

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» RNAi

» *Drosophila*

Received: 17 July 2013

Accepted: 23 January 2014

Published: 11 March 2014

The systematic identification of cytoskeletal genes required for *Drosophila melanogaster* muscle maintenance

Alexander D. Perkins, Michael J.J. Lee and Guy Tanentzapf

Animal muscles must maintain their function and structure while bearing substantial mechanical loads. How muscles withstand persistent mechanical strain is presently not well understood. Understanding the mechanisms by which tissues maintain their complex architecture is a key goal of cell biology. This dataset represents a systematic screen through the *Drosophila melanogaster* cytoskeleton to identify genes that are required to maintain tissue, specifically muscle, architecture. Using RNA interference (RNAi), we knocked down 238 genes in *Drosophila* and assayed for climbing ability with a robust behavioural assay. Here we present the summary of the screen and provide the complete results of the assays. We have uncovered a number of novel hits that would reward further study. The data are easy to use: the raw data are provided to allow researchers to perform their own analysis and analysed results are given indicating whether or not the genes are required for muscle maintenance. This dataset will allow other researchers to identify candidate genes for more detailed study and lead to better understanding of muscle maintenance.

Design Type(s)	time series design • strain comparison design
Measurement Type(s)	developmental lethality screen • behavioural assay
Technology Type(s)	RNAi screening • phenotypic characterization
Factor Type(s)	genotype • developmental stage
Sample Characteristic(s)	<i>Drosophila melanogaster</i> • whole organism

Department of Cellular and Physiological Sciences, University of British Columbia, Life Sciences Institute, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada.

Correspondence and requests for materials should be addressed to G.T. (email: tanentz@mail.ubc.ca)

Background & Summary

Tissues in multicellular organisms are organized into highly ordered, three-dimensional structures that enable diverse and specific functions. Understanding the mechanisms by which tissues develop and maintain this complex architecture is a fundamental goal of cell and developmental biology. Muscles are the most common mechanism for force-generation in animals and provide an excellent model for the study of tissue maintenance as they are both highly structured and long-lived¹. Their highly structured nature is typified by the sarcomere, the contractile apparatus that generates muscle contraction. Longevity is exemplified in vertebrates by cardiac muscles which live for decades² and in invertebrates by fly muscle cells which survive for the entire adult lifespan³.

Genetic analysis of long-term muscle maintenance has been a topic of intense study due to interest in both myodegenerative diseases and ageing. Genes shown to be required for muscle maintenance have come both from studies in animal models and from cloning genes involved in human myodegeneration diseases. Two broad categories of genes have been identified for long-term muscle tissue maintenance: cytoskeletal and sarcomeric genes and muscle-repair genes. Work in the genetic model organisms *C. elegans* and *Drosophila* has been useful in highlighting the importance of several cytoskeletal components of the sarcomere in maintaining the adult musculature^{3,4}. The second category of mutations that affect muscle maintenance cause an increase in the rate of damage accumulation in muscle tissue and reveal themselves when muscle repair and renewal mechanisms are overwhelmed^{3,5}.

Several previous screens have approached the topic of muscle function in depth and from several angles. The most comprehensive of these performed a genome-wide RNAi screen using the Mef2:GAL4 system to identify, and characterize with microscopic assays, 2,785 genes required for muscle function⁶. Muscle maintenance and myogenesis have been studied in *Drosophila* primary cell culture, although not in adult organisms⁷. Within the context of Dystrophin/Dystroglycan-related muscular dystrophies, two important genome-wide interaction screens identified key genetic pathways and specific genes that are involved in maintaining muscle integrity in dystrophic backgrounds^{8,9}.

However, these studies suffered from a key limitation in the context of studying muscle maintenance: gene knockdown occurred throughout muscle development. In many cases it is likely that defects observed in developed muscle are caused by incorrect myogenesis and are only revealed in adulthood^{10–15}. Thus the main problem in studying how adult muscle structure and function is maintained is in describing function in fully formed muscles for genes whose activity is required for muscle development^{16–18}. Importantly, this issue has prevented the systematic identification of genes that are required for the long-term maintenance of muscle function in the adult. We have circumvented this issue by using temporal and tissue-specific expression of RNAi constructs to limit knockdown specifically to adult flies. Given its fundamental role in muscles, we focused our analysis on the cytoskeleton. Our screen is unique as it focuses specifically on *in vivo* gene knockdown in adult flies. Using this approach we were able to systematically analyse the role of the cytoskeleton in *Drosophila* muscle maintenance. Here we present the data from our screen which identified 46 genes required to maintain muscle function in adult flies (Data Records 1 and 2; Data Citations 1 and 2).

Methods

Screen design

We performed an RNAi screen of candidate genes with a particular focus on cytoskeletal proteins. A library of RNAi construct lines was assembled that represented the fly cytoskeleton¹⁹. The library contained RNAi constructs targeting 238 genes. A random set were retested using additional lines so in total 271 RNAi lines were screened in the Mef2:GAL4 screen and 150 RNAi lines in the TARGET screen (Data Record 1; Data Citation 1). The genes targeted by the RNAi constructs included, but were not limited to, actins and their associated proteins; tubulins and their associated proteins; motor proteins such as myosins, kinesins and dyneins; and other cytoskeletal elements including lamins, membrane linkers and septins. Since our primary goal was to identify all the genes in our library that affected adult muscle function—even mildly—we designed a sensitive, comprehensive and robust functional screen. A key feature of our screen was its high time resolution—designed to detect subtle differences between control and knockdown conditions. A previous systematic RNAi knockdown screen in the fly assayed adult muscle function at a single time-point 7 days post-eclosion using a flight assay⁶. Our screen measured climbing ability every 3 days for 30 days after eclosion.

