Data in Brief 43 (2022) 108482

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Gene expression profiling of NUAK kinase overexpression in *Drosophila* larval muscle development

David Brooks, Erika R Geisbrecht*

Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS 66506, USA

ARTICLE INFO

Article history:
Received 24 June 2022
Revised 13 July 2022
Accepted 18 July 2022
Available online 22 July 2022

Dataset link: Gene expression profiling of NUAK kinase overexpression in Drosophila larval muscle development (Original data)

Keywords:
RNA-seq
Drosophila melanogaster
larvae
NUAK
kinase

ABSTRACT

Signal transduction pathways mediated by kinases control diverse biological outputs at the level of cells and tissues to regulate a diverse array of biological and developmental events. To gain insight into how muscle expression of the evolutionarily conserved NUAK kinase regulates the transcriptional landscape during *Drosophila melanogaster* development, we performed high-throughput sequencing of RNA from either whole larvae or dissected muscle fillets at the end of larval development. Raw data was generated using the Illumina HiSeq 4000 platform. After trimming and mapping to the *Drosophila* reference genome, differential gene expression and GO enrichment analysis were completed. Raw data are deposited in the NCBI Gene Expression Ominbus (GEO) repository under GEO accession GSE204894.

DOI of original article: 10.1016/j.ydbio.2022.06.014
* Corresponding author.
E-mail address: geisbrechte@ksu.edu (E.R. Geisbrecht).
Social media: @GeisbrechtE_Lab (E.R. Geisbrecht)

https://doi.org/10.1016/j.dib.2022.108482

Specifications Table

Subject	Biological Sciences
Specific subject area	Developmental Biology; Molecular Biology; Transcriptomics
Type of data	RNA-seq data, Tables, Figures
How the data were acquired	RNA sequencing by Illumina HiSeq 4000 Software: Trimmomatic v.0.36, STAR aligner v.2.5.2b, featureCounts from Subread v.1.5.2, DESeq2, Gorilla
Data format	Raw Analyzed
Description of data collection	Total RNA was isolated from either wandering L3 whole larvae or filleted muscle carcasses from control (*mef2>lacZ*) or NUAK overexpression (*mef2>NUAK 548* or *mef2>NUAK 550*) samples. Three biological replicates were prepared for each genotype. After assessment of RNA quality, libraries were constructed using standard Illumina protocols, and sequenced.
Data source location	Kansas State University, Manhattan, KS
Data accessibility	Raw and analyzed RNA-Seq data were deposited in the NCBI GEO database under GEO accession GSE204894. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE204894)
Related research article	D. Brooks, S. Bawa, A. Bontrager, M. Stetsiv, Y. Guo, E.R. Geisbrecht, Independent pathways control muscle tissue size and sarcomere remodeling, Dev Biol 490 (2022) 1-12. https://doi.org/10.1016/j.ydbio.2022.06.014.

Value of the Data

- These data provide a comprehensive overview of transcriptomic changes that occur as a result of NUAK kinase overexpression in muscle tissue.
- This dataset and accompanying analysis will be useful to *Drosophila* researchers who study kinases and/or the intersection between muscle tissue and larval development.
- Results in these datasets can be further mined to directly identify individual genes and/or pathways that are misregulated in response to aberrant kinase signaling. Since data outputs are either from muscle tissue or whole larvae, the associated gene expression changes may provide information about inter-tissue communication during larval development.

1. Data Description

To investigate transcriptional changes that contribute to a reduction in larval and pupal body size upon expression of the serine/threonine NUAK kinase in muscle tissue [1], we used RNA-seq (Fig. 1). The bipartite UAS/Gal4 system [2] was used to express two independent NUAK transgenes (UAS-*NUAK 548* or UAS-*NUAK 550*) or a control transgene (UAS-lacZ) in developing muscle with *mef2*-Gal4. RNA-seq was performed on either wandering whole larvae or dissected muscle fillets with three biological replicates of each genotype. RNA library preparation with polyA selection and Illumina HiSeq 2 $\times$ 150bp sequencing was performed. The total library size and mapping statistics are provided in Table 1, with more than 94% of total reads uniquely mapped.

