussion

MyoD is known to direct several different sub-programmes of gene expression during myogenesis (Bergstrom et al., 2002; Blais et al., 2005; Soleimani et al., 2012) consistent with its role as an essential determination gene for the proliferative myoblast (Rudnicki et al., 1993). However, MyoD is also a robust initiator of transcriptional targets during myogenic differentiation, distinguishing itself in this way from Myf5 (Conerly et al., 2016). It is an interesting proposition that one transcription factor can activate distinct panels of genes at two different stages of cell lineage specification. Indeed, this notion of promoter swapping is supported by MyoD binding analysis using chromatin immunoprecipitation (ChIP-seq) protocols that have shown that MyoD functions as a transcriptional regulator during both myogenic determination and differentiation by binding and activating distinct sets of genes (Soleimani et al., 2012).

The ability to model myogenic differentiation in cell culture has been exceptionally informative; pointing the way to a detailed understanding of how transcriptional regulators can direct cell fate. Our study investigated genes that require MyoD during the earliest stages of myogenesis *in vivo*, using the diploid frog model *X.tropicalis*. CRISPR/Cas9 gene editing very effectively targeted the MyoD gene in embryos, however not all INDELs result in alleles that would generate a disrupted protein. Approximately 80% of injected embryos are successfully targeted and the penetrance of mutation in each individual is also very high. However, because Cas9 can act on one or both (or neither) alleles in cells as the early embryo divides, and the nature of NHEJ is leads to random INDELs, the resulting F0 embryos are inherently genetically mosaic. This leads to a population of F0s with ill-defined genotype, with less than half of alleles analyses carrying a disruptive mutation. In zebrafish, it is standard practice to outcross founder fish and breed to a known mutant genotype (Li et al., 2016); outcrossing frogs is not as practical as it requires more space and time.

Nevertheless, we have established here that using founder embryos from gene editing protocols in transcriptional analyses is valuable. We show here that CRISPR/Cas9 targeting of MyoD results in a significant reduction of MyoD transcripts overall and a high percentage of these with INDELs (Figure 2); moreover, the known target gene *Rmb24*(*Seb4*) is significantly down regulated in these samples. Our transcriptomic analysis has provided a shortlist of genes that require MyoD, *in vivo*, prior to myogenic differentiation (Table 1).

Our analysis of putative MyoD targets in the context of a published time course of gene expression during *Xenopus tropicalis* development (Tan et al., 2013) provided a way of curating genes on the basis of temporal expression, however spatial restriction of expression is also an important factor to consider. *MyoD* and *Rbm24* (aka *Seb4*) share a very close expression pattern, both temporally and spatially, with the notable exception that *Rbm24* is expressed in the cardiac as well as the skeletal muscle cell lineage (Li et al., 2010; Maguire et al., 2012). The spatial expression pattern of *Rbm20* has not been examined in *Xenopus*, however, in chick embryos it shows very early (yet transient) expression in somites with persistent expression the heart (Geisha.arizona.edu). *Gli2* and *Zeb2* (aka *XSip1*) are expressed in the neurectoderm just after gastrulation (Aguero 2012 and Papin 2002); however, during gastrulation Zeb2 is expressed in the dorsal marginal zone with some mesodermal expression; there is no *in situ* hybridization data for gastrula specific expression of *Gli2*. *FoxC1* shows both early mesodermal and later somitic expression in *Xenopus*, and like *Rbm24, FoxC1* was also identified as a target of MyoD in an earlier alaysis using morpholino oligos (Maguire et al., 2012).

We found that *FoxC1* and *FoxC2* are significant targets in our analyses, and are known to be expressed in the paraxial mesoderm amniotes (Kume et al., 1998), as well as fish and frogs (Köster et al., 1998; Maguire et al., 2012; Topczewska et al., 2001). FoxC1/C2 are essential for somitogenesis (Kume et al., 2001; Topczewska et al., 2001), and identified as transcriptional targets of MyoD in previous studies (Gianakopoulos et al., 2011; Maguire et al., 2012). In the early paraxial mesoderm, *FoxC1/C2* are co-expressed with the early muscle regulator *Pax3*, however later in somitogenesis *FoxC1/2* regulate the endothelial lineage (Lagha et al., 2009; Mayeuf-Louchart et al., 2014) but their expression is nonetheless essential for the normal migration of muscle precursor cells to the limb (Mayeuf-Louchart

et al., 2016). Interestingly, another gene we identified in our screen, *Gli2*, has been found to act upstream of *FoxC1*/2 in the induction of myogenesis in P19 cells (Savage et al., 2010).

