1 Supplemental experimental procedures

1.1 Measurement of inter-individual variability in gene expression

Source code and data files are available at:
`http://www.igh.cnrs.fr/equip/Seitz/Pinzon_et_al_2016_data_and_code/Figure_1.tar.bz2`.

Ten adult (3 month old) male, pathogen-free S/SPF C57BL/6J mice (from Charles River Laboratories; accustomed to our mouse facility for 9 days before the experiment) were sacrificed according to local regulations. Approval for these studies was obtained from the Ethics Committee on Animal Research of the Languedoc-Roussillon region (CE-LR-0505). 800 to 900 μL blood was collected from each mouse and heparin-treated. Five blood samples were analyzed separately ("biological replicates") and the other five samples were pooled, then split into five "technical replicates". All ten samples were then treated identically in a double-blind manner.

Neutrophil RNA was isolated using the anti-mouse Ly6G PE antibody (BD Biosciences cat. # 551461), anti-PE magnetic beads (Miltenyi Biotec cat. # 130-048-801), MS columns (Miltenyi Biotec cat. # 130-042-201) and the QIAamp RNA extraction kit (Qiagen cat. #52304). Neutrophil purity (ranging from 71.5 to 99.5 %) was verified by flow cytometry. Neutrophil lysis and RNA extraction was performed on a QIAcube robot, for better sample-to-sample reproducibility. RNA was precipitated, DNase-treated, further purified using the RNeasy kit (Qiagen cat. # 74104), and handed to the IRB microarray facility (Montpellier, France).

cDNAs were labeled using the Affymetrix 3´ IVT labeling kit with the two-cycle cDNA synthesis protocol, and hybridized on MG-430 PM array strips (Affymetrix cat. # 901570). Hybridization was performed at 45°C and washes were performed between 37.07 and 38.00°C, on an Affymetrix GeneChip station. Image acquisition was performed on a 3000 7G Affymetrix scanner. Intensity values were background-subtracted and normalized using the RMA method. Out of 196 predicted miR-223 targets according to TargetScan mouse v. 6.2, four are not probed by the array (genes 5031414D18Rik, Cdh12, Zfp839 and Fignl2). Baek et al. (2008)'s analysis, as well as ours, was thus restricted to the remaining 192 genes.

For each probeset on the array, we measured technical variability for probesets with similar signal intensities (*i.e.*: the 5 percentiles whose mean intensity is lower than the mean intensity of the probeset of interest, and the 5 percentiles whose mean intensity is higher). For each of the 5 technical replicates of each of these probesets, we computed the ratio: (replicate intensity)/(mean intensity across all 5 technical replicates). These ratios follow approximately a normal distribution (see middle panel of figure 1**B** in the main text for an example).

For each probeset, the probability of each possible underlying biological value for each individual mouse was inferred from such normal distribution (see right panel of figure 1**B** in the main text for an example). The p-values shown in figure 1**D** and in table S1 measure the probability that the median of the miR-223-guided fold-change in gene expression (taken from NCBI's GEO accession #GSE12001, described in Baek et al. (2008)) exceeds the measured inter-individual variability in gene expression. They were estimated by randomly picking 100,000 values in the distribution of possible underlying biological values for the most highly and lowly expressing mice, according to their estimated probability density. The p-value was defined as the frequency of random pairs whose difference is smaller than the median miR-223-guided fold-change. Classification into variable genes ($p < 0.05$) and tightly-regulated genes ($p \geqslant 0.05$) was based on the median p-value across probesets.

Using the Loess normalization method instead of RMA, even more genes appear to have a large inter-individual variability: with a p-value cutoff of 0.01, only 6 genes (*Bai3*, *Fbxo8*, *Srp54a*, *Srp54b*, *Tgfbr3* and *Ube2q2*) had a median p-value above the cutoff (meaning that their inter-individual variability appears to be smaller than miR-223-guided repression).

Because neutrophil purity is not identical across individual mouse samples (see Supplemental Table S1), it is formally possible that contaminating cells contribute variable amounts of mRNA from genes that are mostly expressed in some non-neutrophil cells. Such contamination, which is variable across samples, could generate arbitrarily high artifactual variability. In order to control for this confounding effect, we scanned the probesets on the array to identify which sample exhibits the highest expression level. If indeed variable amounts of contaminating cells generate artifactual variability for non-neutrophil-specific genes, then one would expect the least pure neutrophil preparations to exhibit anomalously high expression levels for these non-neutrophil-specific genes. We thus scanned all 45,141 probesets on the microarray and recorded which of the 5 individual mouse samples exhibits the highest expression level (see Supplemental Table S1): probe level repartition does not seem to depend on neutrophil purity, ruling out the possibility that our observed inter-individual variability is due to variable contamination.

1.2 Comparison of dose-sensitivity predictors with miRNA binding site conservation

Source code and data files are available at:
`http://www.igh.cnrs.fr/equip/Seitz/Pinzon_et_al_2016_data_and_code/Figure_2.tar.bz2`.

The list of known human haplo-insufficient genes was taken from Dang et al. (2008). TargetScan 7.0's aggregate P_{CT} (probability of conserved targeting) scores were downloaded from `http://www.targetscan.org/vert_70/vert_70_data_download/Conserved_Family_Info.txt.zip` and `http://www.targetscan.org/vert_70/vert_70_data_download/Nonconserved_Family_Info.txt.zip` (see Friedman et al. (2009) for the description of the aggregate P_{CT}). For each gene, we considered the miRNA family with the highest aggregate P_{CT} (*i.e.*: miRNAs with the most conserved interaction to that gene). Every known haplo-insufficient gene described by Dang et al. (2008) is a predicted target for human miRNAs according to TargetScan 7.0. Among TargetScan-predicted miRNA targets, the remaining 19,066 genes are not known as haplo-insufficient according to Dang et al. (2008): they were used as a control in Figure 2**A**.

The probability of human genes for being haplo-insufficient was calculated by Huang et al. (2010) for 12,218 genes. 9,520 of these exhibit conserved miRNA binding sites according to TargetScan 7.0. For each gene, we considered the miRNA family with the highest aggregate P_{CT}.

1.3 Absolute quantification of miRNAs and their targets during C2C12 cell differentiation

Source code and data files are available at:
`http://www.igh.cnrs.fr/equip/Seitz/Pinzon_et_al_2016_data_and_code/Figure_3.tar.bz2` and:
`http://www.igh.cnrs.fr/equip/Seitz/Pinzon_et_al_2016_data_and_code/Figure_4.tar.bz2`

C2C12 cells were ordered freshly from ATCC (cat. #CRL-1772) then always kept sub-confluent. Differentiation was induced as described in Sweetman et al. (2008) once cells reached confluency.

1.3.1 RIP-based target identification

Cells were transfected with 20 nM antisense oligonucleotides using Lipofectamine 2000 (Invitrogen) 51 hours after differentiation induction (3 replicates per transfection; seven confluent 150 mm diameter Petri dishes per replicate). 24 hours after antisense oligonucleotide transfection, cells were rinsed with cold PBS then cross-linked at 0 °C with 150 mJ.cm^{-2} 254 nm U.V. light (through 12 ml PBS per 150 mm diameter Petri dish). Cells were collected immediately, cell suspension was divided into two equal parts, pelletted and flash-frozen in liquid nitrogen. One Petri dish out of seven was not cross-linked, but collected by trypsination and analyzed by flow cytometry for transfection efficiency quantification.

