Video Article

Purification of Transcripts and Metabolites from *Drosophila* Heads

Kurt Jensen[1], Jonatan Sanchez-Garcia[1], Caroline Williams[2], Swati Khare[1], Krishanu Mathur[1], Rita M. Graze[3], Daniel A. Hahn[2], Lauren M. McIntyre[3], Diego E. Rincon-Limas[1,4], Pedro Fernandez-Funez[1,4]

[1]Department of Neurology, McKnight Brain Institute, University of Florida

[2]Department of Entomology and Nematology, University of Florida

[3]Genetics Institute, Department of Molecular Genetics and Microbiology, University of Florida

[4]McKnight Brain Institute, Department of Neuroscience, Genetics Institute, Center for Translational Research on Neurodegenerative Diseases, and Center for Movement Disorders and Neurorestoration, University of Florida

Correspondence to: Diego E. Rincon-Limas at diego.rincon@neurology.ufl.edu, Pedro Fernandez-Funez at pedro.fernandez@neurology.ufl.edu

URL: https://www.jove.com/video/50245
DOI: doi:10.3791/50245

Keywords: Genetics, Issue 73, Biochemistry, Molecular Biology, Neurobiology, Neuroscience, Bioengineering, Cellular Biology, Anatomy, Neurodegenerative Diseases, Biological Assay, Drosophila, fruit fly, head separation, purification, mRNA, RNA, cDNA, DNA, transcripts, metabolites, replicates, SCA3, neurodegeneration, NMR, gene expression, animal model

Date Published: 3/15/2013

Citation: Jensen, K., Sanchez-Garcia, J., Williams, C., Khare, S., Mathur, K., Graze, R.M., Hahn, D.A., McIntyre, L.M., Rincon-Limas, D.E., Fernandez-Funez, P. Purification of Transcripts and Metabolites from *Drosophila* Heads. *J. Vis. Exp.* (73), e50245, doi:10.3791/50245 (2013).

Abstract

For the last decade, we have tried to understand the molecular and cellular mechanisms of neuronal degeneration using *Drosophila* as a model organism. Although fruit flies provide obvious experimental advantages, research on neurodegenerative diseases has mostly relied on traditional techniques, including genetic interaction, histology, immunofluorescence, and protein biochemistry. These techniques are effective for mechanistic, hypothesis-driven studies, which lead to a detailed understanding of the role of single genes in well-defined biological problems. However, neurodegenerative diseases are highly complex and affect multiple cellular organelles and processes over time. The advent of new technologies and the omics age provides a unique opportunity to understand the global cellular perturbations underlying complex diseases. Flexible model organisms such as *Drosophila* are ideal for adapting these new technologies because of their strong annotation and high tractability. One challenge with these small animals, though, is the purification of enough informational molecules (DNA, mRNA, protein, metabolites) from highly relevant tissues such as fly brains. Other challenges consist of collecting large numbers of flies for experimental replicates (critical for statistical robustness) and developing consistent procedures for the purification of high-quality biological material. Here, we describe the procedures for collecting thousands of fly heads and the extraction of transcripts and metabolites to understand how global changes in gene expression and metabolism contribute to neurodegenerative diseases. These procedures are easily scalable and can be applied to the study of proteomic and epigenomic contributions to disease.

Video Link

The video component of this article can be found at https://www.jove.com/video/50245/

Introduction

In the last century, *Drosophila* has proven to be an invaluable laboratory tool for investigating profound biological questions, most notably in development, cell biology, genetics, and neurobiology[1]. The reasons for this success are that fruit flies are easy to manipulate, have an ever-expanding toolbox[18, 21, 31], and possess a relatively simple genome with high-quality annotation[26]. These technical advantages, combined with ingenuity and constant innovation within the fly community, have led to broad scientific contributions, as illustrated by the awarding of three Nobel prizes (1933, 1995, 2011). More recently, *Drosophila* has emerged as a relevant model for studying human maladies, primarily developmental disorders, innate immunity, cancer, and neurodegeneration[2]. We are particularly interested in uncovering the cellular and molecular basis of neurodegenerative diseases. These complex and diverse conditions are linked to assemblies, possibly soluble oligomers, of abnormally folded proteins and are, therefore, easily modeled in flies. All of the major neurodegenerative diseases, including Alzheimer's, Parkinson's, and Huntington's disease, amyotrophic lateral sclerosis, several ataxias, tauopathies, prion diseases, and other rare disorders, have been modeled in flies in the last fifteen years[23]. Fly laboratories have contributed to understanding these diseases mainly by exploiting the prowess of *Drosophila* genetics to identify new genes implicated in the neurotoxicity of the pathogenic proteins. Once new genes relevant to the neurotoxic cascade are identified, their effects are typically analyzed by traditional approaches, including histology to ascertain patterns of degeneration, immunofluorescence to determine protein distribution and cellular pathology, and biochemical analyses to assess the quantity and type of abnormal protein conformations. Finally, behavioral analysis serves as a functional readout of disease outcomes. These well-established techniques have been exploited to examine the contribution of one or a few candidate genes to the disease process, including oxidative stress and mitochondrial dysfunction[13], transcriptional dysregulation[9, 27, 30], aberrant axonal transport and synaptic activity[14], abnormal RNA biology[9], dysregulated cell signaling[29], ER dyshomeostasis[6], hindered cellular proteostasis[33], and many others[23]. However, it is not clear how these toxic proteins may interfere simultaneously with multiple interconnected pathways, what is the temporal sequence of these alterations, and what is the relative contribution of each pathway to pathogenesis. Decades of research focused on single gene, hypothesis-driven approaches in both humans and animal models have led to an incomplete, puzzling picture of the cellular mechanisms that cause neurodegeneration. The

current poor understanding of the exact mechanisms by which these toxic protein assemblies cause neuropathology is a key limitation to the development of disease-modifying therapies.

We are now interested in the application of new approaches to understanding how these pathogenic proteins induce global cellular perturbations. The advent of the omics era allows the deep probing of complex biological problems using sophisticated high-throughput technologies, which can lead to effective disease treatments in the near future. Gene expression (transcriptomics) studies were established following the completion of multiple genome sequences since high-quality annotation can predict most transcripts. The recent application of next-generation sequencing to transcript analysis (RNA-seq) has provided new advantages and opportunities compared to microarrays, including an unbiased approach, improved quantitative range, and reduced cost[32]. We want to exploit the advantages of RNA-seq to better understand the most common form of dominantly inherited ataxia, Spinocerebellar ataxia type 3 (SCA3) or Machado-Joseph disease. SCA3 is a monogenic, dominant disease with full penetrance, caused by a CAG trinucleotide expansion in the *Ataxin3* gene (*ATXN3*)[16]. This mutation leads to the production of Atxn3 protein with a long polyglutamine (polyQ) track that makes it prone to aggregate. Since mutant Atxn3 is the neurotoxic agent in SCA3 responsible for all disease-related perturbations, this is an ideal disease for applying these new omic approaches. Additionally, SCA3 was one of the earliest neurodegenerative diseases modeled in flies[34].

While putting in place the procedures for collecting large numbers of fly heads for RNA-seq, we realized that we could use the same material for performing global studies of other informational molecules. Among other emerging disciplines, proteomics, epigenomics, and metabolomics allow the examination of cellular pathology at a depth not previously achieved, although not without challenges. As opposed to mRNA, the catalog of proteins and metabolites is fairly incomplete at this time due to the increased difficulty in predicting all possible splicing variants and the even more enigmatic origin of all normal and pathogenic metabolites. Large efforts are currently underway to catalog proteins in many organisms and in disease conditions. For this, we wanted to focus on the contribution of altered metabolites to SCA3 pathogenesis using a simple organism, such as *Drosophila*. This is a relatively new and promising field, particularly with the recent introduction of nuclear magnetic resonance (NMR), which provides high reproducibility and does not destroy the samples[5, 24]. We propose that metabolite profiling can contribute to identifying biomarkers and disease mechanisms implicated in neurodegeneration.