1.1. The obtained RNA-seq data are appropriate for differential gene expression analysis

PCA analysis of log transformed counts from each genotype generally revealed the expected groupings among replicates within samples and across different sample groups (Fig. 2A). Overexpression of NUAK compared to control samples was confirmed by assessing the relative abundance of transcripts in Reads Per Kilobase Million (RPKM) (Fig. 2B) and verifies the suitability of this data for differential expression analysis. Differentially expressed genes were determined using DESeq2 and the Wald test was used to define significance as p-value $<$ 0.05

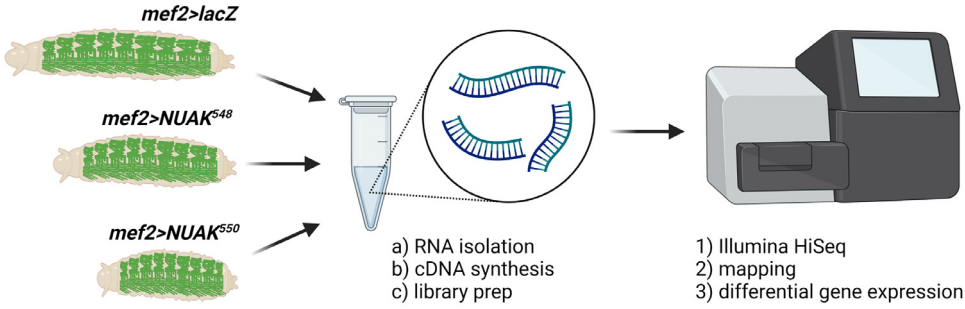

Fig. 1. Schematic overview of experimental design. Either UAS-lacZ, UAS-NUAK 548, or UAS-NUAK 550 were expressed in muscle tissue under control of the mef2 promoter. Total RNA isolated from either whole larvae or dissected muscle carcasses were polyA-selected and used to prepare libraries for Illumina HiSeq sequencing. Mapping of trimmed reads was followed by analysis of relative transcript changes across different genotypes.

Table 1
RNA-Seq library size and mapping statistics.

Sample name	Number of raw reads	Number of uniquely mapped reads	% Unique mapped reads
mef2>lacZ whole larvae replicate 1	23617983	22537806	95.4
mef2>lacZ whole larvae replicate 2	19423407	18561995	95.5
mef2>lacZ whole larvae replicate 3	20747462	19915420	95.9
mef2>NUAK548 whole larvae replicate 1	22136606	21209696	95.8
mef2>NUAK548 whole larvae replicate 2	20880347	19993578	95.7
mef2>NUAK548 whole larvae replicate 3	19026288	18323885	96.3
mef2>NUAK550 whole larvae replicate 1	17322652	16504068	95.2
mef2>NUAK550 whole larvae replicate 2	20806708	19742045	94.9
mef2>NUAK550 whole larvae replicate 3	25760507	24514194	95.1
mef2>lacZ muscle carcass replicate 1	19837827	18899008	95.2
mef2>lacZ muscle carcass replicate 2	22943306	21975867	95.7
mef2>lacZ muscle carcass replicate 3	19958901	19139082	95.7
mef2>NUAK548 muscle carcass replicate 1	20677137	19927007	96.3
mef2>NUAK548 muscle carcass replicate 2	20853004	20076398	96.2
mef2>NUAK548 muscle carcass replicate 3	19108287	18368233	96.1
mef2>NUAK550 muscle carcass replicate 1	21556984	20737964	96.2
mef2>NUAK550 muscle carcass replicate 2	25920057	24780240	95.6
mef2>NUAK550 muscle carcass replicate 3	21055121	20020999	95.0

and absolute log2 fold change $>$ 1. Volcano plots confirm global transcriptional changes between control and NUAK overexpression conditions in both whole larvae and muscle samples (Fig. 2C).

