bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Proteogenomics of adenosine-to-inosine RNA editing in fruit fly

Ksenia G. Kuznetsova[1*], Irina Y. Ilina[1], Alexey L. Chernobrovkin[2], Svetlana E. Novikova[1], Tatyana E. Farafonova[1], Dmitry S. Karpov[3], Mark V. Ivanov[4,5], Olga E. Voronko[1], Ekaterina V. Ilgisonis[1], Anna A. Kliuchnikova[1], Victor G. Zgoda V[1], Roman A. Zubarev[2], Mikhail V. Gorshkov[4,5], Sergei A. Moshkovskii[1,6 *]

[1] Institute of Biomedical Chemistry, Moscow, Russia

[2]Karolinska Institutet, Stockholm, Sweden

[3]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia.

[4]Institute of Energy Problems of Chemical Physics, Russian Academy of Sciences, Moscow, Russia

[5]Moscow Institute of Physics and Technology (State University), Dolgoprudny, Moscow Region, Russia

[6]Pirogov Russian National Research Medical University (RNRMU), Moscow, Russia.

*Corresponding authors: smosh@mail.ru; kuznetsova.ks@gmail.com

Running title: Proteogenomics of A-to-I RNA editing in fruit fly

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Summary

Adenosine-to-inosine RNA editing is one of the most common types of RNA editing, a posttranscriptional modification made by special enzymes. We present a proteomic study on this phenomenon for *Drosophila melanogaster*. Tree deep proteome data sets for Canton-S fruit fly line were used in the study: two taken from public repository and the third one obtained here for the isolated brains. A customized protein sequence database was generated using results of genome-wide adenosine-to-inosine RNA studies in fruit fly and applied for identifying the edited proteins. The total number of 56 edited proteins was found in all data sets, 7 of them being shared between the whole insect, head and brain proteomes. Two edited sites in syntaxin 1A (*Syx1a*) and complexin (*cpx*) belonging to a presynaptic vesicle fusion complex were selected for validation by targeted analysis. The results obtained for two selected peptides using Multiple Reaction Monitoring have shown remarkably constant ratios of unedited-to-edited protein variants in flies raised under different ambient temperatures of 10, 20 and 30°C. Specifically, these ratios were 34.5:1 and 2.1:1 for *Syx1a* and *cpx*, respectively. The work demonstrates the feasibility to identify the RNA editing event at the proteome level using shotgun proteomics and customized edited protein database.

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abbreviations

BCA – bicinchoninic acid

TEABC - triethylammonium bicarbonate

IAM – iodoacetamide

rpm – rounds per minute

HCD – higher-energy collisional dissociation

VAI – Variant Annotation Integrator

RADAR – rigorously annotated database of A-to-I RNA editing

ADAR – adenosine deaminase, RNA specific.

FDR – false discovery rate

PSM – peptide spectrum match

Fmoc - 9-fluorenyl methyloxy carbonyl

OtBu - tert-butyl

DMF – dimethylformamide

HCTU - O-(1H-6-Chloro-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate

TMP - 2,4,6-Trimethylpyridine

TIS – Triisopropylsilane

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

QqQ – triple quadruplole

CV – coefficient of variance

SNP – single nucleotide polymorphism

SNARE – SNAP (Soluble NSF Attachment Protein) REceptor

GO – gene ontology

MRM – multiple reaction monitoring

CE –collision energy

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

RNA editing is a type of posttranscriptional modification made by specific enzymes. Being first described to happen in mitochondrial RNAs of kinetoplastid protozoa (1), it was then observed for various organisms and different kinds of RNA (2). RNA editing includes nucleotide insertion or deletion, as well as deamination of cytosine and adenosine bases. Cytosine gets transformed into uridine by cytidine deaminase (CDA) (3), and adenosine is converted to inosine by adenosine deaminases acting on RNA (ADARs) (4, 5). While the former is described mostly for plant cells (6, 7), although it occurs also during apolipoprotein B synthesis (3), the latter is common for neural and glandular tissues of many invertebrate and vertebrate species (8).

Messenger RNA editing is the most interesting kind of RNA editing for proteomics as it may affect the structure of proteins. At the same time, specifically adenosine to inosine (A-to-I) editing provides the interest to neurobiology, because, reportedly, this type of modification is believed to have a function of rapid fine neuron tuning (9).Protein products of RNA editing can exist in organisms in both variants. The ratio of these variants may possess a functional significance (10, 11).

To date, the phenomenon of RNA editing was studied mostly at the transcriptome level that included a number of works on *Drosophila melanogaster* (12). Yet, the workflows for identification and characterization of the RNA editing products have not been developed. The only study on the RNA editing at the proteome level was focused characterization of all types of proteoforms for the rat liver (13). However, liver is not reported as a tissue of functional A-to-I

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

RNA editing (14), and twenty events of RNA editing identified for rat proteome were simply listed without further discussion (13).

With the introduction of proteogenomic approach as use of customized nucleic acid databases for specific samples (15), the workflow for proteomic investigation of the products of RNA editing became pretty clear. First, a customized proteomic database is made, based on known 'editome' of the organism under study (16). As the editome includes extra variants of the edited mRNA sequences, this database contains both unedited and edited peptide variants. Then, the shotgun proteomic spectra are searched against this database. Finally, information about the edited peptides is extracted from the search results and optionally validated to exclude false discoveries (17).

Generally speaking, the term 'proteogenomics' is not accurate of suing the transcript sequence databases for the peptide search. However, in the field, use of customized transcriptomes for cancer variant identification was considered in scope of this term (18).

A-to-I RNA editing at the transcriptome level has been studied comprehensively for *D. melanogaster* by Hoopengardner et al. using comparative genomic approaches (19). In the other study by Rodriguez et al., the author used nascent RNA sequencing (20). Comparing wild type and the *adar* mutant flies they have shown the critical role the ADAR is playing in RNA editing. A method of cDNA to genomic DNA comparison was used to find RNA editing sites by Stapleton et al. (21).

Recently, a genome-wide analysis of A-to-I RNA editing sites was performed, and an editome of *Drosophila* was thoroughly characterized (22). The analysis revealed 3581 high-confidence editing sites in the whole body of a fruit fly. The authors used a single-molecule

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sequencing method with the introduction of the so-called 'three-letter' alignment procedure to avoid misreading of the A-to-G substitution sites for wild type and *adar*-deficient flies. This allowed increasing the accuracy of the database containing the A-to-I RNA editing sites and provided the most complete editome of *D. melanogaster*. This editome was further used for generating the customized protein database in this work. The summary of the results of previous efforts to study the *D. melanogaster* editome and the evolutional analysis of the function of A-to-I editing in seven *Drosophila* species was also provided last year by Yu et al. (23).

It also has been shown that A-to-I editing happens as a response to environmental changes such as temperature (12), which makes great sense in terms of the purpose of editing versus genomic recoding evolutionally. The authors have described 54 A-to-I editing sites, some of which are demonstrating significant differences in edited-to-unedited transcript ratios in the flies raised at 10, 20, and 30°C. The list of sites consists of various genes including *adar* itself.

From previous works with successful use of customized databases to identify protein-coding genome variants (24), we deduced that similar customized databases may be designed protein coding RNA editome. In this proof-of-principle study, we used a fruit fly with well characterized A-to-I RNA editing (22). The main purpose of the study was identifying the RNA editing events in the proteome followed by validation of selected edited peptides by targeted mass spectrometry. To our knowledge, this is a first attempt to characterize RNA editing in *Drosophila* at the proteome level. Tandem mass spectrometry data were taken from recent shotgun proteomics studies (25, 26) available at ProteomeXchange (http://www.proteomexchange.org/) (27). These results contain data for proteome of *Drosophila*'s whole bodies (25) and whole heads (26). The other data set for *Drosophila*'s brain proteome was obtained here using high-resolution mass spectrometry.

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Experimental Procedures

Experimental design and statistical rationale

The shotgun proteomic analysis was performed using 1 sample consisted of 200 isolated fruit fly brains combined. The number of technical replicates in the shotgun experiment was 3. For the targeted proteomic experiment 4 samples had been prepared. The first one was used for preliminary MRM experiment and had been derived from flies of different age. The other 3 consisted of 80 brains each and had been made of flies raised in 10, 20 and 30°C. During the MRM experiment 5 technical replicates have been done. The details of each experiment are provided below. One sample consisting of 100 brains was used for the genomic sequencing experiment. The summarizing table of all the *Drosophila* samples used in this work is provided in Supplemental file 1.

The data from 3 proteomes were used for the RNA editing sites search. One proteome was obtained experimentally here and the other two were taken from (25) and (26). A thorough schematic explanation of the whole workflow performed is given in Figure 1.

During the shotgun data analysis the peptide identification was held at a 1% false discovery rate as described in details in the corresponding section. In the MRM experiment, the Exact Fisher Test has been performed for the edited-to-unedited peptide ratio changes estimation. For the overall protein expression changes, the ANOVA test has been used. These results are described in the “MRM analysis results section”.

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Drosophila melanogaster culture

Live samples of *Drosophila melanogaster* Canton S line were kindly provided by Dr. Natalia Romanova from Moscow State University, Department of Biology. The flies were kept on Formula 5-24 instant Drosophila medium (Carolina Biological supply Company, USA) in 50 ml disposable plastic test tubes (Orange Scientific, Belgium). The initial temperature for the fly culture was 25°C. The flies had been transferred to a new tube as they reached the adult stage. For general samples that were used for shotgun proteome analysis, as well as for the initial targeted analysis, 200 flies were selected from the culture regardless of their age.

Different experimental procedure was used to study variants produced in response to the temperature changes. The workflow was similar to the one used in previous studies (12) with some modifications according to local laboratory practices. A few flies were put into a new test tube with fresh medium at 25°C and kept there for 24 hours to let the parent flies lay eggs. The parent flies were, then, removed from the test tubes. The test tubes with eggs were kept at 25°C until the first population of fly imago hatched from pupae. As the young flies hatched they were immediately transferred to a new test tube and placed to thermostats set at 10°C, 20°C, and 30°C. The flies were kept at the set temperature for 72 hours and then snap frozen at -80°C.

Brain dissection

Frozen flies were kept on ice in a Petri dish during the whole procedure. Each fly was taken out, and the body was rapidly removed by a needle. The head was placed into 0.01 M PBS at pH 7.4 (recovered from tablets, Sigma-Aldrich, USA), and the head capsule was tore apart by two forceps under visual control through a stereo microscope (Nikon SMZ645, Japan) with 10x1

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

magnification. The extracted brains were collected into the same PBS, and then centrifuged at 6000 g for 15 minutes (Centrifuge 5415R; Eppendorf, Germany). The buffer solution was removed and the brain pellet was frozen at -80°C for the future sample preparation. The photo of the dissected brain was taken with DCM510 Microscope CMOS Camera (Scope Tek, China).

