Adenosine-to-inosine RNA editing in mouse and human brain proteomes

Lev I. Levitsky[1], Anna A. Kliuchnikova[2,3], Ksenia G. Kuznetsova[2], Dmitry S. Karpov[2,4], Mark V. Ivanov[1], Mikhail A. Pyatnitskiy[2,5,6], Olga V. Kalinina[7,8], Mikhail V. Gorshkov[1,9], Sergei A. Moshkovskii[2,3]

[1] V.L. Talrose Institute for Energy Problems of Chemical Physics, N.N. Semenov Federal Research Center of Chemical Physics, Russian Academy of Sciences, Moscow, Russia

[2] Institute of Biomedical Chemistry, Moscow, Russia

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

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

[5] Onco Genotest LLC, Moscow, Russia

[6] National Research University Higher School of Economics, Moscow, Russia

[7] Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarbrücken, Germany

[8] Medical Faculty, Saarland University, Homburg, Germany

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

Received: 08/06/2019; Revised: 25/09/2019; Accepted: 30/09/2019

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/pmic.201900195.

Correspondence: Prof. Sergei A. Moshkovskii, Pirogov Russian National Research Medical University (RNRMU), 1 Ostrovityanova St., Moscow, 117997, Russia

E-mail: smosh@mail.ru

Abbreviations: ADAR, adenosine deaminase, RNA dependent; NSAF, normalized spectral abundance factor; PSM, peptide spectrum match; SNP, single nucleotide polymorphism.

Keywords: Adenosine deaminase / AMPA glutamate receptor/ Brain proteome / Proteogenomics / RNA editing.

Abstract

Proteogenomics is based on the use of customized genome or RNA sequencing databases for interrogation of shotgun proteomics data in search for proteome-level evidence of genome variations or RNA editing. In this work, we identified the products of adenosine-to-inosine RNA editing in human and murine brain proteomes using publicly available brain proteome LC-MS/MS datasets and an RNA editome database compiled from several sources. After filtering of false-positive results, 20 and 37 sites of editing in proteins belonging to 14 and 32 genes were identified for murine and human brain proteomes, respectively. Eight sites of editing identified with high spectral counts overlapped between human and mouse brain samples. Some of these sites were previously reported using orthogonal methods, such as NMDA glutamate receptors, CYFIP2, coatomer alpha, etc. Also, differential editing between neurons and microglia was demonstrated in this work for some of the proteins from primary murine brain cell cultures. Because many edited sites are still not characterized functionally at the protein level, our results provide a necessary background for their further analysis in normal and diseased cells and tissues using targeted proteomic approaches.

Statement of significance of the study

Today, shotgun proteomics is able to identify ten or more thousand proteins from a single specimen with increase in coverage of eukaryotic proteomes almost to maximum. This is true if we consider each protein as a sum of products expressed from a single gene. The next step, which is widely recognized by the community, is to distinguish different products of a gene, the proteoforms, and further elucidate their potential different roles. In this work, we have identified proteoforms originating from the post-transcriptional modification, adenosine-to-inosine RNA editing which occurs in each cell of mammalian organism and especially expressed in the brain tissue. To this end, we reanalyzed major murine and human brain proteomic datasets with a specific database which contained proteoforms predicted from RNA editome data obtained elsewhere by next-generation sequencing. Our findings, the murine and human edited proteins, provide a base for further targeted analyses for better understanding their functional significance in normal and diseased cells and tissues.

1 Introduction

Shotgun proteomics has experienced a rapid development and a swift progress in efficiency in recent years after introduction of high-resolution mass spectrometry, which became routinely available in the proteomic labs worldwide. The last generations of LC-MS/MS instruments and data processing software are capable of identifying and quantifying more than ten thousand protein groups in the course of one experimental run. These improvements allowed achieving a number of milestones within the Human Proteome Project and other global proteomic initiatives[1]. Among the recent achievements is, for example, the identification at the proteome level of most of the protein coding reading frames predicted before from the genome sequence [2].

The hunt for missing proteins primarily considered each of them as a sum of polypeptide products expressed from a single gene. Another challenging task of proteomic studies is distinguishing the individual products of each protein-coding gene, which are called proteoforms [3]. These are produced by events that can be divided into two major groups. The first and more complex one are post-translational modifications (PTMs) of genetically encoded amino acids. Proteins are subjected to dozens of common PTMs which can be studied using a number of approaches [4]. A second source of proteoforms are coding mutations, where amino acid sequences are changed as a result of variation of coding nucleic acid sequences, such as DNA polymorphism, alternative splicing, coding post-transcriptional modifications of RNA and, last but not least, errors in ribosomal translation [3]. In order to determine the changes in protein sequences resulting from nucleic acid variation, multi-omics data analysis approaches are employed, including Next Generation Sequencing and

shotgun proteomics data for a particular sample. The methodology of this multi-omics approach is known as proteogenomics [5].

Thus far, proteogenomic identification of protein sequence variants was mainly focused on cancer genome-associated mutations which could be functional and actionable during the tumor growth [6]. In this area of research, one important and growing field is identification of cancer neoantigens originating from mutated histocompatibility antigen-bound peptides. These peptides may be used for personalized anticancer vaccine development [7]. Some of the workflows aiming at these clinically important applications employ mass spectrometry for direct identification of mutant peptides extracted from the complexes with cancer cell-associated HLA molecules [8].

The focus of the present study was on the type of coding RNA modification known as RNA editing by double-strand RNA-dependent adenosine deaminases (ADARs, EC 3.5.4.37) [9]. This is the most common type of RNA editing, which is often neglected in proteogenomic studies. ADARs in mammals include two enzymatically active isoforms encoded by *ADAR* and *ADARB1* genes, ADAR1 and ADAR2, respectively. Both affect dsRNA sites by deaminating adenosine to inosine. The latter is more affine to cytosine than to uracil, thus re-coding some amino acids if deamination occurs in a codon[10]. ADAR1 is thought to inactivate an excess of dsRNA, which further stimulates cell immunity. Thus, this isoform participates in immune suppressing pathways which may facilitate cancer growth [11]. While ADAR1 reacts with long stretches of dsRNA, the ADAR2, in a yet unknown manner, specifically modifies selected residues in coding regions of mRNAs. Further, the importance of RNA editing that leads to protein sequence variation was shown for serotonin receptor 2C [12] and AMPA glutamate receptor subunits encoded by *GRIA2*,

GRIA3 and *GRIA4* human genes [13]. The editing of these subunits is essential for survival in both mouse and human. In mice, knockout of *ADARB1* gene caused early death from seizures due to excess of glutamate receptor-promoted excitation. This effect is reversed by a genomic change of the edited site [13].

From hundreds to thousands of coding editing sites were reported for human and mouse transcriptomes [14], as well as transcriptomes of invertebrates [15]. However, only several of them have been characterized functionally and detected at the protein level until recently [12,13].

Proteomic identification of RNA editing products requires employing the proteogenomic approaches by combining RNA editome and shotgun proteomic data. A pioneering work has described the RNA editome and the proteome of a cephalopod species called Californian octopus, which uses editing of RNA and further modifying the sequences of the corresponding proteins for adopting its poikilothermic body to environmental changes [16]. Also, RNA editing at the proteome level was characterized for the fruit fly brain, in which about 70 edited sites were identified [17]. Quantitative analysis of selected sites using targeted MS has shown that some of brain proteins were edited in up to 70% of adult insects.

As for the editing in mammalian proteome, the differences between DNA and RNA sequences were revealed by genome-wide analysis of human B-cells followed by the confirmation of protein-coding substitutions using shotgun proteomics [18]. Many of these substitutions were caused by the activity of ADAR enzymes. More recently, the proteins changed by RNA editing in human cancer tissues were identified by re-analysis of proteogenomic data from CPTAC cancer proteomics project, which is an addition to The Cancer Genome Atlas [19]. In the latter work,

thirteen edited sites were validated at the protein level. At least one of those sites, located in COPA vesicle transport protein, was under suspicion for being involved in breast cancer progression [19]. On the contrary, an edited site in the cyclin I protein was reported to induce neoepitope, which elicited cytotoxic T-cell response against tumor [20]. In the latter study, five edited antigenic peptides were only identified in human leukocyte antigen (HLA)-bound peptidome using a proteogenomic approach [20].

Enhanced activity of ADARs in the central neural system has been recognized since the discovery of this type of RNA editing [21] and motivated us to search for editing events mammalian brain proteome in this work. Here, deep proteome datasets obtained recently for mouse [22–24] and human [25–27] brains were selected from publicly available sources. These data were re-analyzed using a genomic database with added proteins bearing amino acid substitutions obtained from ADAR editing results [28,29] and from the analysis of editomes of both human and mouse reported recently [30]. It was also important to understand if the shotgun proteomics can reproduce the findings reached with rigorous studies of selected ADAR-edited translated sites, such as in glutamate receptors [31].

2 Methods

2.1 Composing a database for coding A-to-I editing events

Mouse genomic coordinates of RNA editing were obtained from RADAR [28] and DARNED [29] databases and from the supplementary data for the recent paper [30] where 11103 RNA editing sites in exons were identified in adult and embryonic murine tissues. Changes in protein sequences induced by RNA editing were

annotated against mm9 version of mouse genome using Variant Annotation Integrator (VAI; http://genome.ucsc.edu/cgi-bin/hgVai) [32].

Human genomic coordinates of RNA editing sites were obtained from the same three sources. In the large-scale study[30], 12871 sites in exons from 8,551 human adult and fetal tissue samples from the Genotype-Tissue Expression project were identified. Changes in protein sequences induced by RNA editing were annotated against hg19 version of human genome using VAI.

FASTA files for proteogenomic search finally contained 609 and 1,878 coding edited sites for mouse and human, respectively. In more detail, the contribution of each source into final databases is described in Supporting Information Table 1.

2.2 Benchmark proteomic data

Murine and human brain proteome datasets, three for each species, were selected from ProteomeXchange repository [33]. The datasets represent shotgun proteomes obtained using high-resolution MS/MS with data-dependent acquisition. Surprisingly, not many datasets were available, especially for human proteomes. Some journals in the field, even well-established, do not oblige contributors to provide raw proteome data. There were two major datasets available for murine and human brain, PXD001250 [22] and PXD005445 [25], respectively. These latter were specifically directed to the inventory of brain proteomes and, to our knowledge, represented the deepest brain proteomes available to date. We used these two datasets as core benchmark data, adding the less comprehensive data to validate the findings. Additional murine datasets were taken from models of Huntington disease [23] and neuropathic pain [24], under accession numbers PXD003442 and PXD004087, respectively. Additional human data were taken from cerebrospinal

fluid (CSF) study [26], PXD009646, and from proteogenomic brain study intended to identify protein variants [27], PXD004143. We chose the CSF study to verify if the edited proteins can be identified in this body fluid, which is normally more readily available in clinical conditions than the brain tissues. Numerical characteristics of proteomic data obtained by our re-processing are listed in Table 1.