Pbx2 is also an interesting target as this family of TALE-class homeodomain proteins are associated with myogenesis (Berkes et al., 2004; Maves et al., 2007) and binding sites for Pbx transcription factors are found in regions of the genome associated with MyoD binding (Fong et al., 2015); it is thought that Pbx proteins help 'pioneer' or establish the myogenic programme (Yao et al., 2013). This model fits well with *Pbx* genes being early targets of MyoD *in vivo*. Rbm24 (Seb4) is an RNA binding protein essential for skeletal muscle specific alternative splicing (Yang et al., 2014) and has been shown previously to be a direct target of MyoD (Li et al., 2010). Recently, it has been found that Rbm24 is essential for normal somitogenesis in fish (Maragh et al., 2014) consistent with the findings in our previous publication in *Xenopus* where Rbm24 was identified as a MyoD target gene (Maguire et al., 2012). *Rbm20*, a related gene with similar functions in directing cell specific alternative splicing (Li et al., 2013), was also identified as a target in our analyses. *Zeb2* codes for an E-box binding repressor, which could act like Snail repressors in modulating 'enhancer swapping', where MyoD binds to regulatory sequences in different genes in myoblasts as compared with myotubes (Soleimani et al., 2013). Genes coding for the zinc finger transcription factors Sp8 and Sp5 were also identified as targets in a screen for early targets of Wnt signalling (Nakamura et al., 2016) and as downstream regulators promoting FGF signalling (Branney et al., 2009; Kasberg et al., 2013). Identifying the genes coding for Sp5/8 as early MyoD targets is consistent with the important role for FGF (Fisher, 2002) and Wnt (Hoppler et al., 1996) signalling in activating MyoD in *Xenopus*. The fact that we finding some regulators that are known to act with MyoD (such as Pbx and Sp5/8) as downstream targets of MyoD is not surprising as our analyses are focus on a window of time very early during the specification of the myogenic lineage when transcriptional feed-forward pathways are being established. More surprising is that we were able to detect later targets of MyoD at this early stage: the skeletal muscle specific protocadherins (*pcdh8*) and kinases (*pgk1* and *pygm*), and follistatin (*fstl1*) are expressed in somites (Berti et al., 2015) and are notable as they are identified as MyoD targets at such an early timepoint.

Considering the statistically significant hits identified in our study in the context of their normal developmental expression patterns and with regard to other findings in the published literature, we describe here a set of genes (Table 1) that require MyoD *in vivo*, early during the establishment of the skeletal muscle cell lineage in *Xenopus tropicalis.*

4. Experimental Procedures

4.1 Gene targeting in *Xenopus tropicalis*

gRNA design

The basic domain of MRFs functions as a DNA binding domain, binding at E-box consensus sequences in regulatory regions downstream genes to activate transcription. Therefore, a synthetic guide RNA sequence (gRNA) was designed to disrupt a region within the Xenopus tropicalis MyoD basic domain, located in exon 1 of the coding sequence. The design tool ChopChop (https://chopchop.rc.fas.harvard.edu/) was used to scan the input sequence for suitable Cas9 target sequences including a PAM site. Any off-targets with up to 2 mismatches in the first 20bp of sequence were searched for using the *Xenopus tropicalis* genome version (xenTro3/GCA_000004195.1).

gRNA template synthesis

A sequence spanning 71-89bp in the first exon of the XtMyoD coding sequence (5'-TCGTCGTAGAAGTCATCGG-3') on the reverse strand was selected as no off-targets were predicted for this sequence. We designed the 5' primer to include an increased efficiency promoter for transcription by T7 RNA

Polymerase, and an added 5' G nucleotide to fit requirements of the T7 polymerase. Our resulting forward primer consisted of the sequence
5'- GCAGCTAATACGACTCACTATAGG TCGTCGTAGAAGTCATCGG GTTTTAGAGCTAGAAATA-3'
and the reverse primer is common to all gRNAs is
5'AAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC 3'
(Nakayama et al., 2014).

gRNA transcription
Phusion polymerase was used to amplify the template using 5uM of each primer, annealing at 60C and extending for 15 seconds over 35 cycles. Template was taken directly from the PCR reaction for in vitro transcription using Megashortscript® T7 Transcription Kit (Life Technologies) following the manufacturers guidelines. Incubation at 37˚c overnight was followed by TURBO DNase treatment, and gRNA purification by phenol-chloroform extraction and NH_4OAc/ethanol precipitation. gRNA was resuspended in a final volume of 20ul. RNA quality assessment by gel electrophoresis and nanodrop measurement was also carried out. An optimal concentration of 1.8ug/ul is desired for Cas9 coinjection mixtures.

Cas9 protein production
Cas9 protein was made using the plasmid (Addgene) by expressing in Rosetta-2 cells (Novagen) at 30C in kanamycin/chloramphenicol substituted autoinduction media (Studier F.W., Protein
Expression and Purification 2005). Pelletted cells were resuspended in lysis buffer (buffer A (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10mM imidazole) with cOmplete EDTA-free protease inhibitor cocktail (Roche) and lysed by sonication. GE Healthcare HiTrap Nickel NTA column was equilibrated in 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole) and washed before a gradient of Buffer B (Buffer A + 500 mM Imidazole) was applied to elute the protein. The peak eluted between 7- 30% buffer B. The fractions containing Cas9 were pooled and concentrated before size-exclusion chromatography in 20 mM Tris, 200 mM KCl, 10 mM MgCl 2 (Superdex 200 16/60; Amersham Pharmacia Biotech). The purified Cas9 was concentrated to 50 mg/ml using ultrafiltration in Amicon centrifugation filter units (Millipore). Aliqouts were flash-frozen and stored at −80 °C.