One important consideration with these omic projects is that they require large, uniform samples (200 fly heads) as well as precise purification techniques to minimize experimental variation. We started to collect flies following procedures we had used previously for microarrays and also used recently for RNA-seq[12]. But we soon encountered a number of problems related to misexpressing the SCA3 constructs. First, we had to select a Gal4 driver for these experiments[4], where the target tissue for analysis was the brain. The obvious choice would be a pan-neural driver since SCA3 is a brain disease. However, since we intend to collect mRNA and metabolites from whole heads, this option would produce mixed material from tissues expressing and non-expressing the constructs. To generate homogeneous samples where all the cells express the constructs, we decided to use a weak but ubiquitous driver line, *daughterless (da)-Gal4*. Interestingly, many of the genes implicated in neurodegenerative diseases, including *ATXN3*[20], have a broad distribution, making the choice of ubiquitous expression appropriate. Then, we encountered a second problem: ubiquitous expression of pathogenic Atxn3-78Q caused lethality during development. To bypass this lethality, we incorporated Gal80^{ts} to control the temporal activation of Gal4[19]. This allowed us to collect the experimental flies, age them for up to 20 days, freeze them, and process their lyophilized heads for either RNA (TRIzol) or metabolite (methanol-chloroform) extraction (**Figure 1**). We have already used this protocol for obtaining samples for transcriptomic and metabolomic analyses, but there are other obvious applications for this technique (*e.g.* proteomic or neuro-peptidomic analyses).

Experimental overview: The overall goal of these protocols is to support the collection of consistent samples of fly heads for the extraction of transcripts and metabolites. The specific goal of these procedures is to acquire flies expressing Atxn3-27Q and Atxn3-78Q (non-pathogenic and pathogenic experimental constructs) as well as LacZ (control construct), where a single sample consists of 200 fly heads. Although current technology allows the analysis of very small samples, including single cells[28], we pooled 200 heads to eliminate experimental, biological, and technical variation, thus allowing us to identify the cellular changes associated with the disease process with high statistical confidence. This approach is designed to detect only those changes in gene expression that are consistent in the population, directing us towards the most critical pathways driving degeneration. Since we are interested in age-dependent changes in the heads of these flies, each genotype was obtained at 1, 10, and 20 days of age.

A fundamental aspect of these global analyses is the production of biological replicates that support a robust statistical analysis. Our previous experience with microarrays and RNA-seq suggests that analysis of biological samples in triplicate provide strong statistical significance of the data[11, 12]. We describe in section 1) how to generate a single biological replicate (Rep 1) for each genotype and time point. These procedures can be repeated to generate Reps 2 and 3, or done in parallel, as long as each Rep is properly identified. Although three Reps is the goal, setting a fourth (or even fifth) Rep is highly recommended to cover issues related to low yield, loss of flies during aging, or accidental loss of material (flies, heads, or RNA) in one or more samples.

We found that ten crosses (bottles identified by letters A-J) produced enough progeny to obtain 200 heads / genotype / time point. Extreme care must be applied to avoid confusion, since large numbers of bottles are required to obtain one Rep (10 bottles x 3 genotypes x 3 time points = 90 bottles). For this, we designated each bottle using genotype, time point, and bottle letter. For example, SCA3-27Q 20E refers to the fifth bottle (of ten) of flies expressing Atxn3-27Q aged for 20 days. Once the progeny eclose, the flies are maintained in separate vials labeled with the unique identifier.

We maintained the numbers of flies obtained for each sample, bottle, and collection point (mornings and evenings) in an Excel spreadsheet. We revised the number of flies per vial prior to freezing to take into account lethality and escapees. From those final counts, vials adding 200 flies can be easily located before opening the freezer, thus contributing to the preservation of the samples. Since flies collected from one bottle can only be used in one Rep, we maximized their contribution per biological replicate, although all Reps contained flies from multiple bottles. We reserved the flies from bottles not used in their scheduled Rep for other Reps, as replacement samples, or for other related experiments (western blot, qPCR).

Protocol

1. Generating a Replicate and Freezing the Flies (CAUTION, Liquid Nitrogen)

Objectives: Collect at least 200 females per genotype / time point, and flash-freeze.

1. Genotypes: *Gal80^{ts}; da-Gal4*, strain for ubiquitous expression of Gal4 under the control of the *daughterless* regulatory region and temperature-sensitive Gal80. As a control line, we used *UAS-LacZ*, which has previously been shown to induce no deleterious effects. UAS-*Atxn3-27Q* and *UAS-Atxn3-78Q*, non-pathogenic and pathogenic experimental constructs, respectively. (The *Y, hs-Hid* construction has been used before for a more efficient collection of virgin females, but we decided against it because of the extra time needed to introduce *Y, hs-Hid* with *Gal80^{ts}; da-Gal4* on chromosomes II and III.)
2. Expand all stocks by combining 10 females and 5 males into bottles to prevent overcrowding. All experiments were carried out in Jazz Mix fly media, which is easy to prepare and promotes fast *Drosophila* growth and low infestation with mites.
3. Collect 50 *Gal80^{ts}; da-Gal4* males, and 100 *UAS-Atxn3-78Q* virgins (only one UAS line will be used as an example) for each time point.
4. Make ten crosses per genotype / time point. Anesthetize the males and females on the carbon dioxide (CO_2) sleeper/pad. Place ten virgin *UAS-Atxn3-78Q* flies (no older than five days) and five male *Gal80^{ts}; da-Gal4* into each large bottle spiked with yeast paste (rehydrated dry yeast). Label the bottles with genotypes, time points, and bottle identifiers (letters A-J), and place them at 25 °C.
5. After 2 days, clear the adults from the bottles, add a sprinkle of dry yeast and a squirt of water to encourage optimal larval growth, and place the bottles at 18 °C (Gal80 active, Gal4 repressed). Do **not** copy the crosses, as this may introduce experimental variability in progeny derived from fertilized versus virgin females. Also, the progeny from each bottle represents independent biological replicates (each bottle produces unique progeny), but transferred crosses (copies) would be technical replicates (additional progeny with the same genetic constitution).
6. When the progeny ecloses at 18 °C, anesthetize on the CO_2 sleeper/pad, collect virgins of the desired genotypes, and place them in small vials with (two to twelve flies per vial). Label vials with the bottle identifier. Do **not** pool flies from different bottles. Place the vials back at 18 °C for 24 hr. To maximize progeny, continue collecting on consecutive mornings and evenings.
7. When each vial completes 24 hr at 18 °C, shift them to 29 °C (Gal80 inactive, Gal4 active). This is day 0, the starting point of the experiment, when the transgenes initiate their expression.
8. Transfer the flies to clean vials every 2 days until they reach days 10 and 20.
9. When the flies reach the desired age (1, 10, 20 days), count and transfer them from the food vials into pre-labeled cryovials using a small funnel. Do **not** anesthetize the flies. Invert the vials on the bench.
10. Wait 30 min to allow the flies to recover from the transfer (stress), then flash-freeze the vials by immersion in liquid nitrogen and store in a pre-labeled, pre-chilled freezer box. **Remember**: Freeze the flies in the same order they were transferred to the cryovial, making time spent in the cryovial uniform across samples.