Protein extraction, trypsin digestion, total protein and peptide concentration measurement

Brain pellet containing 200 brains was resuspended in 100 µL lysis solution containing 0.1 % (w/v) Protease MAX Surfactant (Promega, USA), 50 mM ammonium bicarbonate, and 10% (v/v) acetonitrile (ACN). The cell lysate was stirred for 60 min at 550 rpm at room temperature. The mixture was then subjected to sonication by Bandelin Sonopuls HD2070 ultrasonic homogenizer (Bandelin Electronic, Germany) at 30% amplitude using short pulses for 5 min. The supernatant was collected after centrifugation at 15,700 g for 10 min at 20°C (Centrifuge 5415R; Eppendorf, Germany). Total protein concentration was measured using bicinchoninic acid assay (BCA Kit; Sigma-Aldrich, USA).

Two µL of 500 mM dithiothreitol (DTT) in 50 mM triethylammonium bicarbonate (TEABC) buffer were added to the samples to the final DTT concentration of 10 mM followed by incubation for 20 min at 56°C. Thereafter, 2 µL of 500 mM iodoacetamide (IAM) in 50 mM TEABC were added to the sample to the final IAM concentration of 10 mM. The mixture was incubated in the darkness at room temperature for 30 min.

The total resultant protein content was digested with trypsin (Trypsin Gold; Promega, USA). The enzyme was added at the ratio of 1:40 (w/w) to the total protein content and the

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

mixture was incubated overnight at 37°C. Enzymatic digestion was terminated by addition of acetic acid (5% w/v).

After the reaction was stopped, the sample was stirred (500 rpm) for 30 min at 45°C followed by centrifugation at 15,700g for 10 min at 20°C (Centrifuge 5415R; Eppendorf, Germany). The supernatant was then added to the filter unit (10 kDa; Millipore, USA) and centrifuged at 13,400g for 20 min at 20°C in the same centrifuge. After that, 100 µL of 50% formic acid were added to the filter unit and the sample was centrifuged at 13,400 g for 20 min at 20°C. The final peptide concentration was measured using Peptide Assay (Thermo Fisher Scientific, USA) on a NanoDrop spectrophotometer (Thermo Fisher Scientific, USA). Sample was dried up using a vacuum concentrator (Eppendorf, Germany) at 45°C. Dried peptides were stored at -80°C until the LC-MS/MS analysis.

Shotgun proteomic analysis

Chromatographic separation of peptides was achieved using homemade C18 column, 25 cm (Silica Tip 360µm OD, 75µm ID, New Objective, USA) connected to an UltimateTM 3000 RSLCnano chromatography system (Thermo Fisher Scientific, USA). Peptides were eluted at 300 nL/min flow rate for 240 min at a linear gradient from 2% to 26% ACN in 0.1% formic acid. Eluted peptides were ionized with electrospray ionization and analyzed on Orbitrap QExactive Plus mass spectrometer (Thermo Fisher Scientific, USA). The survey MS spectrum was acquired at the resolution of 60,000 in the range of m/z 200-2000.

MS/MS data for 20 most intense precursor ions, having charge state of 2 and higher, were obtained using higher-energy collisional dissociation (HCD) at a resolution of 15,000. Dynamic

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

exclusion of up to 500 precursors for 60 seconds was used to avoid repeated analysis of the same peptides.

Proteomic data obtained in this work were deposited in the public repository ProteomeXchange (http://www.proteomexchange.org/) (27) under the accession number PXD004949.

Customized database generation

Fly genomic coordinates of RNA editing sites mapped to exons were obtained from RADAR (1328 sites) (28) and genome-wide studies performed by St Laurent et al. (22). We used two lists of RNA editing sites from previous works (22). The first one, named TIER1, contains 645 exonic non-synonymous high-confident editing sites with high validation rate of >70%. The second list, named TIER2, contains 7986 less confident sites with expected validation rate of 9%. Genomic coordinates obtained from these sources were converted to the coordinates of the recent *Drosophila* genome assembly Dm6 using FlyBase (http://flybase.org/). Changes in protein sequences induced by RNA editing were annotated for all three lists (RADAR, TIER1, and TIER2) using Variant Annotation Integrator (VAI) (http://genome.ucsc.edu/cgi-bin/hgVai). The input was prepared using Python script developed in-house (Supplemental file 2). The VAI output was used to create VAI protein databases containing original and edited fly proteins using another Python script (Supplemental file 3). These protein databases were used to generate the edited protein databases for MaxQuant and X!Tandem searches using another in-house developed Python script (Supplemental file 4). The databases were named RADAR, TIER1, and TIER2 and can be found in Supplemental file 5, 6 and 7, respectively. The search was held

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

against the combination of the databases such as reference+RADAR+TIER1 and reference+RADAR+TIER1+TIER2. The *Drosophila melanogaster* proteome taken from UniProt (released by the September of 2015) was used as a reference database.

MaxQuant and X!Tandem search parameters.

Peptide identification for all data sets except the one published by Aradska et al. (26) were performed using MaxQuant 1.5.4.1 (29) with Andromeda as a search engine as well as with X!tandem, version 2012.10.01.1 (30). For the X!tandem search a peacklist-generation procedure has been held using the MSConvert as a part of ProteoWizard 3.0.8990 software. This resulted in the conversion of the RAW files into the .mgf file format. Multi-parameter algorithm MPscore (31) was used for post-search validation. The data from Aradska et al. were processed using X!tandem only due to the lack of RAW files on ProteomeXchange. All searches were performed against customized databases described in the previous section. MaxQuant's analyses included an initial search with a precursor mass tolerance of 20 ppm, the results of which were used for mass recalibration. In the main searches, precursor and the fragment mass tolerances were set to 4.5 ppm and 20 ppm, respectively. The search included variable modifications, such as methionine oxidation and N-terminal acetylation, as well as carbamidomethylation of cysteine as a fixed modification. The minimal peptide length was set to seven amino acids and up to 1 missed cleavage was allowed.

X!tandem searches were performed using 5 ppm and 0.01 Da mass tolerances set for precursor and fragment ions, respectively. Up to 1 missed cleavage was allowed. Carbamidomethylation of cysteine was used as fixed modification. Methionine oxidation and

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

deamidation of glutamine and asparagines were used as variable modifications. The false discovery rate (FDR) was set to 1%. The group-specific FDR method, which provides separate FDR for the variant peptides as described elsewhere (24), was employed. A target-decoy approach was used to calculate FDR according to the following equation (31) FDR = (number of variant decoy PSMs+1)/number of variant target PSMs.

Variant peptides extraction

In order to analyze the MaxQuant output results, peptide database was generated using built in-house Python script (Supplemental file 8). The script takes the protein database of wild-type fly proteins combined with the database of common protein contaminants (Supplemental file 9) and the VAI protein database as the FASTA database for the search. Pyteomics library was used to develop data processing tools including generation of tryptic peptides *in silico*. (32). Theoretical peptides considered for the analysis were up to 60 amino acids in length and no more than one missed cleavage was allowed. A list of variant peptides generated *in silico* for edited proteins was separated from the unedited peptides of the corresponding original protein. These peptides formed a list of editing-specific variant peptides. At the next step, the list of the unedited peptides was cleaned from the variant peptides to build the editing-specific unedited peptides. Both lists corresponded to the coordinate of the genomic substitution, which leads to changes in the protein sequence. Thus, both lists of peptides characterize the same substitution in the coding sequence of a given protein and allow quantifying the ratio of edited protein to the original protein for a given sample.

Peptide standard synthesis

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Peptides were synthesized by solid phase method using amino acid derivatives with 9-fluorenyl methyloxy carbonyl (Fmoc) protected α-amino groups (Novabiochem). The procedure was performed as described elsewhere (33). Stable isotope-containing leucine (Fmoc-Leu-OH-$^{13}C_6$, ^{15}N, Cambridge Isotope Laboratories) was applied for labeling 11-membered peptides from *cpx* protein (NQMETQVNE$\boldsymbol{L^h}$K and NQIETQVNE$\boldsymbol{L^h}$K). A resin with attached stable isotope-labeled lysine (L-Lys (Boc) ($^{13}C_6$, 99%; $^{15}N_2$, 99%) 2-Cl-Trt, Cambridge Isotope Laboratories) was used for synthesis of two 20-membered peptides of *Syx1A* protein (IEYHVEHAMDYVQTATQDT$\boldsymbol{K^h}$and IEYHVEHAVDYVQTATQDT$\boldsymbol{K^h}$). Further steps of synthesis were also preceded as described (33).

For quality synthesis control an easy LC-MS analysis was held using a chromatographic Agilent ChemStation 1200 series connected to an Agilent - 1100 series LC/MSD Trap XCT Ultra mass spectrometer (Agilent, USA). Since our peptides contained methionines, the quality control also included manual inspection of the MS and MS/MS spectra for possible presence of the peaks produced by oxidized compounds. No such peaks were found in our case.

Concentrations of synthesized peptides were determined using conventional amino acid analysis with their orthophtalic derivatives according to standard amino acid samples.

Multiple Reaction Monitoring experiments

Each sample was analyzed using Dionex UltiMate 3000 RSLC nano System Series (Thermo Fisher Scientific, USA) connected to a triple quadrupole mass spectrometer TSQ Vantage (Thermo Fisher Scientific, USA) in five technical replicates. Generally, 1μl of each

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sample containing 2µg of total native peptides and 100 fmol of each standard peptide was loaded on a precolumn, Zorbax 300SB-C18 (5 µm, 5 × 0.3 mm) (Agilent Technologies, USA) and washed with 5% acetonitrile for 5 min at a flow rate of 10 µl/min before separation on the analytical column. Peptides were separated using RP-HPLC column, Zorbax 300SB-C18 (3.5 µm, 150mm × 75 µm) (Agilent Technologies, USA) using a linear gradient from 95% solvent A (0.1% formic acid) and 5 % solvent B (80% acetonitrile, 0.1% formic acid) to 60% solvent A and 40% solvent B over 25 minutes at a flow rate of 0.4µl/minute.