4.2 Analysis of X.tropicalis embryos
Genotyping embryos:
To assess the efficiency of Cas9 targeting, genomic DNA was extracted from single embryos and the target region was amplified by PCR, cloned into pGEM T-easy and sequenced. Single embryos at NF Stage 25 were transferred to 0.5ml PCR tubes containing 200ul of lysis buffer (50mM Tris pH7.0, 50mM NaCl, 5mM EDTA, 0.5% SDS, 10% Chelex, fresh 250ug/ml Proteinase K) and incubated at 55˚c for 1hour followed by 95˚c for 15minutes to deactivate the Proteinase K. After centrifugation, the DNA is diluted 1:10 and amplified by PCR reaction using primers flanking the Cas9 target site (forward: TTA CTT TGC GCC GTT GCT AT and reverse: GTT GCG CAA AAT CTC CAC TT). PCR products were cloned into the pGEM®-T Easy vector system as per manufacturer's guidelines and transformed into *E.coli*. Minipreps of individual clones sequenced by GATC Biotech using an SP6 primer.
Sequences from 3-10 clones from each embryo were aligned alongside the wild type amplicon sequence using DNAStar SeqMan to identify INDELs.

Expression analysis of targeted embryos
To assess the level of MyoD transcripts and that of its known target Rmb24 (Seb4) in CRISPR/Cas9 targeted embryos prior to RNA-Seq analysis, control and experimental embryos were collected at NF stage 11.5 and snap frozen on dry ice. RNA was extracted using Sigma TRI Reagent® together with Zymo RNA Clean & Concentrator™-5 column, and purified as per the manufacturer's instructions. For qPCR, cDNA was synthesised using Thermo Fisher Scientific

Superscript IV Reverse Transcriptase according to manufactures guidelines. Primers for MyoD, Rmb24, and normalisation gene Dicer were used with Fast SYBR Green 2X master mix in triplicate for three biological repeats.

RNAseq

mRNA libraries were prepared for sequencing using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina, using the NEBNext Poly(A) mRNA Magnetic Isolation Module to isolate poly(A) mRNA from total RNA. HiSeq3000 2 x 150 bp paired end sequencing was performed by the University of Leeds Next Generation Sequencing Facility.

4.3 Computational analysis of RNAseq data

Illumina deep sequencing resulted in ~440million reads across the 6 samples. Raw RNA-Seq reads were mapped using the *Xenopus tropicalis* genome version 9.0 (Xenbase). FPKM (fragments per kilobase of transcript per million mapped reads) values were calculated to normalise the number of reads per fragment to the length of the fragment in order to avoid bias towards longer fragments. FPKM values for three biological replicates were analysed by pairwise t-tests comparing expression in control and MyoD CRISPR-targeted samples. 655 genes showed differential expression based on paired t-tests ($p<0.05$). Expression data for target genes was extracted from RNA-seq data available (Tan et al 2013) and uploaded to https://software.broadinstitute.org/morpheus/ creating a heatmap of expression and hierarchical clustering.

Acknowledgments

We are very grateful to Olga Moroz in the York Structural Biology labs for protein production of Cas9. We also thank the Genomics lab in University of York Biology Department technology facility especially Sally James and Katherine Newling for technical and bioinformatic support. CM is funded by a White Rose BBSRC DTP studentship (BB/M011151/1).

References

Bergstrom, D.A., Penn, B.H., Strand, A., Perry, R.L., Rudnicki, M.A., and Tapscott, S.J. (2002). Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. Mol Cell *9*, 587-600.

Berkes, C.A., Bergstrom, D.A., Penn, B.H., Seaver, K.J., Knoepfler, P.S., and Tapscott, S.J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol Cell *14*, 465-477.

Berti, F., Nogueira, J.M., Wohrle, S., Sobreira, D.R., Hawrot, K., and Dietrich, S. (2015). Time course and side-by-side analysis of mesodermal, pre-myogenic, myogenic and differentiated cell markers in the chicken model for skeletal muscle formation. J Anat *227*, 361-382.

Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., and Dynlacht, B.D. (2005). An initial blueprint for myogenic differentiation. Genes Dev *19*, 553-569.

Blum, R., Vethantham, V., Bowman, C., Rudnicki, M., and Dynlacht, B.D. (2012). Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. Genes Dev *26*, 2763-2779.

Branney, P.A., Faas, L., Steane, S.E., Pownall, M.E., and Isaacs, H.V. (2009). Characterisation of the fibroblast growth factor dependent transcriptome in early development. PLoS ONE *4*, e4951.

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., MacQuarrie, K.L., Davison, J., Morgan, M.T., Ruzzo, W.L*., et al.* (2010). Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Dev Cell *18*, 662-674.