The screen was conducted in two phases. A preliminary screen where the RNAi construct was expressed using the Mef2:GAL4 driver²⁰ throughout life was designed to identify all of the constructs whose expression gave rise to phenotypes either during development or in the adult (Fig. 1a and c). In the second phase all the RNAi constructs whose expression under the Mef2:GAL4 system resulted in a phenotype were passed to the TARGET screen (Fig. 1b and c). The TARGET system limited RNAi-construct expression to only the adult muscle²¹. This part of the screen specifically identified lines required for muscle maintenance rather than muscle development. In total, we carried out over 113,000 individual negative geotaxis assays.

The library of RNAi constructs was assembled from the Vienna *Drosophila* RNAi Center (VDRC, <http://stockcenter.vdrc.at>; ref. 22), the Harvard Transgenic RNAi Project (TRiP, <http://www.flyrnai.org>; ref. 23) and the National Institute of Genetics (NIG-Fly, <http://www.shigen.nig.ac.jp/fly/nigfly/>). Detailed information regarding the design of the RNAi constructs is given at the stock centre websites and the

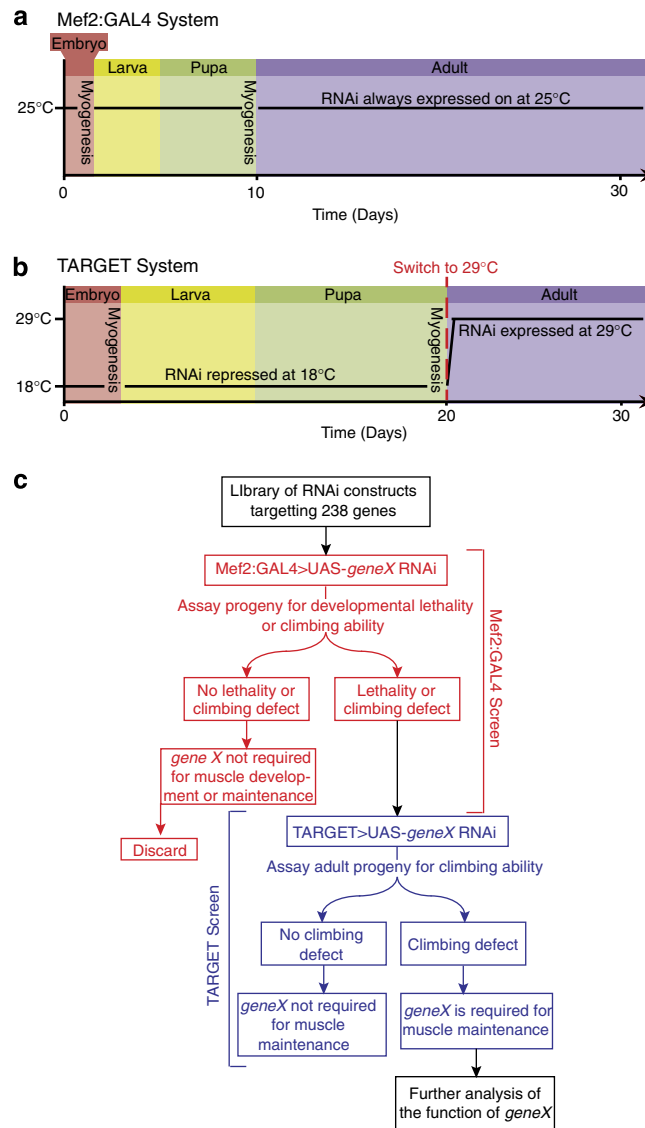


Figure 1. Design of a screen to identify cytoskeletal genes required for muscle maintenance. **(a)** The Mef2:GAL4 system. The Mef2:GAL4 system allows for muscle-specific expression of RNAi constructs using the GAL4/UAS system. **(b)** The TARGET system uses the Mef2:GAL4 system but includes a temperature-sensitive allele of the GAL4 repressor GAL80. At 18 °C, RNAi-construct expression is turned off. At 29 °C, the RNAi construct is expressed. To limit RNAi-mediated knockdown to adults, flies were kept at 18 °C until they eclosed, at which point they were transferred to 29 °C to induce RNAi expression. **(c)** Flowchart of two-step screening protocol. In the first part of the screen, RNAi constructs were expressed with the Mef2:GAL4 system. Lines that gave rise to phenotypes were then tested in the secondary screen using the TARGET system to assay for adult-specific defects.

associated publications. All RNAi lines expressed short hairpin constructs under UAS control. For all TARGET experiments males from each UAS-RNAi line were crossed to females of the genotype *P[tubP-GAL80ts]9/FM7;;Mef2:GAL4/TM3*. The F1 progeny were raised at 18 °C until eclosion, at which point virgin females were collected and transferred to 29 °C to induce RNAi-construct expression. For the TARGET control line, the progeny of Oregon-R males and *P[tubP-GAL80ts]9/FM7;;Mef2:GAL4/TM3* females were raised and assayed at the same temperatures as the experimental line. For the Mef2:GAL4 screen, UAS-RNAi males were crossed to females with the genotype *y1,w*;;P[GAL4-Mef2.R]3*. As a control, Oregon-R males were crossed to *y1,w*;;P[GAL4-Mef2.R]* females. Flies in the Mef2:GAL4 screen were kept at 25 °C throughout development and adulthood.