MRM analysis was performed using QqQ TSQ Vantage (Thermo Scientific, USA) equipped with a nano-electrospray ion source. A set of transitions used for the analysis is shown in Supplemental file 10. Capillary voltage was 2100 V, isolation window was set to 0.7 Da for the first and the third quadrupole, and the cycle time was 3 s. Fragmentation of precursor ions was performed at 1.0 mTorr using collision energies calculated by Skyline 3.1 software (MacCoss Lab Software, USA) (https://skyline.ms/project/home/software/Skyline/begin.view)software (Supplemental file 10). Quantitative analysis of MRM data was also performed using Skyline 3.1 software. Quantification data were obtained from the "total ratio" numbers calculated by Skyline - weighted mean of the transition ratios, where the weight is the area of the internal standard. 5 transitions were used for each peptide including the isotopically labeled standard peptide. Isotopically labeled peptide counterparts were added at the concentration of 1 mg/ml. Each MRM experiment was repeated in 5 technical runs. The results were inspected using Skyline software to compare chromatographic profiles of endogenous peptide and stable-isotope labeled peptide. CV of transition intensity did not exceed 30% in technical runs.

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

All the MRM spectra can be downloaded from Passel (http://www.peptideatlas.org/passel/) (34) under the accession number PASS00946.

Genomic sequencing

The DNA was extracted from 100 *Drosophila* heads (sample #6 in Supplemental file 1) using the standard phenol-chloroform method described elsewhere (35).

The polymorphic sites of seven *D. melanogaster* genes (M244V in *Syx1A*, K398R in Atx2, Y390C in Atpalpha, R489G in CG4587, I125M in cpx, K137E in EndoA and Q1700R in alpha-Spec) were genotyped using Sanger sequencing on Applied Biosystems 3500xL genetic analyzer and SeqScape® software (Thermo Fisher Scientific, USA).

Initial PCRs were performed in a 25 µL volume containing 50 ng genomic DNA template, 10x PCR buffer, 0.5 U of HS Taq DNA Polymerase, 0.2 mM dNTPs (all from Evrogen, Russia), and 80 pmol of each primer. The PCR cycling conditions were the same for all SNPs and were as follows: 95°C for 5 minutes followed by 35 cycles of 94°C for 15 seconds, 59 ^{o}C for 20 seconds, 72°C for 20 seconds and final elongation at 72°C for 6 minutes. Primers were designed using PerlPrimer free software (http://perlprimer.sourceforge.net/) and primer's sequences were as follows: Syx1AFor 5'-ATATAGATCGGGTCTGGATGAG-3' and Syx1ARev 5'-GGATACAGCGTCAACTGGA-3', Atx2For 5'-GGACGCGATCGTGACA-3' andAtx2Rev 5'-GTAGGAGTATTGACTCGGCAT-3', AtpalphaFor-5'-AGAACTGTCTGGTGAAGAATCT-3'andAtpalphaRev 5'-CAGAGCCAGTTCCATGCA-3', CG4587For 5'-GTCGATGTACTGGTTGGCA-3' and CG4587Rev 5'-TCGTTCAGATCAACGATTACGA-3', cpxFor 5'-ACATAACAGTTACAGCTACAGTAGA-

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3' and cpxRev 5'-GCTATGTTATCAGTATTACACGTGT-3', EndoAFor 5'-GCGGTCAAGGGCATCT-3' and EndoARev 5'-GGAGCGATTCACCGAACT-3', alpha-SpecFor 5'-AGATCGCGACCATAGTCGT-3' and alpha-SpecRev 5'-CACCTATATCGCTGCTGTCA-3'. The same primers were used for sequencing.

PCR products were then cleaned up by incubation with the mix of 1U of ExoI and 1U of SAP (both enzymes from Thermo Fisher Scientific, USA) at 37°C for 30 minutes, followed by 80°C for 15 minutes. The sequencing reactions with following EDTA/ethanol purification were carried out using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA) according to manufacturer's instructions.

Results

Search for RNA editing sites in deep *Drosophila* proteome

Tandem mass spectrometry data were taken from recent shotgun proteomics studies (25, 26) available at ProteomeXchange. These results contain data for proteome of *Drosophila*'s whole bodies (25) and whole heads (26). The other data set for *Drosophila*'s brain proteome was obtained here using high-resolution mass spectrometry. Note also that the proteome characterized by Aradska et al. (26) contains only membrane proteins as they were intentionally extracted during the sample preparation. In total, three data sets, representing proteomes of the whole body, the head, and the brain of *Drosophila* were available for the analysis in this study

As noted above, the search for the RNA editing sites was performed using the proteogenomics approach (36). Following this approach the standard fruit fly proteome FASTA database was extended with addition of the edited protein variants found from the transcriptomic

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

data. Three FASTA files containing protein databases with edited sequences derived from transcriptome sequencing results (22) and FlyBase were generated as described in Method section and named TIER1, TIER2, and RADAR. These FASTA files were used in two combinations separately: a stricter TIER1+RADAR and its expanded version TIER1+TIER2+RADAR (Supplemental file 5 and 6). Figure 1 shows schematically the workflow used in this work. The edited peptide identifications are listed in Table1. The genomic coordinates, unedited sequences and UniProt IDs of the peptides are listed in Supplemental file 11.

In all the data studied, 56 peptides corresponding to the RNA editing events were identified. These peptides represent 54 proteins, 2 proteins were represented by two edited peptides each. Note that according to transcriptomic data, a protein can carry several editing sites (22). However, because a mass-spectrometry provides partial protein sequence coverage only, some of the editing events for a protein are missing in the shotgun proteomics data.

All the edited peptides found in this work and listed in Table 1 can be divided into five groups based on the confidence level. Group I contains 7peptides found in all three data sets. Group II contains 7 peptides found in two data sets. Group III contains 7 peptides found in one data set by both search engines. Group IV consists of the peptides found in one data set by one search engine only and contains 11 peptides. Finally, the last group V consists of one peptide found in one data set by one search engine against one database. As expected, this group of 24 peptides includes false-positive identifications.

Genotyping of genomic DNA sites corresponding to found RNA editing events

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The polymorphic sites of seven selected *Drosophila* genes (M244V in *Syx1A*, K398R in *Atx2*, Y390C in *Atpalpha*, R489G in CG4587, I125M in *cpx*, K137E in *EndoA* and Q1700R in*alpha-Spec*) were genotyped, and no traces of genetically encoded A-to-G substitutions were found. This fact confirms the assumption that the substitutions have happened post transcriptionally and the found peptides are the products of RNA editing.

Functional features of edited proteins

To bare light on the purpose of RNA editing, all the peptides found to undergo RNA editing (all the peptides from Table 1) were analyzed with the system of functional protein interactions(STRING, version 10.0 http://string-db.org/). Figure 2 shows the STRING analysis results. There are three groups of proteins with highly confident interactions. These groups were selected and named by manual curation based on Gene Ontology biological process analysis.

First ("synapse signaling" in the Fig. 3) group contains the following proteins: syntaxin (*Syx1A*), synaptotagmin (*Syt1*), complexin (*cpx*), adaptin (*AP-2 alpha*), endophilin A (*endoA*), stoned protein B (*StnB*), calcium-dependent secretion activator (*Cadps*), nervous wreck (*nwk*), and Ataxin-2 (*Atx-2*). These proteins play a key role in synaptic transmission. Particularly, syntaxin is a component of a SNARE complex that provides fusion of synaptic vesicles with the presynaptic membrane (37). Synaptotagmin and stoned protein B also interact and plays a role in synaptic vesicle endocytosis (38). Complexin is described as binding the SNARE proteins and, thus, acting on synaptic transmission as well (39). Adaptins play a role in a process of synaptic vesicle recycling (40). Endophilin (*endoA*), nervous wreck (*nwk*), and thickveins (*tkv*) are acting in the process of vesicle endocytosis in neuromuscular junction (41–44). Calcium-dependent

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

secretion activator (*Cadps*) and frequenin (*Frq1*) are Ca^{2+}-dependent factors of vesicle endocytosis (45, 46). The synapse signaling group contains 4 peptides from confidence Group I, i. e. the most confidently identified edited peptides, and 3 peptides from confidence Group II described above.

The second group ("cytoskeleton" in the Fig. 3) of proteins consists of non-muscular myosin (*zip*), alpha-Spectrin (*alpha-Spec*), titin (*sls*), Dynein (*Dhc64C*), polychaetoid (*pyd*), paramyosin (*Prm*), and the products of *Zasp52*, *Unc-89*and *CG11148* gens. All proteins from this group are either components of cytoskeleton or interact with them and take a part in cell transport processes. This group includes only one peptide from confidence Group I.

The third clearly associated group ("RNA helicase activity" shown in Fig. 3) consists of the protein named probable ATP-dependent RNA helicase pitchoune (*pit*), ribosome biogenesis protein WDR12 homolog (CG6724), a product of the gene CG10077, and ATP-dependent RNA helicase p62 (*Rm62*). All these proteins act on RNA. RNA helicase p62 is involved in RNA interference (47). The pitchhoune gene of probable RNA helicase is reportedly required for cell growth and proliferation. It is believed to play a role in ribosome genesis and, thus, affect the protein synthesis (48).

There is also one protein not assigned to any of the three groups, yet found being edited in all tree data sets. The product of the gene *CG4587* is a protein engaged in Ca^{2+}- dependent nociception response (49). This protein shares about 30% identity with its important paralog, straitjacket (*stj*), which is another Ca^{2+}-dependent channel, also involved in nociception (50). Straitjacket is not in our list of the proteins corresponding to the RNA editing event, but it was found in the fruit fly proteome and identified by search engines in our data. Theoretically, *stj*

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

may correspond to the editing event, because its transcript is edited and the protein was in our customized database (supplemental files 5 and 6). The fact that *stj* did not listed among the identified edited proteins can be explained by its underrepresentation in the spectra compared to *CG4587*. The head and brain proteomes are of the most interest in terms of the expression of the Ca^{2+}-dependent channels. In the brain proteome, the relative intensity of *stj* calculated by MaxQuant search engine is 34.7 times lower compared with *CG4587*. Label-free quantification (LFQ) obtained for X!tandem search results has shown the following results for *CG4587*/*stj* LFQ ratios: 12.4, 4.7, and 6.6 for SIN (51), NSAF (52), and emPAI (53) LFQ algorithms, respectively. For the whole head proteome the *CG4587*/*stj* ratios were 5.4, 4.4, and 8.5 for these LFQ algorithms, respectively. Therefore, even though straitjacket is an important and well characterized protein of nervous tissue, its paralog, *CG4587* gene product has significantly higher abundance in the proteomes of isolated brain as well as the whole head. According to our observations, it also undergoes RNA editing that warrants further studies.