Carvajal, J.J., and Rigby, P.W. (2010). Regulation of gene expression in vertebrate skeletal muscle. Exp Cell Res *316*, 3014-3018.

Conerly, M.L., Yao, Z., Zhong, J.W., Groudine, M., and Tapscott, S.J. (2016). Distinct Activities of Myf5 and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and Differentiation. Dev Cell *36*, 375-385.
Emerson, C.P., Jr. (1993). Embryonic signals for skeletal myogenesis: arriving at the beginning. Curr Opin Cell Biol *5*, 1057-1064.
Fisher, M.E., Isaacs, H.V. and Pownall, M.E. (2002). eFGF is required for the activation of XmyoD in the myogenic cell lineage of Xenopus laevis. Development *129*, 1307-1315.
Fong, A.P., Yao, Z., Zhong, J.W., Johnson, N.M., Farr, G.H., 3rd, Maves, L., and Tapscott, S.J. (2015). Conversion of MyoD to a neurogenic factor: binding site specificity determines lineage. Cell reports *10*, 1937-1946.
Gianakopoulos, P.J., Mehta, V., Voronova, A., Cao, Y., Yao, Z., Coutu, J., Wang, X., Waddington, M.S., Tapscott, S.J., and Skerjanc, I.S. (2011). MyoD directly up-regulates premyogenic mesoderm factors during induction of skeletal myogenesis in stem cells. J Biol Chem *286*, 2517-2525.
Hoppler, S., Brown, J.D., and Moon, R.T. (1996). Expression of a dominant-negative Wnt blocks induction of MyoD in Xenopus embryos. Genes Dev *10*, 2805-2817.
Hopwood, N.D., Pluck, A., and Gurdon, J.B. (1989). MyoD expression in the forming somites is an early response to mesoderm induction in Xenopus embryos. EMBO J *8*, 3409-3417.
Hopwood, N.D., Pluck, A., Gurdon, J.B., and Dilworth, S.M. (1992). Expression of XMyoD protein in early Xenopus laevis embryos. Development *114*, 31-38.
Kasberg, A.D., Brunskill, E.W., and Steven Potter, S. (2013). SP8 regulates signaling centers during craniofacial development. Dev Biol *381*, 312-323.
Kaul, A., Koster, M., Neuhaus, H., and Braun, T. (2000). Myf5 revisited: Early loss of myotome does not lead to a rib phenotype in homozygous myf5 mutant mice. Cell *102*, 17 -19.
Köster, M., Dillinger, K., and Knöchel, W. (1998). Expression pattern of the winged helix factor XFD-11 during Xenopus embryogenesis. Mech Dev *76*, 169-173.
Kume, T., Deng, K.Y., Winfrey, V., Gould, D.B., Walter, M.A., and Hogan, B.L. (1998). The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. Cell *93*, 985-996.
Kume, T., Jiang, H., Topczewska, J.M., and Hogan, B.L. (2001). The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. Genes Dev *15*, 2470-2482.
Lagha, M., Brunelli, S., Messina, G., Cumano, A., Kume, T., Relaix, F., and Buckingham, M.E. (2009). Pax3:Foxc2 reciprocal repression in the somite modulates muscular versus vascular cell fate choice in multipotent progenitors. Dev Cell *17*, 892-899.
Li, H.Y., Bourdelas, A., Carron, C., and Shi, D.L. (2010). The RNA-binding protein Seb4/RBM24 is a direct target of MyoD and is required for myogenesis during Xenopus early development. Mech Dev.
Li, M., Zhao, L., Page-McCaw, P.S., and Chen, W. (2016). Zebrafish Genome Engineering Using the CRISPR-Cas9 System. Trends Genet *32*, 815-827.
Li, S., Guo, W., Dewey, C.N., and Greaser, M.L. (2013). Rbm20 regulates titin alternative splicing as a splicing repressor. Nucleic Acids Res *41*, 2659-2672.
Liu, Z., Cheng, T.T., Shi, Z., Liu, Z., Lei, Y., Wang, C., Shi, W., Chen, X., Qi, X., Cai, D*., et al.* (2016). Efficient genome editing of genes involved in neural crest development using the CRISPR/Cas9 system in Xenopus embryos. Cell Biosci *6*, 22.
Maguire, R.J., Isaacs, H.V., and Pownall, M.E. (2012). Early transcriptional targets of MyoD link myogenesis and somitogenesis. Dev Biol *371*, 256-268.
Maragh, S., Miller, R.A., Bessling, S.L., Wang, G., Hook, P.W., and McCallion, A.S. (2014). Rbm24a and Rbm24b are required for normal somitogenesis. PLoS ONE *9*, e105460.
Maves, L., Waskiewicz, A.J., Paul, B., Cao, Y., Tyler, A., Moens, C.B., and Tapscott, S.J. (2007). Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. Development *134*, 3371-3382.
Mayeuf-Louchart, A., Lagha, M., Danckaert, A., Rocancourt, D., Relaix, F., Vincent, S.D., and Buckingham, M. (2014).