2. Collection and Lyophilization of Fly Heads (Perform in Cold Room, Pre-chill All Equipment, CAUTION, Liquid Nitrogen and Dry Ice)

Objective: Fast separation, collection, and freeze-drying of tissue.

1. Assemble the sieve (largest mesh in upper chamber to collect bodies, second largest in lower chamber to collect heads, lid on the bottom to collect small parts) and pre-chill at -80 °C for at least 30 min.
2. Collect 200 flies from cryovials by maximizing the contribution of flies originated from the same bottle. To compensate for differences between flies collected in the morning or evening, ensure that every sample uses an identical morning / evening ratio (we used 47% morning / 53% evening). **Note:** the resulting RNA from two 100-head samples can be combined later to generate the equivalent of 200 flies in a single replicate.
3. Chill a piece of parafilm (~9 x 14 cm) on dry ice for 5 min.
4. Empty the cryovials with the complete flies onto the parafilm, a few at a time; count flies using paintbrush, and store in another cryovial on dry ice until reaching 100 flies.
5. Dip the sieve in the liquid nitrogen; add the 100 flies to the upper chamber. Shake the sieve vigorously for 1 min. Dip it in the liquid nitrogen again, and shake for 1 min.
6. Open the sieve, collect heads from the lower chamber in a microtube, and place at -80 °C.
7. Open the microtubes, wrap the tops in parafilm, and make small holes.
8. Place the samples in the lyophilizer for a minimum of 2 days.

3. Total RNA Extraction and mRNA Purification (Treat All Surfaces with RNaseZap; Change Gloves Frequently)

Objective: Homogenize tissue to maximize yield, extract total RNA, and purify mRNA.

1. Remove the samples from the freeze-dryer and add three small, sterile, steel grinding balls to each microtube. Place in the Geno/Grinder; run at 1,500 strokes/min for 1 min.
2. Open the tubes, add 1 ml TRIzol, and seal the tubes with parafilm. Grind 1 min; tap the tubes upside-down to dislodge the steel balls if necessary. Grind 1 min.
3. Do **not** centrifuge the tube at this point (the tube will be brittle). Transfer the mixture (except the steel balls) into a new RNase-free microtube. Pellet the cellular debris in a tabletop microcentrifuge at 4 °C for 10 min at 12,000 x g.

4. Transfer the supernatant to a new microtube. Add 0.1 volumes of 1-bromo-3-chloropropane (BCP); shake vigorously for 15 sec. Incubate at room temperature for 3 min. Spin in a microcentrifuge at 4 °C for 15 min at 12,000 x g. **Note**: many RNA extraction protocols use chloroform at this stage; BCP is intended to maximize the RNA yield by increasing the efficiency of phase separation. Keeping the samples at 4 °C during the centrifugation step will also increase the efficiency of phase separation and reduce contamination of protein and genomic DNA into the aqueous phase.
5. Transfer the top aqueous layer to a new microtube. Precipitate the RNA by adding 0.5 volumes of ice-cold isopropanol; invert six times to mix. Incubate at room temperature for 10 min. Spin in a microcentrifuge at 4 °C for 10 min at 12,000 x g.

Note: In addition to recovering RNA, proteins can also be extracted from the organic layer and interphase of TRIzol, so both proteomics and transcriptomics analyses can be performed on the same samples. Although we did not perform this step in our analyses, TRIzol extraction yields excellent representation of soluble proteins and greater representation of membrane and nuclear proteins than most buffer/detergent extractions. Lee and Lo[17] describe a procedure to extract proteins using TRIzol appropriate for proteomics analyses, including 2-D gel and LC-MS/MS.

6. Remove the supernatant. Wash the RNA by adding 1 ml ice-cold 70% ethanol (use DEPC-treated water), and mix with a short vortex. Spin in a microcentrifuge at 4 °C for 5 min at 12,000 x g.
7. Remove the supernatant. Air dry the pellet for at least 15 min. Add 80 µl of DEPC-water to the RNA pellet and place it at 55 °C for 10 min. Leave at 4 °C overnight.
8. Determine the RNA concentration using the NanoDrop spectrophotometer. The ratio of absorbance at 260 nm/280 nm should be between 1.8 and 2.1, with 2.1 being optimal.
9. Take 30 µg bulk RNA, add 4 U DNase I, 60 U RNasin, 10 µl 10x reaction buffer, and DEPC-treated water to bring the total volume up to 100 µl. Incubate at 37 °C for 20 min.
10. Purify the RNA using the RNeasy Mini kit. Follow the manufacturer's instructions, except perform all centrifugations for 30 sec and include all "optional" steps.
11. Determine the RNA concentration using the NanoDrop, and assess the RNA quality using the "total RNA pico" assay on a BioAnalyzer.
12. For mRNA purification, transfer 30 µl Dynabeads to an RNase-free microfuge tube.
13. Place the tube on a magnet for 1-2 min and remove supernatant. Resuspend the beads in 15 µl of Binding Buffer. Repeat.
14. Starting with 5 µg of total RNA, adjust the volume to 15 µl using DEPC-water. Heat at 65 °C for 2 min, and place on ice.
15. Add 15 µl of total RNA to the 15 µl pre-washed beads. Mix thoroughly by pipetting every 5 min for 30 min to anneal the RNA to the beads. Place on the magnet for 1-2 min and remove all of the supernatant.
16. Add 35 µl of Washing Buffer B. Mix by careful pipetting. Place on the magnet and carefully remove the supernatant. Repeat the wash twice.
17. Elute the mRNA by adding 15 µl cold 10 mM Tris-HCl (pH 7.5) and heating to 80 °C for 2 min. Immediately place the tube on the magnet, transfer the supernatant to an RNase-free tube, and freeze at -80 °C. Quantify as in 3.8). Yield should be ~1-5% of total RNA.

For metabolomic analyses, perform protocols 1 and 2), and proceed with 4) below:

4. Methanol-chloroform Extraction of Metabolites

Objective: Separate polar and non-polar metabolites.

1. Add the antioxidant butylated hydroxytoluene (BHT) to the chloroform at a concentration of 0.1 mg/ml to prevent auto-oxidation of lipids during extraction. Use this BHT-chloroform for all subsequent steps.
2. Transfer 200 heads into chilled, pre-weighed conical screw-cap polypropylene tubes, freeze at -80 °C, and lyophilize. Determine the dry weight.
3. Put 5-7 zirconia beads in the tubes and homogenize in a bead-beater for two cycles of 20 sec each at 800 Hz.
4. Based on the dry weight of the heaviest sample, add water, methanol, and chloroform in the following ratios: 11.74 x methanol; 10.59 x water; 11.74 x chloroform, where x is the weight of the heaviest sample in grams and the product is the volume in ml. Add all the methanol as calculated and 45% of the total water to the bead beater tube containing the lyophilized heads.
5. Homogenize in the bead-beater for two more cycles of 20 sec each.
6. Combine the volumes of chloroform (100%) and the rest of the water (55%) in a conical glass vial on ice. Then, transfer the tissue mixture (with beads) to the glass vial. Shake for 10 min. Incubate for 10 min on ice.
7. Spin in a centrifuge at 4 °C for 5 min at 2,000 x g. Remove the polar (upper) phase with a Pasteur pipette (avoid plastics) and place in a second microcentrifuge tube.
8. Save the non-polar (lower) phase in a small glass vial and freeze at -80 °C. Dry under a stream of nitrogen gas; make sure that the tubing running from the nitrogen tank is high-purity Tygon to avoid plasticizers in the extract.
9. Rehydrate the non-polar phase with 620 µl deuterated chloroform. Transfer 600 µl into labeled NMR tubes. Store in an explosion-proof -80 °C freezer.
10. Place polar phase in the Centrivap and run until the tube is completely dry at 30 °C (~7 hr).
11. Rehydrate the polar phase with 620 µl 0.1 M sodium phosphate buffer (pH 7.3) in deuterium oxide with 1 mM TMSP [3-(trimethylsilyl)propionic-2,2,3,3-d4 acid] as an internal standard. Vortex for 30 sec. Spin in a microcentrifuge at 4 °C for 5 min at 10,000 rpm to precipitate pigments and other large molecules, which would interfere with the NMR spectra. Transfer 600 µl of the supernatant to labeled NMR tubes.