1.2. GO enrichment

Significant differentially expressed genes were further analyzed for GO classifications using the GOrilla online analysis software. GO terms featuring biological processes are shown in Table 2 for whole larvae and Table 3 for muscle carcass samples.

Table 2
Enriched GO terms categorized by biological process in whole larvae RNA seq samples.

Genotype	ID	GO term	Count	P value
Whole larvae				
mef2>NUAK 548 vs *mef2>lacZ*	GO:0042398	cellular modified amino acid biosynthetic process	4	6.32E-06
	GO:0005975	carbohydrate metabolic process	12	6.38E-06
	GO:0006575	cellular modified amino acid metabolic process	7	3.03E-05
	GO:0009066	aspartate family amino acid metabolic process	4	2.29E-04
	GO:0006528	asparagine metabolic process	2	7.04E-04
	GO:0009109	coenzyme catabolic process	2	7.04E-04
mef2>NUAK 550 vs *mef2>lacZ*	GO:0003341	cilium movement	22	3.82E-13
	GO:0044782	cilium organization	36	5.25E-08
	GO:0007283	spermatogenesis	44	1.07E-06
	GO:0035082	axoneme assembly	16	1.75E-06
	GO:0070286	axonemal dynein complex assembly	12	2.91E-06
	GO:0000003	reproduction	41	8.79E-06
	GO:0120031	plasma membrane bounded cell projection assembly	33	2.15E-05
	GO:0006936	muscle contraction	10	1.03E-04
	GO:0072522	purine-containing compound biosynthetic process	40	1.14E-04
	GO:0060285	cilium-dependent cell motility	7	2.93E-04
	GO:0070585	protein localization to mitochondrion	22	5.97E-04
	GO:0043648	dicarboxylic acid metabolic process	14	8.02E-04
	GO:0044281	small molecule metabolic process	179	8.59E-04
	GO:0003012	muscle system process	21	8.81E-04
	GO:0006096	glycolytic process	24	8.82E-04

Table 3
Enriched GO terms categorized by biological process in muscle carcass RNA seq samples.

Genotype	ID	GO term	Count	P value
Muscle carcass				
mef2>NUAK 548 vs *mef2>lacZ*	GO:0007594	puparial adhesion	7	8.73E-10
	GO:1901605	alpha-amino acid metabolic process	16	2.02E-07
	GO:0006566	threonine metabolic process	4	2.11E-06
	GO:0009066	aspartate family amino acid metabolic process	8	6.70E-06
	GO:0006520	cellular amino acid metabolic process	16	1.08E-04
	GO:0044282	small molecule catabolic process	15	1.91E-04
	GO:0019752	carboxylic acid metabolic process	25	3.44E-04
	GO:1901607	alpha-amino acid biosynthetic process	7	4.30E-04
	GO:0009081	branched-chain amino acid metabolic process	4	5.61E-04
	GO:0006082	organic acid metabolic process	25	6.34E-04
mef2>NUAK 550 vs *mef2>lacZ*	GO:0042335	cuticle development	25	4.25E-09
	GO:0030497	fatty acid elongation	8	2.99E-06
	GO:0032504	multicellular organism reproduction	15	8.55E-06
	GO:0006959	humoral immune response	9	3.27E-05
	GO:0003012	muscle system process	7	5.30E-05
	GO:0050830	defense response to Gram-positive bacterium	9	5.55E-05
	GO:0006633	fatty acid biosynthetic process	11	7.99E-05
	GO:0030148	sphingolipid biosynthetic process	9	9.05E-05
	GO:0072330	monocarboxylic acid biosynthetic process	12	3.64E-04
	GO:0009074	aromatic amino acid family catabolic process	4	4.90E-04
	GO:0009617	response to bacterium	20	6.80E-04
	GO:0044281	small molecule metabolic process	52	9.77E-04