It is important to note several limitations inherent in the design of our screen. The screen was designed to achieve high time-resolution analysis of muscle function in *Drosophila* adults so as to identify

both dramatic and subtle defects in muscle function. However, in achieving this goal, several compromises were necessary. First, the behavioural assay used tested only the performance of leg muscles. While other assays have been used to test flight muscle function, these assays typically result in the death of the tested flies at each time-point. Thus, it would not have been practical to perform the repeated measurements required for the sensitive analysis. However, different muscle sets have different gene expression patterns²⁴. Therefore, the results of this screen should be interpreted with the caveat that only leg muscle function was tested. Secondly, we did not directly confirm mRNA transcript knockdown using qPCR. As such it is possible that some RNAi constructs caused off-target effects. This could potentially lead to false-negative or false-positive results. To address this, we tested a random selection of genes with multiple RNAi constructs in both the Mef2:GAL4 and TARGET components of the screen (see Technical Validation).

Developmental lethality

In the Mef2:GAL4 screen, the RNAi construct was expressed throughout development (Fig. 1a). While the majority of lines tested resulted in adult viable flies, which were then assayed for climbing ability, many caused developmental lethality. We scored lethal phenotypes as either 'Embryo Lethal' or 'Pupal Lethal' depending on the stage at which development failed.

Negative geotaxis assay

The negative geotaxis assays were used to identify adult climbing defects in both the Mef2:GAL4 screen and the TARGET screen. Assays were carried out as previously described^{13,25} with modifications. Flies were collected immediately following eclosion and separated into batches of 10 in vials containing standard food media. For the climbing assay, the flies from one vial were transferred to an empty vial with a line drawn 7.5 cm from the bottom. Flies were tapped to the bottom to induce an innate climbing response. The flies were scored as a 'Success' if they climbed above the 7.5 cm line or a 'Failure' if they did not. Five technical replicates were performed to ensure an accurate reading for each vial at each time-point. The total number of 'Success' or 'Failure' events from the 5 technical replicates was recorded as the raw data (Data Record 2; Data Citation 2). Approximately five biological replicate vials for each experimental line were tested at each time-point. Control flies with the appropriate expression system (Mef2:GAL4 or TARGET) but lacking an RNAi construct were assayed in parallel with the experimental line. Flies were assayed every three days from eclosion to 30 days post-eclosion. However, when analysing the raw data, only the time-points up to Day 21 were used as too much variability was observed in the later time-points.

Data processing and statistical analysis

The raw climbing assay data (Data Record 2; Data Citation 2) were analysed in following manner. For each experimental fly line, and the control line, the observed data represented a Success/Failure ratio for each vial as given in the Raw Data.xlsx file in Data Record 2 (see also Data Citation 2). We designed a model, based on these raw data, which predicted the odds of a fly crossing the line at any given time-point:

$$\text{Odds} = e^{[(x_{con} + x_{exp}) + (y_{con} + y_{exp})t + (z_{con} + z_{exp})t^2]} \quad (1)$$

This approach allowed us to compare the entire set of data over the 21-day time-course instead of comparing only individual time-points. To determine the parameters of the model, the values for the x_{con} , x_{exp} , y_{con} , y_{exp} , z_{con} and z_{exp} coefficients in equation (1), we used the Generalized Estimating Equations (GEE) approach in the statistical analysis software 'R'^{26,27}. The GEE approach uses the observed Success/Failure data to determine the values for the x_{con} , x_{exp} , y_{con} , y_{exp} , z_{con} and z_{exp} coefficients that will allow the model in equation (1) to correctly predict the observed data. The coefficients are unique to each RNAi construct. The specific coefficient values for all the RNAi constructs are provided in Data Record 1, Coefficients.xlsx. t is time in days post-eclosion. Equation (1) calculates the odds of a fly crossing the line at any time. For example, to calculate the odds at Day 3 for a fly expressing the RNAi construct *bent* 46253 in the TARGET screen we use the relevant coefficient values in Data Record 1, Coefficients.xlsx and enter them into equation (1):

$$\text{Odds for bent 46253, Day 3} = e^{(2.4487 + 0.1378) + (-0.1392 - 0.2929)3 + (-0.0005 + 0.0036)3^2} \quad (2)$$

Which, when solved, gives the following odds value:

$$\text{Odds for bent 46253, Day 3} = 3.74 \quad (3)$$

Thus, the predicted odds for a *bent* RNAi 46253 fly climbing above the experimental line at Day 3 are 3.74 to 1. Expressed as a probability this would be 3.74/4.74 or 78% (Fig. 2b). By the same method, the control odds at this same timepoint can be calculated by replacing the x_{con} , x_{exp} , y_{con} , y_{exp} , z_{con} and z_{exp} values in equation (1) with the relevant values from Coefficients.xlsx in Data Record 1. As this is the control line, the x_{exp} , y_{exp} and z_{exp} are set to '0':

$$\text{Control odds, Day 3} = e^{(2.4487) + (-0.1392)3 + (-0.0005)3^2} \quad (4)$$