2.3 Proteogenomic search

The databases for mouse and human searches were constructed in the following way. All tryptic peptides containing exactly two missed cleavage sites and exactly one modified RNA editing site were listed (35878 mouse peptides and 14234 human peptides) and annotated using original protein ID, as well as genomic and proteomic coordinates of the editing site. Decoy versions of these peptides were constructed by reversing the peptide sequences while keeping the C-terminal residue in place. Canonical protein sequences from Mouse UniProt (May 2014, 57260 entries) and Human SwissProt (October 2018, 20411 entries), respectively, were added along with their decoy counterparts obtained by reversing the protein sequence while keeping the N-terminal residue in place.

The data were searched against the constructed databases using X!Tandem Cyclone [34] with the following principal search parameters: 10 ppm and 0.01 Da for precursor and fragment ion mass errors; one allowed missed cleavage site; tryptic digestion with proline exception; carbamidomethylation of cysteine (fixed), oxidation of methionine (variable). Additionally, a variable modification of 229.162932 Da was applied to lysines and N-termini in PXD004143, which was labeled with TMT. Then, peptide-spectrum matches (PSMs) corresponding to peptides with RNA editing sites

were extracted, ranked by Hyperscore and filtered to 1% FDR using target-decoy approach (group-specific FDR) as implemented in Pyteomics library [35].

Search engine results were merged for biological and technical replicates, as well as for the fractions corresponding to the same type of tissue prior to filtering.

2.4 Heuristic rules for manual filtering of edited peptides

The edited peptides identified automatically were then manually inspected to filter out suspicious hits. The following items were excluded: (i) Asn-to-Asp substitutions, because editing events may be mimicked by chemical deamidation[36]; (ii) miscleaved peptides lacking their cleaved counterparts (for explanation, cf. Results section); (iii) for human samples, editing events corresponding to known genomic SNPs, except cancer mutations (NextProt[37] knowledgebase, release 2016-08-25, used for curation with embedded references to the dbSNP database[38]); (iv) chemical modifications mimicking substitutions which were found during manual inspection of mass-spectra.

2.5 Label-free quantitation of proteins of interest

For the identified peptides bearing RNA editing substitutions, the canonical protein forms were quantified based on normalized spectral abundance factor (NSAF) [39] by processing the X!Tandem output with Scavager [40] and using the scav2nsaf script to calculate the NSAF values over pooled fraction results. For comparison between different samples, all NSAF values were converted to percentiles.

3 Results and discussion

3.1 Filtering out dubious identifications of A-to-I edited sites

In proteogenomics, discoveries produced by the search engine must be carefully verified at the PSM level. For example, some amino acid substitutions are often mimicked by artefact or natural chemical modifications [5,41]. In case of A-to-I RNA editing and in contrast with the genomic variation, the number of possible amino acid substitutions is limited to A-to-G changes in codons. Despite this, all findings of edited peptides reported from brain proteomes by automatic output of the search engines were inspected to filter out all dubious peptide hits.

One of the most common substitutions predicted from A-to-I editome was the Asp-to-Asn change. Unfortunately, chemical deamidation is often observed to occur naturally and, thus, we had to exclude all such peptides from the results.

Manual inspection of the mass spectra of variant peptides was also performed for expert validation of findings as suggested recently [42]. As one of the striking results obtained for the murine data, an edited peptide VTIAQGGVLP**S**IQAVLLPK with one acetylation site from the common histone HIST2H2AC was reported with high spectral counts. The high extent of editing of a highly abundant histone protein could potentially be a significant discovery. However, manual inspection of the corresponding mass spectra evidently showed a wild-type sequence, VTIAQGGVLP**N**IQAVLLPK, with ammonia addition. Next, for human data, some A-to-I edited sites were also polymorphic in the genome with A-to-G substitutions, which led to the same protein variants. We excluded such peptides from the results if they contained germline SNPs, but not cancerous mutations from COSMIC catalog

[43]. The presence of cancer-related mutations in brain tissues used in the datasets can be neglected.

Peptides with edited sites which were identified with a trypsin miscleavage without identification of their fully cleaved counterparts were also excluded. Using the target-decoy method, we estimated FDR separately for the lonely miscleaved peptides and those miscleaved peptides which had their pair or even two partners. As a result, in one exemplary dataset, the FDR for the lonely miscleaved peptides was about 12%. In contrast, the pairs of overlapping miscleaved and cleaved products were almost absent among decoys, with an estimated FDR of less than 0.01% for those with a single cleaved counterpart and zero for those with two cleaved parts identified in the same dataset (Fig.1). Thus, miscleaved peptides without cleaved counterparts are enriched by false-positive results. Working at the peptide level, we chose to exclude this fraction from our report in this project. Expectedly, the paired overlapping peptides detected independently are very unlikely to occur as false positives.

The complete, unfiltered list of peptides returned by the search engines is provided in Supporting Information Tables 2 and 3.

3.2 A-to-I RNA editing in the murine brain proteome

After applying the exclusion criteria described above, 20 edited sites which belong to products of 14 genes were identified for murine brain data (Table 2). Of those, 15 sites were characterized by spectral counts above two. Among identified edited peptides, some have been previously reported and validated by orthogonal molecular methods[13]. Expectedly, the higher the levels of editing (expressed quantitatively as spectral counts), the better studied are the corresponding proteins.

Table 2 also shows that profiles of identified edited peptides from the lesser datasets [23,24] are generally similar to the largest dataset [22]. In the former sets, the most represented edited sites were also identified.

Glutamate receptor subunits GRIA2 (GluA2) and GRIA3 (GluA3) are classical examples of coding editing events [44]. For them, both *flip* and *flop* splice isoforms with R/G editing were identified (Table 2). It was recently shown that this type of editing accelerated opening and desensitization of GRIA2-formed ion channels, at least for *flop* isoforms [45]. The better recognized Q/R editing site on GRIA2, which was shown to be completely edited in adult brains [9], was also identified with lesser spectral counts than the R/G site, most likely due to the chemical properties of the corresponding peptide. This latter contains a glutamine residue at the N-terminus, which often leads to unpredictable chemical changes during fragmentation [46]. RNA editing in glutamate receptor subunits was thoroughly studied on different molecular levels [47]. For proteins, the main instrument was to create recombinant, forcibly edited receptors in cell and organism models. With targeted proteomics based on our findings, one can monitor dynamics of the edited isoform in intact tissues, including clinical samples [45].

The cytoplasmic fragile X mental retardation 1 interacting protein 2 (CYFIP2) is the most extensively edited protein found in murine datasets with a major edited K/E site at the 320th position. CYFIP2 is an important component of actin regulatory WAVE complex [48], which reorganizes actin filaments in response to external stimuli with lamellipodia formation. Functionally, the complex was shown to participate in spinal morphogenesis and presynaptic modulation [49]. Germline mutations of CYFIP2 are responsible for early-onset epileptic encephalopathy [50] and binge eating

[51]. Despite the accumulating knowledge about the roles of this protein in various pathways, the exact function of its K/E editing remains unclear.

The alpha-filamin (FLNA) is a large protein with multiple immunoglobulin-like repeats which provides the spatial organisation of actin filaments, as well as binds multiple ligands, including cytoplasmic domains of transmembrane receptors [52]. Its edited site, Q/R, located in 22th filamin repeat and identified here, was described before [53]. The functional meaning of this substitution remains unclear. It was hypothesised that editing can modify binding of FLNA with its partner protein [53].

The coatomer subunit alpha (COPA) is a component of specific coating for vesicles (COPI) that provide a retrograde transport of proteins from the Golgi complex to the endoplasmic reticulum [54]. This subunit of the complex is a well-known target for RNA editing. Interestingly, the search performed earlier for large-scale cancer proteomic data [19] found the same I/V substitution which was identified herein. Despite the conservative nature of the substitution, it was shown that, at least in the breast cancer cell line, this editing event facilitates the cancer growth through an unknown mechanism [19].

The calcium-dependent secretion activator 1 (CADPS) is shown to participate in calcium-dependent release of synaptic vesicles [55]. This was the only protein from our list whose ortholog was also edited in the fruit fly proteome [17]. The set of edited proteins in mammals and fruit fly is substantially different, with components of presynaptic release mechanisms mostly edited in the insect. However, in murine and insect brains, CADPS was edited in different, non-homologous sites.

Among the previously described targets of ADAR editing, the insulin-like growth factor binding protein 7 (IGFBP7) should be mentioned with its R/G site identified in

corpora striati dataset [53]. The meaning of edited residues in this protein also remains unclear.

A number of editing events identified in this work were never discussed in literature and, at the same time, they have lower spectral counts than the established edited sites. Two ion channels should be mentioned, the metabotropic glutamate receptor 4 (GRM4) and the voltage-dependent T-type calcium channel subunit alpha-1G (CACNA1G). Notably, the latter is functionally related to a fruit fly protein, CG4587, which was shown to be extensively edited in previous work [17]. Unexpectedly, among the edited proteins of murine brain, two enzymes with mitochondrial localisation were found, specifically, ACOT10 and ME3.

From omics experiments without use of functional models, one can only state that the coding events took place. The high sensitivity of molecular techniques may result in identification of sites that have no functional significance and arise from side effects of ADAR enzymatic reactions. Levanon et al [53] also suggested that the coding editing events could be a side-effect of alternative splicing regulation by ADAR, e.g. in the filamin alpha, as it was shown before for the glutamate receptor[56].