150 μL Dynabeads protein G suspension (Novex; 30 mg.mL^{-1}) were washed 3 times in 1 mL PBS + 0.02% Tween 20, resuspended in 300 μL PBS + 0.02% Tween 20, then incubated overnight at 4°C under gentle agitation with 3 μL anti-mammalian Ago 2A8 ascites (a kind gift of Prof. Z. Mourelatos, University of Pennsylvania; $\approx$ 15 μg.μL^{-1}). Beads were then washed 3 times in 1 mL PBS + 0.02% Tween 20. Immunoprecipitation was performed on one of the two frozen cell pellets for each replicate. Cells were disrupted in 3 mL lysis buffer[1] per cell pellet, then incubated on ice for 10 min. Lysates were DNase-treated[2], then ultra-centrifuged 20 min at 35,000 g at 4 °C. The supernatant (after sampling 350 μL for input control) was incubated with 2A8-bound Dynabeads for 1h30 at 4°C under gentle agitation. 350 μL of supernatant were kept for control, and beads were successively washed with 2 mL of the following buffers at 4°C:

- lysis buffer[1];
- high salt buffer[3]
- high stringency buffer[4]
- low salt buffer[5]

[1] PBS with 1% Empigen, 40 U.mL^{-1} RNasin (Promega), Protease inhibitor cocktail (Roche) (1 tablet per 10 mL).
[2] 36 μL RQ1 DNase (Promega; 1 U.μL^{-1}) for 3 mL of lysate, incubated 8 min at 37 °C.
[3] PBS with 1% Empigen, supplemented with 50 g.L^{-1} NaCl.
[4] 15 mM Tris-HCl (pH=7.5), 5 mM EDTA (pH=8.0), 2.5 mM EGTA (pH=8.0), 1% Empigen, 120 mM NaCl, 25 mM KCl.
[5] 15 mM Tris-HCl (pH=7.5), 5 mM EDTA (pH=8.0).

- PNK buffer[6]

then optionally, 33 μL of bead suspension in final wash were kept for Western blotting control. Washed beads were resuspended in 100 μL proteinase K buffer[7] and 2.5 μL proteinase K (New England Biolabs; 20 μg.μL^{-1}), incubated at 37°C for 30 min then extracted in 100 μL phenol/chloroform/isoamyl alcohol (25/24/1) and RNA was precipitated with 50 μg glycogen, 10 μL NaAcO (3 M, pH=5.2) and 300 μL ethanol. Recovered RNA quantities ranged from 184 to 289 ng.

RNA samples were handed to Beijing Genomics Institute for library preparation (poly(A)-independent protocol, fragmenting RNAs prior to reverse-transcription; 49 nt reads, single-end sequencing; adapter sequence: 5´-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3´). After adapter trimming, reads shorter than 20 nt (*i.e.*: 0.14% to 1.96% of each library) were discarded.

Differential mRNA abundance between anti-miR-1a/miR-206 and anti-∅ libraries, and between anti-miR-133 and anti-∅ libraries was identified using HTseq (`http://www-huber.embl.de/HTSeq/doc/overview.html`) and EdgeR (`https://bioconductor.org/packages/release/bioc/html/edgeR.html`) (normalizing read counts with the TMM method). "Experimentally identified targets" were defined as mRNAs which are significantly (adjusted p-value < 0.05) less abundant in anti-miRNA transfection libraries than in anti-∅ libraries.

1.3.2 Absolute RNA quantification

At each time point (every 24 h, from 0 to 6 days after differentiation induction), cells from three indepent Petri dishes (150 mm diameter) were collected in 10 mL DPBS. 500 μL of the cell suspension was mixed with 4 μL Thiazole Orange, 10 μL of suspension were then monitored by microscopy to measure cell size. The remaining 494 μL were mixed with 50 μL Counting Beads (BD Biosciences cat. #349480, 1,043 beads.μL^{-1}) and analyzed by flow cytometry to determine cell number, following the manufacturer's intructions.

The remaining 9.5 mL of initial cell suspension was used for RNA extraction with Trizol (measuring aqueous phase loss after centrifugation by weighting). 60 to 300 μg RNA was obtained for each replicate.

20 μg total RNA were used for Northern blotting for miRNA quantification. A mix of 27 *in vitro* transcribed and polyadenylated spike-ins (not matching the mouse genome) was prepared and verified by Northern blot. Transcript concentration in the mix span a 100-fold concentration range. Spike-in mix was added to $\approx$ 10 μg total RNA from day 0, 3 and 6 of differentiation then handed to the Beijing Genomics Institute for RNA-seq library preparation and sequencing using the poly(A)-dependent protocol.

RNA abundance was calculated from RNA-seq data using the TopHat v.2.0.10 (`http://ccb.jhu.edu/software/tophat/index.shtml`) and Cufflinks v.2.1.1 (`http://cufflinks.cbcb.umd.edu/`) programs using the -b and -G options.

1.3.3 *Tmsb4x* miR-1a/miR-206 site mutagenesis

Homology donor sequences were prepared by PCR on C2C12 genomic DNA, with the left homology arm containing either the wild-type (`CAUUCC`) or a mutated (`UCCAUC`) seed match for miR-1a/miR-206 in the *Tmsb4x* 3´ UTR (nt 30 to 35 of the UTR) (see Figure 4**A** for a schematic map). The left homology arm is 986 bp long and the right homology arm is 861 bp long.

Undifferentiated C2C12 cells were co-transfected using Lipofectamine 2000 with:

- sgRNA-expressing plasmid: pmU6-gRNA (Addgene #53187), see section 1.6 for insert oligo sequences;
- High-fidelity-Cas9-expressing plasmid: VP12 (Addgene #72247);
- Homology donor plasmid: pmirGLO (Promega cat.#E1330) containing wt or miR-1a/miR-206 site mutant left homology arm (cloned between *Bgl*II and *Mlu*I sites) and right homology arm (cloned between *Cla*I and *Bam*HI sites). In both constructs, two miR-1a/miR-206 binding sites were cloned in the pmirGLO multiple cloning site (the miR-1a/miR-206 sites from *Tpm4* and *Tppp* 3´ UTRs, each embedded in 94 nt of genomic context; we chose these two genes among our list of experimentally-identified targets [for *Tpm4*] and among TargetScan-predicted targets that we did not detect experimentally [for *Tppp*]), to make the luc2 reporter sensitive to miR-1a/miR-206 activity. That plasmid also contains a fusion gene between Renilla luciferase and the neomycin

[6] 20 mM Tris-HCl (pH=7.5), 10 mM $MgCl_2$, 0.2% Tween 20.
[7] 100 mM Tris-HCl (pH=7.5), 50 mM NaCl, 10 mM EDTA (pH=8.0).

resistance gene ("hRluc-neo^R fusion") which provides both a normalizing luciferase activity and an antibiotics resistance gene for stable cell line selection.

Stable polyclonal lines were selected by growing subconfluently in 200 to 600 μg.mL^{-1} G418 for 8 weeks. Five independently-transfected wild-type lines and four independently-transfected mutant lines were finally isolated and used for further analysis. The frequency of wild-type and mutant alleles in each line was verified both by qPCR (see Supplemental Figure S5) and low-throughput sequencing of their genomic *Tmsb4x* 3´ UTR (not shown). The nine cell lines were differentiated for 6 days using the protocol described by Sweetman et al. (2008) then luc2 and hRluc activities were measured using the Dual luciferase reporter assay system (Promega cat.#E1910).