Notch regulation of myogenic versus endothelial fates of cells that migrate from the somite to the limb. Proc Natl Acad Sci U S A *111*, 8844-8849.
Mayeuf-Louchart, A., Montarras, D., Bodin, C., Kume, T., Vincent, S.D., and Buckingham, M. (2016). Endothelial cell specification in the somite is compromised in Pax3-positive progenitors of Foxc1/2 conditional mutants, with loss of forelimb myogenesis. Development *143*, 872-879.
Nakamura, Y., de Paiva Alves, E., Veenstra, G.J., and Hoppler, S. (2016). Tissue- and stage-specific Wnt target gene expression is controlled subsequent to beta-catenin recruitment to cis-regulatory modules. Development *143*, 1914-1925.
Pownall, M., Gustaffson, M., and Emerson, C. (2002). Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. Annual Review of Cell and Developmental Biology *18*, 747-783.
Rudnicki, M.A., Schnegelsberg, P.N., Stead, R.H., Braun, T., Arnold, H.H., and Jaenisch, R. (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. Cell *75*, 1351-1359.
Savage, J., Voronova, A., Mehta, V., Sendi-Mukasa, F., and Skerjanc, I.S. (2010). Canonical Wnt signaling regulates Foxc1/2 expression in P19 cells. Differentiation *79*, 31-40.
Shigeta, M., Sakane, Y., Iida, M., Suzuki, M., Kashiwagi, K., Kashiwagi, A., Fujii, S., Yamamoto, T., and Suzuki, K.T. (2016). Rapid and efficient analysis of gene function using CRISPR-Cas9 in Xenopus tropicalis founders. Genes Cells *21*, 755-771.
Soleimani, V.D., Palidwor, G.A., Ramachandran, P., Perkins, T.J., and Rudnicki, M.A. (2013). Chromatin tandem affinity purification sequencing. Nat Protoc *8*, 1525-1534.
Soleimani, V.D., Yin, H., Jahani-Asl, A., Ming, H., Kockx, C.E., van Ijcken, W.F., Grosveld, F., and Rudnicki, M.A. (2012). Snail regulates MyoD binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. Mol Cell *47*, 457-468.
Tan, M.H., Au, K.F., Yablonovitch, A.L., Wills, A.E., Chuang, J., Baker, J.C., Wong, W.H., and Li, J.B. (2013). RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. Genome Res *23*, 201-216.
Tapscott, S.J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development *132*, 2685-2695.
Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H., and Lassar, A.B. (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science *242*, 405-411.
Thayer, M.J., Tapscott, S.J., Davis, R.L., Wright, W.E., Lassar, A.B., and Weintraub, H. (1989). Positive autoregulation of the myogenic determination gene MyoD1. Cell *58*, 241-248.
Topczewska, J.M., Topczewski, J., Shostak, A., Kume, T., Solnica-Krezel, L., and Hogan, B.L. (2001). The winged helix transcription factor Foxc1a is essential for somitogenesis in zebrafish. Genes Dev *15*, 2483-2493.
Weinberg, E.S., Allende, M.L., Kelly, C.S., Abdelhamid, A., Murakami, T., Andermann, P., Doerre, O.G., Grunwald, D.J., and Riggleman, B. (1996). Developmental regulation of zebrafish MyoD in wild-type, no tail and spadetail embryos. Development *122*, 271-280.
Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc Natl Acad Sci U S A *86*, 5434-5438.
Yang, J., Hung, L.H., Licht, T., Kostin, S., Looso, M., Khrameeva, E., Bindereif, A., Schneider, A., and Braun, T. (2014). RBM24 is a major regulator of muscle-specific alternative splicing. Dev Cell *31*, 87-99.
Yao, Z., Farr, G.H., 3rd, Tapscott, S.J., and Maves, L. (2013). Pbx and Prdm1a transcription factors differentially regulate subsets of the fast skeletal muscle program in zebrafish. Biol Open *2*, 546-555.

Figure Legends

Figure 1. Assessment of CRISPR/Cas9 targeting efficiency of Myod1 by genotyping.

Embryos at the one-cell stage were co-injected with 1ng Cas9 protein and 300pg of MyoD gRNA. At NF stage 25 genomic DNA was extracted and a 432bp region including the predicted CRISPR target site was amplified by PCR and cloned into pGEM T-Easy. 3-15 clones per embryo were sequenced and a total of 153 sequences were analysed. (A) The proportions of mutated vs wild type sequences identified in individual embryos is shown as a bar graph. 88.6% of injected embryos have at least one clone mutated at the target sequence. (B) The overall proportion of mutated sequences identified in targeted embryos. An average of 78% of clones per embryo were mutated. (C) Characterisation of mutation types shows the frequencies of insertions/missense as compared to deletions. (D) 10 sequences from a single Cas9 targeted embryo indicates the level of mosaicism present in F0 individuals. Sequences were aligned to the predicted wild type amplicon sequence (bottom row) and the Cas9 PAM sequence is indicated in red underline. (E) Individual sequences were assessed for deletions, insertions and missense mutations. Mutations causing INDELs in multiples of 3 were categorised as in-frame deletions/insertions. 25.9% sequences returned were confirmed wild type, 35.1% showed frame shift deletions, 8.6% showed frame shift insertions, 24.1% showed in frame deletions, 2.3% showed in frame insertions and 4% showed missense mutations. (Blue indicates mutation or disruption; orange indicates wildtype sequence or silent mutation)

Figure 2. Analysis of transcripts in CRISPR/Cas9 targeted embryos.