3.3 Coding RNA editing events are differentially distributed between murine brain cells

Having deep proteomes of different cells and tissues of murine brain [22], we performed label-free quantitation analysis for the identified edited peptides using spectral counting. For the analysis, we selected the samples where the number of PSMs for edited peptides of each protein of interest was at least 5, according to the heuristic statistical rule, which, for example, is used in Fisher's exact test (for PSMs by samples [22], see Supporting Information Table 2). These were primary cell

cultures of microglia from young and adult mice, cerebellar granule cells and cortical neurons. Expectedly, the cell cultures provided better sensitivity towards edited sites because RNA editing is known mostly for cellular proteins. In tissue samples containing many stromal extracellular components, edited sites were identified with low PSM counts (Supporting Information Table 2). Of all tissues characterized by Sharma et al [22], the whole brain proteome analyzed by fractions provided higher PSM counts for edited peptides.

Using five datasets with higher abundance of edited sites, we aimed to correlate the extent of editing with the protein concentration. In other words, if we have a high PSM count for edited peptides of the glutamate receptor in cortical neurons and a low count in microglia, does this mean that this receptor is highly expressed in the neurons? To answer this question, we recalculated label-free quantitative parameters for the proteins in five datasets of interest, namely, NSAF percentiles [39], and plotted them against PSM counting for the proteins that reached five or more PSMs of edited peptides in at least one of the datasets of interest (Fig.2).

In some cases, the simple logic of 'no protein - no editing' appears to work. For example, microglia almost lacked GRIA2 and GRIA3 glutamate receptors and, thus, no edited sites were identified in this type of cells. For other proteins, a differential editing was apparently observed at comparable expression levels. For example, GRIA2 editing was, probably, specific for cortical neurons with no editing in cerebellar granule neurons (Fig.2), although the expression level of this receptor was significantly higher in cortical neurons in comparison to other cell cultures (p-value of 0.037 calculated here and below with the pairwise Wilcoxon test with FDR as multiple testing correction). GRIA3 had low levels of editing in both types of neurons, despite high levels for the whole brain, which, more likely, means that it is

edited in other neuronal type not covered specifically by the dataset [22]. At the same time, its levels estimated by NSAF were significantly lower only for cortical neurons, with no difference between cerebellar cells and whole brain ($p = 0.037$, which is a minimal possible value generated by the statistical method when three replicates are compared pairwise).

CYFIP2 appeared to be differentially edited in neuronal cells with low levels of editing in microglia, despite relatively high abundance in all cell types. More specifically, its levels are significantly higher in cortical neurons in comparison with microglia, but not in cerebellar granular cells, where it is highly edited . Hypothetically, this may reflect the difference between these cell types in response of actin structures to external stimuli. In contrast, the filamin-A (FLNA) and the coatomer subunit alpha (COPA) are edited mostly in microglia. The former's level is significantly higher in microglia in comparison with cortical neurons, but not with cerebellar cells. Coatomer alpha protein has a high level expressed elsewhere, except the whole brain dataset. Remarkably, COPA's editing was characteristic for microglia from newborn mice and, at the same time, this type of editing was shown to promote cancer growth *in vitro* [19]. Probably, growing cells have some specificity in turnover of Golgi transport proteins which may be facilitated by specific I/V editing of the coatomer alpha as identified here and in other works.

Our results strongly suggest that ADAR targets in different brain cell types are differentially edited. Most likely, this differential pattern is caused by ADAR2, which is characterized by target specificity, not identified yet [57]. Differential RNA editing patterns on the protein level should be further studied, e.g. by targeted proteomics, to elucidate their possible role in disease. For example, microglia, which is obviously

a subject of specific RNA editing, is shown to trigger neuronal death in amyotrophic lateral sclerosis.

Proteomic findings of differential editing in different tissues are generally supported by the recent data of single-cell RNAseq analysis where similar events were also observed[58]. The REDIportal database, which may be used to check A-to-I RNA editing levels deduced from RNAseq data for various human tissues, does not contain data on glial cells yet[14]. Thus, we could not directly compare the proteomic findings with mRNA data. Notably, in a comparison of editing levels of sites of interest between human brain and blood vessel tissues, significant difference can be found, e.g. for glial-specific edited proteins. The major non-synonymous edited site of FLNA (Q/R) was characterized by average editing frequency of 0.36 and 0.88 in brain tissues and blood vessels, respectively. Similarly, the major non-synonymous site of COPA (I/V) has 0.19 and 0.56 in brain and blood vessel tissues ($p<0.01$ in Wilcoxon test for both sites). However, neuron-specific edited sites also tend to be more frequently edited in arteria, such as 0.77 in brain vs.0.94 in arteria for GRIA3[14]. Moreover, blood vessels are not so closely relate to glial cells to draw any conclusions without further experimentation.

While we could identify a number of “classical” examples of RNA-edited proteins, such as glutamate receptors, some others were not detected. The most recognized case is a serotonin receptor, more specifically, 5-hydroxytryptamine receptor 2C (HTR2C) [59]. Unfortunately, one of the main edited sites for this receptor, I156V, is located in a short tryptic peptide which is too lightweight to be detected in typical proteome analysis. Other sites, however, could be identified. At the same time, the HTR2C protein itself was not identified in cell line datasets discussed above, as well as in fractions of whole brain. Most likely, shotgun

proteomic analysis, such as the one performed to obtain the datasets of interest, did not provide enough sensitivity to identify this receptor, which was specific for selected neuron types and brain regions, e.g. amygdala [60].

3.4 A-to-I RNA editing in the human brain proteome

In order to identify edited sites in the human brain proteome, we used the same study design as for murine data. Similarly, there was a main dataset that yielded the majority of identified edited peptides [25]. Even the authors of the original paper by Carlyle et al stated that their work was a human analogue of the major murine study by Sharma et al [22]. However, the data was collected from more subjects and, at the same time, lacked the depth of proteome coverage reached in the murine brain study [22]. Moreover, there were no cell cultures analyzed, which, in case of the murine brain proteome, yielded better coverage for edited products (see Section 3.3). While analysing human data, one should also take into account genome polymorphism which cannot be observed in inbred animals and, in turn, may sometimes overlap with the editing events. As mentioned above, the human editome contained more than 1,800 coding events, three times the size of the murine one. It is still unclear if this fact reflects biological mechanisms or is due to the human editome being better studied.

In total, 37 edited sites which belong to products of 32 genes were identified from human brain data, with 18 sites being characterized by the spectral count above two (Table 3). This tier of sites with higher number of spectra is highly correlated with murine data. The findings common for two species included a major site of CYFIP2, three sites from glutamate receptors GRIA2 and GRIA3, sites of COPA, FLNA and CADPS, as described below (eight interspecies sites are shadowed in

Tables 2 and 3). For GRIA2 receptor subunit, a R/G edited flop isoform was also identified, in contrast to murine data.

In this part of the study, we deliberately involved a dataset of cerebrospinal fluid (CSF) proteome [26]. This fluid is the most available specimen from the human brain which can, inter alia, be withdrawn without complicated operative intervention. Importantly, the edited sites detectable in CSF may potentially serve as biomarkers. Expectedly, we could not find classical sites, such as from glutamate receptors, in this fluid. Instead, a Q/R site from the prostaglandin-H2 D-isomerase (PTGDS) was found. PTGDS is a multifunctional enzyme and neuromodulator, and is also known as CSF biomarker for liquorrhea, a CSF leakage from nose and ear after trauma [61]. In addition, the IGFBP7 editing site shared with the murine proteome of corpora striati [23], as mentioned above, was found in CSF.

One more dataset [27], obtained in the proteogenomic study of prefrontal cortex, was not deep enough and yielded a single minor edited site from the polymerase delta-interacting protein 3 (Table 3). Even the deepest human dataset obtained by Carlyle et al [25] did not provide the sensitivity towards edited sites similar to that of the murine dataset obtained by Sharma et al [22]. Thus, with the human data we could not confirm or refute the finding of differential RNA editing for different cells similar to the one shown for murine brain in Section 3.3.

The question arises which sites of identified here are novel. The novelty is relative because all of them are taken from RNA data reported before. For human, the most comprehensive list of edited sites identified by proteomics was reported for cancers[19]. Of them, COPA and IGFBP7 sites shared between murine and human editomes were also found in this work (Tables 2-3). Sites of GRIA2-4 glutamate

receptor subunits were found to be important at the protein level with transgene technology. Other sites were predicted to exist only from RNA analysis, which means that technically they are novel at the proteome level.

3.5 A-to-I RNA editing sites identified in proteomes of murine and human brains and their editing frequencies measured by RNA sequencing

After identification of edited sites in murine and human brain proteomes, it was relevant to know the extent of their editing on the transcript levels. Publicly available resources provide the editing frequency number which is usually estimated by ratio between quantities of edited and non-edited sequences using RNAseq reads or output of targeted PCR methods[14]. It would be preferable to calculate corresponding ratios for edited sites found in shotgun proteomes. However, it is not technically possible in the scope of the present study due to potentially different behavior of genome-encoded and edited tryptic peptides in the mass spectrometer. Accordingly, the parameters of label-free quantitation between genomic and edited are supposed to be incomparable. The problem may be solved using targeted mass-spectrometric methods with isotopically labeled peptide standards, as shown before for RNA editing in the fruit fly[17]. In this work, which based on analysis of shotgun proteomic big data, the only measure of editing abundance remains spectral counting for edited peptides as listed in the Tables 2-3.

For human brain, editing frequencies for most identified sites were available from REDIportal RNA editing database[14] accompanied with the numbers of RNAseq runs where corresponding edited sites were detected as a measure of analytical reproducibility. Murine editing sites were characterized in the major study

by Tan et al[30] (Supplementary File 5). In that paper, intact mice were compared to ADAR2 knock-outs. As a result, the contribution of each active ADAR isoform, ADAR1 and ADAR2, to editing of each site of interest could be additionally estimated. RNA editing parameters for the sites reliably identified in the major murine[22] and human[25] brain proteomes and in the brain transcriptomes as described above are summarized in Table 4.

Despite the use of spectral counts of edited peptides instead of real editing frequencies, there is a connection between transcriptomic and proteomic data listed in Table 4. Sites with higher spectral counts are also characterized by higher editing frequencies. For murine sites characterized by both proteomic and transcriptomic parameters, Pearson and Spearman correlation coefficients between them are 0.66 and 0.61, respectively. For those human sites which were also detected at the transcript level in more than 20 RNAseq runs, these correlation coefficients are 0.6 and 0.8, respectively. Although it should be further verified, the results suggest that the major transcripts with nonsynonymous RNA editing have more chances to be translated.

The majority of edited sites found in brain proteomes originate from ADAR2 isoform activity, in correspondence with previous art[62]. Interestingly, glutamate receptor transcripts also are subject of ADAR1 action, especially, the GRIA3 site with 60% of contribution of this isoform.