Representative Results

To bypass the lethality induced by ubiquitous expression of expanded Atxn3 under the control of *da-Gal4*, we introduced temporal regulation of Gal4 with Gal80^{ts}. But before proceeding with the collection and freezing of large numbers of flies, we wanted to determine the expression dynamics of our transgene under the control of Gal80^{ts} and *da-Gal4*. To do this, we generated the crosses, left the progeny at 18 °C, collected the adult flies, and maintained them for 24 hr at 18 °C. Then, we placed the flies at the restrictive temperature (29 °C) and incubated them for 0, 6, 12, 24, 36, 48, and 72 hr. Next, we detected the accumulation of the expanded polyQ allele by western blot from single flies following

published protocols[10]. Expression was first detected, but faint 6 hr after temperature shift and continued to increase until 48 hr, when it reached saturation levels (**Figure 2A**). Interestingly, a high molecular weight band of insoluble polyQ appeared after 48 hr, signaling the time it takes for the protein to form large aggregates. We also dissected the fly brains at each time point to determine the distribution of the expanded polyQ allele by immunofluorescence (**Figure 2B**). The protein was first detected 24 hr after the temperature shift, although with uneven distribution. Between 24 and 48 hr the levels of protein continued to rise, making several brain centers clearly visible, including the antennal lobes and the mushroom body projections (**Figure 2B**). Between 48 and 72 hr the expression of the expanded polyQ allele reached maximal levels, suggesting that the Gal4 system was fully activated at this time. Based on these results, we selected 24 hr after the temperature shift as the first time point (day 1) for subsequent experiments. The collections of flies 10 and 20 days old were also measured starting from the temperature shift (time 0), which is the time when the experiment starts with the new expression of the pathogenic and control transgenes.

Once we verified that Atxn3 was detected after 24 hr under the control of Gal80^{ts} and *da-Gal4* at 29 °C, we wondered whether these flies would show signs of neurodegeneration over 20 days. Since these flies start expressing the pathogenic proteins as adults, we feared that they might need a long incubation time to show neurodegenerative phenotypes. To determine the toxicity of Atxn3-78Q in these conditions, we collected flies expressing LacZ, Atxn3-27Q, and Atxn3-78Q, and recorded their longevity. Whereas flies expressing LacZ and Atxn3-27Q displayed over 90% viability after 20 days, flies expressing Atxn3-78Q showed low survival at day 20 (30%) (**Figure 3**). In fact, Atxn3-78Q flies started to die by day 10 (7% mortality) and by day 15 the loss was significant (20%), which accelerated in the last five days. Thus, these results indicated that adult expression of a pathogenic gene resulted in shortened lifespan, a key sign of the neurodegenerative phenotypes induced by Atxn3-78Q.

One of the most critical aspects of this protocol was the handling and preservation of specimens after freezing to ensure the quality of the material extracted. We extracted between 15-80 µg of total RNA per 200 heads (depending on age and genotype) before DNase treatment according to NanoDrop. Approximately one third (6-25 µg) were recovered as highly pure RNA after DNase treatment by the ratio of absorbance at 260 nm/280 nm. However, we found that the best way to determine the quality of RNA for the preparation of cDNA libraries for RNA-seq was to analyze total RNA on the BioAnalyzer. Fortunately, only a small amount of RNA (5 ng) was used on the BioAnalyzer, so it did not impact the ability to produce several libraries per sample. As an example, we ran two samples of total RNA before DNase treatment on a BioAnalyzer chip, one of them suspected of being degraded (**Figure 4A-C**). The high quality RNA (sample 1) produced three sharp RNA bands, two just below 2,000 nucleotides (nt) in size and a smaller band below 200 nt, representing the ribosomal RNAs (**Figure 4A**). The two bands around 2,000 nt correspond to the 18S rRNA (smaller peak) and two fragments from the 28S rRNA (larger peak), which is a unique aspect of rRNA biology in *Drosophila* (see **Figure 4B**). These qualities were also observed in the BioAnalyzer traces (**Figure 4B**). In contrast, a degraded RNA sample (sample 2) did not produce the sharp peaks of the ribosomal RNAs, and accumulated multiple bands of less than 500 nt (**Figure 4A**). This size distribution was readily distinguished in the BioAnalyzer traces (**Figure 4C**), in which broad peaks of shorter size were noted. Also important to note was the difference in the units of the y-axis between the two RNA samples, which demonstrated that the degraded sample had less RNA. Only RNA displaying maximum quality (NanoDrop ratio of absorbance at 260 nm/280 nm between 1.8 and 2.1, with 2.1 being optimal, see protocol step 3.8) will be used for the generation of libraries.

For the extraction of metabolites, we followed a classic water / methanol / chloroform (2.0:2.0:1.8) extraction procedure[3] split into two steps to maximize extraction of both polar and non-polar metabolites[35]. After separating the polar and non-polar phases, we reconstituted them in deuterated water or deuterated chloroform, respectively, and added internal standards for analysis by NMR. NMR spectra obtained using this extraction method were well resolved with flat baselines and narrow lines (**Figure 5**), indicating the purification of high-quality extracts, suitable for both metabolic profiling and the identification of large numbers of metabolites. **Figure 5** shows the identification of a few prominent peaks corresponding to highly abundant metabolites, such as glucose and sucrose, and two key metabolic acids, lactate and acetate, according to chemical shifts in the literature[7]. The chemical shift (in parts per million) represents the electromagnetic shielding of a molecule relative to a standard molecule (TMS, first peak from the right). More shieldedmolecules have higher electron density around the nucleus and will appear to the right of the spectrum, including alkyls. Deshielded molecules such as amide and aromatic hydrogens will appear to the left of the spectrum. The busy middle section (3-4 ppm) is enriched for sugar molecules.