Experimental and control embryos were collected at NF stage 11.5. (A) qPCR analysis shows the expression of *Myod* and the known Myod target gene *Rbm24.* Pair-wise t-tests for the mean relative expression for Cas9 only and Cas9 plus gRNA injected sets for each gene. Error bars represent SEM, * = $p<0.05$, ** = $p<0.01$. (B) Myod RNA sequences returned from mapping raw RNA-Seq reads to the *Xenopus tropicalis myod1* gene. A total of 1200 reads were extrapolated across the three biological replicates and proportions of reads showing each mutation type were calculated. The mutated sequences are shown aligned to the sequence for wild type MyoD. The total proportions for each of the mutations shown are as follows: D1= 175 (15%), D2= 13.5 (1%), D3= 219 (18%), D4= 70 (6%), D6= 69.5 (6%), D8= 40 (3%).

Figure 3. Overview of results from RNA-Seq

RNA-Seq reads were mapped using the Xenopus tropicalis genome version 9.0 (Xenbase). FPKM (fragments per kilobase of transcript per million mapped reads) values were calculated to avoid bias towards longer genes by normalising the number of reads per fragment to the length. FPKM values for three biological replicates were analysed by pairwise t-tests comparing expression in Cas9 only control embryos and MyoD CRISPR-targeted samples. A volcano plot showing t test significance value (-log10 p-value) vs fold change (log2) was constructed in Python. Genes in blue indicate a fold change of greater than 1, genes in yellow indicate a fold change greater than 1 and a p value of <0.05.

Figure 4. Hierarchical clustering of Myod target genes.

Genes identified as significantly downregulated in targeted samples were further analysed using expression data from (Tan et al., 2013) to create a heatmap. Relative expression is shown as a scale of low (blue) to high (orange). Euclidean distance was used as the metric for hierarchical clustering of complete samples, which resulted in 5 clusters showing distinct expression profiles. (*) indicates genes which have known or predicted roles in myogenic or pre-myogenic cells. *Myod* and *α-actin* (*actc1*) were included in the analysis to highlight relevant clusters.

Figure 5. MyoD target gene cluster analysis.

Expression data from (Tan et al., 2013) was used in a time course analysis of whole embryonic development. Target genes with developmentally relevant expression profiles were identified from the initial heatmap. After hierarchical

clustering, profiles of relative gene expression over time for each cluster reveals distinct expression profiles between clusters. Genes in clusters 1 and 2 show similar expression profile to MyoD are better candidates for target genes. Mean relative expression is shown for each cluster along with expression profiles for each gene. (A-E) represent clusters 1-5 respectively.

Supplemental Data

Gene symbol	ENSEMBL ID	mean control	mean experimental	fold change	p-value
bmpr1b	ENSXETG00000019220	8.01	5.45	0.68	0.02
myod1	ENSXETG00000001320	90.10	64.29	0.71	0.02
ndufaf3	ENSXETG00000024755	15.31	11.02	0.72	0.03
tmem141	ENSXETG00000003949	8.53	6.25	0.73	0.04
pgp	ENSXETG00000016097	6.12	4.52	0.74	0.03
trdmt1	ENSXETG00000027671	5.01	3.72	0.74	0.04
rassf8	ENSXETG00000023490	11.13	8.34	0.75	0.04
gbx2.2	ENSXETG00000003293	42.10	31.57	0.75	0.01
nodal		6.12	4.62	0.76	0.03
sp8	ENSXETG00000030115	12.27	9.28	0.76	0.05
ndufb4		16.41	12.42	0.76	0.01
nkx6-2	ENSXETG00000023614	20.42	15.51	0.76	0.02
tsfm	ENSXETG00000009653	5.08	3.96	0.78	0.05
decr2-like	ENSXETG00000010329	12.93	10.08	0.78	0.04
rbm20	ENSXETG00000025245	7.04	5.51	0.78	0.04
zeb2	ENSXETG00000000237	18.67	14.88	0.80	0.01
mespa		9.87	7.87	0.80	0.05
znf638	ENSXETG00000015780	6.74	5.38	0.80	0.02
sod1	ENSXETG00000007350	28.27	22.64	0.80	0.05
foxc2	ENSXETG00000016387	80.68	65.42	0.81	0.03
babam1	ENSXETG00000025571	9.14	7.47	0.82	0.02
gli2	ENSXETG00000011189	12.23	10.04	0.82	0.01
arl10		9.70	7.98	0.82	0.01
ccne2	ENSXETG00000006660	72.36	59.78	0.83	0.02
mrpl15	ENSXETG00000008103	12.68	10.49	0.83	0.05
sp5	ENSXETG00000025407	64.04	53.11	0.83	0.03
lrrn1-like.1		6.57	5.46	0.83	0.05
mthfs	ENSXETG00000024293	17.77	14.80	0.83	0.01
pmm2	ENSXETG00000004549	45.85	38.21	0.83	0.04
c4orf32		26.81	22.39	0.83	0.04
endog	ENSXETG00000025614	9.73	8.14	0.84	0.04
pygm	ENSXETG00000034136	123.57	103.71	0.84	0.04