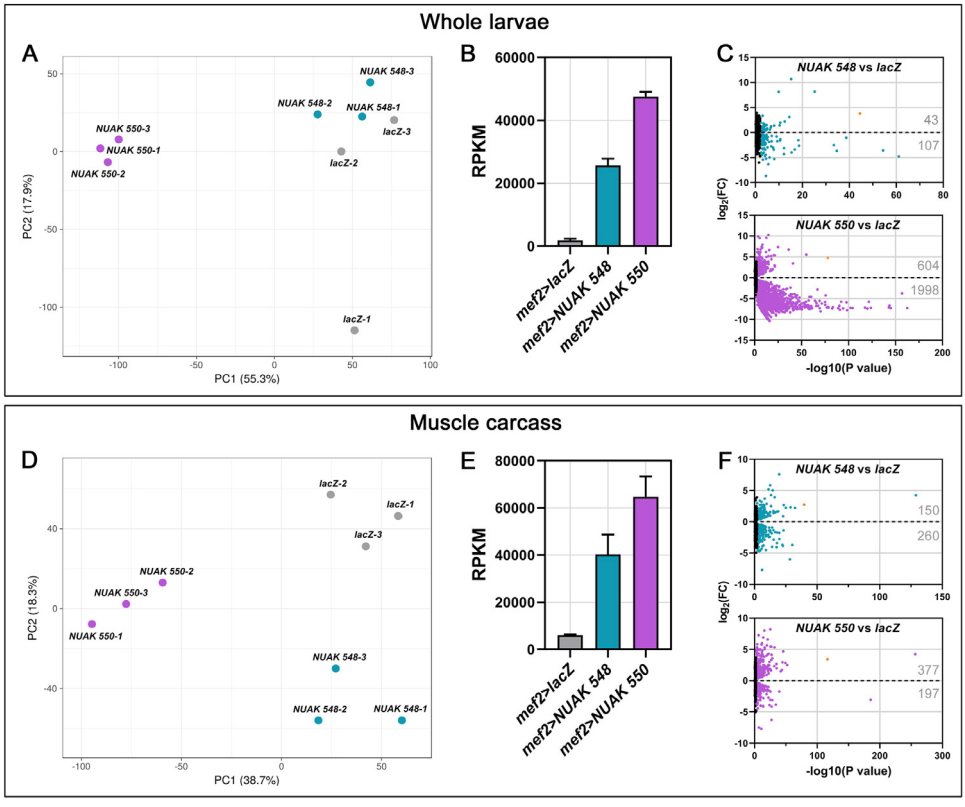

Fig. 2. Assessment of inter- and intragroup variability for gene expression analysis. (A-C) Whole larvae samples. (D-F) Muscle carcass samples. (A,D) Principal component analysis (PCA) for three replicates of control (mef2>lacZ) or three replicates for each NUAK transgene (mef2>NUAK 548 or mef2>NUAK 550). X and Y axes correspond to the first two principle components. (B,E) Barplots depicting NUAK 548 or NUAK 550 transcript expression counts as Reads Per Kilobase Million (RPKM). (C,F) Volcano plots of differential gene expression analysis for the indicated genotypes. Each data point in the scatter plot represents a gene. The log2 fold change of each gene is represented on the y-axis and the log10 of its adjusted p-value is on the x-axis. Numbers in grey above or below the black dotted line correspond to the number of target genes upregulated or downregulated, respectively. Orange circle corresponds to NUAK.

2. Experimental Design, Materials and Methods

2.1. Drosophila genetics

Flies were maintained on standard cornmeal-yeast-agar medium at 25°C. *mef2*-Gal4 females from the Bloomington Stock Center (BL27390) were mated to males of the following genotypes: UAS-*lacZ* (BL3956), UAS-*NUAK 548* [3], or UAS-*NUAK 550* [3] for RNA isolation.