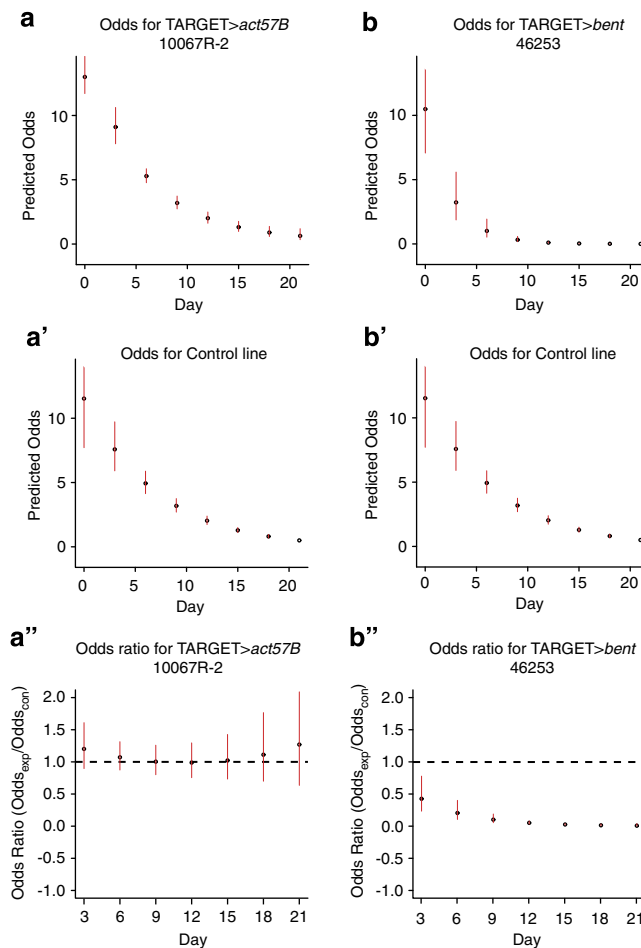


Figure 2. Sample graphs from statistical analysis. **(a)** Predicted odds for flies expressing the RNAi construct *act57B* 10067R-2 under the control of the TARGET system crossing the line over 21 days. **(a')** Predicted odds for flies from the control line crossing the line over 21 days. **(a'')** Odds ratio for the *act57B* 10067R-2 RNAi-construct-expressing flies. Dotted line indicates a ratio of 1. This line was scored as 'None' as the odds ratio never significantly differs from 1. The time-point Day 0 is excluded for simplicity as it was never observed to be significantly higher or lower than 1. **(b)** Predicted odds for flies expressing the RNAi construct *bent* 46253 under the control of the TARGET system crossing the line over 21 days. **(b')** Predicted odds for flies from the control line crossing the line over 21 days. **(b'')** Odds ratio for the *bent* 46253 RNAi-construct-expressing flies. Dotted line indicates a ratio of 1. This line was scored as 'Climbing Defect' as the odds ratio is significantly lower than 1 over the whole time-course ($P \leq 0.05$). The time-point Day 0 is excluded for simplicity as it was never observed to be significantly higher or lower than 1. Red lines are 95% confidence intervals in all panels.

Which, when solved, gives the following odds value:

$$\text{Control odds, Day 3} = 7.59 \quad (5)$$

Thus, the predicted odds for a control fly climbing above the experimental line at Day 3 are 7.59 to 1. Expressed as a probability this would be 7.59/8.59 or 88% (Fig. 2a' and b'). To compare the odds calculated for both the control line and the experimental line we calculated an odds ratio (OR):

$$\text{Odds Ratio} = \frac{\text{odds}_{\text{exp}}}{\text{odds}_{\text{con}}} \quad (6)$$

Using the experimental odds value calculated in equations (2) and (3), and the control odds value calculated in equations (4) and (5), we can find the OR for *bent* 46253 RNAi construct expressing flies

compared to the control flies at Day 3:

$$\text{Odds Ratio, bent RNAi 46253, Day 3} = \frac{3.74}{7.59} \quad (7)$$

When equation (7) is solved, an OR of 0.49 is obtained (Fig. 2b'').

The OR was interpreted as follows. An OR not significantly different from '1' indicates that there is no significant difference in climbing ability between the RNAi-construct-expressing line and the control line. An OR significantly lower than 1 indicates that the RNAi-construct-expressing line has a significantly lower climbing ability than the control line. The significance level used was $P \leq 0.05$. The advantage of analysing our data in this way was that it allowed us to include all the data for a given RNAi knockdown line when calculating the probability for a specific time-point as the GEE approach uses the entire dataset when generating the x_{con} , x_{exp} , y_{con} , y_{exp} , z_{con} and z_{exp} coefficients for equation (1). Thus, for lines with lower replicate Ns a higher statistical strength could be achieved. Furthermore, this analysis approach was optimized for data in which the decline in the odds over time is gradual so as to best fit our data. Finally, given the size of the datasets we needed to analyse the OR assessed visually, as in Figure 2, to expedite the phenotypic classification.

Criteria for phenotypic classification

Phenotypic classifications were made as follows. For the Mef2:GAL4 screen, vials containing RNAi-construct-expressing animals were visually inspected for developmental lethality. RNAi lines that did not progress beyond the embryonic stage were classified as 'Embryo Lethal'. Similarly, RNAi lines that did not progress beyond the pupal stage were classified as 'Pupal Lethal'. For RNAi-construct-expressing lines that survived until adulthood phenotypic classification was made based on climbing ability. This was done the same way for both the Mef2:GAL4 and TARGET screens. As described above, an odds ratio between the experimental line and control line at each time-point was calculated. These ORs were graphed and visually inspected to determine if the OR became significantly less than 1 at any time. If the OR became significantly lower than 1 for the remainder of the time-course, the fly line containing the experimental RNAi construct was classified as having a 'Climbing Defect' (Fig. 2b''). However, if the OR did not become less than 1, or became less than 1 but subsequently returned to being equal to or greater than 1, the experimental line was classified as 'None' (Fig. 2a'').

Examples of the predicted odds and OR for RNAi-construct-expressing animals that were scored as 'None' (*act57B 10067R-2*; Fig. 2a and a'') or 'Climbing Defect' (*bent 46253*; Fig. 2b and b'') in the TARGET screen are given in Figure 2.