We have used GOrilla (http://cbl-gorilla.cs.technion.ac.il/) (54) to check the Gene Ontology (GO) (55) biological process enrichment of the proteins undergoing editing against the whole database of potentially edited proteins based on transcriptomic data. One of the most enriched processes is synaptic signaling (GO:0099536, q-value $6.95 \cdot 10^{-10}$, which is a P-value corrected for multiple testing, for more details see Supplemental file 12), which is in good agreement with the STRING pathway analysis results. The complete Gorilla output data corresponding to biological processes are shown in Supplemental file 12.

Peptide selection for targeted analysis

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

After shotgun analysis of a fruit fly editome on proteomic level a targeted analysis of the most interesting sites was held as a reasonable continuation of the study. The main purpose of this analysis was validation of the shotgun proteomic results. Because temperature was reported as one of the environmental factors affecting RNA editing (12, 56), the quantitative study on the changes in the RNA editing pattern with temperature was performed in this work. The peptides of syntaxin (*Syx1A*) and complexin (*cpx*) were selected for MRM analysis in samples from flies raised at different temperature environment for the following reasons: first, these peptides were edited in all analyzed data sets. Secondly, these proteins act together with the SNARE complex and have a strong functional link between each other (37). Thirdly, the complexin (*cpx*) was reported in the transcriptome studies as undergoing RNA editing which depends on the temperature (12). For the above reasons, the conditions, in which the flies were raised in this work, were exactly the same as in previous transcriptome studies. Also, we have verified that these two edited sites had no nucleotide substitutions at the genome level as described above.

Results of MRM analysis

Currently, there are three tier of assays for targeted MS analyses, as described by Carr et al. (58), and we defined present study as Tier 2 one. In our work, first, we study modified peptides in non-human samples using MRM method. Also, it is not a clinical research, but it uses heavy isotope labeled synthetic internal standards that, did not undergo purification. Despite the lack of purification step, the standard peptides were analyzed using LC-MS prior to the MRM analysis that is also required for Tier 2 experiments according to (57).

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Four samples were prepared for MRM analysis as shown in Supplemental file 1 and described in the "Experiment design and statistical rationale" section.

After the trypsin digestion step, the total peptide concentration was measures for the samples (see Supplemental file 1) followed by MRM measurements for selected peptides. Precursor ions and transitions for monitoring were selected manually using Skyline. The information about the spiked-in labeled synthetic peptides and the targeted peptides is provided in the Supplemental file 10 including *m/z* of precursors, *m/z* of monitored fragment ions, collision energy (CE), and a type of fragment ions.

The selected peptides were IEYHVEHA**M**DYVQTATQDTK and its edited variant IEYHVEHA**V**DYVQTATQDTK from syntaxin 1A (*Syx1A*, Uniprot ID Q24547, sequence positions 122-132), as well as NQIETQVNELK and its edited variation NQMETQVNELK from complexin (*cpx*, Uniprot ID Q8IPM8, sequence positions 382-398). The results of concentration measurements for these peptides are shown in Table2.

Temperature effect on editing

The diagrams representing the relationship between the edited and unedited peptide concentrations in the samples are shown in Figure 3. Even though the ratio of edited to unedited peptide variant concentrations does not change with the temperature, the Exact Fisher Test was performed to statistically confirm this visual assumption. The test has confirmed that there are no statistically significant changes in the ratio with the temperature. This observation contradicts to the transcriptome behavior, which demonstrated a slight increase in editing rate for *cpx* at 20°C, and its decrease at 30°C (12).

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Although, an eye-catching difference in overall expression had to be statistically tested as well and resulted in some interesting observation. Temperature dependency of the expression of the selected peptides was estimated with the use of ANOVA test counting P-value <0.01 as a significant result. The total concentration of *Syx1A* depended significantly on the temperature (ANOVA P-value $6.766e^{-8}$) (Fig. 2). Particularly it stays equal at 10 and 30°C, but drops at 20 ^{o}C by 2 nmol/g of total protein. At the same time, *cpx* also shows temperature dependency (ANOVA P-value $1.215e^{-10}$), but its concentration is equal at 20 and 30 ^{o}C, but is lower at 10 ^{o}C by the about 3.5 nmol/g.

Discussion

Shotgun proteomics recently gained a success in identification and quantification of proteins as gene products. Today, LC-MS/MS analysis using high-resolution mass spectrometry provides deep proteome covering about 50% of human genome from a single sample (58, 59). This ratio is even higher in model organisms, which may be characterized by more open genomes with higher numbers of expressed proteins (60). Correspondingly, a next aim of proteomics is to catalogue proteoforms of each protein, i.e. multiple protein species originated from one gene (61, 62).

A major field in proteoform profiling is posttranslational modification of proteins, which is out of scope of the present study and has its own approaches and limitations. Nucleic acid alterations coding amino acid replacements are included in the emerging field of proteogenomics (15). Among such nucleic acid changes, DNA sequence variants and alternative splicing are most studied types of recoding events (36). Many works describe profiling of proteoforms using

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

customized nucleic acid data on DNA mutations (63, 64) and alternative splicing (36, 65). We hypothesized here that the same approach with construction of customized databases for the search in shotgun proteome would be useful to identify ADAR-mediated RNA editing events in proteins. The workflow was implemented on RNA-editome data of *Drosophila melanogaster* (22, 28) accompanied by proteomes of whole body (25), head (26) and brain of this insect (Figure 1).

Shotgun proteomics data search regardless of data acquisition method conventionally uses consensus genomic database of organism under analysis. In proteogenomics, specific customized sequences may be added to this database. In case of single amino acid substitution, they may be easily mocked by chemical modification of peptides (66). That is why for better validation of ADAR edited hits we used two search engines, where it was possible. Moreover, using X!Tandem with in-home post-search algorithm (67), we also calculated group-specific FDR for edited peptides (24).

In this study, we performed searches of three fruit fly proteome data sets against three RNA editome databases. These searches have resulted in identification of 56 edited sites, 21 of them being validated by identification in at least two data sets and/or by two search engines. Also, the study included validation of two edited sites along with its unedited counterparts using MRM assay. The results of this study have shown that shotgun proteomics can be used for identification of RNA editing sites in proteins employing a workflow similar to the one applied previously for identifying genetically encoded single amino acid polymorphisms at the proteome level (36).

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The other problem considered in this work was revealing the biological meaning of observed RNA editing events. First, the ADAR enzymes affect directly the RNA function (68). It means that some amino acid coding events of editing may carry no sense, representing translational noise or having almost no effect on the protein structure. However, at least several examples of RNA editing important on the protein level are known, such as Q/R substitution in mammalian glutamate receptor subunits (69). Further quantitative and functional studies will elucidate a role of each edited site found.

As described above, the edited proteins may be classified into three functional groups (Fig. 3). Of them, RNA helicase activity was predicted for the group of four proteins. Interestingly, the ADAR activity itself contributes to the structure of double-stranded RNA (70). A putative modulation of function of RNA helicases represents an additional mechanism controlling dsRNAs.

Two other groups of proteins where edited sites are identified include interactors of SNARE complex ("synapse signaling" group) and the protein components of cytoskeleton. As shown in Fig. 3, some members of these functional groups may also interact with each other. Indeed, SNARE proteins provide an ensemble for calcium-dependent fusion of synaptic vesicles to release neurotransmitters (71). Components of SNARE are expectedly attached to the elements of plasma membrane internal lining, such as the non-muscular myosin *zip* (Fig. 3, String protein interaction database). Our data further support the evidence that ADAR-mediated RNA editing extensively regulates the synaptic release and reuptake (19).

In order to validate findings of edited sites from shotgun proteome analysis, we performed quantitative measurements using targeted MS approach for tryptic peptides from two proteins of

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

synaptic SNARE complex, which contained edited sites. These two proteins, *Syx1A* and *cpx*, are reportedly interacting physically during performance of the complex which provides synaptic vesicle fusion with plasma membrane to release content of vesicles into synaptic cleft (39). Upon validation of edited sites, it was tempting to measure the level of editing and its variation with the environmental changes, such as the temperature. Previous transcriptome studies reported that the level of ADAR-mediated RNA editing was temperature dependent (12).

Three temperature points were used for fruit fly raising as described in the transcriptome study (12), two of them representing extreme condition for fruit fly propagation. For two sites of editing, in *Syx1A* and *cpx* proteins, no temperature dependence of the RNA editing level was observed. Overall expression of both proteins as measured by our peptide-based assays was changed with temperature, although not dramatically. Interestingly enough that the level of editing in the exemplary site of *Syx1A* protein was low in comparison with unedited form for all samples under study. However, the biological significance of this site is not obvious. In contrast, Ile384Met editing in complexin occurred for one third of these proteins with remarkably constant ratio from sample-to-sample (Fig. 2). From this data, a role of *cpx* RNA editing at the protein level may be hypothesized for functioning of presynaptic part in fruit fly brain.

Conclusion

RNA editing mediated by the ADAR enzymes is a widespread posttranscriptional modification shown to be important in many pluricellular organisms from worms to human. This type of editing converts adenosine residue of RNA to inosine, thus, changing the amino acid code if the conversions happens in a coding region. Using the customized protein database

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

generated from RNA sequencing data and containing proteins corresponding to the ADAR-mediated RNA editome, we found 56 edited sites in the proteins for three proteomic data sets. Two of these proteins, which belong to components of the SNARE presynaptic complex, were validated and quantified using targeted MS assay. Contrary to the expectations, the measurements of the ratio between edited and unedited sequences of these two sites have shown no dependency on fly growth temperature. This ratio was remarkably constant for all samples studied and was low in syntaxin 1A protein (2-5%), but much higher in complexin (about 30%). Such a high and constant level of editing, at least, in the latter species, indicates its importance for the protein functioning.

Acknowledgements

The work was funded by Federal Agency of Scientific Organizations, Russia, Topic #1 “Search for postgenomic biomarkers” to Institute of Biomedical Chemistry.

Authors thank Dr. Natalia Romanova from the Department of Genetics of Biological Faculty of Moscow State University for providing the *Drosophila melanogaster* Canton S stock.