foxc1	ENSXETG00000000594	73.17	61.58	0.84	0.06
sowahc		5.89	4.96	0.84	0.01
fstl1	ENSXETG00000018009	25.29	21.31	0.84	0.02
sema4c	ENSXETG00000001251	22.06	18.66	0.85	0.04
pex16	ENSXETG00000001027	8.44	7.15	0.85	0.05
fahd1-like		8.00	6.79	0.85	0.05
ephb1	ENSXETG00000013722	7.89	6.71	0.85	0.03
slc13a4	ENSXETG00000008163	18.80	15.99	0.85	0.03
AK6	ENSXETG00000018174	14.39	12.24	0.85	0.03
rbm24	ENSXETG00000024618	30.00	25.56	0.85	0.03
tmem18	ENSXETG00000027985	42.69	36.39	0.85	0.01
hunk	ENSXETG00000007352	41.66	35.63	0.86	0.00
cdr2	ENSXETG00000016662	6.64	5.69	0.86	0.01
dctpp1	ENSXETG00000018900	12.09	10.36	0.86	0.03
pgk1	ENSXETG00000007447	19.96	17.15	0.86	0.02
mrps30	ENSXETG00000017716	11.89	10.23	0.86	0.04
dgcr14	ENSXETG00000022387	6.43	5.54	0.86	0.02
jam3		5.05	4.36	0.86	0.05
stx18	ENSXETG00000016051	14.61	12.60	0.86	0.02
cables2	ENSXETG00000002013	24.42	21.07	0.86	0.00
zbtb8a.1		5.58	4.82	0.86	0.04
cinp	ENSXETG00000010255	7.06	6.10	0.86	0.03
bicd2-like.1		38.26	33.10	0.87	0.04
flvcr2	ENSXETG00000027282	8.63	7.47	0.87	0.04
cdx1	ENSXETG00000010282	77.18	66.87	0.87	0.03
vps25	ENSXETG00000024599	15.21	13.19	0.87	0.05
mib1	ENSXETG00000003146	14.11	12.25	0.87	0.03
sept6-like		12.70	11.03	0.87	0.02
pdlim7	ENSXETG00000007240	13.77	11.97	0.87	0.01
msi1	ENSXETG00000012216	55.14	47.91	0.87	0.00
znf219	ENSXETG00000016157	19.87	17.28	0.87	0.05
rnf7	ENSXETG00000014753	99.16	86.80	0.88	0.04
rint1	ENSXETG00000023226	9.88	8.66	0.88	0.02
hmg20b	ENSXETG00000022092	12.48	10.94	0.88	0.04
pcdh8.2	ENSXETG00000008792	73.31	64.29	0.88	0.01
lrpap1	ENSXETG00000005500	11.86	10.41	0.88	0.02
ext1	ENSXETG00000019136	47.42	41.62	0.88	0.03
wipf2-like		24.34	21.36	0.88	0.02
nr6a1	ENSXETG00000008578	71.16	62.60	0.88	0.05
herpud2	ENSXETG00000013111	5.58	4.91	0.88	0.04
plekhg4		10.65	9.42	0.88	0.04
cenpf	ENSXETG00000023124	34.80	30.79	0.89	0.00
rnf34	ENSXETG00000020965	7.57	6.71	0.89	0.04
dhx32-like		17.74	15.73	0.89	0.05

bri3bp	ENSXETG00000001207	16.21	14.38	0.89	0.01
ggnbp2		24.95	22.16	0.89	0.05
ell2	ENSXETG00000013296	18.61	16.58	0.89	0.05
vdac3		17.56	15.66	0.89	0.05
ephb3	ENSXETG00000017293	42.52	37.95	0.89	0.02
rhog-like.1		14.00	12.52	0.89	0.05
pttg1ip.2		5.42	4.86	0.90	0.05
rfc2	ENSXETG00000018234	11.70	10.49	0.90	0.01
rhoa.2	ENSXETG00000009241	445.54	399.57	0.90	0.02
rnf157	ENSXETG00000019548	5.75	5.16	0.90	0.05
prr13-like		65.55	58.85	0.90	0.04
mybl2	ENSXETG00000012125	51.14	45.92	0.90	0.01
gigyf1	ENSXETG00000018415	23.73	21.32	0.90	0.03
ilvbl		6.74	6.07	0.90	0.05
epn1	ENSXETG00000022662	42.43	38.19	0.90	0.04
nacc1	ENSXETG00000005594	44.47	40.05	0.90	0.03
irf6.2	ENSXETG00000018661	13.55	12.22	0.90	0.02
rhpn2	ENSXETG00000018271	9.65	8.71	0.90	0.02
fmnl3-like		15.30	13.80	0.90	0.02
slc30a1		26.69	24.09	0.90	0.04
dnajc24	ENSXETG00000008179	14.14	12.76	0.90	0.03
usp28	ENSXETG00000022958	10.14	9.16	0.90	0.03
nsd1-like		7.39	6.68	0.90	0.02
cdc42se2-like.1		6.09	5.51	0.90	0.01
pbx2	ENSXETG00000005223	169.56	158.17	0.93	0.02
actc1	ENSXETG00000012911	-	-	-	-