Data Records

The data produced by this screen have been deposited at two online repositories. The phenotypic, analysed data have been uploaded as structured data to GenomeRNAi. The raw data from both the negative geotaxis assays as well as the data showing which genes were tested with multiple RNAi lines were uploaded to FigShare. Links to the data depositions are provided in the Data Citations. The format and column-by-column content for both depositions are described below.

Data record 1 – phenotypic data

The structured data are contained in tables indicating the screened genes and the resulting phenotypes were deposited as structured data with GenomeRNAi (Data Citation 1). The data in GenomeRNAi are organized by phenotype. More information regarding the column headings can be found at GenomeRNAi.

File 1 – Mef2:GAL4 screen for genes required for muscle development and maintenance

Gene ID. Each *Drosophila* gene is assigned a unique and permanent identifier: the CG number. The CG number is the best way to search this dataset or FlyBase for a gene of interest.

Gene symbol. The gene name is given as either the full name or the abbreviation as listed on FlyBase (flybase.org). Given the variability and inconsistencies in the names of *Drosophila* genes, the most reliable method for searching this dataset for a specific gene is by Gene ID or CG number (Column A).

Reagent ID. Each of the RNAi stock centres listed above has a unique identifier for each RNAi hairpin construct they generate and insert into the fly. These numbers enable the community to order the specific RNAi construct that we used to generate a phenotype. This is important as different RNAi constructs targeting the same gene can produce a range of phenotypes dependent on tissue, expression system and targeted sequence. We obtained the RNAi lines from three sources: the NIG-Fly Stock Centre (Kyoto, Japan; <http://www.shigen.nig.ac.jp/fly/nigfly/index.jsp>), the Vienna *Drosophila* RNAi Centre (VDRC, Vienna, Austria; <http://stockcenter.vdrc.at/>) and the Transgenic RNAi Project (TRiP, Boston, USA; <http://www.flyrnai.org/>).

Score. The score that was used to choose hits. The abbreviation 'np' indicates that the score was not provided. In our screen, it was not possible to give a single score for each line as hits were determined either by a qualitative description based on the point at which the flies died during development or on climbing ability. Furthermore, climbing ability was assayed and analysed over a 30-day time-course.

More detailed information is included in the file 'Raw Data.xlsx' at FigShare, which contain the raw data for the Mef2:GAL4 screen and TARGET screen respectively (Data Citation 2).

Phenotype. As outlined above, we conducted a preliminary screen to eliminate the RNAi lines that had no effect on muscle health at any stage (Mef2:GAL4 screen). We expressed RNAi constructs under the control of Mef2, a muscle-specific promoter. Phenotypes were classified as 'Embryo Lethal', 'Pupal Lethal', 'Climbing Defect' or 'None'.

Conditions. Description of conditions under which flies were tested. This column is included to conform to the requirements of GenomeRNAi. For our screens, all lines were tested in the same manner (see Methods).

Follow up. Indicates whether or not the RNAi construct was analysed in the TARGET screen.

Comments. We obtained the RNAi lines from three sources: the NIG-Fly Stock Centre (Kyoto, Japan; <http://www.shigen.nig.ac.jp/fly/nigfly/index.jsp>), the Vienna *Drosophila* RNAi Centre (VDRC, Vienna, Austria; <http://stockcenter.vdrc.at/>) and the Transgenic RNAi Project (TRiP, Boston, USA; <http://www.flyrnai.org/>).

File 2 – TARGET screen for genes required for muscle maintenance

Columns are the same as in Data Record 1, File 1 apart from 'Phenotype'.

Phenotype. RNAi lines were classified as either 'Climbing Defect' or 'None' based on the outcome of the negative geotaxis assays.

Data record 2 – raw data, duplicate line data and coefficient files

The data deposited with FigShare are formatted as three downloadable spreadsheets (Data Citation 2). The column details are given below.

File 1. raw data.xlsx

This data record contains four sheets, two containing the raw data (Mef2:GAL4 Screen and TARGET Screen) and two containing information regarding the stock centre source of each tested line (Mef2:GAL4 Stock centre Source and TARGET Stock centre Source).

The column descriptions for the raw data sheets are as follows:

Column A – gene symbol. The gene name is given as either the full name or the abbreviation as listed on FlyBase (flybase.org). Given the variability and inconsistencies in the names of *Drosophila* genes, the most reliable method for searching this dataset for a specific gene is by Gene ID or CG number (Column A).

Column B – gene ID. Each *Drosophila* gene is assigned a unique and permanent identifier: the CG number. The CG number is the best way to search this dataset or FlyBase for a gene of interest.

Column C – reagent ID. Each of the RNAi stock centres listed above has a unique identifier for each RNAi hairpin construct they generate and insert into the fly. These numbers enable the community to order the specific RNAi construct that we used to generate a phenotype. This is important as different RNAi constructs targeting the same gene can produce a range of phenotypes dependent on tissue, expression system and targeted sequence. We obtained the RNAi lines from three sources: the NIG-Fly Stock Centre (Kyoto, Japan; <http://www.shigen.nig.ac.jp/fly/nigfly/index.jsp>), the Vienna *Drosophila* RNAi Centre (VDRC, Vienna, Austria; <http://stockcenter.vdrc.at/>) and the Transgenic RNAi Project (TRiP, Boston, USA; <http://www.flyrnai.org/>).

Column D – analysis ID. For the purposes of the analysis, each replicate for each RNAi construct analysed is given a unique identifier. This allows the analysis script to differentiate between biological replicates of one RNAi construct.