References

1. Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof P., Van Boom, J.H., and Tromp, M.C. (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46, 819–26

2. Gott, J. M., and Emeson, R. B. (2000) Functions and mechanisms of RNA editing. *Annu.*

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Rev. Genet. 34, 499–531

3. Blanc, V., and Davidson, N. O. (2003) C-to-U RNA editing: mechanisms leading to genetic diversity. *J. Biol. Chem.* 278, 1395–8

4. Bass, B. L., and Weintraub, H. (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* 55, 1089–1098

5. Kim, U., Wang, Y., Sanford, T., Zeng, Y., and Nishikura, K. (1994) Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc. Natl. Acad. Sci. U. S. A.* 91, 11457–61

6. Shikanai, T. (2015) RNA editing in plants: Machinery and flexibility of site recognition. *Biochim. Biophys. Acta* 1847, 779–785

7. Takenaka, M., Verbitskiy, D., Zehrmann, A., Härtel, B., Bayer-Császár, E., Glass, F., and Brennicke, A. (2014) RNA editing in plant mitochondria -connecting RNA target sequences and acting proteins. *Mitochondrion* 19, 191–197

8. Seeburg, P. H. (2002) A-to-I editing: New and old sites, functions and speculations. *Neuron* 35, 17–20

9. He, T., Wang, Q., Feng, G., Hu, Y., Wang, L., and Wang, Y. (2011) Computational detection and functional analysis of human tissue-specific A-to-I RNA editing. *PLoS One* 6, 1–7

10. Vissel, B., Royle, G. A., Christie, B. R., Schiffer, H. H., Ghetti, A., Tritto, T., Perez-Otano, I., Radcliffe, R. A., Seamans, J., Sejnowski, T., Wehner, J. M., Collins, A. C., O'Gorman, S., and Heinemann, S. F. (2001) The role of RNA editing of kainate receptors

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in synaptic plasticity and seizures. *Neuron* 29, 217–27

11. Egebjerg, J., and Heinemann, S. F. (1993) Ca2+ permeability of unedited and edited versions of the kainate selective glutamate receptor GluR6. *Proc. Natl. Acad. Sci. U. S. A.* 90, 755–9

12. Rieder, L. E., Savva, Y. a, Reyna, M. a, Chang, Y.-J., Dorsky, J. S., Rezaei, A., and Reenan, R. a (2015) Dynamic response of RNA editing to temperature in Drosophila. *BMC Biol.* 13,

13. Low, T. Y., van Heesch, S., van den Toorn, H., Giansanti, P., Cristobal, A., Toonen, P., Schafer, S., Hübner, N., vanBreukelen, B., Mohammed, S., Cuppen, E., Heck, A. J. R., and Guryev, V. (2013) Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. *Cell Rep.* 5, 1469–1478

14. Picardi, E., Manzari, C., Mastropasqua, F., Aiello, I., D'Erchia, A. M., and Pesole, G. (2015) Profiling RNA editing in human tissues: towards the inosinome Atlas. *Sci. Rep.* 5, 14941

15. Nesvizhskii, A. I. (2014) Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* 11, 1114–1125

16. Zhao, H., Zhang, P., Gao, H., He, X., Dou, Y., Huang, A. Y., Liu, X., Ye, A. Y., Dong, M., and Wei, L. (2015) Profiling the RNA editomes of wild-type C. elegans and ADAR mutants. *Genome Res.* 25, 66–75

17. Chernobrovkin, A. L., Kopylov, A. T., Zgoda, V. G., Moysa, A. A., Pyatnitskiy, M. A., Kuznetsova, K. G., Ilina, I. Y., Karpova, M. A., Karpov, D. S., Veselovsky, A. V, Ivanov,

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

M. V, Gorshkov, M. V, Archakov, A. I., and Moshkovskii, S. A. (2015) Methionine to isothreonine conversion as a source of false discovery identifications of genetically encoded variants in proteogenomics. *J. Proteomics* 120, 169–178

18. Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M. C., Zimmerman, L. J., Shaddox, K. F., Kim, S., Davies, S. R., Wang, S., Wang, P., Kinsinger, C. R., Rivers, R. C., Rodriguez, H., Townsend, R. R., Ellis, M. J. C., Carr, S. A., Tabb, D. L., Coffey, R. J., Slebos, R. J. C., and Liebler, D. C. (2014) Proteogenomic characterization of human colon and rectal cancer. *Nature* 513, 382–7

19. Hoopengardner, B., Bhalla, T., Staber, C., and Reenan, R. (2003) Nervous system targets of RNA editing identified by comparative genomics. *Science* 301, 832–836

20. Rodriguez, J., Menet, J. S., and Rosbash, M. (2012) Nascent-Seq Indicates Widespread Cotranscriptional RNA Editing in Drosophila. *Mol. Cell* 47, 27–37

21. Stapleton, M., Carlson, J. W., and Celniker, S. E. (2006) RNA editing in Drosophila melanogaster: New targets and functional consequences. *RNA* 12, 1922–32

22. St Laurent, G., Tackett, M. R., Nechkin, S., Shtokalo, D., Antonets, D., Savva, Y. a, Maloney, R., Kapranov, P., Lawrence, C. E., and Reenan, R. A. (2013) Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in Drosophila. *Nat. Struct. Mol. Biol.* 20, 1333–9

23. Yu, Y., Zhou, H., Kong, Y., Pan, B., Chen, L., Wang, H., Hao, P., and Li, X. (2016) The Landscape of A-to-I RNA Editome Is Shaped by Both Positive and Purifying Selection. *PLOS Genet.* 12, e1006191

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

24. Lobas, A. A., Karpov, D. S., Kopylov, A. T., Solovyeva, E. M., Ivanov, M. V, Ilina, I. I., Lazarev, V. N., Kuznetsova, K. G., Ilgisonis, E. V, Zgoda, V. G., Gorshkov, M. V, and Moshkovskii, S. A. (2016) Exome-based proteogenomics of HEK-293 human cell line: Coding genomic variants identified at the level of shotgun proteome. *Proteomics*, 1–12

25. Xing, X., Zhang, C., Li, N., Zhai, L., Zhu, Y., Yang, X., and Xu, P. (2014) Qualitative and quantitative analysis of the adult Drosophila melanogaster proteome. *Proteomics* 14, 286–290

26. Aradska, J., Bulat, T., Sialana, F. J., Birner-Gruenberger, R., Erich, B., and Lubec, G. (2015) Gel-free mass spectrometry analysis of *Drosophila melanogaster* heads. *Proteomics* 15, 3356–3360

27. Vizcaíno, J., Deutsch, E. E. W., Wang, R., Vizcaino, J. A., Deutsch, E. E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P.-A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H.-J., Albar, J. P., Martinez-Bartolome, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotech* 32, 223–226

28. Ramaswami, G., and Li, J. B. (2014) RADAR: A rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* 42, 109–113

29. Neuhauser, N., Nagaraj, N., McHardy, P., Zanivan, S., Scheltema, R., Cox, J., and Mann, M. (2013) High performance computational analysis of large-scale proteome data sets to assess incremental contribution to coverage of the human genome. *J. Proteome Res.* 12, 2858–2868

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

30. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466–1467

31. Levitsky, L. I., Ivanov, M. V., Lobas, A. A., and Gorshkov, M. V. (2016) Unbiased False Discovery Rate Estimation for Shotgun Proteomics Based on the Target-Decoy Approach. *J. Proteome Res.*, Article ASAP

32. Goloborodko, A. A., Levitsky, L. I., Ivanov, M. V., and Gorshkov, M. V. (2013) Pyteomics - A python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrom.* 24, 301–304

33. Hood, C. A., Fuentes, G., Patel, H., Page, K., Menakuru, M., and Park, J. H. (2008) Fast conventional Fmoc solid-phase peptide synthesis with HCTU. *J. Pept. Sci.* 14, 97–101

34. Farrah, T., Deutsch, E. W., Kreisberg, R., Sun, Z., Campbell, D. S., Mendoza, L., Kusebauch, U., Brusniak, M. Y., Hüttenhain, R., Schiess, R., Selevsek, N., Aebersold, R., and Moritz, R. L. (2012) PASSEL: The PeptideAtlas SRMexperiment library. *Proteomics* 12, 1170–1175

35. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., pp. 1-50, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

36. Sheynkman, G. M., Shortreed, M. R., Cesnik, A. J., and Smith, L. M. (2016) Proteogenomics: Integrating Next-Generation Sequencing and Mass Spectrometry to Characterize Human Proteomic Variation. *Annu. Rev. Anal. Chem.* 9, 1-545

37. Wang, Z., and Chapman, E. R. (2010) Rat and Drosophila synaptotagmin 4 have opposite effects during SNARE-catalyzed membrane fusion. *J. Biol. Chem.* 285, 30759–30766

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

38. Fergestad, T., and Broadie, K. (2001) Interaction of stoned and synaptotagmin in synaptic vesicle endocytosis. *J. Neurosci.* 21, 1218–27

39. Hu, K., Carroll, J., Rickman, C., and Davletov, B. (2002) Action of complexin on SNARE complex. *J. Biol. Chem.* 277, 41652–41656

40. González-Gaitán, M., and Jäckle, H. (1997) Role of drosophila α-adaptin in presynaptic vesicle recycling. *Cell* 88, 767–776

41. Guichet, A., Wucherpfennig, T., Dudu, V., Etter, S., Wilsch-Bräuniger, M., Hellwig, A., González-Gaitán, M., Huttner, W. B., and Schmidt, A. A. (2002) Essential role of endophilin A in synaptic vesicle budding at the Drosophila neuromuscular junction. *EMBO J.* 21, 1661–1672

42. Verstreken, P., Kjaerulff, O., Lloyd, T. E., Atkinson, R., Zhou, Y., Meinertzhagen, I. A., and Bellen, H. J. (2002) Endophilin mutations block clathrin-mediated endocytosis but not neurotransmitter release. *Cell* 109, 101–112

43. Coyle, I. P., Koh, Y. H., Lee, W. C. M., Slind, J., Fergestad, T., Littleton, J. T., and Ganetzky, B. (2004) Nervous Wreck, an SH3 Adaptor Protein that Interacts with Wsp, Regulates Synaptic Growth in Drosophila. *Neuron* 41, 521–534

44. Li, W., Yao, A., Zhi, H., Kaur, K., Zhu, Y. C., Jia, M., Zhao, H., Wang, Q., Jin, S., Zhao, G., Xiong, Z. Q., and Zhang, Y. Q. (2016) Angelman Syndrome Protein Ube3a Regulates Synaptic Growth and Endocytosis by Inhibiting BMP Signaling in Drosophila. *PLoS Genet.* 12, e1006062

45. Renden, R., Berwin, B., Davis, W., Ann, K., Chin, C. T., Kreber, R., Ganetzky, B.,

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Martin, T. F. J., and Broadie, K. (2001) Drosophila CAPS is an essential gene that regulates dense-core vesicle release and synaptic vesicle fusion. *Neuron* 31, 421–437