3.6 Potential influence of identified A-to-I RNA editing sites on the protein structure

We have computationally characterized the reliably identified edited sites (cf. top sections of Table 2 and 3) with respect to their potential effect on protein three-dimensional structure, function, and interactions. First, we predicted their damaging effect using SIFT[63]. This method predominantly relies on conservation of the wildtype and mutant amino acids in related proteins. SIFT predicts 9 amino acid substitutions of 23 in the joined list of edited sites from murine and human brain as possibly damaging (Supporting Information Table 4). Then we further characterized structural environment of the affected positions with StructMAn[64]. This tool assesses whether a particular amino acid residue lies in protein core, on its surface, or on an interaction interface with another protein, nucleic acid, or ligand in a complex of the protein in question or of its homolog. We could obtain structural annotation for 11 of 18 substitutions in the human samples and for 10 of 16 in the mouse samples. For human, N227S in GLUD2 is located in a NADPH-binding pocket in a crystal structure of an orthologous bovine protein (PDB[65] ID 3ETE; Fig.3A). T313A in CASP1 lies on a protein-protein interaction interface (PDB ID 3D6H), and homologous residues R764G in GRIA2, R775G in GRIA3, and R765G in GRIA4 are located very close to a protein-protein interaction interface in a structure of a homolog GluA2 (PDB ID 6DM0). In both cases these are interaction interfaces between subunits in a homooligomeric complex. These substitutions in glutamate receptors may modulate subunit interactions despite SIFT predicting this mutation as tolerable. I164V in COPA lies in the protein core and can potentially disrupt the protein stability (Fig.3B). We additionally tested this with FoldX[66], but the small estimated change on the free energy between the wildtype and the mutant structures

(1.33 kcal/mol) argues against this hypothesis. At the same time, a possible functional importance of this site in cancer was mentioned above[19]. Additionally, from the mouse samples, we find that I75V in ME3 lies on a protein-protein interaction interface in a homologous tetramer of human ME1 (PDB ID 3WJA). Q118R in the glutamine receptor GRM4 (in the structure of a human homolog GRM5, PDB ID 6N4X) and R769G ID GRIA3, in accordance with human data, lie close to such an interface. Q607R in GRIA2 lies in the pocket that binds channel blockers and can influence the sensitivity of the protein to their toxicity. This finding corresponds with background knowledge that this editing site in GRIA2 substantially decreases the calcium flux through the ion channel with important functional consequences[67].

4 Concluding remarks

In this work, we revealed for the first time the consequences of adenosine-to-inosine RNA editing at the scale of brain proteomes obtained previously for mouse and human. While this type of RNA editing is known as a ubiquitous pathway which takes place literally in each cell of mammalian organism[14,68], only tens of thousands of theoretical coding sites could be identified at the proteome level. A whole collection of cancer proteomes accumulated by CPTAC consortium has recently yielded a mere thirteen validated ADAR edited sites [19]. We were able to find more in the selected data, with thirty seven potential edited sites in the human brain and twenty of them in the murine brain, in the deepest proteomes publicly available to date. This gain is supposed to be explained by ADAR2 enzyme activity characteristic for neural tissues[67].

The question arises, why the number of edited sites found in proteomes is so low in comparison to those observed at the transcript level. We think that it is not due to the low detection limit for minor protein isoforms although the method limitations may contribute. When translation of multiple mRNA transcripts originating from alternative splicing was studied in deep proteome data, it was shown that, for most genes, one principal isoform survived as a protein[69,70]. We may expect the same situation with edited mRNAs, many of which may originate as a side-effect of ADAR1 isoform activity towards potentially toxic double-strand RNA sites[71]. These unnecessary edited transcripts may be translated with a lower yield than wild-type or principally important transcripts. It was also shown that inosine residue in the codon can slow down the ribosome followed by formation of truncated polypeptides[10]. Thus, among the diversity of edited transcripts, only few of them are likely to lead to protein translation and these latter can be essential, such as edited glutamate receptors[72].

Interspecies comparison of results showed a good correlation between human and murine data, with eight common editing sites among the hits with higher abundance. In general, many edited sites identified from shotgun proteomes with higher spectral counts are also identified by orthogonal techniques, among them being the edited glutamate receptor subunits, the CYFIP2 protein, the coatomer alpha subunit and the filamin-A. At the same time, many of them still are not functionally characterized, except RNA editing of AMPA glutamate receptor subunits. Thus, we provided here a list of targets for further quantitative analysis of edited proteins by targeted proteomic methods for the cells in normal and disease states, such as amyotrophic lateral sclerosis, neural tumors, seizures, etc.

The study was supported by the Russian Scientific Foundation, grant #17-15-01229 to S.M. We are grateful to Sebastian Keller (Max Planck Institute for Informatics, Saarbrücken, Germany) and Alexander Gress (HIPS, Saarbrücken, Germany) for the help with the StructMAn tools. We also thank Shamsudin Nasaev (IBMC, Moscow, Russia) for his technical assistance.

The authors have declared no conflict of interest.

5 References

[1] G. S. Omenn, L. Lane, C. M. Overall, F. J. Corrales, J. M. Schwenk, Y.-K. Paik, J. E. Van Eyk, S. Liu, M. Snyder, M. S. Baker, E. W. Deutsch, *J. Proteome Res.* **2018**, *17*, 4031.

[2] M. S. Baker, S. B. Ahn, A. Mohamedali, M. T. Islam, D. Cantor, P. D. Verhaert, S. Fanayan, S. Sharma, E. C. Nice, M. Connor, S. Ranganathan, *Nat. Commun.* **2017**, *8*, 14271.

[3] A. Lisitsa, S. Moshkovskii, A. Chernobrovkin, E. Ponomarenko, A. Archakov, *Expert Rev. Proteomics* **2014**, *11*, 121.

[4] Y. Zhao, O. N. Jensen, *Proteomics* **2009**, *9*, 4632.

[5] A. I. Nesvizhskii, *Nat. Methods* **2014**, *11*, 1114.

[6] M. A. Karpova, D. S. Karpov, M. V Ivanov, M. A. Pyatnitskiy, A. L. Chernobrovkin, A. A. Lobas, A. V Lisitsa, A. I. Archakov, M. V Gorshkov, S. A. Moshkovskii, *J. Proteome Res.* **2014**, *13*, 5551.

[7] P. A. Ott, Z. Hu, D. B. Keskin, S. A. Shukla, J. Sun, D. J. Bozym, W. Zhang, A. Luoma, A. Giobbie-Hurder, L. Peter, C. Chen, O. Olive, T. A. Carter, S. Li, D. J. Lieb, T. Eisenhaure, E. Gjini, J. Stevens, W. J. Lane, I. Javeri, K. Nellaiappan, A. M. Salazar, H. Daley, M. Seaman, E. I. Buchbinder, C. H. Yoon, M. Harden, N. Lennon, S. Gabriel, S. J. Rodig, D. H. Barouch, J. C. Aster, G. Getz, K. Wucherpfennig, D. Neuberg, J. Ritz, E. S. Lander, E. F. Fritsch, N. Hacohen, C. J. Wu, *Nature* **2017**, *547*, 217.

[8] M. Bassani-Sternberg, E. Bräunlein, R. Klar, T. Engleitner, P. Sinitcyn, S.

Audehm, M. Straub, J. Weber, J. Slotta-Huspenina, K. Specht, M. E. Martignoni, A. Werner, R. Hein, D. H Busch, C. Peschel, R. Rad, J. Cox, M. Mann, A. M. Krackhardt, *Nat. Commun.* **2016**, *7*, 13404.

[9] B. L. Bass, *Annu. Rev. Biochem.* **2002**, *71*, 817.

[10] K. Licht, M. Hartl, F. Amman, D. Anrather, M. P. Janisiw, M. F. Jantsch, *Nucleic Acids Res.* **2019**, *47*, 3.

[11] J. J. Ishizuka, R. T. Manguso, C. K. Cheruiyot, K. Bi, A. Panda, A. Iracheta-Vellve, B. C. Miller, P. P. Du, K. B. Yates, J. Dubrot, I. Buchumenski, D. E. Comstock, F. D. Brown, A. Ayer, I. C. Kohnle, H. W. Pope, M. D. Zimmer, D. R. Sen, S. K. Lane-Reticker, E. J. Robitschek, G. K. Griffin, N. B. Collins, A. H. Long, J. G. Doench, D. Kozono, E. Y. Levanon, W. N. Haining, *Nature* **2019**, *565*, 43.

[12] K. Iwamoto, M. Bundo, T. Kato, *RNA Biol.* **2009**, *6*, 248.

[13] M. Higuchi, S. Maas, F. N. Single, J. Hartner, A. Rozov, N. Burnashev, D. Feldmeyer, R. Sprengel, P. H. Seeburg, *Nature* **2000**, *406*, 78.

[14] E. Picardi, A. M. D'Erchia, C. Lo Giudice, G. Pesole, *Nucleic Acids Res.* **2017**, *45*, D750.

[15] A. L. Yablonovitch, P. Deng, D. Jacobson, J. B. Li, *PLoS Genet.* **2017**, *13*, e1007064.

[16] N. Liscovitch-Brauer, S. Alon, H. T. Porath, B. Elstein, R. Unger, T. Ziv, A. Admon, E. Y. Levanon, J. J. C. Rosenthal, E. Eisenberg, *Cell* **2017**, *169*, 191.

[17] K. G. Kuznetsova, A. A. Kliuchnikova, I. U. Ilina, A. L. Chernobrovkin, S. E.

Novikova, T. E. Farafonova, D. S. Karpov, M. V Ivanov, A. O. Goncharov, E. V Ilgisonis, O. E. Voronko, S. S. Nasaev, V. G. Zgoda, R. A. Zubarev, M. V Gorshkov, S. A. Moshkovskii, *J. Proteome Res.* **2018**, *17*, 3889.

[18] M. Li, I. X. Wang, Y. Li, A. Bruzel, A. L. Richards, J. M. Toung, V. G. Cheung, *Science* **2011**, *333*, 53.

[19] X. Peng, X. Xu, Y. Wang, D. H. Hawke, S. Yu, L. Han, Z. Zhou, K. Mojumdar, K. J. Jeong, M. Labrie, Y. H. Tsang, M. Zhang, Y. Lu, P. Hwu, K. L. Scott, H. Liang, G. B. Mills, *Cancer Cell* **2018**, *33*, 817.