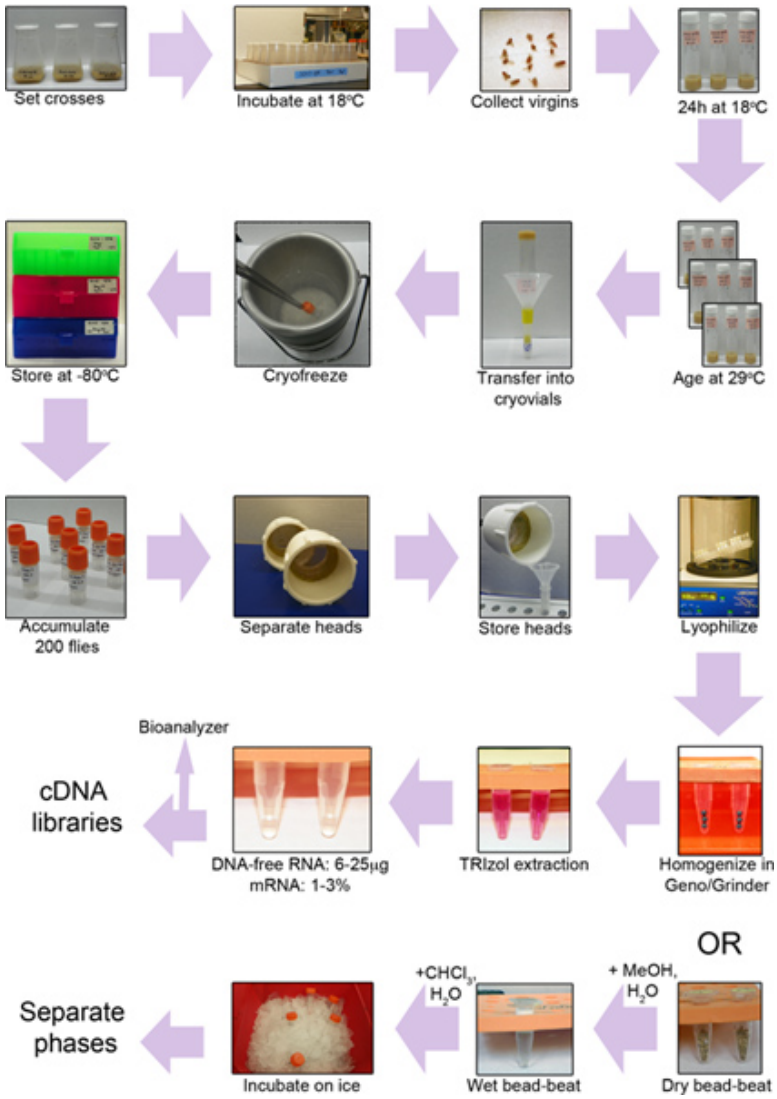

Figure 1. Experimental workflow. First, we set crosses to collect flies of the desired genotypes (top row). Then, we shift virgins to 29 °C to activate gene expression, age them for 1, 10, and 20 days, and flash-freeze them (second row). We next collect frozen heads using a microsieve (third row). The heads are freeze-dried, ground up, and treated for the extraction of RNA (fourth row) or metabolites (bottom row).

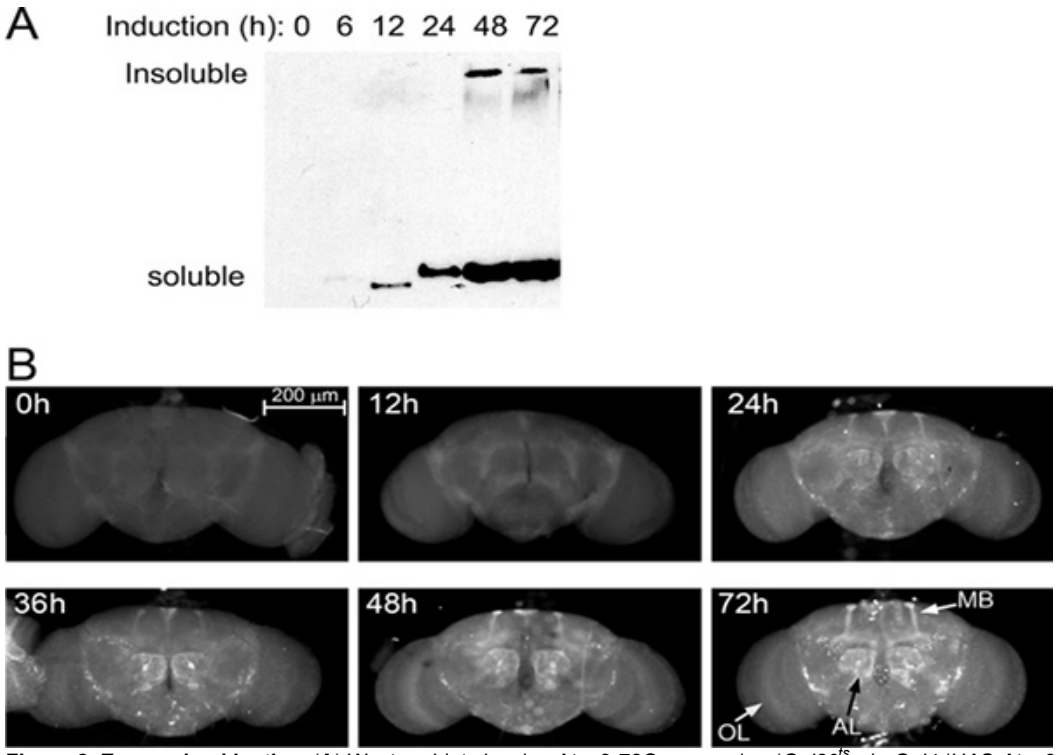

Figure 2. Expression kinetics. (**A**) Western blot showing Atxn3-78Q expression (*Gal80^{ts}; da-Gal4 /UAS-Atxn3-78Q*) in flies incubated at 29 °C from 0 to 72 hr. A faint band is first detected at 6 hr and saturation is reached by 48 hr after induction. (**B**) Immunofluorescence of whole brains in the same flies. Brains were dissected, fixed, and stained with an antibody that recognizes the expanded polyQ protein. The protein is first detected in the mushroom body (MB) projections and the antennal lobes (AL). Between 48 and 72 hr, the expression reaches maximum levels and is highly visible in brain centers with abundant innervation. The expression level per neuron is weak as seen in the optical lobe (OL), except in a few large neurons between the OL and the central brain.

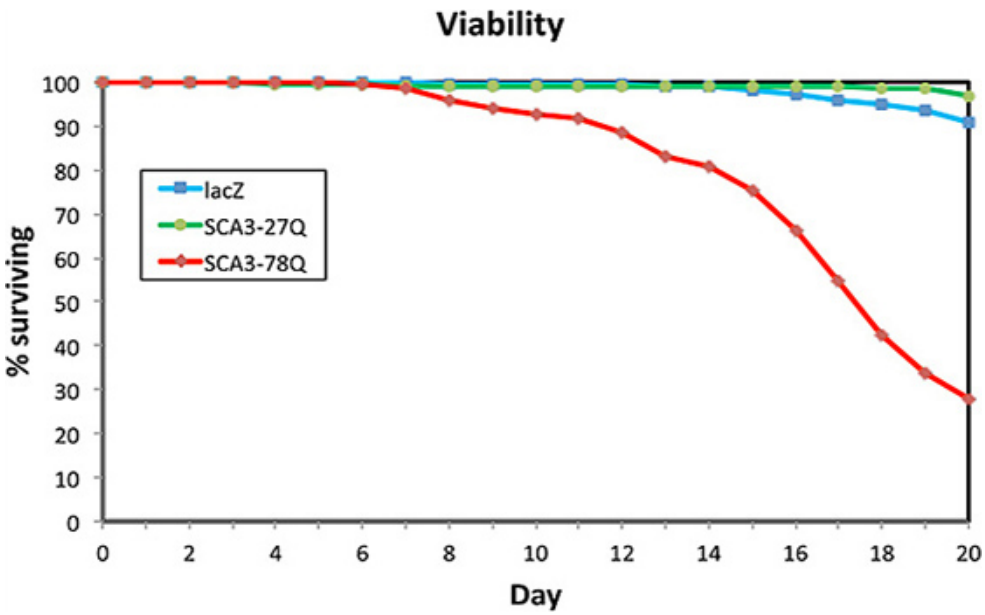

Figure 3. Atxn3-78Q reduces lifespan. Flies expressing *LacZ*, *Atxn3-27Q*, and *Atxn3-78Q* ubiquitously in adult stages (*Gal80^{ts}; da-Gal4*) were monitored for their longevity at 29 °C. Flies expressing *LacZ* (blue) and *Atxn3-27Q* (green) showed low levels of mortality by day 20. In contrast, flies expressing *Atxn3-78Q* (red) displayed a pronounced mortality starting at day 10. By day 15 they showed a 20% reduction in viability, which accelerated over the next five days, reaching a 70% mortality by day 20.