46. Pongs, O., Lindemeier, J., Zhu, X. R., Theil, T., Engelkamp, D., Krah-Jentgens, I., Lambrecht, H., Koch, K. W., Schwemer, J., Rivosecchi, R., Mallart, A., Galceran, J., Canal, I., Barbas, J. A., and Ferrús, A. (1993) Frequenin-A novel calcium-binding protein that modulates synaptic efficacy in the drosophila nervous system. *Neuron* 11, 15–28

47. Ishizuka, A., Siomi, M. C., and Siomi, H. (2002) A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev.* 16, 2497–2508

48. Zaffran, S., Chartier, A., Gallant, P., Astier, M., Arquier, N., Doherty, D., Gratecos, D., Séméria, M., Acharya, J. K., Jalink, K., Hardy, R. W., Hartenstein, V., Zuker, C. S., Amati, B., Land, H., Blackwood, E. M., Eisenman, R. N., Brand, A., Perrimon, N., Brown, N. H., Kafatos, F. C., Burgess, S., Coulo, J. R., Guthrie, C., Cerini, C., Semeriva, M., Gratecos, D., Chen, J. H., Lin, R. J., Chou, M. M., Blenis, J., Chou, T. B., Perrimon, N., Company, M., Arenas, J., Abelson, J., Courey, A. J., Tjian, R., Dalbadie-McFarland, G., Abelson, J., Dingwall, C., Laskey, A., Eberl, D. E., Lorenz, L. J., Melnick, M. B., Sood, V., Lasko, P., Perrimon, N., Edgar, B. A., Lehner, C. F., Evan, G. I., Lewis, G. K., Ramsay, G., Bishop, J. M., Evan, G. I., Littlewood, T. D., Foe, V. E., François, V., Solloway, M., O'Neil, J., Emery, J., Bier, E., Frasch, M., Glover, D. M., Saumweber, H., Gallant, P., Shio, Y., Cheng, P. F., Parkhurst, S. M., Eisenman, R. N., Garzino, V., Periera, A., Laurenti, P., Graba, Y., Levis, R. W., Parco, Y. Le, Pradel, J., Ghisolfi, L., Kharrat, A., Joseph, G., Amalric, F., Erard, M., Graba, Y., Laurenti, P., Perrin, L., Aragnol, D., Pradel, J., Grandori, C., Mac, J., Siëbelt, F., Ayer, D. E., Eisenman, R. N.,

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gurujaran, R., Perry-O'Keefe, H., Melton, D. A., Weeks, D. L., Hirling, H., Scheffner, T., Restle, T., Stahl, H., Iggo, R., Jamieson, D., MacNeill, S., Southgate, J., MacPheat, J., Lane, D., Jamieson, D., Beggs, J., Jones, R. M., Branda, J., Johnston, K. A., Polymenis, M., Gadd, M., Rustgi, A., Callanan, L., Schmidt, E. V., Kim, Y., Nirenberg, M., Lasko, P. F., Ashburner, M., Liang, L., Diehl-Jones, W., Lasko, P., Liang, W. Q., Clark, J. A., Fournier, M. J., Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P., Nishi, K., Schnier, J., Slonimski, P. P., Luo, L., Liao, J., Jan, Y., Jan, N., Melese, T., Xue, Z., Metz, A. M., Timmer, R. T., Browning, K. S., Morata, G., Ripoll, P., Nakajima, T., Uchida, C., Anderson, S. F., Lee, C.-G., Hurwitz, J., Parvin, J. D., Montminy, M., Norbury, C., Nurse, P., O'Day, C., Chavanikamannil, F., Abelson, J., Pardee, A. B., Pause, A., Sonenberg, N., Perrimon, N., Noll, E., McCall, K., Brand, A., Polymenis, M., Schmidt, E. V., Ripmaster, T., Vaughn, G., Woolford, J., Roussel, D., Bennett, K., Rubin, G., Spradling, A., Ruohola, A., Bremer, K. A., Baker, D., Swedlow, J. R., Jan, L. Y., Jan, Y. N., Sachs, A., Davis, R., Savitsky, K., Rotman, G., Ziv, Y., Bar-Shira, A., Gilad, S., Sartiel, A., Shiloh, ., Schmid, S., Linder, P., Schreiber-Argus, N., Stein, D., Chen, K., Goltz, J. S., Stevens, L., DePinho, R. A., Schwer, B., Guthrie, C., Serrano, N., O'Farrell, P. H., Simpson, P., Morata, G., Slusarski, D. C., Corces, V. G., Moon, R. T., Smith, A., Orr-Weaver, T., Stewart, M. J., Thomas, G., Strauss, E., Guthrie, C., Thummel, C., Pirrotta, V., Venema, J., Bousquet-Antonelli, C., Gelugne, J. P., Caizergues-Ferrer, M., Tollervey, D., Weaver, P., Sun, C., Chang, T., Xu, T., Rubin, G., Zaffran, S., Astier, M., Gratecos, D., and Semeriva, M. (1998) A Drosophila RNA helicase gene, pitchoune, is required for cell growth and proliferation and is a potential target of d-Myc. *Development* 125, 3571–84

49. Neely, G. G., Keene, A. C., Duchek, P., Chang, E. C., Wang, Q. P., Aksoy, Y. A.,

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Rosenzweig, M., Costigan, M., Woolf, C. J., Garrity, P. A., and Penninger, J. M. (2011) TrpA1 regulates thermal nociception in Drosophila. *PLoS One* 6, e24343

50. Neely, G. G., Hess, A., Costigan, M., Keene, A. C., Goulas, S., Langeslag, M., Griffin, R. S., Belfer, I., Dai, F., Smith, S. B., Diatchenko, L., Gupta, V., Xia, C. ping, Amann, S., Kreitz, S., Heindl-Erdmann, C., Wolz, S., Ly, C. V, Arora, S., Sarangi, R., Dan, D., Novatchkova, M., Rosenzweig, M., Gibson, D. G., Truong, D., Schramek, D., Zoranovic, T., Cronin, S. J. F., Angjeli, B., Brune, K., Dietzl, G., Maixner, W., Meixner, A., Thomas, W., Pospisilik, J. A., Alenius, M., Kress, M., Subramaniam, S., Garrity, P. A., Bellen, H. J., Woolf, C. J., and Penninger, J. M. (2010) A Genome-wide Drosophila screen for heat nociception identifies α2δ3 as an evolutionarily conserved pain gene. *Cell* 143, 628–638

51. Griffin, N. M., Yu, J., Long, F., Oh, P., Shore, S., Li, Y., Koziol, J. A., and Schnitzer, J. E. (2010) Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* 28, 83–89

52. Paoletti, A. C., Parmely, T. J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R. C., Conaway, J. W., Florens, L., and Washburn, M. P. (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci U S A* 103, 18928–18933

53. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell. Proteomics* 4, 1265–1272

54. Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009) GOrilla: a tool for

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48

55. Blake, J. A., Dolan, M., Drabkin, H., Hill, D. P., Ni, L., Sitnikov, D., Bridges, S., Burgess, S., Buza, T., McCarthy, F., Peddinti, D., Pillai, L., Carbon, S., Dietze, H., Ireland, A., Lewis, S. E., Mungall, C. J., Gaudet, P., Chisholm, R. L., Fey, P., Kibbe, W. A., Basu, S., Siegele, D. A., McIntosh, B. K., Renfro, D. P., Zweifel, A. E., Hu, J. C., Brown, N. H., Tweedie, S., Alam-Faruque, Y., Apweiler, R., Auchinchloss, A., Axelsen, K., Bely, B., Blatter, M.-C., Bonilla, C., Bougueleret, L., Boutet, E., Breuza, L., Bridge, A., Chan, W. M., Chavali, G., Coudert, E., Dimmer, E., Estreicher, A., Famiglietti, L., Feuermann, M., Gos, A., Gruaz-Gumowski, N., Hieta, R., Hinz, U., Hulo, C., Huntley, R., James, J., Jungo, F., Keller, G., Laiho, K., Legge, D., Lemercier, P., Lieberherr, D., Magrane, M., Martin, M. J., Masson, P., Mutowo-Muellenet, P., O'Donovan, C., Pedruzzi, I., Pichler, K., Poggioli, D., Porras Millan, P., Poux, S., Rivoire, C., Roechert, B., Sawford, T., Schneider, M., Stutz, A., Sundaram, S., Tognolli, M., Xenarios, I., Foulger, R., Lomax, J., Roncaglia, P., Khodiyar, V. K., Lovering, R. C., Talmud, P. J., Chibucos, M., Gwinn Giglio, M., Chang, H.-Y., Hunter, S., McAnulla, C., Mitchell, A., Sangrador, A., Stephan, R., Harris, M. A., Oliver, S. G., Rutherford, K., Wood, V., Bahler, J., Lock, A., Kersey, P. J., McDowall, M. D., Staines, D. M., Dwinell, M., Shimoyama, M., Lauledderkind, S., Hayman, T., Wang, S.-J., Petri, V., Lowry, T., D'Eustachio, P., Matthews, L., Balakrishnan, R., Binkley, G., Cherry, J. M., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hitz, B. C., Hong, E. L., Karra, K., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., Weng, S., Wong, E. D., Berardini, T. Z., Li, D., Huala, E., Mi, H., Thomas, P. D., Chan, J., Kishore, R., Sternberg, P., Van Auken, K., Howe, D., and

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Westerfield, M. (2013) Gene ontology annotations and resources. *Nucleic Acids Res.* 41, D530-5

56. Garrett, S. C., and Rosenthal, J. J. C. (2012) A role for A-to-I RNA editing in temperature adaptation. *Physiology* 27, 362–369

57. Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., Domon, B., Deutsch, E. W., Grant, R. P., Hoofnagle, A. N., Hüttenhain, R., Koomen, J. M., Liebler, D. C., Liu, T., MacLean, B., Mani, D. R., Mansfield, E., Neubert, H., Paulovich, A. G., Reiter, L., Vitek, O., Aebersold, R., Anderson, L., Bethem, R., Blonder, J., Boja, E., Botelho, J., Boyne, M., Bradshaw, R. A., Burlingame, A. L., Chan, D., Keshishian, H., Kuhn, E., Kinsinger, C., Lee, J. S. H., Lee, S., Moritz, R., Oses-Prieto, J., Rifai, N., Ritchie, J., Rodriguez, H., Srinivas, P. R., Townsend, R. R., Van Eyk, J., Whiteley, G., Wiita, A., and Weintraub, S. (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* 13, 907–17

58. Mann, M., Kulak, N. A., Nagaraj, N., and Cox, J. (2013) The coming age of complete, accurate, and ubiquitous proteomes. *Mol. Cell* 49, 583–90

59. Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L., and Gygi, S. P. (2015) A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat. Biotechnol.* 33, 743–9

60. Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* 11, M111.013722

61. Toby, T. K., Fornelli, L., and Kelleher, N. L. (2016) Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* 9, 499–519

62. Lisitsa, A., Moshkovskii, S., Chernobrovkin, A., Ponomarenko, E., and Archakov, A. (2014) Profiling proteoforms: promising follow-up of proteomics for biomarker discovery. *Expert Rev. Proteomics* 11, 121–129

63. Mertins, P., Mani, D. R., Ruggles, K. V, Gillette, M. A., Clauser, K. R., Wang, P., Wang, X., Qiao, J. W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatza, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., Huang, K., Lin, C., McLellan, M. D., Yan, P., Davies, S. R., Townsend, R. R., Skates, S. J., Wang, J., Zhang, B., Kinsinger, C. R., Mesri, M., Rodriguez, H., Ding, L., Paulovich, A. G., Fenyö, D., Ellis, M. J., Carr, S. A., and NCI CPTAC (2016) Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 534, 55–62

64. Karpova, M. A., Karpov, D. S., Ivanov, M. V, Pyatnitskiy, M. A., Chernobrovkin, A. L., Lobas, A. A., Lisitsa, A. V, Archakov, A. I., Gorshkov, M. V, and Moshkovskii, S. A. (2014) Exome-driven characterization of the cancer cell lines at the proteome level: the NCI-60 case study. *J. Proteome Res.* 13, 5551–60

65. Panwar, B., Menon, R., Eksi, R., Li, H.-D., Omenn, G. S., and Guan, Y. (2016) Genome-Wide Functional Annotation of Human Protein-Coding Splice Variants Using Multiple Instance Learning. *J. Proteome Res.* 15, 1747–1753

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

66. Chernobrovkin, A. L., Kopylov, A. T., Zgoda, V. G., Moysa, A. a, Pyatnitskiy, M. A, Kuznetsova, K. G., Ilina, I. Y., Karpova, M. a, Karpov, D. S., Veselovsky, A. V, Ivanov, M. V, Gorshkov, M. V, Archakov, A. I., and Moshkovskii, S. A. (2015) Methionine to isothreonine conversion as a source of false discovery identifications of genetically encoded variants in proteogenomics. *J. Proteomics* 120, 169–78

67. Ivanov, M. V., Levitsky, L. I., Lobas, A. A., Panic, T., Laskay, Ü. A., Mitulovic, G., Schmid, R., Pridatchenko, M. L., Tsybin, Y. O., and Gorshkov, M. V. (2014) Empirical Multidimensional Space for Scoring Peptide Spectrum Matches in Shotgun Proteomics. *J. Proteome Res.* 13, 1911–1920

68. Nishikura, K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* 79, 321–49

69. Seeburg, P. H., Higuchi, M., and Sprengel, R. (1998) RNA editing of brain glutamate receptor channels: mechanism and physiology. *Brain Res. Rev.* 26, 217–229

70. Reenan, R. a, Hanrahan, C. J., and Ganetzky, B. (2000) The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na+ channel transcript in a region of RNA editing. *Neuron* 25, 139–149

71. Sollner, T., Whiteheart, S. W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., and Rothman, J. E. (1993) SNAP receptors implicated in vesicle targeting and fusion. *Nature* 362, 318–324

Table 1. Edited peptides found in Xing, Zhang et al., 2013, in Aradska et al., 2015 and in the own shotgun data by MaxQuant (MQ) and X!tandem (XT) against two databases: RADAR+TIER1 and RADAR+TIER1+TIER2. The top 2 peptides highlighted were selected for the MRM experiment. The top 7 highlighted peptides were found edited in all tree datasets, the top 14 were found in two out of tree datasets. The 24 palely highlighted peptides are the ones found in only one dataset by only one search engine.

Edited sequence	Xing, Zhang et al., 2013			Aradska et al., 2015			This work					
	MQ TIER1+ RADAR	MQ TIER1+ TIER2+ RADAR	XT TIER1+ RADAR	XT TIER1+ TIER2+ RADAR	XT TIER1+ RADAR	XT TIER1+ TIER2+ RADAR	MQ TIER1+ RADAR	MQ TIER1+ TIER2+ RADAR	XT TIER1+ RADAR	XT TIER1+ TIER2+ RADAR	aa substitution	gene name
IEYHVEHAVDYVQTATQDT**K**	√	√	√	√	√	√	√	√	√	√	M>V	Syx1A
NQ**M**ETQVNELK	√	√	√	√	√		√	√	√	√	I>M	cpx
MTVAHMWFDNQIIEADTTEDQSGVQ**C**DR	√	√	√		√	√	√	√	√	√	Y>C	Atpalpha
LVTTVSTPVFD**G**R	√	√			√	√	√	√	√	√	R>G	CG4587
ADSLVESGQFDTAGIQEK			√	√	√	√			√	√	Q>R	alpha-Spec
L**V**SVLESTLSK	√	√	√	√	√					√	M>V	Cadps
YSLDDNI**E**QNFLEPLHHMQTK			√	√	√				√	√	K>E	EndoA
GVGPAPSANASADSSS**R**	√	√	√	√			√	√	√	√	K>R	Atx2
GLTNSIAVVVATSTCP**R**	√								√	√	Q>R	hig

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

YLLQIWPQPQAQH**Q**R					√	√	√	√	√	√	Q>R	RhoGAP100F
LQQSSV**G**PGDPVR					√				√	√	S>G	brp
CTLNPYYNESFSFEVPFEQ**M**QK	√	√	√		√	√					I>M	Syt1
AIA**G**VLTPLNPLWALR			√	√	√	√					R>G	CG1882
AIHALAAGAS**S**SAEQDGAYNPWPHFR		√		√		√					N>S	CG13630
LAASQQVAVVHQHPSDQA**G**ESGHK							√	√	√	√	R>G	sif
NTCVFGGAP**R**									√	√	K>R	Rm62
DHCIAMVQC**R**	√	√	√	√							K>R	VhaSFD
FIYVSE**E**ELLAVAK	√	√	√	√							K>E	CG5862
IIHLLNDQHMGVVTAA**A**SLIDALVK	√	√	√	√							T>A	AP-2alpha
LAVHEIVD**A**NGAGDAFVGGFLSQFVQGK	√	√	√	√							T>A	AdenoK
LHAGDDEEADDPI**R**	√	√	√	√							Q>R	Calx
HDG**G**VIQVDDR	√	√		√							E>G	Unc-89
EDIQANQLVMGEFEVSDVPGQIIDY**V**AR	√	√									I>V	bai
EQLA**T**LMDTLR	√	√									K>T	zip
FVSTSQDQTAMLWQW**H**VGSNSVECVSVC	√	√									N>H	CG6724

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

K												
GAL**S**SAGYHLNNR	√	√									N>S	CalpB
LAIVEYMPALAGQLGQEFFDQK	√	√									I>V	Pp2A-29B
QTDGQWDLTSYHLVTLGD**R**	√	√									N>T	syd
SLLESSLH**R**			√	√							Q>R	Prm
LLGSS**G**VDSASDTSAK					√	√					S>G	eag
SSQDMNLNLNQNPSK					√	√					Y>S	pyd
G**G**FASNLDGHK					√	√					S>G	Mp
TVGETETFLT**G**NEITIK										√	S>G	CG10077
EETQAAAQT**R**PK		√									K>R	nwk
ESPKEKTPNIK		√									K>E	CG18178
EVV**C**VDASPCFEQVK		√									Y>C	CG15543
EV**V**EEIEEIVEEK		√									I>V	sls
QWIIFDGDV**G**PEWVENLNSVLDDNK		√									D>G	Dhc64C
RSSAFAAMQAFQPSR		√									K>R	Zasp52
SAQKQEPPKQNGNKPSK		√									N>S	pit
SQTLPALS**S**PEK		√									N>S	CG11148
V**C**LTQPEHNHPPK		√									Y>C	mod(mdg4)

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

VCPIAGDCLDPDLGISQSDR**R**		√									Q>R	CG17560
VLGDITESE**V**SSDLVER		√									I>V	Usp7
MALTSYDQIPSELAPYAFVEFTMPATQVSH**A**TVR	√										T>A	stnB
AAALSGMEMGSGPGSEGYEDADNE**R**						√					K>R	tkv
DR**S**GLCQLCAPNEPGVFIGK						√					N>S	Fatp
LPVNL**G**LSSILVPHGVDLDEPDVK						√					S>G	CG4587
SLMSTPNFPTQLITDAD**R**						√					K>R	cdm
MDITWFPFDDQHCEMK					√						I>M	nAChRalpha6
LPVNL**R**					√						R>G	CG4587
MVQALSLGGG					√						I>M	Frq
SQKG**P**NPPSQQVQPK					√						Q>P	CG31064
MSRPNDCLDILPEQCEQGDEEELMY**R**PFPDCCPVYCNLK					√						K>R	CG1552
MDPMIY**M**FPR					√						I>M	shakB
EVAQK**S**YAMKLASAGTKPGVSSAAAAYTGLTLNK					√						N>S	mbl

bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Table 2. Concentrations (C, nmol/g of total protein) of the selected peptides measured by MRM and their standard deviations (st.d.) in 4 samples (samples #2-#5 for MRM analysis, for all the samples see supplementary file 1 table 1) raised in different temperatures: sample #2 – preliminary analysis, raised in 25°C with the age undefined; samples #3-#5 raised for 72 hours in 10, 20 and 30°C. For each peptide one isotopically labeled standard peptide with 5 transitions was used. Each MRM experiment was repeated in 5 technical runs. CV of transition intensity did not exceed 30% in technical runs.