Results of the luciferase assay were modeled by a mixed-effects model, wherein the genotype confers a fixed effect and the cell line introduces a random effect:

$$log(y_{ijk}) = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad (1)$$

where:

- y_{ijk} is the response (luc2/hRluc ratio) in genotype $\#i$, cell line $\#j$ and replicate $\#k$;
- μ is the grand mean;
- α_i is the effect of genotype $\#i$;
- β_j is the effect of cell line $\#j$; β_j coefficients are assumed to be normally distributed, centered on 0 and with a variance that we will note τ^2;
- ε_{ijk} is the error, assumed to be normally distributed, centered on 0 and with a specific variance for each cell type (let's note the variance: σ_j^2).

then the significance of the effect of genotype was assessed by the likelihood ratio test, that compares the informational content of two models: the full model as described above, and a model where the effect of genotype is purposedly omitted (that incomplete model thus only tests the effect of cell line identity).

1.4 Correlation analysis between mRNA abundance and miRNA binding site conservation

Source code and data files are available at:
`http://www.igh.cnrs.fr/equip/Seitz/Pinzon_et_al_2016_data_and_code/Figure_5.tar.bz2`

For each murine miRNA family (as defined in miRBase v17), conservation of its interaction with predicted targets was estimated by the aggregate probability of conserved targeting of its miRNA binding sites (Friedman et al., 2009). For each probeset on the array, mRNA abundance was defined as the mean array signal across available biological replicates. Kendall's τ correlation coefficient was then calculated between these two datasets, for every miRNA family and every tissue.

1.5 Comparison between seed match conservation and seed conservation

Source code and data files are available at:
`http://www.igh.cnrs.fr/equip/Seitz/Pinzon_et_al_2016_data_and_code/Figure_6.tar.bz2`.

miRNA seed match coordinates were extracted from the 100-species whole genome alignment centered on the human genome (`http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz100way/maf/`). 3´ UTR coordinates were retrieved from the UCSC Genome Browser RefSeq gene database (`http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz`). For each human 3´ UTR seed match, the list of vertebrate species having that seed match conserved in the genome alignment was extracted.

For vertebrate species listed in miRBase (release 21), the existence of each seed in the miRNA repertoire of a given species was extracted from miRBase. It is likely that some miRNAs have escaped experimental identification in some species: in order to avoid such false negatives, and also to determine whether species not recorded in miRBase express miRNAs with a given seed, we searched these species' genomes with HMMer for potential orthologous pre-miRNAs of known vertebrate miRNAs with the seed of interest:

1. known vertebrate miRNAs with that seed were collected, and their pre-miRNAs were aligned with `blast`;
2. HMMer profiles were built for each group of mutually similar pre-miRNA sequences (pairwise `blast` E-value<0.001); if the group contains only 1 pre-miRNA, no HMMer profile could be built: homology search was then performed with `blast`;
3. each genome outside the clade of interest was searched for homologous loci using `nhmmer` with these profiles (when possible), or with `blast` otherwise;
4. each homologous locus (with an $E-value < 0.001$) was folded using `RNAsubopt` to predict transcript secondary structure; loci predicted to be transcribed in unbranched hairpins (whose predicted ΔG was at most 75% of the most stable structure's ΔG) were selected;
5. potential pre-miRNA orthologs were considered if their miRNA seed is identical to the seed being searched.

Seeds for which we found potential pre-miRNA orthologs in a genome outside the clade of interest were discarded.

For comparison with miRNA prediction programs, predicted miRNA binding sites were downloaded from:

- `http://cbio.mskcc.org/microrna_data/human_predictions_S_C_aug2010.txt.gz` and `http://cbio.mskcc.org/microrna_data/human_predictions_S_0_aug2010.txt.gz` (miRanda aug2010);
- `http://www.targetscan.org/vert_70/vert_70_data_download/Conserved_Site_Context_Scores.txt.zip` and `http://www.targetscan.org/vert_70/vert_70_data_download/Nonconserved_Site_Context_Scores.txt.zip` (TargetScan 7.0);
- `http://diana.imis.athena-innovation.gr/DianaTools/data/microT_CDS_data.tar.gz` (microT 5.0);
- and `http://dorina.mdc-berlin.de/` (selecting the PicTar2 mammalian miRNA:targets in 3´ UTRs) for PicTar2.

Prediction programs do not have the exact same requirements for target predictions, but they all tend to favor miRNA binding sites with extensive pairing to the seed. In order to make every analysis comparable, they were restricted to perfect seed matches in 3´ UTRs, excluding seed matches that overlap exon-exon junctions.

1.6 Oligonucleotides used in this study

RNA and 2´*O*-methyl oligos:

	Description	Sequence (5´ to 3´)
anti-miRNAs	anti-miR-1a	`Cy5-AmCmGmAmUmAmCmAmUmAmCmUmUmCmUmUmUmAmCmAmUmUmCmCmAmAmCmGm-ddC`
	anti-miR-206	`Cy5-AmCmGmCmCmAmCmAmCmAmCmUmUmCmCmUmUmAmCmAmUmUmCmCmAmAmCmGm-ddC`
	anti-miR-133	`Cy5-AmCmGmCmAmGmCmUmGmGmUmUmGmAmAmGmGmGmAmCmCmAmAmAmCmGm-ddC`
	anti-∅	`Cy5-AmCmGmAmUmAmAmCmGmUmAmCmGmUmAmCmGmUmAmCmGmUmAmCm-ddC`
miRNAs	miR-1a	`P-UGGAAUGUAAAGAAGUAUGUAU`
	miR-206	`P-UGGAAUGUAAGGAAGUGUGUGG`
	miR-133	`P-UUUGGUCCCCUUCAACCAGCUG`

DNA oligos:

	Description	Sequence (5´ to 3´)
Tmsb4x mutagenesis	Left homology arm cloning (forward)	ttacgcgtTGATGAATATGGCCTGCAAA
	Left homology arm cloning (reverse)	ttagatctCTACCCCTTCATTCCACAGC
	Right homology arm cloning (forward)	ttatcgatGCTGTGGAATGAAGGGGTAG
	Right homology arm cloning (reverse)	ttggatccAAAGGCCTGTTCACTGATGG
	sgRNA cloning (top strand)	P-TTGTTTGCAAGCTGGCGAATCGTAATG
	sgRNA cloning (bottom strand)	P-AAACCATTACGATTCGCCAGCTTGCAA
qPCR	Forward primer for qPCR (common to wt and mutant allele)	AGCGAGGCTGCTATGTGTCT
	Reverse primer for qPCR on wt allele	AAGGCAATGCTCGTGGAAT
	Reverse primer for qPCR on mutant allele	AAGGCAATGCTCGTGATGG
miR-1 reporter	*Tpm4* miR-1a/miR-206 site cloning (forward)	aagagctcTGGCTTTGTAGTTTTCCTTTCTC
	Tpm4 miR-1a/miR-206 site cloning (reverse)	aactcgagTCAGGACTGTAAACTTGAGTTGG
	Tppp miR-1a/miR-206 site cloning (forward)	aactcgagCACTATAGGTGGCAGGCACA
	Tppp miR-1a/miR-206 site cloning (reverse)	aatctagaGCCTAGTGGAGGTGCATTCT

Cy5: Cyanine5; ddC: dideoxy C; Nm: 2´-*O*-methylated nucleotide; P: phosphate. Lower case letters: extragenomic sequences, for restriction site addition.