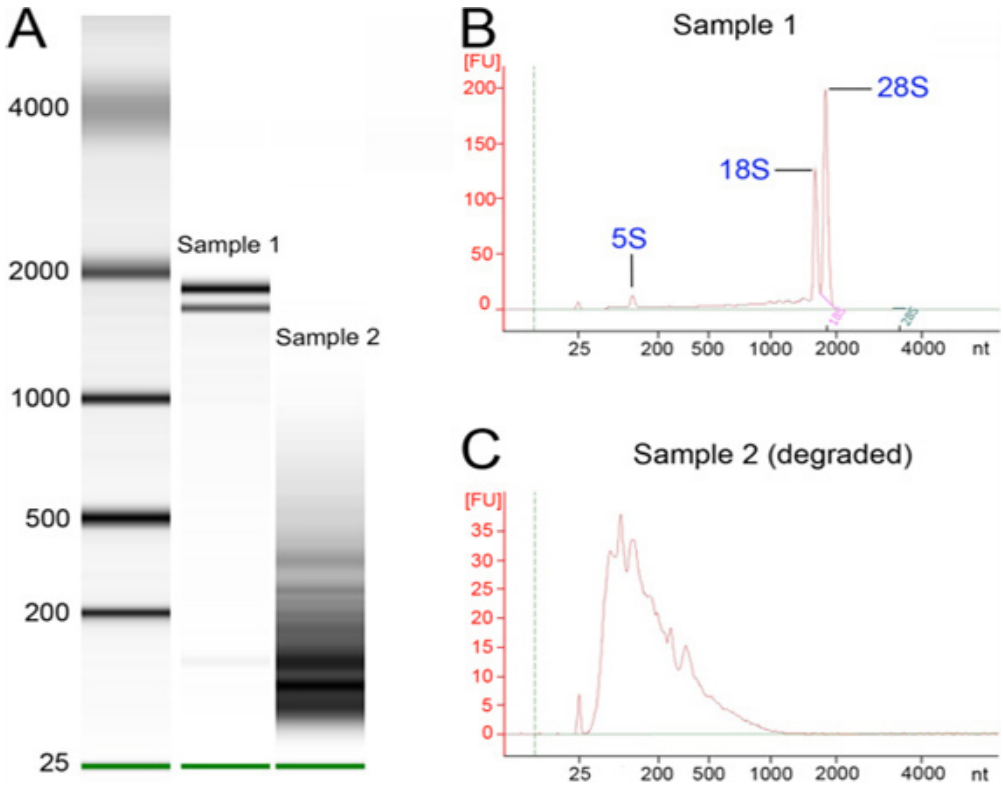

Figure 4. RNA quality assessment. RNA from high quality and degraded samples was run on a BioAnalyzer chip. (**A**) BioAnalyzer run with a high quality RNA (sample 1) and a degraded RNA (sample 2) next to a size marker (left lane). Note the three sharp peaks corresponding to ribosomal RNA in the central lane. (**B**) BioAnalyzer tracing for sample 1, with two characteristic strong peaks just below 2,000 nt and another below 200. (**C**) BioAnalyzer tracing for sample 2, which consists of mostly degraded RNA. Note the difference in units on the y-axis between panels **B** and **C**.

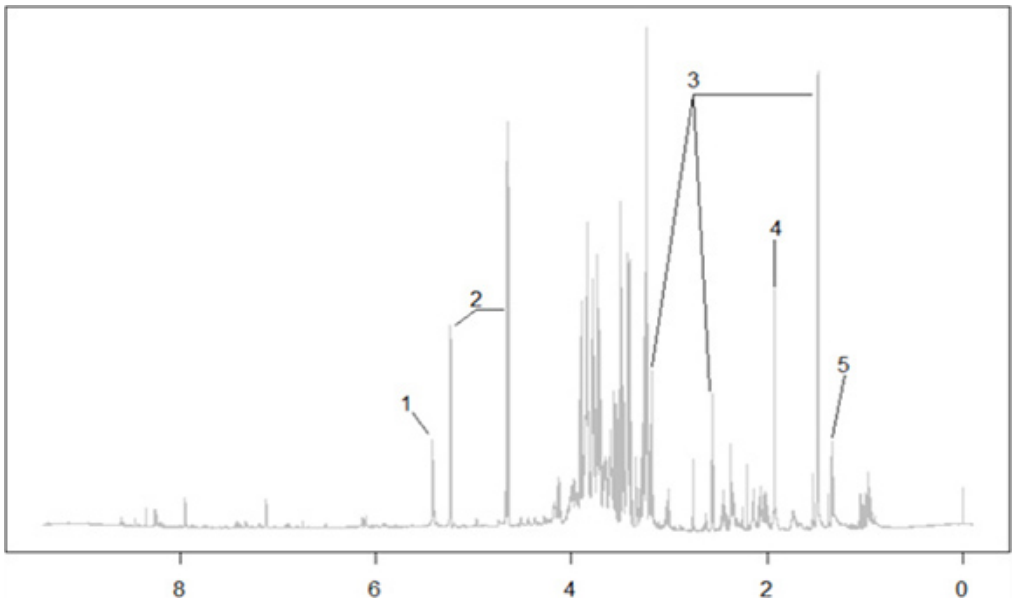

Figure 5. Representative 1D NMR spectra of polar metabolites. Representative 1D proton NMR spectrum of polar metabolites from *Drosophila* and complementary 2D COSY (correlation spectroscopy) and HSQC (heteronuclear single quantum correlation) experiments for peak assignment. Metabolites identified according to known chemical shifts- 1: Sucrose; 2: Glucose; 3: Beta-alanine; 4: Acetate; 5: Lactate.

Discussion

Building on our previous experience in transcriptomics[11, 12], metabolism[15], and neurodegenerative diseases[6, 8], we present here a new protocol for collecting thousands of fly heads with adult-specific expression of pathogenic genes, and purifying mRNA and metabolites for RNA-seq and NMR spectroscopy, respectively. The nature of these experiments required the incorporation of several innovations prior to the fly collection,

including the selection of a ubiquitous Gal4 line and the use of temporal regulation of gene expression. Regarding Gal4, ubiquitous expression of the transgenes was crucial for two reasons. First, a large number of genes involved in neurodegenerative diseases, including *ATXN3*, *Huntingtin*, *Amyloid precursor protein*, and *Superoxide dismutase 1*, among others, are widely expressed in neuronal and glial cells as well as in other cell types outside the nervous system. We wanted to correctly model this aspect of the biology of these pathogenic genes by analyzing the consequences of ubiquitous expression rather than focusing only on neural expression. Second, it is not feasible to collect fly brains in large numbers, but we can do so with fly heads (over 200 in triplicate per condition)[11, 12]. Since homogenization of whole heads mixes neural and non-neural tissues, ubiquitous transgene expression becomes critical for obtaining uniform RNA and metabolites from populations of cells expressing the same transgenes. It is important to note that *da-Gal4* was chosen for its fairly even distribution, not because of its high expression level, although a recent report suggested that *da-Gal4* may not be totally ubiquitous[25]. We had previously considered other ubiquitous Gal4 lines, including *arm*-, *Tub*-, and *Act-Ga4*, but they all showed complex expression patterns in the adult CNS and muscles, making them less adequate for this study. In fact, the immunofluorescence in adult brains showed that *da*-Gal4 induces widespread but weak expression, which only accumulated at high levels in brain centers comprised of a few thousand axons and in a few large neurons (**Figure 2**). Although the expression levels of the constructs were low, these conditions dramatically reduced survival in a polyQ-dependent fashion. This expression dynamics fulfilled our needs while keeping expression levels low, which was important for observing the slow, progressive cellular alterations induced by neurotoxic proteins.