	Sample #2, 25°C, age undefined		Sample #3, 10°C, 72 h		Sample #4, 20°C, 72 h		Sample #5, 30°C, 72 h	
Peptide	C, nmol/g	st.d.	C, nmol/g	st.d.	C, nmol/g	st.d.	C, nmol/g	st.d.
Syx1A, unedited IEYHVEHAMDYVQTATQDTK	14.28	0.91	13.48	0.38	11.52	0.23	13.56	0.23
Syx1A, edited IEYHVEHAVDYVQTATQDTK	0.76	0.05	0.40	0.01	0.31	0.01	0.42	0.01
cpx, unedited NQIETQVNELK	12.43	0.21	7.50	0.15	10.43	0.19	10.68	0.35
cpx, edited NQMETQVNELK	4.73	0.30	3.46	0.11	4.75	0.22	5.34	0.29

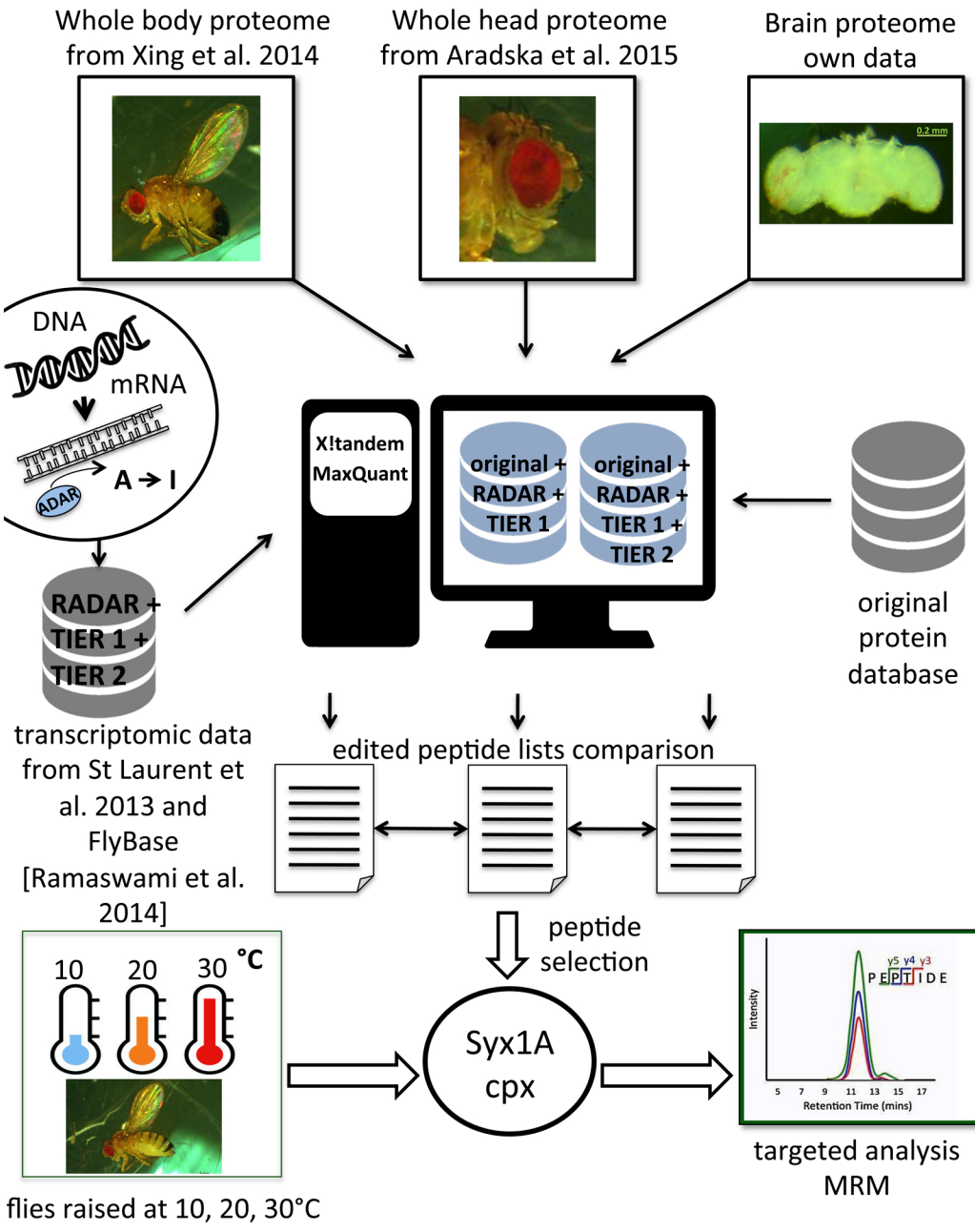
bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

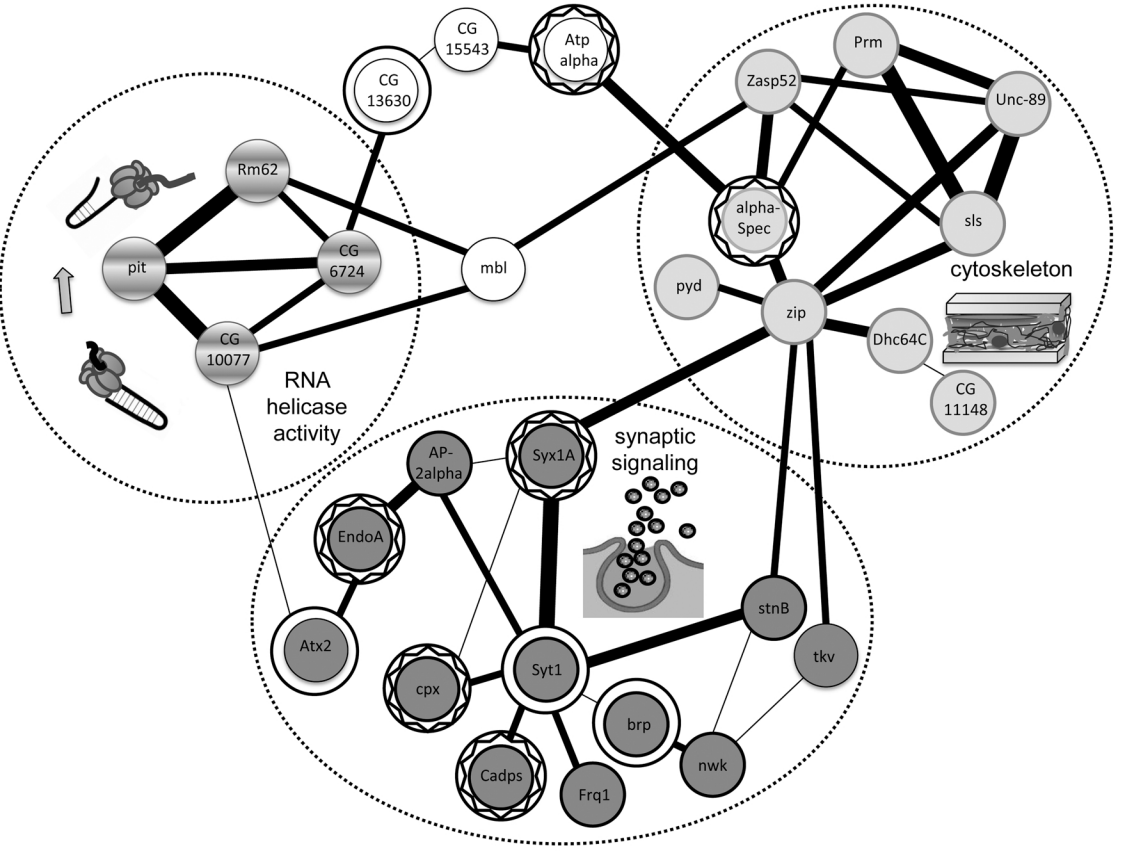
bioRxiv preprint doi: https://doi.org/10.1101/101949; this version posted January 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

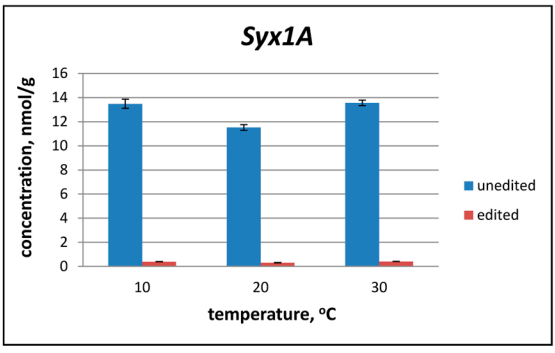
Figure 1. The workflow performed in the work. Three datasets obtained from proteomes of the whole body, the head and the brain of *Drosophila melanogaster* have been analyzed. Two of them had been taken elsewhere via ProteomeXchange, the third one was obtained experimentally. The data were processed with MaxQuant and X!tandem against two customized databases. Two peptides, *Syx1A* and *cpx*, were selected for MRM analysis and for this purpose samples of flies raised in three different temperatures had been made made.

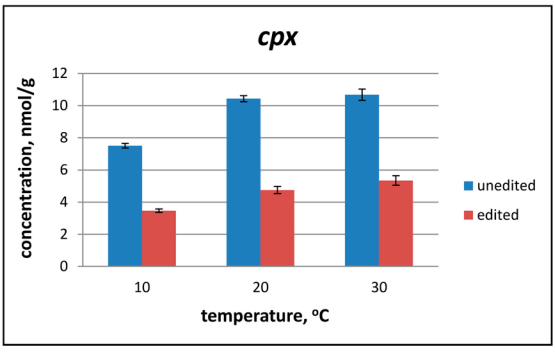
Figure 2. Diagrams showing the relationship of the concentrations (nmol/g of total protein) of edited and unedited forms of the selected peptides *Syx1A* and *cpx* in the samples, raised for 72 hours in 10, 20 and 30°C (samples #3-5).

Figure 3. Interacting proteins according to STRING among the proteins found to undergo RNA editing. The proteins with their gene names are united in the groups according to their functions. The highlighted by double circles proteins are the ones belonging to Group 1, i.e. identified in all 3 datasets. The asterix and circles mark the ones belonging to Group 2, i.e. identified in 2 out of 3 datasets.


Whole body proteome from Xing et al. 2014
Whole head proteome from Aradska et al. 2015
Brain proteome own data
0.2 mm
DNA
mRNA
A → I
ADAR
RADAR + TIER 1 + TIER 2
transcriptomic data from St Laurent et al. 2013 and FlyBase [Ramaswami et al. 2014]
X!tandem
MaxQuant
original + RADAR + TIER 1
original + RADAR + TIER 1 + TIER 2
original protein database
edited peptide lists comparison
peptide selection
Syx1A cpx
10
20
30
°C
flies raised at 10, 20, 30°C
y5 y4 y3
PEPTIDE
Intensity
5 7 9 11 13 15 17
Retention Time (mins)
targeted analysis MRM


CG 15543
Atp alpha
CG 13630
Prm
Zasp52
Unc-89
Rm62
alpha-Spec
sls
pit
CG 6724
mbl
cytoskeleton
pyd
zip
Dhc64C
CG 10077
RNA helicase activity
CG 11148
synaptic signaling
AP-2alpha
Syx1A
EndoA
stnB
Atx2
tkv
cpx
Syt1
brp
Cadps
nwk
Frq1


Syx1A
concentration, nmol/g
16
14
12
10
8
6
4
2
0
10
20
30
temperature, °C
unedited
edited


cpx
concentration, nmol/g
12
10
8
6
4
2
0
10
20
30
temperature, °C
unedited
edited