2 Supplemental figures

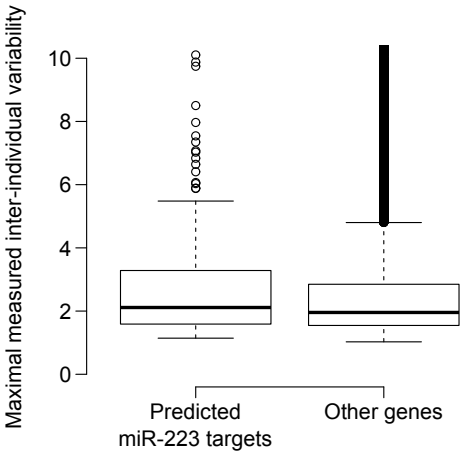

Figure S1: **Inter-individual variability in gene expression in mouse neutrophils.** Measured inter-individual variability in gene expression for the most accurately quantified genes (mean of the 5 individual measured $\log_2$(signal) $> 2\times$standard deviation of $\log_2$(signal) of probesets with similar microarray signal). Each point represents one gene. The y-axis displays the maximal measured inter-individual variability (signal in the most strongly expressing mouse / signal in the least strongly expressing mouse). miR-223 targets were predicted using TargetScan mouse (v. 6.2). For graphical clarity, only the genes for which variability is less than 10-fold are represented (that is: 177 predicted targets and 20,542 other genes). Variability for the remaining 12 predicted targets and 947 other genes reaches 46-fold and 197-fold, respectively (not shown).

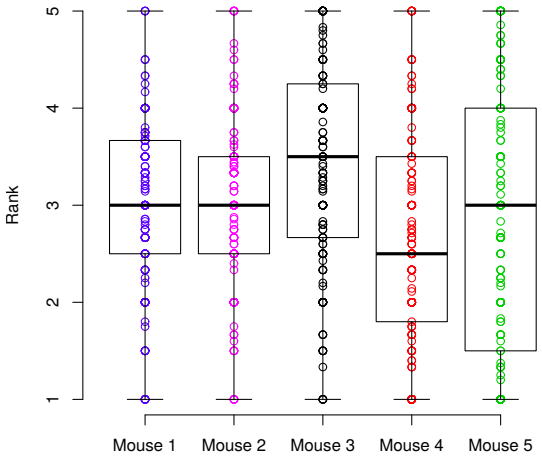

Figure S 2: **Uncoordinated expression of predicted miR-223 targets among individual mice.** For each predicted miR-223 target, we ranked the 5 mice according to that gene's expression. For genes matched by several probesets on the array, we considered the mean of the ranks for every probeset. Each point on the plot represents a predicted miR-223 target. Rank distribution is rather uniform across the 5 mice, indicating that predicted targets are not coordinately up- or down-regulated from one individual to the next.

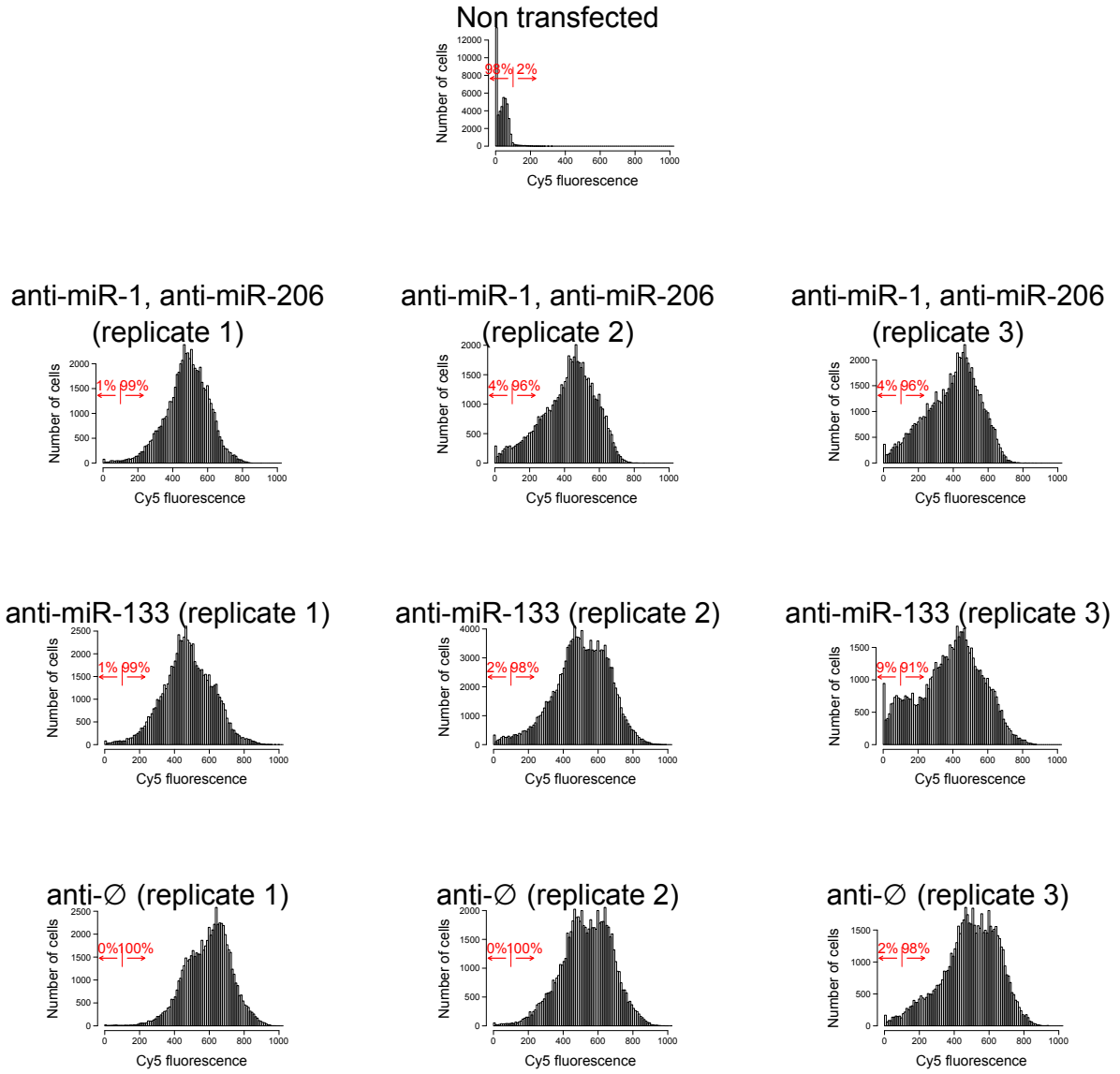

Figure S3: **Anti-miRNA oligonucleotide transfection control.** Flow cytometry was performed with the same acquisition settings for every sample. Recorded events with a high FSC (forward-scattering light; FSC cutoff=850 with our settings) were flagged as "cells" and selected for fluorescence analysis. Percentages of cells with a Cy5 fluorescence higher or lower than our cutoff (100 with our settings) are indicated in red on each histogram (rounded to the closest integer).