A predictable consequence of inducing ubiquitous expression of neurotoxic proteins was that it caused lethality prior to adult eclosion. Fortunately, this complication was bypassed with the use of Gal80^{ts}. As discussed above, gene expression reached maximal levels approximately 48 hr after the temperature shift, which fits well with the time frame of the experiment. One additional advantage of using Gal80^{ts} was that, in contrast to experiments where the neurotoxic proteins are constitutively expressed pan-neurally, there was no expression during the development of the nervous system. Thus, the use of Gal80^{ts} created a "clean" background where the nervous system was not exposed to the toxic activities of these neurotoxic proteins during sensitive developmental stages. In our view, activation of gene expression in adults resulted in a better model for this class of adult-onset pathologies. However, Gal80^{ts} also introduced a few complications associated with the temperature changes. One was that growing flies at low temperatures (18-20 °C) delayed the life cycle, making the experiments significantly longer. Also, it is well known that growing flies at high temperatures (29-31 °C) accelerates their metabolism, as illustrated by the shortened lifespan compared to flies aged at 25 °C. Since the transcriptional and metabolic studies will be performed in flies aged at high temperature, we may expect to see important changes in cellular stress pathways and energy metabolism. This is why it was so critical to introduce two controls to the experiment, one with the non-pathogenic Atnx3-27Q and an unrelated control expressing LacZ. These controls will provide a baseline for gene expression and metabolic changes associated with high temperature so that we can identify those changes directly linked to the expression of toxic Atxn3. Overall, we have introduced several modifications to the collection of flies that provide numerous advantages - and a few disadvantages (temperature issues and next paragraph) - for studying adult-onset, progressive diseases.

Although the temporal regulation with Gal80^{ts} was a helpful addition, it also introduced an unexpected complication. While generating the flies carrying both *Gal80^{ts}* and *da-Gal4* constructs in a stable stock, we noticed that flies doubly homozygous for both constructs were viable but sterile. Since homozygous *Gal80^{ts}* and *da-Gal4* strains were viable and fertile by themselves, it is not clear why the combination displayed low fertility. It has been known for some time that Gal4 is slightly toxic to *Drosophila* tissues[22]. It is possible, then, that the heterologous expression of the yeast transcriptional repressor Gal80 contributed to the toxicity of Gal4 by interfering, potentially, with the endogenous transcriptional machinery. This toxicity may be exacerbated by the ubiquitous distribution of Gal4 under the control of *da*, a gene that plays a key role in early development and sexual determination. Regardless of the underlying causes of the sterility, we expanded the *Gal80^{ts}; da-Gal4* flies by maintaining a balancer chromosome over *da-Gal4* while keeping the homozygosity for *Gal80^{t}s*, suggesting that Gal4 is a more critical contributor to the sterility. This solution was key because we used hundreds of these males every other week to cross with each of the UAS transgenes. However, crosses carrying a balancer created additional work during the selection of the appropriate progeny. Only one fourth (25%) of the progeny was collected (females, non-balancer), which added selection time, reduced the yield per bottle, and increased the number of bottles needed to obtain at least 200 flies per genotype / replicate / time point. Overall, although the collection of flies became time consuming due to the large sets needed, we are confident that studying cellular changes only a few hours after turning on the expression of pathogenic genes ubiquitously will identify novel and interesting processes implicated in progressive neuronal loss.

Once the genetic design was in place, we paid special attention to the experimental manipulations that could introduce biological variability. First, we only collected virgin females to ensure the homogeneity of the samples, although this meant throwing away males and checking for progeny twice a day. During aging, no more than 12 flies were kept in the same vial to avoid overcrowding and stress. Also, before freezing, flies were transferred to cryovials without anesthesia and were allowed to recover from the transfer for 30 min. These seemingly trivial factors can influence gene expression and metabolism, introducing variability in the final analysis that would not be relevant to the experiment.

To ensure the quality of the frozen materials (heads, RNA, and metabolites), we paid special attention to each detail that can contribute to tissue degradation. Since flies are transferred to cryovials in small numbers (up to 12 flies per vial), we had to retrieve many vials for the collection of 200 heads. These manipulations were done with extreme care in a cold room with dry ice and liquid nitrogen. The collection of flies, separation of heads, and purification of informational molecules is too time consuming to discover that the material was degraded at the end of this long process. In addition to ensuring that the material retains its quality, this protocol describes several steps that maximize the yield of RNA and metabolites, including freeze-drying and bead beating. These steps are not required for normal extractions for RT-PCR or quantitative PCR from whole flies, but they are critical for consistent purification from small samples such as fly heads. We determined that other steps affect the yield of RNA. For instance, older flies produced significantly less RNA than younger flies, regardless of genotype. This observation suggested that aging at high temperature induces dramatic changes. Also, extracting RNA from 100 heads was actually more efficient than from 200 heads, so we proceeded to purify in two batches of 100 per sample, only to combine them after verification of quality with the BioAnalyzer. Also, columns for mRNA purification seemed to saturate easily, so the amount of total RNA used per column needs to be optimized.

Metabolites are a diverse mix of small molecules, so quality control is more difficult than with DNA, RNA, or proteins. Unfortunately, there is no complete catalog of metabolites for any animal model at this time, something that may be changing in the next few years thanks to the application of several technological innovations, including NMR spectroscopy and the expanded use of liquid chromatography - Mass Spectrometry (LC-MS). Although the NMR is less sensitive than MS, it shows many promising advantages, including no destruction of samples, no analytical procedures that introduce bias (LC), and high reproducibility that allows statistical comparison of Reps and several experimental

conditions[24]. Thus, the sample can be retested to verify observations or re-analyzed by LC-MS to validate NMR data. Here, we showed preliminary NMR data from flies that we are using to compile a catalogue of metabolites in *Drosophila*. This information will be critical for determining how expression of pathogenic genes alters normal metabolism, opening a new window to further understanding the cellular mechanisms underlying neurodegenerative diseases.

Newly emerging omic technologies offer the best opportunity to study neurodegenerative diseases at a level of detail not previously achieved. Perhaps the biggest obstacle to overcome in these high-throughput studies is the faithful collection of reproducible samples that can be analyzed with highly sensitive methods. The procedures introduced here provide a way to achieve this consistency, and will lead to results that can be easily compared across datasets. Although our primary research interests involved examining changes of gene expression and metabolites, the flies generated could also be subjected to proteomic or epigenomic analysis. Proteomic and epigenomic changes occurring during neurodegeneration are also likely to be of paramount importance to the disease process. When the scientific community ultimately identifies the full spectrum of molecular changes impacting the disease process, a true systems level understanding of the disease will be possible. At this level, disease-modifying therapies will finally emerge as a reality.

Disclosures

The authors declare no competing financial interests.

Acknowledgements

We are thankful to Janice Santiago, Gabriel Figueroa, and Jose Herrera for technical assistance and the Bloomington *Drosophila* Stock Center at Indiana University for fly stocks. DH and CW were supported by funds from NSF-IOS-1051890. PF-F and LMM were supported by an Opportunity Seed Fund from the University of Florida.