Column E – day. Days of RNAi-construct expression on which flies were assayed for climbing ability. If a time-point was missed no data are given for that time-point as the analysis script is not able to process non-numeric values.

Column F – flies. The number of flies tested for each replicate. If a fly escaped between time-points, which occasionally happened while flies were transferred between the assay vial and food vial, the number of flies was reduced accordingly. However, if a fly died, then the number was not changed. Dead flies are interpreted as non-climbers for the purposes of the climbing assay.

Column G – ratio. The ratio of successes to failures in the climbing assay expressed as a proportion of 1, where 1 is a 100% success rate and 0 is a 100% failure rate. With this value, climbing ability can be quickly and easily graphed.

Columns H and I – success and failure. Raw data representing the number of flies that succeeded, or failed, to climb 7.5 cm in 8 s. Data are the combination of the 5 technical replicates performed for each

biological replicate vial at each time-point. The column descriptions for the stock centre source data sheets are as follows:

Column A – gene symbol. The gene name is given as either the full name or the abbreviation as listed on FlyBase (flybase.org). Given the variability and inconsistencies in the names of *Drosophila* genes, the most reliable method for searching this dataset for a specific gene is by Gene ID or CG number (Column A).

Column B – gene ID. Each *Drosophila* gene is assigned a unique and permanent identifier: the CG number. The CG number is the best way to search this dataset or FlyBase for a gene of interest.

Column C – reagent ID. Each of the RNAi stock centres listed above has a unique identifier for each RNAi hairpin construct they generate and insert into the fly. These numbers enable the community to order the specific RNAi construct that we used to generate a phenotype. This is important as different RNAi constructs targeting the same gene can produce a range of phenotypes dependent on tissue, expression system and targeted sequence.

Column D – stock centre. We obtained the RNAi lines from three sources: the NIG-Fly Stock Centre (Kyoto, Japan; <http://www.shigen.nig.ac.jp/fly/nigfly/index.jsp>), the Vienna *Drosophila* RNAi Centre (VDRC, Vienna, Austria; <http://stockcenter.vdrc.at/>) and the Transgenic RNAi Project (TRiP, Boston, USA; <http://www.flyrnai.org/>).

File 2. duplicate line data.xlsx

This data record contains two sheets, one for the Mef2:GAL4 screen and one for the TARGET screen. The column descriptions are the same for both sheets. For the comparison of the secondary lines, ‘Embryo Lethal’, ‘Pupal Lethal’ and ‘Climbing Defect’ were considered the same outcome in that the expression of the RNAi construct had caused the phenotype, albeit one of varying severity. In each sheet, the lines that did not produce similar results are highlighted in red.

Column A – gene ID. Each *Drosophila* gene is assigned a unique and permanent identifier: the CG number. The CG number is the best way to search this dataset or FlyBase for a gene of interest.

Column B – gene symbol. The gene name is given as either the full name or the abbreviation as listed on FlyBase (flybase.org). Given the variability and inconsistencies in the names of *Drosophila* genes, the most reliable method for searching this dataset for a specific gene is by Gene ID or CG number (Column A).

Column C – reagent ID. Each of the RNAi stock centres listed above has a unique identifier for each RNAi hairpin construct they generate and insert into the fly. These numbers enable the community to order the specific RNAi construct that we used to generate a phenotype. This is important as different RNAi constructs targeting the same gene can produce a range of phenotypes dependent on tissue, expression system and targeted sequence.

Column D – source. We obtained the RNAi lines from three sources: the NIG-Fly Stock Centre (Kyoto, Japan; <http://www.shigen.nig.ac.jp/fly/nigfly/index.jsp>), the Vienna *Drosophila* RNAi Centre (VDRC, Vienna, Austria; <http://stockcenter.vdrc.at/>) and the Transgenic RNAi Project (TRiP, Boston, USA; <http://www.flyrnai.org/>).

Column E – phenotype. Phenotype caused by expression of the RNAi construct using either the Mef2:GAL4 or TARGET system.

File 3. coefficients.xlsx

This data record contains the coefficient values used in equation (1) to solve for the odds of a fly crossing the experimental line. These values are predicted based on the experimentally observed probabilities listed in File 1 of this Data Record (Raw Data.xlsx). There are two sheets in this data record, one for the Mef2:GAL4 screen data and one for the TARGET data. Column descriptions are the same for both sheets.

Column A – gene ID. Each *Drosophila* gene is assigned a unique and permanent identifier: the CG number. The CG number is the best way to search this dataset or FlyBase for a gene of interest.

Column B – gene symbol. The gene name is given as either the full name or the abbreviation as listed on FlyBase (flybase.org). Given the variability and inconsistencies in the names of *Drosophila* genes, the most reliable method for searching this dataset for a specific gene is by Gene ID or CG number (Column A).

Column C – reagent ID. Each of the RNAi stock centres listed above has a unique identifier for each RNAi hairpin construct they generate and insert into the fly. These numbers enable the community to order the specific RNAi construct that we used to generate a phenotype. This is important as different RNAi constructs targeting the same gene can produce a range of phenotypes dependent on tissue, expression system and targeted sequence.

Column D – source. We obtained the RNAi lines from three sources: the NIG-Fly Stock Centre (Kyoto, Japan; <http://www.shigen.nig.ac.jp/fly/nigfly/index.jsp>), the Vienna *Drosophila* RNAi Centre (VDRC,

Vienna, Austria; <http://stockcenter.vdrc.at/>) and the Transgenic RNAi Project (TRiP, Boston, USA; <http://www.flyrnai.org/>).