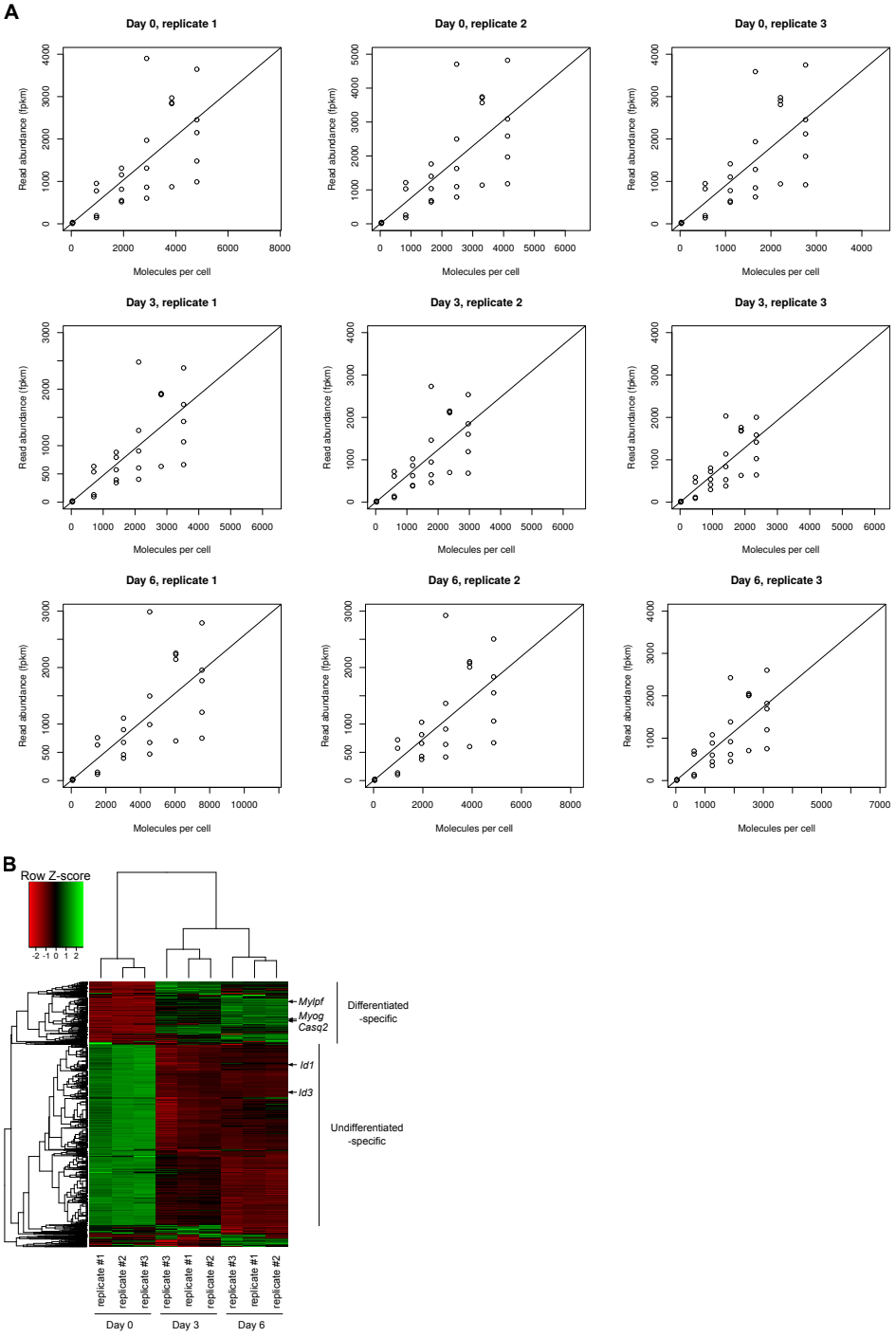

Figure S 4: **A. *In vitro* transcribed spike-ins in calibrated RNA-seq.** In each of the 9 RNA-seq libraries, 27 polyadenylated *in vitro* transcripts were introduced in controlled amounts for internal calibration (each *in vitro* transcript is represented by a circle). For each transcript, between 3×10^{-17} and 3×10^{-15} mol were added in each RNA sample (RNA samples were prepared from ≈243,000 to ≈780,000 cells each). Linear regression of spike-in data, forcing a null y-intercept, then allowed us to convert $fpkm$ RNA-seq results into numbers of molecules per cell, with a precision of ≈2-fold (the linear regression line is shown as a plain black line). Please note the data is not log-transformed. **B. C2C12 differentiation control.** Unsupervised clustering of RNA-seq data (for mRNAs with more than 100 fpkm in at least one library) reveals the similarity of biological replicates at each differentiation time point (columns) and identifies groups of differentially regulated genes during C2C12 differentiation (rows). Distance between features was measured by (1 - Kendall's τ correlation coefficient) and clustering was performed using the McQuitty method. Very similar results were obtained with alternative distance definitions and clustering procedures (not shown). *Id1* and *Id3* are known markers of the undifferentiated state of C2C12 cells (Jen et al., 1992; Mohamed et al., 2013) while *Mylpf*, *Myog* and *Casq2* are specifically expressed upon differentiation (Arai et al., 1992; Yuasa et al., 2009; Wu et al., 2014).

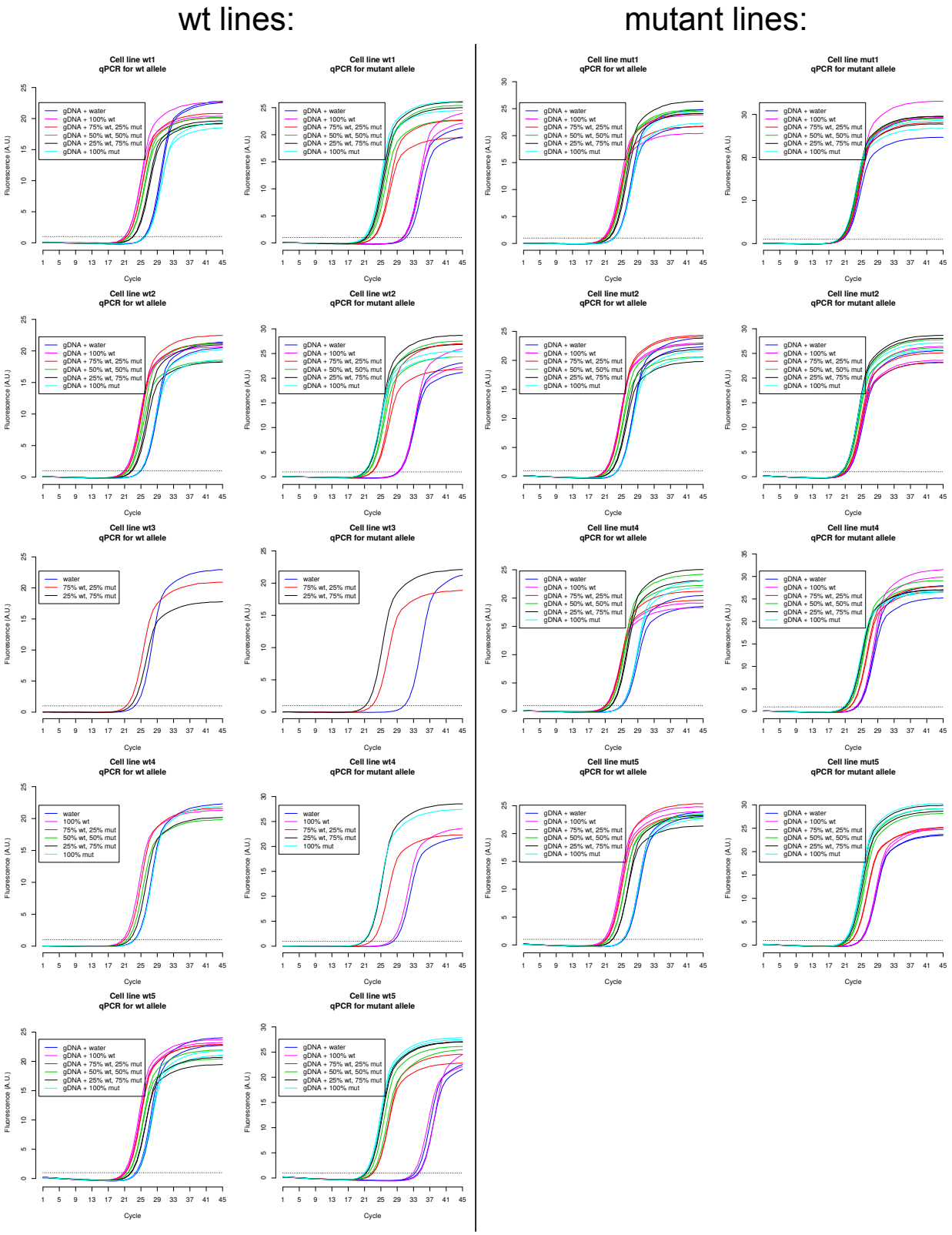

Figure S 5: **qPCR-based quantification of wt and mutant *Tsmb4x* alleles in the established polyclonal lines.** qPCR reactions were performed on 100 ng genomic DNA, supplemented with either water or $4{\times}10^{-13}$ g of various plasmid mixes. Plasmids were 4 kb long and contained either the *Tmsb4x* 3′ UTR miR-1a/miR-206 seed match or the mutant version (CAUUCC⟶UCCAUC); they were mixed in various proportions, with 0, 25, 50, 75 or 100% of either version. Each qPCR was performed as two technical replicates, except for lines "wt3" and "wt4".