References

1. Bellen, H.J., Tong, C., & Tsuda, H. 100 years of *Drosophila* research and its impact on vertebrate neuroscience: a history lesson for the future. *Nat. Rev. Neurosci.* **11**, 514-522 (2010).
2. Bier, E. *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat. Rev. Genet.* **6**, 9-23 (2005).
3. Bligh, E.G. & Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911-917 (1959).
4. Brand, A.H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* **118**, 401-415 (1993).
5. Bundy, J.G., Papp, B., Harmston, R., Browne, R.A., Clayson, E.M., Burton, N., Reece, R.J., Oliver, S.G., & Brindle, K.M. Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling. *Genome Res.* **17**, 510-519 (2007).
6. Casas-Tinto, S., Zhang, Y., Sanchez-Garcia, J., Gomez-Velazquez, M., Rincon-Limas, D.E., & Fernandez-Funez, P. The ER stress factor XBP1s prevents amyloid-beta neurotoxicity. *Hum. Mol. Genet.* **20** , 2144-2160 (2011).
7. Fan, T.W.M. Metabolite Profiling by One and Two-Dimensional NMR Analysis of Complex Mixtures. *Proress in Nuclear Magnetic Resonance Spectroscopy.* **28**, 161-219 (1996).
8. Fernandez-Funez, P., Casas-Tinto, S., Zhang, Y., Gomez-Velazquez, M., Morales-Garza, M.A., Cepeda-Nieto, A.C., Castilla, J., Soto, C., & Rincon-Limas, D.E. *In vivo* generation of neurotoxic prion protein: role for hsp70 in accumulation of misfolded isoforms. *PLoS Genet.* **5**, e1000507 (2009).
9. Fernandez-Funez, P., Nino-Rosales, M.L., de Gouyon, B., She, W.C., Luchak, J.M., Martinez, P., Turiegano, E., Benito, J., Capovilla, M., Skinner, P.J., McCall, A., Canal, I., Orr, H.T., Zoghbi, H.Y., & Botas, J. Identification of genes that modify ataxin-1-induced neurodegeneration. *Nature.* **408**, 101-106 (2000).
10. Fernandez-Funez, P., Zhang, Y., Casas-Tinto, S., Xiao, X., Zou, W.Q., & Rincon-Limas, D.E. Sequence-dependent prion protein misfolding and neurotoxicity. *J. Biol. Chem.* **285** 36897-36908 (2010).
11. Graze, R.M., McIntyre, L.M., Main, B.J., Wayne, M.L., & Nuzhdin, S.V. Regulatory divergence in *Drosophila melanogaster* and *D. simulans*, a genomewide analysis of allele-specific expression. *Genetics.* **183**, 547-561, 541SI-521SI (2009).
12. Graze, R.M., Novelo, L.L., Amin, V., Fear, J.M., Casella, G., Nuzhdin, S.V., & McIntyre, L.M. Allelic imbalance in *Drosophila* hybrid heads: exons, isoforms, and evolution. *Mol. Biol. Evol.* **29**, 1521-1532 (2012).
13. Greene, J.C., Whitworth, A.J., Kuo, I., Andrews, L.A., Feany, M.B., & Pallanck, L.J. Mitochondrial pathology and apoptotic muscle degeneration in *Drosophila* parkin mutants. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4078-4083 (2003).
14. Gunawardena, S. & Goldstein, L.S. Disruption of axonal transport and neuronal viability by amyloid precursor protein mutations in *Drosophila. Neuron.* **32**, 389-401 (2001).
15. D.A. Hahn, & Denlinger, D.L. Energetics of insect diapause. *Annu. Rev. Entomol.* **56**, 103-121 (2011).
16. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., *et al.* CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.* **8** 221-228 (1994).
17. Lee, F.W. & Akiguchi, S.C. The use of Trizol reagent (phenol/guanidine isothiocyanate) for producing high quality two-dimensional gel electrophoretograms (2-DE) of dinoflagellates. *J. Microbiol. Methods.* **73**, 26-32 (2008).
18. Matthews, K.A., Kaufman, T.C., & Gelbart, W.M. Research resources for *Drosophila*: the expanding universe. *Nat. Rev. Genet.* **6**, 179-193 (2005).
19. 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).
20. Paulson, H.L., Das, S.S., Crino, P.B., Perez, M.K., Patel, S.C., Gotsdiner, D., Fischbeck, K.H., & Pittman, R.N. Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. *Ann. Neurol.* **41**, 453-462 (1997).
21. Pfeiffer, B.D., Ngo, T.T., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W., & Rubin, G.M. Refinement of tools for targeted gene expression in *Drosophila. Genetics.* **186**, 735-755 (2010).

22. Rezaval, C., Werbajh, S., & Ceriani, M.F. Neuronal death in *Drosophila* triggered by GAL4 accumulation. *Eur. J. Neurosci.* **25**, 683-694 (2007).
23. Rincon-Limas, D., Jensen, K., & Fernandez Funez, A. *Drosophila* models of proteinopathies: the little fly that could. *Curr. Pharm. Des.* **18**, 1108-1122 (2012).
24. Robinette, S.L., Ajredini, R., Rasheed, H., Zeinomar, A., Schroeder, F.C., Dossey, A.T., & Edison, A.S. Hierarchical Alignment and Full Resolution Pattern Recognition of 2D NMR Spectra: Application to Nematode Chemical Ecology. *Anal. Chem.* (2011).
25. Simon, A.F., Daniels, R., Romero-Calderon, R., Grygoruk, A., Chang, H.Y., Najibi, R., Shamouelian, D., Salazar, E., Solomon, M., Ackerson, L.C., Maidment, N.T., Diantonio, A., & Krantz, D.E. *Drosophila* vesicular monoamine transporter mutants can adapt to reduced or eliminated vesicular stores of dopamine and serotonin. *Genetics.* **181**, 525-541 (2009).
26. Smith, C.D., Shu, S., Mungall, C.J., & Karpen, G.H. The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science.* **316**, 1586-1591 (2007).
27. Steffan, J.S., Bodai, L., Pallos, J., Poelman, M., McCampbell, A., Apostol, B.L., Kazantsev, A., Schmidt, E., Zhu, Y.Z., Greenwald, M., Kurokawa, R., Housman, D.E., Jackson, G.R., Marsh, J.L., & Thompson, L.M. Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in *Drosophila. Nature.* **413**, 739-743 (2001).
28. Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K., & Surani, M.A. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* **5**, 516-535 (2010).
29. Tare, M., Modi, R.M., Nainaparampil, J.J., Puli, O.R., Bedi, S., Fernandez-Funez, P., Kango-Singh, M., & Singh, A. Activation of JNK signaling mediates amyloid-ss-dependent cell death. *PLoS One.* **6**, e24361 (2011).
30. Taylor, J.P., Taye, A.A., Campbell, C., Kazemi-Esfarjani, P., Fischbeck, K.H., & Min, K.T. Aberrant histone acetylation, altered transcription, and retinal degeneration in a *Drosophila* model of polyglutamine disease are rescued by CREB-binding protein. *Genes Dev.* **17**, 1463-1468 (2003).
31. Venken, K.J. & Bellen, H.J. Transgenesis upgrades for *Drosophila melanogaster. Development.* **134**, 3571-3584 (2007).
32. Wang, Z., Gerstein, M., & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57-63 (2009).
33. Warrick, J.M., Chan, H.Y., Gray-Board, G.L., Chai, Y., Paulson, H.L., & Bonini, N.M. Suppression of polyglutamine-mediated neurodegeneration in *Drosophila* by the molecular chaperone HSP70. *Nat. Genet.* **23**, 425-428 (1999).
34. Warrick, J.M., Paulson, H.L., Gray-Board, G.L., Bui, Q.T., Fischbeck, K.H., Pittman, R.N., Bonini, N.M. Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in *Drosophila. Cell.* **93**, 939-949 (1998).
35. Wu, H., Southam, A.D., Hines, A., & Viant, M.R. High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal. Biochem.* **372**, 204-212 (2008).