[20] M. Zhang, J. Fritsche, J. Roszik, L. J. Williams, X. Peng, Y. Chiu, C.-C. Tsou, F. Hoffgaard, V. Goldfinger, O. Schoor, A. Talukder, M. A. Forget, C. Haymaker, C. Bernatchez, L. Han, Y.-H. Tsang, K. Kong, X. Xu, K. L. Scott, H. Singh-Jasuja, G. Lizee, H. Liang, T. Weinschenk, G. B. Mills, P. Hwu, *Nat. Commun.* **2018**, *9*, 3919.

[21] S. Tomaselli, F. Locatelli, A. Gallo, *Cell Tissue Res.* **2014**, *356*, 527.

[22] K. Sharma, S. Schmitt, C. G. Bergner, S. Tyanova, N. Kannaiyan, N. Manrique-Hoyos, K. Kongi, L. Cantuti, U.-K. Hanisch, M.-A. Philips, M. J. Rossner, M. Mann, M. Simons, *Nat. Neurosci.* **2015**, *18*, 1819.

[23] P. Langfelder, J. P. Cantle, D. Chatzopoulou, N. Wang, F. Gao, I. Al-Ramahi, X. H. Lu, E. M. Ramos, K. El-Zein, Y. Zhao, S. Deverasetty, A. Tebbe, C. Schaab, D. J. Lavery, D. Howland, S. Kwak, J. Botas, J. S. Aaronson, J. Rosinski, G. Coppola, S. Horvath, X. W. Yang, *Nat. Neurosci.* **2016**, *19*, 623.

[24] C. Altmann, S. Hardt, C. Fischer, J. Heidler, H.-Y. Lim, A. Häussler, B. Albuquerque, B. Zimmer, C. Möser, C. Behrends, F. Koentgen, I. Wittig, M. H.

H. Schmidt, A. M. Clement, T. Deller, I. Tegeder, *Neurobiol. Dis.* **2016**, *96*, 294.

[25] B. C. Carlyle, R. R. Kitchen, J. E. Kanyo, E. Z. Voss, M. Pletikos, A. M. M. Sousa, T. K. T. Lam, M. B. Gerstein, N. Sestan, A. C. Nairn, *Nat. Neurosci.* **2017**, *20*, 1787.

[26] C. Macron, L. Lane, A. Núnez Galindo, L. Dayon, *J. Proteome Res.* **2018**, *17*, 4113.

[27] T. S. Wingo, D. M. Duong, M. Zhou, E. B. Dammer, H. Wu, D. J. Cutler, J. J. Lah, A. I. Levey, N. T. Seyfried, *J. Proteome Res.* **2017**, *16*, 3336.

[28] G. Ramaswami, J. B. Li, *Nucleic Acids Res.* **2014**, *42*, 109.

[29] A. M. Kiran, J. J. O'Mahony, K. Sanjeev, P. V Baranov, *Nucleic Acids Res.* **2013**, *41*, D258.

[30] M. H. Tan, Q. Li, R. Shanmugam, R. Piskol, J. Kohler, A. N. Young, K. I. Liu, R. Zhang, G. Ramaswami, K. Ariyoshi, A. Gupte, L. P. Keegan, C. X. George, A. Ramu, N. Huang, E. A. Pollina, D. S. Leeman, A. Rustighi, Y. P. S. Goh, F. Aguet, K. G. Ardlie, B. B. Cummings, E. T. Gelfand, G. Getz, K. Hadley, R. E. Handsaker, K. H. Huang, S. Kashin, K. J. Karczewski, M. Lek, X. Li, D. G. MacArthur, J. L. Nedzel, D. T. Nguyen, M. S. Noble, A. V. Segrè, C. A. Trowbridge, T. Tukiainen, N. S. Abell, B. Balliu, R. Barshir, O. Basha, A. Battle, G. K. Bogu, A. Brown, C. D. Brown, S. E. Castel, L. S. Chen, C. Chiang, D. F. Conrad, N. J. Cox, F. N. Damani, J. R. Davis, O. Delaneau, E. T. Dermitzakis, B. E. Engelhardt, E. Eskin, P. G. Ferreira, L. Frésard, E. R. Gamazon, D. Garrido-Martín, A. D. H. Gewirtz, G. Gliner, M. J. Gloudemans, R. Guigo, I. M.

Hall, B. Han, Y. He, F. Hormozdiari, C. Howald, H. Kyung Im, B. Jo, E. Yong Kang, Y. Kim, S. Kim-Hellmuth, T. Lappalainen, G. Li, X. Li, B. Liu, S. Mangul, M. I. McCarthy, I. C. McDowell, P. Mohammadi, J. Monlong, S. B. Montgomery, M. Muñoz-Aguirre, A. W. Ndungu, D. L. Nicolae, A. B. Nobel, M. Oliva, H. Ongen, J. J. Palowitch, N. Panousis, P. Papasaikas, Y. Park, P. Parsana, A. J. Payne, C. B. Peterson, J. Quan, F. Reverter, C. Sabatti, A. Saha, M. Sammeth, A. J. Scott, A. A. Shabalin, R. Sodaei, M. Stephens, B. E. Stranger, B. J. Strober, J. H. Sul, E. K. Tsang, S. Urbut, M. van de Bunt, G. Wang, X. Wen, F. A. Wright, H. S. Xi, E. Yeger-Lotem, Z. Zappala, J. B. Zaugg, Y.-H. Zhou, J. M. Akey, D. Bates, J. Chan, L. S. Chen, M. Claussnitzer, K. Demanelis, M. Diegel, J. A. Doherty, A. P. Feinberg, M. S. Fernando, J. Halow, K. D. Hansen, E. Haugen, P. F. Hickey, L. Hou, F. Jasmine, R. Jian, L. Jiang, A. Johnson, R. Kaul, M. Kellis, M. G. Kibriya, K. Lee, J. B. Li, Q. Li, X. Li, J. Lin, S. Lin, S. Linder, C. Linke, Y. Liu, M. T. Maurano, B. Molinie, S. B. Montgomery, J. Nelson, F. J. Neri, M. Oliva, Y. Park, B. L. Pierce, N. J. Rinaldi, L. F. Rizzardi, R. Sandstrom, A. Skol, K. S. Smith, M. P. Snyder, J. Stamatoyannopoulos, B. E. Stranger, H. Tang, E. K. Tsang, L. Wang, M. Wang, N. Van Wittenberghe, F. Wu, R. Zhang, C. R. Nierras, P. A. Branton, L. J. Carithers, P. Guan, H. M. Moore, A. Rao, J. B. Vaught, S. E. Gould, N. C. Lockart, C. Martin, J. P. Struewing, S. Volpi, A. M. Addington, S. E. Koester, A. R. Little, L. E. Brigham, R. Hasz, M. Hunter, C. Johns, M. Johnson, G. Kopen, W. F. Leinweber, J. T. Lonsdale, A. McDonald, B. Mestichelli, K. Myer, B. Roe, M. Salvatore, S. Shad, J. A. Thomas, G. Walters, M. Washington, J. Wheeler, J. Bridge, B. A. Foster, B. M. Gillard, E. Karasik, R. Kumar, M. Miklos, M. T. Moser, S. D. Jewell, R. G. Montroy, D. C. Rohrer, D. R. Valley, D. A. Davis, D. C. Mash, A. H. Undale, A. M. Smith, D. E. Tabor, N. V. Roche, J. A. McLean, N.

Vatanian, K. L. Robinson, L. Sobin, M. E. Barcus, K. M. Valentino, L. Qi, S. Hunter, P. Hariharan, S. Singh, K. S. Um, T. Matose, M. M. Tomaszewski, L. K. Barker, M. Mosavel, L. A. Siminoff, H. M. Traino, P. Flicek, T. Juettemann, M. Ruffier, D. Sheppard, K. Taylor, S. J. Trevanion, D. R. Zerbino, B. Craft, M. Goldman, M. Haeussler, W. J. Kent, C. M. Lee, B. Paten, K. R. Rosenbloom, J. Vivian, J. Zhu, A. Chawla, G. Del Sal, G. Peltz, A. Brunet, D. F. Conrad, C. E. Samuel, M. A. O'Connell, C. R. Walkley, K. Nishikura, J. B. Li, *Nature* **2017**, *550*, 249.

[31] M. Higuchi, S. Maas, F. N. Single, J. Hartner, A. Rozov, N. Burnashev, D. Feldmeyer, R. Sprengel, P. H. Seeburg, *Nature* **2000**, *406*, 78.

[32] A. S. Hinrichs, B. J. Raney, M. L. Speir, B. Rhead, J. Casper, D. Karolchik, R. M. Kuhn, K. R. Rosenbloom, A. S. Zweig, D. Haussler, W. J. Kent, *Bioinformatics* **2016**, *32*, 1430.

[33] J. A. Vizcaíno, E. W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J. A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P.-A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R. J. Chalkley, H.-J. Kraus, J. P. Albar, S. Martinez-Bartolomé, R. Apweiler, G. S. Omenn, L. Martens, A. R. Jones, H. Hermjakob, *Nat. Biotechnol.* **2014**, *32*, 223.

[34] R. Craig, R. C. Beavis, *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2310.

[35] L. I. Levitsky, J. A. Klein, M. V Ivanov, M. V Gorshkov, *J. Proteome Res.* **2019**, *18*, 709.

[36] H. Yang, R. A. Zubarev, *Electrophoresis* **2010**, *31*, 1764.

[37] P. Gaudet, P.-A. Michel, M. Zahn-Zabal, A. Britan, I. Cusin, M. Domagalski, P.

D. Duek, A. Gateau, A. Gleizes, V. Hinard, V. Rech de Laval, J. Lin, F. Nikitin, M. Schaeffer, D. Teixeira, L. Lane, A. Bairoch, *Nucleic Acids Res.* **2017**, *45*, D177.

[38] T. Tatusova, *Methods Mol. Biol.* **2016**, *1415*, 3.

[39] B. L. Zybailov, L. Florens, M. P. Washburn, *Mol. Biosyst.* **2007**, *3*, 354.

[40] M. V. Ivanov, L. I. Levitsky, J. A. Bubis, M. V. Gorshkov, *Proteomics* **2019**, *19*, 1800280.