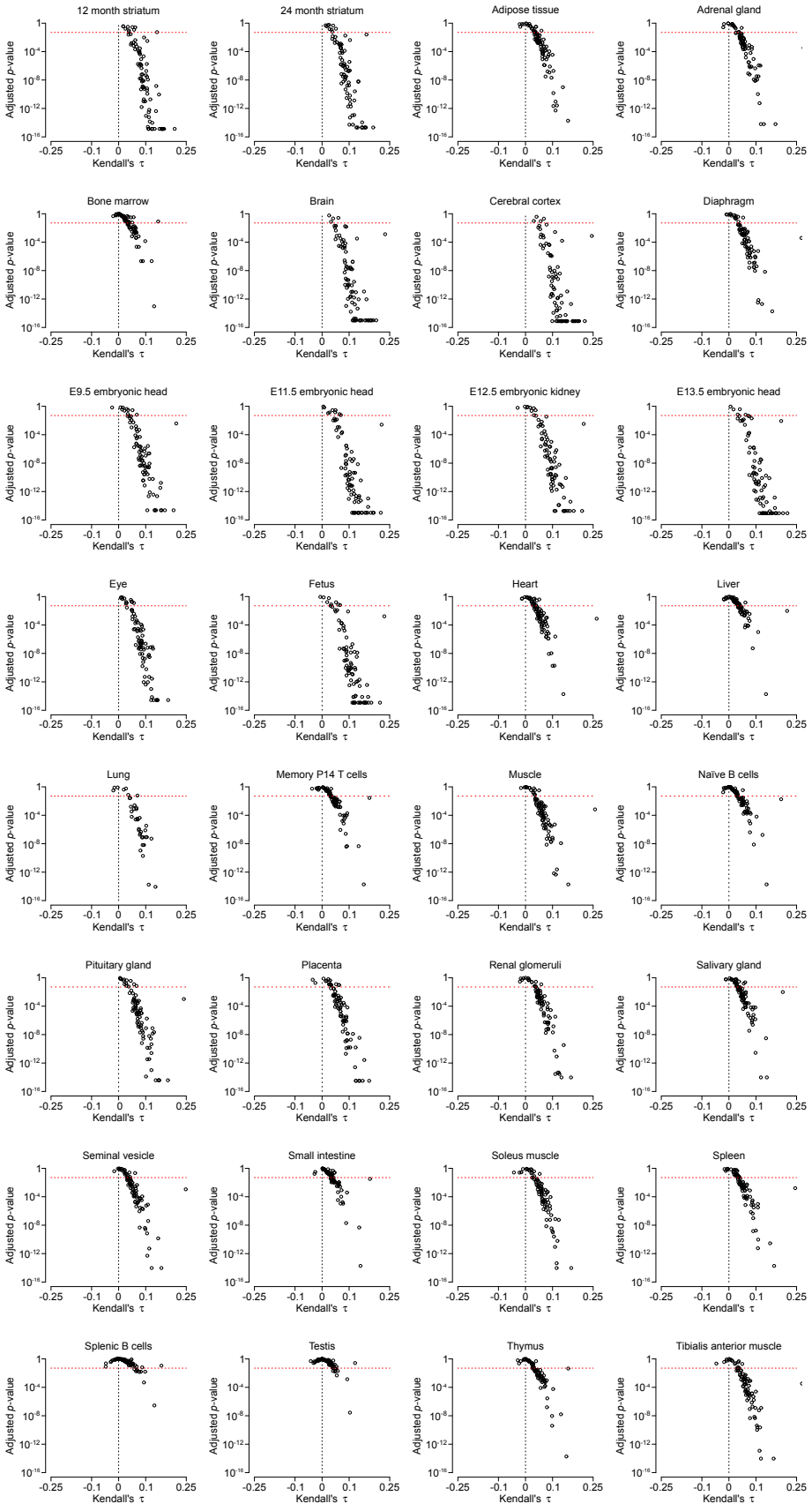

Figure S6: **The most highly expressed genes tend to bear the most conserved miRNA binding sites.** Same conventions as in Figure 5**A** in the main text.

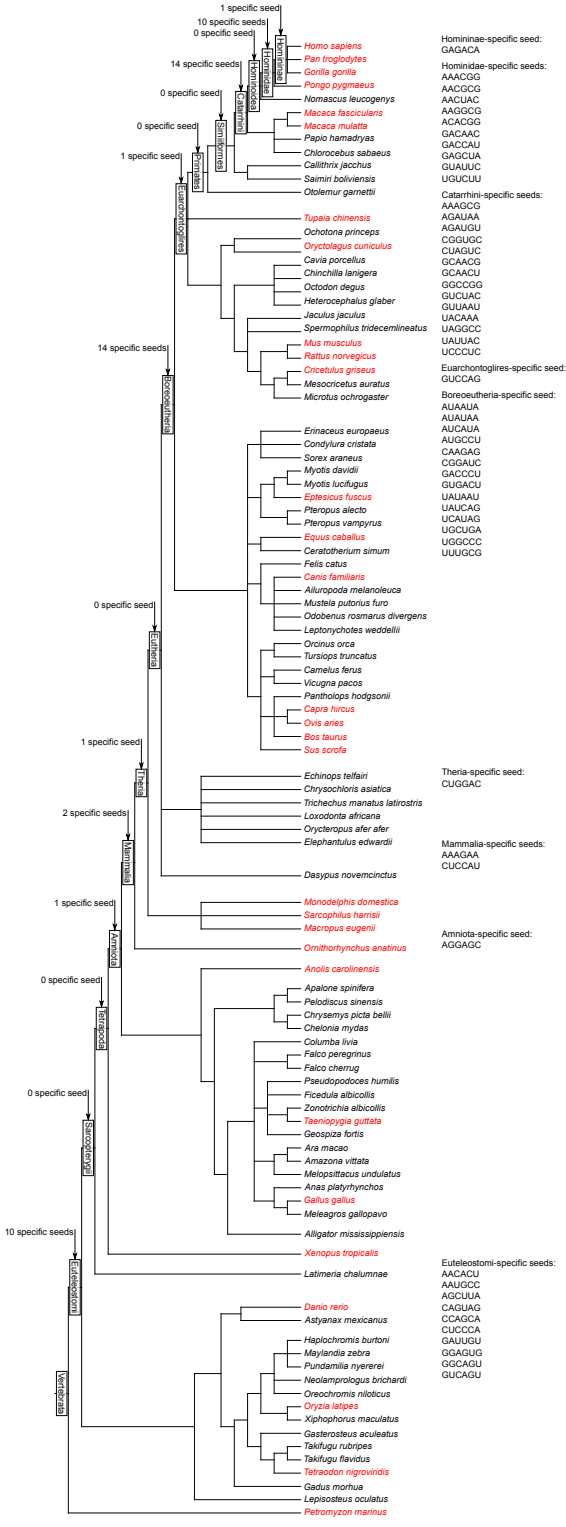

Figure S 7: **Identification of clade-specific miRNA seeds.** 100 vertebrate species are used in the human-centered whole genome alignment (Rosenbloom et al., 2015); species listed in miRBase release 21 (Kozomara and Griffiths-Jones, 2014) are shown in red. For each clade containing *Homo sapiens* we identified miRNA seeds that are present in at least 75% of the species in the clade, while being absent in every species outside the clade. It is likely that some miRNAs have escaped experimental identification in some species: in order to avoid such false negatives, these species' genomes were searched for potential orthologous pre-miRNAs of known vertebrate miRNAs with that seed. We excluded every seed for which potential pre-miRNA orthologs were found in genomes outside the clade of interest. Only the clades for which $\geqslant 10$ clade-specific seeds could be identified were analyzed in Figure 6.