Column E – X-value. Value of coefficient ‘X’ as predicted by the GEE approach from experimental data for use in equation (1).

Column F – Y-value. Value of coefficient ‘Y’ as predicted by the GEE approach from experimental data for use in equation (1).

Column G – Z-value. Value of coefficient ‘Z’ as predicted by the GEE approach from experimental data for use in equation (1).

Technical Validation

Validity of the screen

The methodology presented in this dataset has been validated in several ways. First, negative controls were performed to ensure that the genetic elements required to induce RNAi expression did not cause lethality or climbing defects in the absence of the RNAi. As expected, these flies were able to reach adulthood, had lifespans comparable to completely wild-type flies and had no defect in climbing ability. Secondly, multiple RNAi lines from various RNAi stock centres were analysed for a randomized selection of genes. We saw a high frequency of phenotype confirmation in both the Mef2:GAL4 screen (93%) and the TARGET screen (88%) indicating that off-target RNAi effects were not a significant issue for our screen, although the potential for false negatives or false positives should be noted (Data Record 2; Data Citation 2). The most comprehensive *in vivo* RNAi screen undertaken in *Drosophila* for muscle defects had a false-negative rate of 5% and a false-positive rate of 1.3% indicating that *in vivo* RNAi screening in *Drosophila* is a reliable method for identifying genes required for muscle function⁵. We previously tested the TARGET system to confirm its ability to restrict and induce gene expression. GFP transgene expression under the TARGET system was 11 times higher at the permissive temperature (29 °C) compared to the restrictive temperature (18 °C)¹⁰. Additionally we have previously shown that induction of RNAi in adult flies via the TARGET system causes a significant reduction in both mRNA transcript levels and protein levels¹⁰.

Validity of the statistical model

Linear regression analysis (Fig. 3) was performed using the predicted and observed probability values from 50 RNAi construct lines over the 21 day time-course. To demonstrate that the model we designed was able to provide an accurate analysis of our data, we tested the hypothesis that the predicted and observed probabilities fit the linear relationship $y = x$ (Fig. 3). That is, the predicted probability value at a given time-point is the same as the observed probability at that time-point for flies of the same genotype. Linear regression analysis of the fit gave an R^2 -value of 0.82. The null hypothesis, that the predicted and observed values did not fit the linear equation $y = x$, was rejected with a P -value of < 0.005 , thus confirming the validity of our model (Fig. 3).

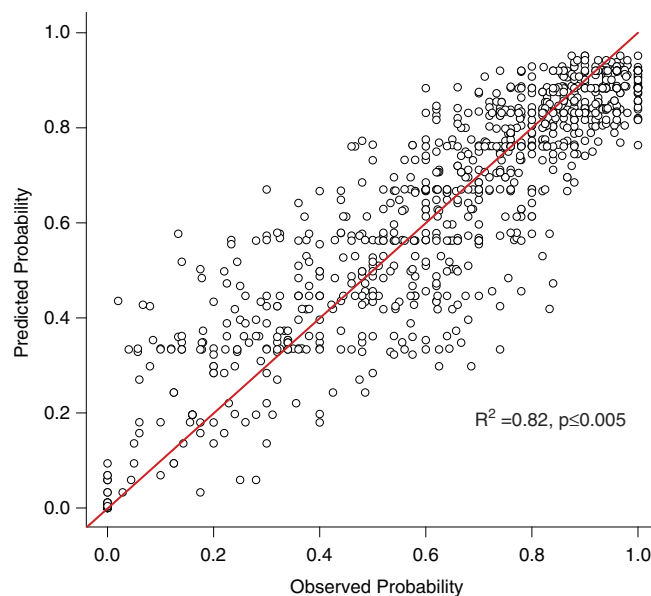


Figure 3. Validation of the statistical model. Linear fit of the predicted probability to the observed probability of a fly crossing the line. Each dot represents a given RNAi construct at a given time-point. The red line shows the line $y = x$. Note clustering of points around the $y = x$ line.

Usage Notes

There are several potential uses for this dataset. First, it suggests candidate genes for further, more detailed analysis. Secondly, it offers the chance to compare the set of genes required for muscle development to those required for muscle maintenance (Data Record 1). The overlap or differences between the two could provide insight into the underlying mechanisms of both processes. Lastly, it could be analysed to reveal which complexes or gene-products are key to muscle maintenance. This analysis can be performed using the software package Cytoscape. Cytoscape is an open source software program that allows for complex network analysis based on user-defined parameters. The *Drosophila* Interaction Database (www.DroID.org), a comprehensive protein and gene interaction database, codes and supports a Cytoscape plug-in to assist in the analysis of *Drosophila* genomics and proteomics. By using this plug-in it is possible to organize the dataset by any number of criteria. For example, annotated terms from the Gene Ontology (www.geneontology.org) can be combined with the DroID plug-in in Cytoscape to organize the dataset by biological process, molecular function or specific roles within the cell. Cytoscape is publicly available at www.cytoscape.org. The DroID plug-in is available from <http://www.droidb.org/CytoscapePluginHelp.jsp>.