[41] A. L. Chernobrovkin, A. T. Kopylov, V. G. Zgoda, A. A. Moysa, M. A. Pyatnitskiy, K. G. Kuznetsova, I. Y. Ilina, M. A. Karpova, D. S. Karpov, A. V Veselovsky, M. V Ivanov, M. V Gorshkov, A. I. Archakov, S. A. Moshkovskii, *J. Proteomics* **2015**, *120*, 169.

[42] A. A. Lobas, D. S. Karpov, A. T. Kopylov, E. M. Solovyeva, M. V. Ivanov, I. Y. Ilina, V. N. Lazarev, K. G. Kuznetsova, E. V. Ilgisonis, V. G. Zgoda, M. V. Gorshkov, S. A. Moshkovskii, *Proteomics* **2016**, *16*, 1980.

[43] S. A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, C. Y. Kok, M. Jia, T. De, J. W. Teague, M. R. Stratton, U. McDermott, P. J. Campbell, *Nucleic Acids Res.* **2015**, *43*, D805.

[44] H. Lomeli, J. Mosbacher, T. Melcher, T. Höger, J. R. Geiger, T. Kuner, H. Monyer, M. Higuchi, A. Bach, P. H. Seeburg, *Science* **1994**, *266*, 1709.

[45] W. Wen, C.-Y. Lin, L. Niu, *Sci. Rep.* **2017**, *7*, 13654.

[46] B. Godugu, P. Neta, Y. Simón-Manso, S. E. Stein, *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1169.

[47] A. Filippini, D. Bonini, L. La Via, A. Barbon, *Mol. Neurobiol.* **2017**, *54*, 6795.

[48] Z. Chen, D. Borek, S. B. Padrick, T. S. Gomez, Z. Metlagel, A. M. Ismail, J. Umetani, D. D. Billadeau, Z. Otwinowski, M. K. Rosen, *Nature* **2010**, *468*, 533.

[49] K. Han, H. Chen, V. A. Gennarino, R. Richman, H.-C. Lu, H. Y. Zoghbi, *Hum. Mol. Genet.* **2015**, *24*, 1813.

[50] M. Nakashima, M. Kato, K. Aoto, M. Shiina, H. Belal, S. Mukaida, S. Kumada, A. Sato, A. Zerem, T. Lerman-Sagie, D. Lev, H. Y. Leong, Y. Tsurusaki, T. Mizuguchi, S. Miyatake, N. Miyake, K. Ogata, H. Saitsu, N. Matsumoto, *Ann. Neurol.* **2018**, *83*, 794.

[51] S. L. Kirkpatrick, L. R. Goldberg, N. Yazdani, R. K. Babbs, J. Wu, E. R. Reed, D. F. Jenkins, A. F. Bolgioni, K. I. Landaverde, K. P. Luttik, K. S. Mitchell, V. Kumar, W. E. Johnson, M. K. Mulligan, P. Cottone, C. D. Bryant, *Biol. Psychiatry* **2017**, *81*, 757.

[52] D. T. Loo, S. B. Kanner, A. Aruffo, *J. Biol. Chem.* **1998**, *273*, 23304.

[53] E. Y. Levanon, *Nucleic Acids Res.* **2005**, *33*, 1162.

[54] L. Orci, M. Stamnes, M. Ravazzola, M. Amherdt, A. Perrelet, T. H. Söllner, J. E. Rothman, *Cell* **1997**, *90*, 335.

[55] R. T. O'Neil, X. Wang, M. V. Morabito, R. B. Emeson, *Mol. Brain* **2017**, *10*, 11.

[56] E. Bratt, M. Öhman, *RNA* **2003**, *9*, 309.

[57] J. M. Eggington, T. Greene, B. L. Bass, *Nat. Commun.* **2011**, *2*, 319.

[58] E. Picardi, D. S. Horner, G. Pesole, *RNA* **2017**, *23*, 860.

[59] C. M. Burns, H. Chu, S. M. Rueter, L. K. Hutchinson, H. Canton, E. Sanders-Bush, R. B. Emeson, *Nature* **1997**, *387*, 303.

[60] E. A. Hackler, D. C. Airey, C. C. Shannon, M. S. Sodhi, E. Sanders-Bush, *Neurosci. Res.* **2006**, *55*, 96.

[61] G. Bachmann, H. Petereit, *Acta Neurol. Scand.* **2004**, *110*, 337.

[62] M. Higuchi, F. N. Single, M. Köhler, B. Sommer, R. Sprengel, P. H. Seeburg, *Cell* **1993**, *75*, 1361.

[63] R. Vaser, S. Adusumalli, S. N. Leng, M. Sikic, P. C. Ng, *Nat. Protoc.* **2016**, *11*, 1.

[64] A. Gress, V. Ramensky, J. Büch, A. Keller, O. V. Kalinina, *Nucleic Acids Res.* **2016**, *44*, W463.

[65] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235.

[66] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, *Nucleic Acids Res.* **2005**, *33*, W382.

[67] A. Gallo, D. Vukic, D. Michalík, M. A. O'Connell, L. P. Keegan, *Hum. Genet.* **2017**, *136*, 1265.

[68] L. M. Bowe, C. W. DePamphilis, *Mol. Biol. Evol.* **1996**, *13*, 1159.

[69] M. L. Tress, F. Abascal, A. Valencia, *Trends Biochem. Sci.* **2017**, *42*, 98.

[70] B. C. Carlyle, R. R. Kitchen, J. Zhang, R. S. Wilson, T. T. Lam, J. S. Rozowsky, K. R. Williams, N. Sestan, M. B. Gerstein, A. C. Nairn, *J. Proteome Res.* **2018**, *17*, 3431.

[71] K. Pestal, C. C. Funk, J. M. Snyder, N. D. Price, P. M. Treuting, D. B. Stetson, *Immunity* **2015**, *43*, 933.

[72] A. Barbon, S. Barlati, *Biochem.* **2011**, *76*, 882.

Table 1. Parameters of datasets used for edited peptide identification

ProteomeXchange ID	Species	Biomaterial	# of PSMs, total	# of PSMs, 1% FDR
PXD001250 [22]	Mouse	Various brain tissues and primary cell cultures of healthy inbred mice	15,949,908	8,633,512
PXD003442 [23]		Striata from knock-in mice with increasing CAG repeat length (Huntington disease model)	9,625,663	4,641,359
PXD004087 [24]		Prefrontal cortex of aged mice with progranulin deficiency or neuronal progranulin overexpression	1,099,253	701,336
PXD005445 [25]	Human	Autopsies of various brain tissues of 16 individuals without obvious pathology	6,623,202	3,969,816
PXD009646 [26]		Cerebrospinal fluid sample from a pool of 96 healthy donors	558,627	60,764
PXD004143 [27]		Dorsolateral prefrontal cortex from two subjects	64,876	27,521

Table 2. ADAR-edited coding sites identified in murine brain proteomes [22–24]. The sites with two or more spectral counts are listed in the top of the table, and those with a single spectrum detected are in the bottom. The protein sites that are also edited in the human brain are shadowed.

Protein name	Peptide sequence	Uniprot ID	Amino acid substitution	Gene name	Total spectral count		
					Sharma et al [22]	Langenfelder et al [23]	Altmann et al [24]
Acyl-coenzyme A thioesterase 10, mitochondrial	VHSEVFSLGSREHMTTNVFHFTFMSEK	Q32MW3	D381G	ACOT10	14	4	
Calcium-dependent secretion activator 1 (Calcium-dependent activator protein for secretion 1) (CAPS-1)	VNGEMYIER, *dk*VNGEMYER[a)]	Q80TJ1	E1252G	CADPS	17	2	7
Calcium-dependent secretion activator 1 (Calcium-dependent activator protein for secretion 1) (CAPS-1)	YVDVPEPGMDVADAYVTFVR	Q80TJ1	K1226E	CADPS	8		
Coatomer subunit alpha	VWDVSGLR	F8WHL2	I164V	COPA	27		
Cytoplasmic FMR1-interacting protein 2	YIETSAHYEENK, YIETSAHYEENK*sk*	Q5SQX6	K320E	CYFIP2	88	72	35
Cytoplasmic FMR1-interacting protein 2	IQFPIEMGMPWILTDHILETK*epsmmeyvlypldlyndsayyaltk*, *r*IQFPIEMGMPWILTDHILETK	Q5SQX6	S337G	CYFIP2	2		
Filamin-A	LTVSSLR, *r*LTVSSLR	B7FAU9	Q2333R	FLNA	15	1	

Glutamate receptor 2	QGCDISPR	E9QKC0	Q607R	GRIA2	13	10	
Glutamate receptor 3, flip isoform	GSALGTPVNLAVLK	Q9Z2W9	R769G	GRIA3	26	39	20
Glutamate receptor 3, flop isoform	GSALGNAVNLAVLK	F2Z488	R769G	GRIA3	9	22	10
Glutamate receptor 2; Glutamate receptor 4, flip isoforms	GSSLGTPVNLAVLK	E9QKC0; Q9Z2W8	R765G, R764G	GRIA2; GRIA4	41	22	18
Insulin-like growth factor-binding protein 7	GEGEPCGGGAAGGGHCAPGMECVK	E9Q5D9	R78G	IGFBP7		3	
Metabotropic glutamate receptor 4 (mGluR4)	SLTFVQALIEK	G3XA00	Q118R	GRM4	8		
NADP-dependent malic enzyme, mitochondrial	LSNHVFVFQGAGEAAMGVAHLLVMALEK	Q8BMF3	I75V	ME3	7	2	
Voltage-dependent T-type calcium channel subunit alpha	VSRHIR[b)]	Q5SUF9	K2078R	CACNA1G	3		
Calcium-activated potassium channel subunit alpha-1	EEVVAAEVGWMTGVK	J3QMT8	S36G	KCNMA1	1		
Cytoplasmic FMR1-interacting protein 2	VMGFGLYLMDGNVSNIYR	Q5SQX6	K283R	CYFIP2	1		
Protein Son	AAELSVVSASVISEQSEQPMPGMLEPSMTK	H9KV15	T216A	SON	1		

SH3 domain-binding protein 2	HPCGYAGPR	Q06649	Y597C	SH3BP2	1
Voltage-dependent T-type calcium channel subunit alpha	DPQGTR	Q5SUF9	E2099G	CACNA1G	1

[a] In miscleaved peptides, additions to edited peptides are shown by italic lowercase.

[b] The miscleaved peptide is reported because its cleaved counterparts cannot be seen due to the search conditions.