3 Supplemental tables

Sample	Neutrophil purity (%)	Usage	Probesets with highest signal among individuals	Probesets with lowest signal among individuals
1	90.5	Individual mouse	5,631	4,987
2	88.8	Pooled blood	–	–
3	85.0	Individual mouse	3,296	7,009
4	91.6	Pooled blood	–	–
5	71.5	Individual mouse	9,979	8,544
6	92.7	Pooled blood	–	–
7	89.8	Pooled blood	–	–
8	88.1	Pooled blood	–	–
9	99.5	Individual mouse	5,562	12,606
10	88.4	Individual mouse	20,673	11,995

Table S 1: **Neutrophil purity in the analyzed samples.** Purity was assessed by flow cytometry using APC anti-mouse Ly6G antibody. "Individual mice" were used for measurement of inter-individual variability; "pooled blood" was used for measurement of technical inaccuracies. For "individual mouse" samples, we recorded the number of probesets for which the highest or lowest signal among individual mouse samples was found in that sample for quality control (see section 1.1).

	Tightly-regulated genes		Variable genes			
Tarbase-supported interactions	*Arfip1*	*Cbx5*	*A930038C07Rik*	*Abhd13*	*Ankrd17*	*Atp1b1*
	Rasa1	*Sept8*	*Atp2b1*	*F3*	*Fa2h*	*Fam176a*
	Srsf12		*Foxp1*	*Hlf*	*Igf1r*	*Itgb1*
			Mbnl1	*Mef2c*	*Mga*	*Mtpn*
			Myh10	*Myst3*	*Nfia*	*Nfib*
			Otud4	*Pds5b*	*Phf20l1*	*Ptbp2*
			Rap2a	*Rhob*	*Slc37a3*	*Sox6*
			Srgap3	*Stk39*	*Ypel1*	
Other interactions	*Aco1*	*Acvr2a*	*Acsl3*	*Adcy7*	*Aebp2*	*Armc1*
	Anks1b	*Arpp19*	*Armcx1*	*Atp7a*	*Atxn1*	*Bai3*
	Brpf3	*Ctsl*	*Cd2ap*	*Cdk17*	*Celf2*	*Cep68*
	Fam199x	*Fam46a*	*Cnot2*	*Cntln*	*Cops2*	*Cpne4*
	Fbxo25	*Fbxo8*	*Crim1*	*Csnk1g1*	*D19Wsu162e*	*Ddit4*
	Fgfr2	*Frmd4a*	*Dennd5b*	*Dlc1*	*Dusp2*	*E330021D16Rik*
	Gpr22	*Lelp1*	*Ebf3*	*Eif1ad*	*Eif2c3*	*Eif5b*
	Mospd1	*Nlrp3*	*Elf5*	*Elk1*	*Ept1*	*Fam120c*
	Orc4	*Pik3c2a*	*Fam168a*	*Fam5c*	*Fat1*	*Fbxw7*
	Plce1	*Rc3h1*	*Flrt1*	*Galnt7*	*Gna13*	*Gpm6b*
	Rcn2	*Scn1a*	*Gpr155*	*Hhex*	*Hsp90b1*	*Il6st*
	Sh3pxd2b	*Slc23a2*	*Inpp4a*	*Inpp5b*	*Jmy*	*Kif21b*
	Slc35f1	*Slc39a1*	*Lass6*	*Lif*	*Lmo2*	*Mafb*
	Smarcd1	*Srp54a*	*Mll3*	*Mmp16*	*Mon2*	*Mpz*
	Srp54b	*Tet3*	*Msi2*	*Mtap4*	*Naa50*	*Nfat5*
	Tgfbr3	*Tmem170*	*Nlgn2*	*Nutf2*	*Olfm1*	*Pde4d*
	Tmprss11e	*Ube2q2*	*Phactr4*	*Phip*	*Pkn2*	*Pknox1*
	Ulk2	*Zcchc14*	*Plagl2*	*Plekhh1*	*Prpf39*	*Rab22a*
	Zxdc		*Rab8b*	*Rabgap1*	*Rabl2*	*Rad17*
			Ralgps2	*Rbm16*	*Rbm20*	*Rbpj*
			Rnf34	*Rnpc3*	*Rorb*	*Rps6kb1*
			Rras2	*Scn2a1*	*Scn3a*	*Sept6*
			Sgms2	*Slain2*	*Slc24a2*	*Slc4a4*
			Slc8a1	*Smoc1*	*Smurf2*	*Sox11*
			Sp3	*Spata13*	*Srp54c*	*Srpk2*
			Styx	*Syncrip*	*Tmem229a*	*Tmem47*
			Tmem64	*Tmtc2*	*Tnrc6b*	*Treml2*
			Trps1	*Tshz3*	*Ube2a*	*Uqcc*
			Vamp2	*Vav3*	*Wdr62*	*Wwtr1*
			Zfhx3	*Zfp238*	*Zfx*	

Table S 2: **Tightly-regulated genes and variable genes among predicted miR-223 targets.** "Tightly-regulated genes" are the predicted miR-223 targets whose inter-individual variations in expression are lower than miR-223-guided repression ($p < 0.05$, with p-value defined as in Figure 1**C**). "Variable genes" are the other predicted targets. TarBase data was downloaded from `http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index`. Among the 42 tightly-regulated genes, 5 are listed in TarBase; among the 150 variable genes, 31 are listed in TarBase (proportions not significantly different: Fisher's exact test p-value=0.37).

Differentiation day	Replicate name	Number of reads	Genome matching reads	Transcriptome matching reads	Number of detected mRNAs	Median abundance of detected mRNAs (fpkm)
0	1	343,336,245	336,133,200	326,332,352	17,481	3.46836
0	2	329,549,176	323,584,543	313,815,004	17,356	3.29692
0	3	339,119,201	332,961,581	322,807,963	17,475	3.25078
3	1	325,133,660	319,030,091	314,639,678	17,898	4.121915
3	2	272,084,347	267,228,824	264,513,001	17,785	4.19869
3	3	319,687,481	313,691,691	309,703,151	17,846	4.162125
6	1	313,253,641	307,337,296	303,106,144	18,061	4.09468
6	2	339,291,429	332,544,825	327,073,727	18,059	4.07273
6	3	298,424,803	292,694,151	288,940,732	18,115	4.1594

Table S3: **RNA-seq statistics.** The number of transcriptome-matching reads was rounded from Cufflinks' "normalized map mass" for each library. "Detected mRNAs" are the murine mRNAs for which read abundance > 0 fpkm.