References

- Clark, K. A., McElhinny, A. S., Beckerle, M. C. & Gregorio, C. C. Striated muscle cytoarchitecture: an intricate web of form and function. *Annu. Rev. Cell. Dev. Biol.* **18**, 637–706 (2002).
- Bergmann, O *et al.* Evidence for cardiomyocyte renewal in humans. *Science* **324**, 98–102 (2009).
- Augustin, H. & Partridge, L. Invertebrate models of age-related muscle degeneration. *Biochim. Biophys. Acta.* **1790**, 1084–1094 (2009).
- Sparrow, J, Hughes, S. M. & Segalat, L. Other model organisms for sarcomeric muscle diseases. *Adv. Exp. Med. Bio.* **642**, 192–206 (2008).
- Rando, T. A. Oxidative stress and the pathogenesis of muscular dystrophies. *Am. J. Phy. Med. Rehabil.* **81**, S175–186 (2002).
- Schnorrer, F *et al.* Systematic genetic analysis of muscle morphogenesis and function in *Drosophila*. *Nature* **464**, 287–291 (2010).
- Bai, J *et al.* RNA interference screening in *Drosophila* primary cells for genes involved in muscle assembly and maintenance. *Development* **135**, 1439–1449 (2008).
- Kucherenko, M. M, Marrone, A. K, Rishko, V. M, Magliarelli, Hde, F. & Shcherbata, H. R. Stress and muscular dystrophy: a genetic screen for dystroglycan and dystrophin interactors in *Drosophila* identifies cellular stress response components. *Dev. Biol.* **352**, 228–242 (2011).
- Kucherenko, M. M *et al.* Genetic modifier screens reveal new components that interact with the *Drosophila* dystroglycan-dystrophin complex. *PLoS One* **3**, e2418 (2008).
- Allikian, M. J *et al.* Reduced life span with heart and muscle dysfunction in *Drosophila* sarcoglycan mutants. *Hum. Mol. Genet.* **16**, 2933–2943 (2007).
- Bassett, D. I. & Currie, P. D. The zebrafish as a model for muscular dystrophy and congenital myopathy. *Hum. Mol. Genet.* **12** (Spec No 2): R265–270 (2003).
- Pai, A. C. Developmental Genetics of a Lethal Mutation, Muscular Dysgenesis (Mdg), in the Mouse. I. Genetic Analysis and Gross Morphology. *Dev. Biol.* **11**, 82–92 (1965).
- Perkins, A. D *et al.* Integrin-mediated adhesion maintains sarcomeric integrity. *Dev. Biol.* **338**, 15–27 (2010).
- Thornhill, P, Bassett, D, Lochmuller, H, Bushby, K. & Straub, V. Developmental defects in a zebrafish model for muscular dystrophies associated with the loss of fukutin-related protein (FKRP). *Brain* **131**, 1551–1561 (2008).
- Udaka, J *et al.* Disuse-induced preferential loss of the giant protein titin depresses muscle performance via abnormal sarcomeric organization. *J. Gene Physiol.* **131**, 33–41 (2008).
- Barton, B, Ayer, G, Maughan, D. W. & Vigoreaux, J. O. Site directed mutagenesis of *Drosophila* flightin disrupts phosphorylation and impairs flight muscle structure and mechanics. *J. Muscle Res. Cell Motil.* **28**, 219–230 (2007).
- Contompasis, J. L, Nyland, L. R, Maughan, D. W. & Vigoreaux, J. O. Flightin is necessary for length determination, structural integrity, and large bending stiffness of insect flight muscle thick filaments. *J. Mol. Biol.* **395**, 340–348 (2010).
- Kronert, W. A *et al.* Defects in the *Drosophila* myosin rod permit sarcomere assembly but cause flight muscle degeneration. *J. Mol. Biol.* **249**, 111–125 (1995).
- Goldstein, L. S. & Gunawardena, S. Flying through the *Drosophila* cytoskeletal genome. *J. Cell. Biol.* **150**, F63–68 (2000).
- Ranganayakulu, G, Elliott, D. A, Harvey, R. P. & Olson, E. N. Divergent roles for NK-2 class homeobox genes in cardiogenesis in flies and mice. *Development* **125**, 3037–3048 (1998).
- McGuire, S. E, Le, P. T, Osborn, A. J, Matsumoto, K. & Davis, R. L. Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302**, 1765–1768 (2003).
- Dietzl, G *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
- Ni, J. Q *et al.* A *Drosophila* resource of transgenic RNAi lines for neurogenetics. *Genetics* **182**, 1089–1100 (2009).
- Elgar, S. J, Han, J. & Taylor, M. V. *mef2* activity levels differentially affect gene expression during *Drosophila* muscle development. *Proc. Nat. Acad. Sci. USA* **105**, 918–923 (2008).
- Leal, S. M. & Neckameyer, W. S. Pharmacological evidence for GABAergic regulation of specific behaviors in *Drosophila melanogaster*. *J. Neurobiol.* **50**, 245–261 (2002).
- Prentice, R. L. & Zhao, L. P. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**, 825–839 (1991).
- Zeger, S. L. & Liang, K. Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121–130 (1986).

Data Citations

- Perkins, A. D., Lee, M. & Tanentzapf, G. GenomeRNAi GR00238-S (2014).
- Perkins, A. D., Lee, M. & Tanentzapf, G. Figshare <http://dx.doi.org/10.6084/m9.figshare.806269> (2014).

Acknowledgements

We are indebted to Dr John Petkau and Yanling Cai from the Department of Statistics, University of British Columbia, for their assistance in developing the statistical model.

Author Contributions

ADP designed the screens, did the experiments, analysed the data and prepared the manuscript. MJJL assisted with the design of the screens, the experiments and the data analysis. GT designed the screens, supervised the work and prepared the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Perkins, A. D. *et al.* The systematic identification of cytoskeletal genes required for *Drosophila melanogaster* muscle maintenance. *Sci. Data*. 1:140002 doi: 10.1038/sdata.2014.2 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Metadata associated with this Data Descriptor is available at <http://www.nature.com/sdata/> and is released under the CC0 waiver to maximize reuse.