Table 3. ADAR-edited coding sites identified in human brain proteomes [25–27]. The sites with two or more spectral counts are listed in the top of the table, and those with a single spectrum detected are in the bottom. All but two sites shown are found in the deep brain proteome by Carlyle et al [25]. The values in the right column are summarized peptide spectral counts of each edited site in all samples analyzed there, unless otherwise specified. The protein sites that are also edited in the murine brain are shadowed.

Protein name	Peptide sequence	Uniprot ID	Amino acid substitution	Gene name	Total spectral count
Ankyrin repeat and death domain-containing protein 1A	SQLTAASASR	Q495B1-5	T507A	ANKDD1A	5
ATPase family AAA domain-containing protein 2B	GLSVTSEQINPR	Q9ULI0	H1092R	ATAD2B	2
Calcium-dependent secretion activator 1	AGGGRPSGPSPSVVSEK	Q9ULU8	S91G	CADPS	5
Calcium-dependent secretion activator 1	VNGEMYIER, *dk*VNGEMYIER[a)]	Q9ULU8	E1241G	CADPS	3
Caspase-1	DFIAFCSSAPDNVSWR	P29466	T313A	CASP1	2
Coatomer subunit	VWDVSGLR	P53621	I164V	COPA	3

alpha					
Cytoplasmic FMR1-interacting protein 2	YIETSAHYEENK	Q96F07	K124E	CYFIP2	73
Filamin-A	LTVSSLR, *r*LTVSSLR	P21333	Q2341R	FLNA	10
Glutamate dehydrogenase 2, mitochondrial	GFIGPGVDVPAPDMSTGER	P49448	N227S	GLUD2	2
Glutamate receptor 3 (GluR-3), flop isoform	GSALGNAVNLAVLK	P42263	R775G	GRIA3	26
Glutamate receptor 3 (GluR-3), flip isoform	GSALGTPVNLAVLK	P42263	R775G	GRIA3	10
Glutamate receptor 2 (GluR2), Glutamate receptor 4 (GluR4), flop isoforms	GSSLGNAVNLAVLK	P48058; P42262	R765G (Gria4); R764G (Gria2)	GRIA4; GRIA2	13
Glutamate receptor 2 (GluR2), Glutamate receptor	GSSLGTPVNLAVLK	P48058; P42262	R765G (Gria4); R764G (Gria2)	GRIA4; GRIA2	43

4 (GluR4), flip isoforms					
Heat shock 70 kDa protein 1-like	QTQIFTTYSDSQPGVLIQVYEGER	P34931	N436S	HSPA1L	4
Insulin-like growth factor-binding protein 7 (IGFBP-7)[b)]	GEGEPCGGGGAGGGYCAPGMECVK	Q16270	R78G	IGFBP7	4
Nesprin-1	DIQQTEQTIEQR	Q8NF91	K4121R	SYNE1	4
Prostaglandin-H2 D-isomerase[c)]	AAPEAQVSVQPNFQQDK	P41222	Q21R	PTGDS	3
Zinc finger protein 587B	TSLGNMVK	E7ETH6	I396M	ZNF587B	4
40S ribosomal protein S18	VITIMQSPR	P62269	N73S	RPS18	1
ADP-ribosylation factor-like protein 1	IGISESELVAMLEEEELR	P40616	K104E	ARL1	1
Ankyrin repeat and death domain-containing protein 1A	SQLAATSASR	Q495B1-5	T505A	ANKDD1A	1

Coiled-coil domain-containing protein 40	LAGILNR	Q4G0X9	S336G	CCDC40	1
Cyclin-I	QLLHCVACNQLLQFR	Q14094	M189V	CCNI	1
Dapper homolog 3	QRPPDASPSPGGARPAR	Q96B18	S298G	DACT3	1
Filamin-B	VNQPAGFAIR	O75369	S2283G	FLNB	1
Glutamate receptor 3 (GluR-3), flip isoform	LSEQGILDKLENK	P42263	K795E	GRIA3	1
Histone demethylase UTY	SPVTVIR	O14607	N957S	UTY	1
Polyadenylate-binding protein 3	APPSGYFMTAVPQTQNR	Q9H361	H413R	PABPC3	1
Polymerase delta-interacting protein 3[d)]	ALANMSR	Q9BY77	T268A	POLDIP3	1
Probable global transcription activator SNF2L1	EEMLQVIR	P28370	M656V	SMARCA1	1
Propionyl-CoA carboxylase alpha chain,	LLIEEFIDNPR	P05165	K236E	PCCA	1

mitochondrial					
Propionyl-CoA carboxylase alpha chain, mitochondrial	LSSQEAASGFGDDR	P05165	S252G	PCCA	1
Protein numb homolog (h-Numb)	VTAATEQAER	P49757	T199A	NUMB	1
Sorbin and SH3 domain-containing protein 1	LECSGTVIAHCSLR	Q9BX66	K457R	SORBS1	1
Sorting nexin-14	IPVFCIDVGR	Q9Y5W7	E545G	SNX14	1
Zinc finger protein 432	TLESNLSLVNQNK	O94892	K138E	ZNF432	1
Zinc finger protein 714	SGVQDRPGQHGK	Q96N38	Q530R	ZNF714	1

[a] In miscleaved peptides, additions to edited peptides are shown by italic lowercase.

[b] In human, this site was detected only in the dataset by Macron et al [26] (cerebrospinal fluid). At the same time, it was also found in the murine dataset by Langfelder et al [23] (*corpora striati*).

[c] The same site was detected in the dataset by Macron et al [26] (cerebrospinal fluid) with the spectral count of 2.

[d] The site was detected only in the dataset by Wingo et al [27].

Table 4. RNA editing in proteome and transcriptome of murine and human brain. Spectral counts for edited sites reliably identified in deep mouse [22] and human [25] proteomes are supplemented by RNA editing frequencies of the same sites originated from RNA sequencing data. For murine brain, RNA data are derived from the major RNA editing study [30] (Supplementary File 5). The information from ADAR2 knockout mice is added which illustrates a ration between editing provided by ADAR1 and ADAR2 enzymes. For human brain, editing frequencies are taken from REDIportal database [14] as averages from all brain RNAseq runs listed for each site. The protein sites that are edited in both murine and human brains are shadowed.

Protein name	**Amino acid substitution**	**Gene name**	**Spectral count in murine brain** [22]	**Editing frequency in murine brain (% editing provided by ADAR2)** [30]	**Spectral count in human brain** [25]	**Editing frequency in human brain (the number of RNAseq runs in database)** [14]
Acyl-coenzyme A thioesterase 10, mitochondrial	*m:*D381G	ACOT10	14	N/A		
Ankyrin repeat and death domain-containing protein 1A	*h*:T507A	ANKDD1A			5	0.09 (218)
ATPase family AAA domain-containing protein 2B	*h*:H1092R	ATAD2B			2	N/A
Calcium-dependent secretion activator 1	*h*:S91G h	CADPS			5	0.08 (9)
Calcium-dependent	*h*:E1241G/	CADP	17	0.27 (68	3	0.32 (302

secretion activator 1	*m:*E1252G	S		%)		)
Calcium-dependent secretion activator 1	*m:*K1226E	CADPS	8	N/A		
Caspase-1	*h:*T313A	CASP1			2	0.14 (1)
Coatomer subunit alpha	*h/m*:I164V	COPA	27	0.06 (98 %)	3	0.19 (307)
Cytoplasmic FMR1-interacting protein 2	*h:*K124E / *m:*K320E	CYFIP2	88	0.79 (91%)	73	0.5 (275)
Cytoplasmic FMR1-interacting protein 2	*m:*S337G	CYFIP2	2	N/A		
Filamin-A	*h:*Q2341R / *m:*Q2333R	FLNA	15	0.33 (99 %)	10	0.36 (213)
Glutamate dehydrogenase 2, mitochondrial	*h:*N227S	GLUD2			2	0.24 (6)
Glutamate receptor 2 (GluR2)	*m:*Q607R	GRIA2	10	N/A		
Glutamate receptor 2 (GluR2), Glutamate receptor 4 (GluR4), flip and flop isoforms [b)]	*h/m*:R764G (Gria2); *h/m*:R765G (Gria4)	GRIA2, GRIA4	41	0.85 (74 %)	56	0.49 (586)
Glutamate receptor 3 (GluR-3), flip and flop isoforms [a)]	*h:*R775G / *m:*R769G	GRIA3	35	0.91 (39 %)	36	0.77 (232)
Heat shock 70 kDa protein 1-like	*h:*N436S	HSPA1L			4	N/A
Metabotropic glutamate receptor 4 (mGluR4)	*m:*Q118R	GRM4	8	0.07 (92 %)		
NADP-dependent malic enzyme, mitochondrial	*m:*I75V	ME3	7	N/A		
Nesprin-1	*h:*K4121R	SYNE1			4	N/A
Prostaglandin-H2 D-isomerase	*h:*Q21R	PTGDS			3	0 (135)
Voltage-dependent T-type calcium channel	*m:*K2078R	CACN	3	N/A		

subunit alpha		A1G		
Zinc finger protein 587B	*h*:I396M	ZNF587B	4	0.64 (20)

[a)] Flip and flop edited isoforms of GRIA3 are distinguished by shotgun proteomics, but not by RNAseq. For this analysis, spectral counts for both proteomic sites were summarized.

[b)] Edited sites of GRIA2 and GRIA4 are distinguished by RNAseq, but not by proteomics. For this analysis, spectral counts for corresponding proteomic sites were summarized, and RNAseq-derived editing ratios were averaged.