Table S1: After paired t-test analysis, 1165 genes were identified as showing significant differences in expression in MyoD targeted samples as compared to Cas9 only injected controls. Of these genes, 100 were shortlisted as potential target genes of Myod showing a mean fold change of 0.68-0.91 and an average control FPKM of >5. In addition, *foxc1* and *pbx2* were manually added to the list despite although just outside of the criteria cutoffs due to their known interactions with Myod. The differentiation gene *actc1* is a known target of MyoD and, while not identified in our screen, was included in our cluster analysis as a reference gene.

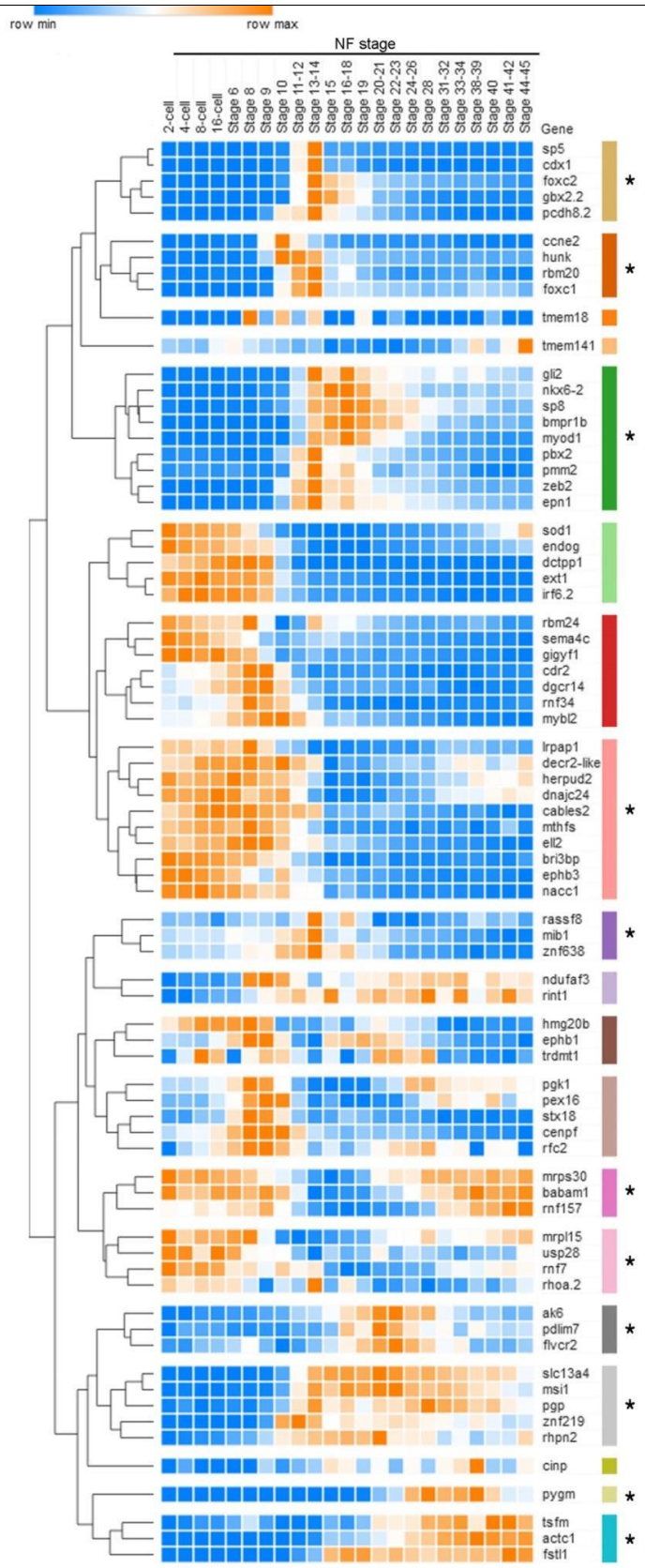

Figure S1: Hierarchical clustering using Euclidean distance of developmental expression profiles of 73 target genes shortlisted from the initial RNA-Seq results filtering. Genes included are those which had ENSEMBL IDs and were included in the dataset from (Tan et al., 2013). Myod1 and actc1 were also included as reference genes. Clusters were manually assessed to shortlist developmentally relevant clusters for further analysis (*).