NCBI's GEO sample	NCBI's GEO series	Organ or tissue
GSM461846	GSE18551	12 month striatum
GSM461847		
GSM461848		
GSM461849		
GSM461850	GSE18551	24 month striatum
GSM461852		
GSM461855		
GSM461859		
GSM252093	GSE9954	adipose tissue
GSM252094		
GSM252095		
GSM252086	GSE9954	adrenal gland
GSM252087		
GSM252088		
GSM252089	GSE9954	bone marrow
GSM252090		
GSM252091		
GSM252092		
GSM252077	GSE9954	brain
GSM252078		
GSM252079		
GSM149511	GSE6514	cerebral cortex
GSM149512		
GSM149513		
GSM149514		
GSM149515		
GSM252064	GSE9954	diaphragm
GSM252065		
GSM252066		
GSM200700	GSE8091	E11.5 embryonic head
GSM200701		
GSM200703		
GSM200704		
GSM243346	GSE9629	E12.5 embryonic kidney
GSM243347		
GSM243348		
GSM200706	GSE8091	E13.5 embryonic head
GSM200707		
GSM200708		
GSM200709		
GSM200711		
GSM200713		
GSM200695	GSE8091	E9.5 embryonic head
GSM200696		
GSM200697		
GSM200698		
GSM200699		
GSM200716		
GSM252122	GSE9954	ES cells
GSM252123		
GSM252124		

(to be continued)

NCBI's GEO sample	NCBI's GEO series	Organ or tissue
GSM252119	GSE9954	eye
GSM252120		
GSM252121		
GSM252131	GSE9954	fetus
GSM252132		
GSM252133		
GSM252113	GSE9954	heart
GSM252114		
GSM252115		
GSM149576	GSE6514	hypothalamus
GSM149577		
GSM149578		
GSM149599		
GSM149600		
GSM252083	GSE9954	kidney
GSM252084		
GSM252085		
GSM228786	GSE9012	liver
GSM228787		
GSM228788		
GSM228789		
GSM228790		
GSM252074	GSE9954	
GSM252075		
GSM252076		
GSM252080	GSE9954	lung
GSM252081		
GSM252082		
GSM426400	GSE17812	memory P14 T cells
GSM426401		
GSM252070	GSE9954	muscle
GSM252071		
GSM252072		
GSM252073		
GSM94741	GSE4142	naïve B cells
GSM94744		
GSM94745		
GSM252128	GSE9954	ovary
GSM252129		
GSM252130		
GSM252096	GSE9954	pituitary gland
GSM252097		
GSM252098		
GSM252099		
GSM252100		
GSM252125	GSE9954	placenta
GSM252126		
GSM252127		

(to be continued)

NCBI's GEO sample	NCBI's GEO series	Organ or tissue
GSM458136	GSE18358	renal glomeruli
GSM458140		
GSM458141		
GSM458142		
GSM458143		
GSM252101	GSE9954	salivary gland
GSM252102		
GSM252103		
GSM252104	GSE9954	seminal vesicle
GSM252105		
GSM252106		
GSM252116	GSE9954	small intestine
GSM252117		
GSM252118		
GSM261510	GSE10347	soleus muscle
GSM261511		
GSM261512		
GSM252067	GSE9954	spleen
GSM252068		
GSM252069		
GSM61837	GSE2826	splenic B cells
GSM61838		
GSM252110	GSE9954	testis
GSM252111		
GSM252112		
GSM252107	GSE9954	thymus
GSM252108		
GSM252109		
GSM261516	GSE10347	tibialis anterior muscle
GSM261517		
GSM261518		

Table S4: **Transcriptomic datasets used for the analysis of correlation between mRNA abundance and miRNA binding site conservation.**

Target prediction program	Total predicted sites	Predicted sites in 3´ UTRs		
		Overlapping exon-exon junctions	Not on exon-exon junctions	
			No perfect seed match	With perfect seed match
microT 5.0	32,400,362	444,545	nt 2–7: 15,322,673 nt 2–8: 23,309,068	nt 2–7: 12,240,601 nt 2–8: 4,254,206
miRanda aug2010	4,417,884	0	nt 2–7: 1,266,616 nt 2–8: 1,966,976	nt 2–7: 3,151,268 nt 2–8: 2,450,908
PicTar2	1,287,398	0	nt 2–7: 479,513 nt 2–8: 754,240	nt 2–7: 550,906 nt 2–8: 276,179
TargetScan 7.0	14,542,205	21,157	nt 2–7: 496,637 nt 2–8: 4,328,678	nt 2–7: 9,473,770 nt 2–8: 5,641,729

Table S 5: **Predicted miRNA binding sites characteristics.** Origin of data: `http://diana.imis.athena-innovation.gr/DianaTools/data/microT_CDS_data.tar.gz` for microT 5.0 predictions; `http://cbio.mskcc.org/microrna_data/human_predictions_S_C_aug2010.txt.gz` and `http://cbio.mskcc.org/microrna_data/human_predictions_S_0_aug2010.txt.gz` for miRanda aug2010 predictions; `http://dorina.mdc-berlin.de/` for PicTar2 predictions; `http://www.targetscan.org/vert_70/vert_70_data_download/Conserved_Site_Context_Scores.txt.zip` and `http://www.targetscan.org/vert_70/vert_70_data_download/Nonconserved_Site_Context_Scores.txt.zip` for TargetScan 7.0. Predicted binding site location was compared to 3´ UTR exon coordinates from the UCSC Genome Browser RefSeq gene data file (`http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz`).

References

Arai, M., Otsu, K., MacLennan, D. H., and Periasamy, M., 1992. Regulation of sarcoplasmic reticulum gene expression during cardiac and skeletal muscle development. *Am J Physiol*, **262**(3 Pt 1):C614–20.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P., 2008. The impact of microRNAs on protein output. *Nature*, **455**(7209):64–71.

Dang, V. T., Kassahn, K. S., Marcos, A. E., and Ragan, M. A., 2008. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur J Hum Genet*, **16**(11):1350–1357.

Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*, **19**(1):92–105.

Huang, N., Lee, I., Marcotte, E. M., and Hurles, M. E., 2010. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet*, **6**(10):e1001154.

Jen, Y., Weintraub, H., and Benezra, R., 1992. Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. *Genes Dev*, **6**(8):1466–1479.

Kozomara, A. and Griffiths-Jones, S., 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, **42**(Database issue):D68–73.

Mohamed, J. S., Lopez, M. A., Cox, G. A., and Boriek, A. M., 2013. Ankyrin repeat domain protein 2 and inhibitor of DNA binding 3 cooperatively inhibit myoblast differentiation by physical interaction. *J Biol Chem*, **288**(34):24560–24568.

Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., *et al.*, 2015. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res*, **43**(Database issue):D670–681.

Sweetman, D., Goljanek, K., Rathjen, T., Oustanina, S., Braun, T., Dalmay, T., and Münsterberg, A., 2008. Specific requirements of MRFs for the expression of muscle specific microRNAs, miR-1, miR-206 and miR-133. *Dev Biol*, **321**(2):491–499.

Wu, Y., Long, Q., Zheng, Z., Xia, Q., Wen, F., Zhu, X., Yu, X., and Yang, Z., 2014. Adipose induces myoblast differentiation and mediates TNFα-regulated myogenesis. *Biochim Biophys Acta*, . In press.

Yuasa, K., Masuda, T., Yoshikawa, C., Nagahama, M., Matsuda, Y., and Tsuji, A., 2009. Subtilisin-like proprotein convertase PACE4 is required for skeletal muscle differentiation. *J Biochem*, **146**(3):407–415.