Figure legends

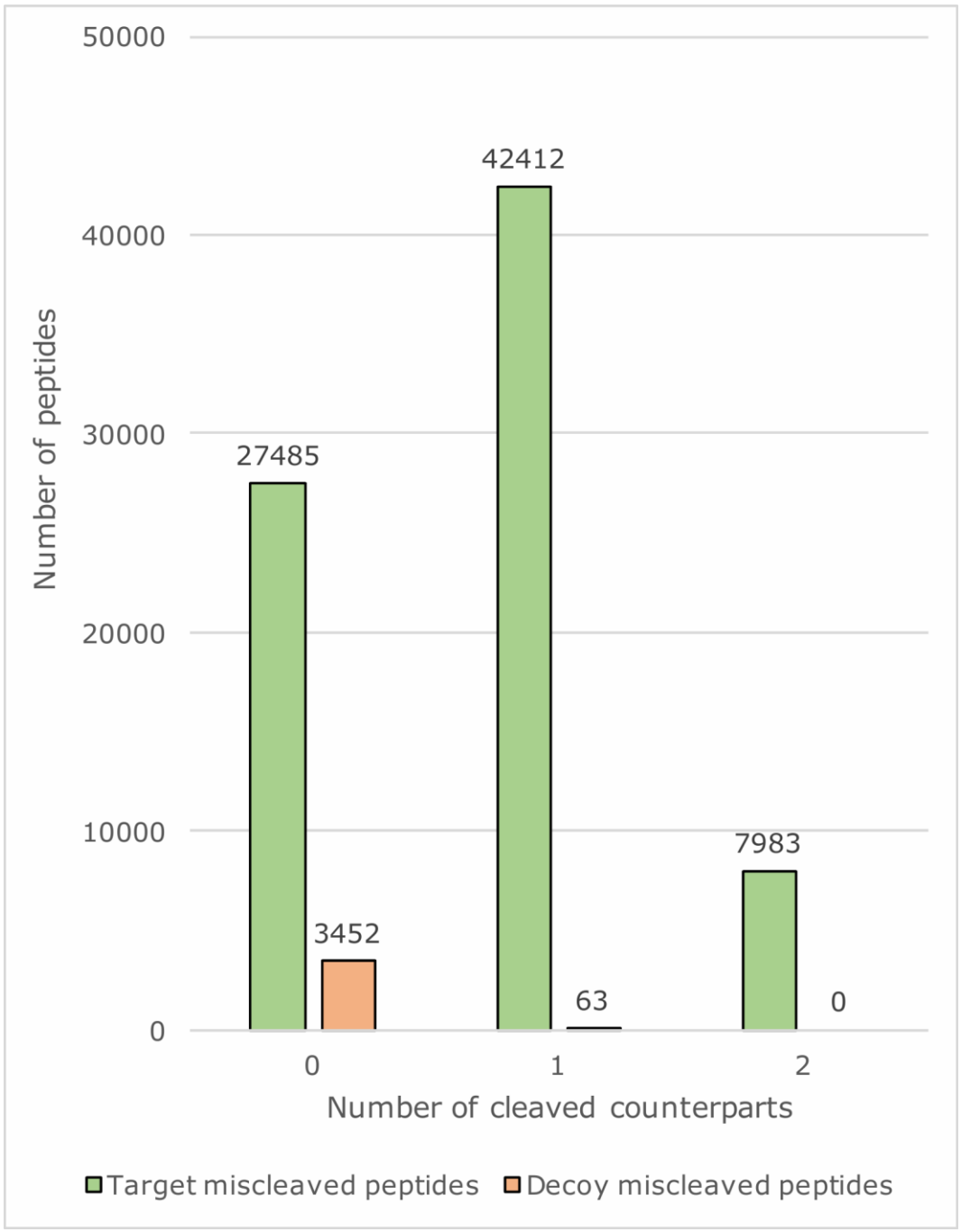

Fig. 1. Counts of target and decoy peptides containing a miscleavage identified in dataset PXD003442 [23] **with 1% PSM-level FDR.** Overall peptide-level FDR was estimated at 3.5%; FDR among peptides with miscleavages was 4.5%. As can be seen from the figure, these false positives are distributed unevenly depending on the existence of identified fully cleaved counterparts. The estimated FDR values are 12.6%, 0.15% and 0% for peptides with no identified fully cleaved parts, and with one and two such peptides identified, respectively.

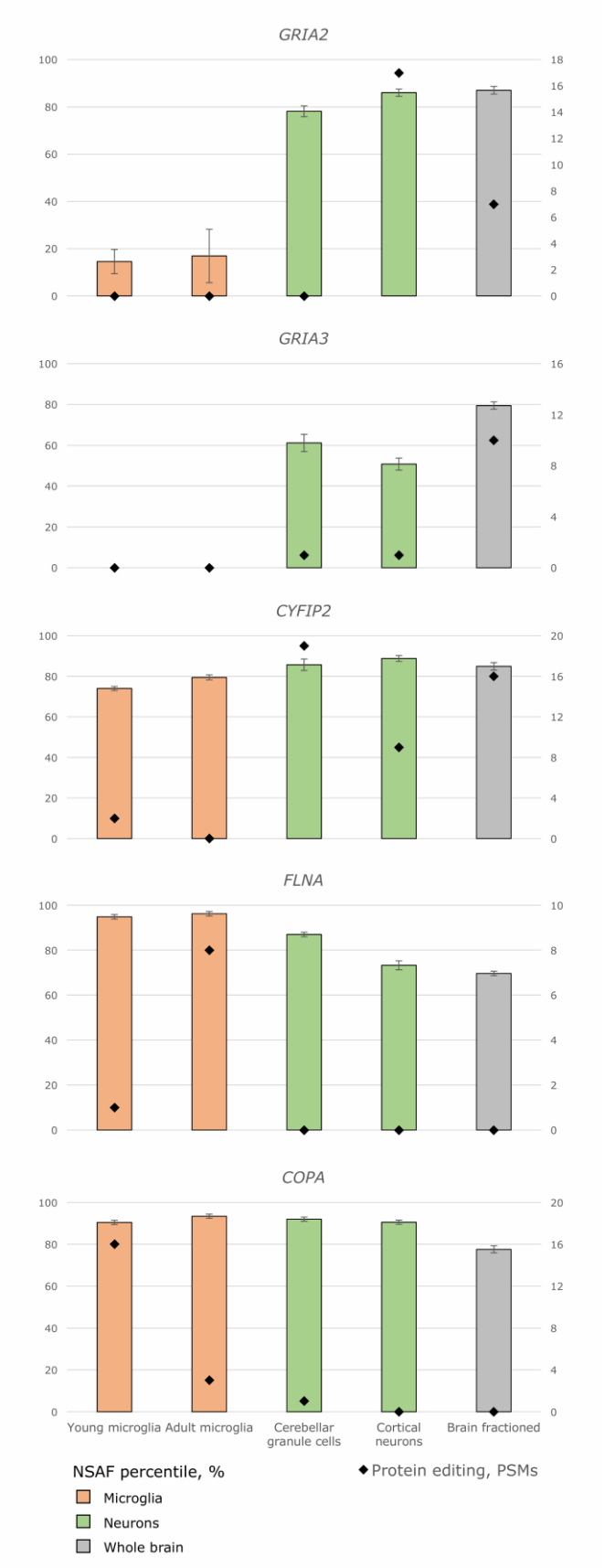

Fig. 2. Editing extent vs. abundance for selected proteins in murine brain subproteomes of young and adult microglia (YM and AM), cerebellar granule (CG) and cortical neurons (CN), all primary cultures, and the whole brain fractionated (BF). The editing extent is expressed in PSMs of edited peptides summarized for each protein, if it had more than one editing site. The abundance is expressed as NSAF percentiles [39] with standard error bars shown for three biological repeats. Datasets from Sharma et al [22] were selected if at least one of edited sites identified had more than five PSMs. GRIA2 glutamate receptor has a shared edited site with GRIA4, however, the editing is reported primarily for the former protein. Significance of differential NSAF percentiles was estimated by pairwise between-tissue comparisons for all proteins using Wilcoxon test, with FDR utilized as multiple testing correction. Minimal p-values were 0.037, which related to YM vs. CN, AM vs. CN and CG vs. CN for GRIA2, BF vs. CN and CG vs. CN for GRIA3, YM vs. CN and AM vs. CN for CYFIP2 and FLNA (with higher levels in neurons for the former and in glia for the latter) and BF vs CN for COPA. Note that GRIA3 was not detected in glial cells.

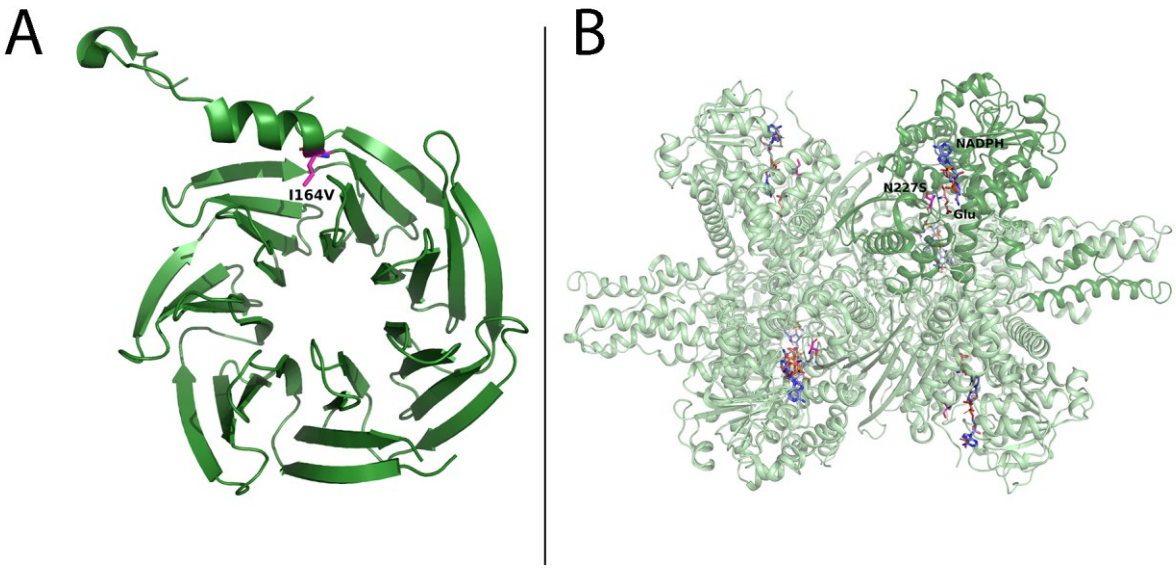

Fig.3. Selected sites of ADAR-mediated editing identified at the proteome level mapped into the spatial structures of homologous molecules. A. Mapping of the N227S substitution into a three-dimensional hexameric structure of the bovine GLUD2 homolog (PDB ID 3ETE). The substitution site is shown in magenta, NADPH molecules are shown in blue, substrate glutamates are shown in pink. One monomer is shown in a darker color for clarity. B. Mapping of the I164V substitution into a three-dimensional structure of the COPA homolog from baker's yeast (PDB ID 4J87). The substitution site is shown in magenta.