2.2. RNA isolation and library preparation

Both male and female third instar larvae (L3) were combined for total RNA extraction according to manufacturer protocol using the RNeasy Mini Kit (Qiagen, Germnay). For each genotype (*mef2>lacZ, mef2>NUAK 548*, or *mef2>NUAK 550*), RNA was isolated from either three whole larvae or ten dissected muscle fillets for each biological replicate. Muscle fillets were prepared by pinning L3 larvae on Sylgard plates followed by the removal of fat body and other internal

organs in Phosphate Buffered Saline (PBS). The quality of RNA prepared from three individual biological replicates of each genotype was analyzed using the Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). High quality RNA ($\geq$2 µg; OD260/280 – 1.8-2.2) was sent to Genewiz for further quality assessment using the Qubit RNA Assay to measure RNA concentration ($\geq$50 ng/µl) and the Agilent Bioanalyzer to assess RNA quality (RIN$\geq$6.0). The NEBNext Ultra II RNA Library Prep Kit was used to prepare the RNA-seq according to standard Illumina polyA selection protocols.

2.3. Sequencing, raw data processing, and data analysis

The RNA-seq libraries were sequenced using the Illumina HiSeq 4000 (2 $\times$ 150 bp sequencing). Poor quality regions and adapter sequences were trimmed with Trimmomatic v.0.36 [4]. The trimmed reads were aligned with the *Drosophila melanogaster* BDGP6 reference genome available on ENSEMBL using the STAR aligner v.2.5.2b [5] to generate .bam files.

PCA plots were generated using log transformed counts imported into the Clustvis software (https://biit.cs.ut.ee/clustvis/) [6]. For either whole larvae or muscle carcass plots, unit variance scaling was applied to rows and SVD with imputation was used to calculate principal components. The X and Y axes show principal component 1 and principal component 2, with the total variance listed in the axes. N = 9 data points for each.

2.4. Differential gene expression and Gene ontology (GO) analysis

Unique gene hit counts, calculated using featureCounts from the Subread package v.1.5.2, were used for downstream differential expression analysis. DESeq2 was used to compare gene expression between experimental and control samples. P-values and log2 fold changes were calculated using the Wald test. Adjusted p-value $<$ 0.05 and absolute log2 fold change $>$ 1 were used as cut-offs for differentially expressed genes. Volcano plots were generated in GraphPad Prism 9.2.0 using genes designated as significant after differential gene expression analysis.

Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla) (http://cbl-gorilla.cs.technion.ac.il) was used to identify significantly enriched GO terms featuring biological processes with P-values less than 0.05 [7].

Ethics Statements

This work does not contain any experiments with humans, animals, or social media platforms.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Gene expression profiling of NUAK kinase overexpression in Drosophila larval muscle development (Original data) (NCBI GEO).

CRediT Author Statement

David Brooks: Conceptualization, Methodology, Investigation, Writing – review & editing; **Erika R Geisbrecht:** Conceptualization, Software, Data curation, Supervision, Writing – original draft, Funding acquisition.

Acknowledgments

Funding: This work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) of the National Institutes of Health (NIH) under award number RO1AR060788 and the Supplements to Advance Research (STAR) program, both awarded to E.R.G.

References

[1] D. Brooks, S. Bawa, A. Bontrager, M. Stetsiv, Y. Guo, E.R. Geisbrecht, Independent pathways control muscle tissue size and sarcomere remodeling, Dev Biol 490 (2022) 1–12, doi:10.1016/j.ydbio.2022.06.014.
[2] A.H. Brand, N. Perrimon, Targeted gene expression as a means of altering cell fates and generating dominant phenotypes, Development 118 (2) (1993) 401–415, doi:10.1242/dev.118.2.401.
[3] D. Brooks, F. Naeem, M. Stetsiv, S.C. Goetting, S. Bawa, N. Green, C. Clark, A. Bashirullah, E.R. Geisbrecht, Drosophila NUAK functions with Starvin/BAG3 in autophagic protein turnover, PLoS Genet 16 (4) (2020) e1008700, doi:10.1371/journal.pgen.1008700.
[4] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114–2120, doi:10.1093/bioinformatics/btu170.
[5] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (1) (2013) 15–21, doi:10.1093/bioinformatics/bts635.
[6] T. Metsalu, J. Vilo, ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap, Nucleic Acids Res 43 (W1) (2015) W566–W570, doi:10.1093/nar/gkv468.
[7] E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists, BMC Bioinformatics 10 (2009) 48, doi:10.1186/1471-2105-10-48.