REDIportal: toward an integrated view of the A-to-I editing

Pietro D'Addabbo[1], Roni Cohen-Fultheim[2,3,†], Itamar Twersky[2,3,†], Adriano Fonzino[1], Domenico Alessandro Silvestris[1], Ananth Prakash[4], Pietro Luca Mazzacuva[5], Juan Antonio Vizcaino [4], Andrew Green[4], Blake Sweeney [4], Andy Yates [4], Yvonne Lussi[4], Jie Luo[4], Maria-Jesus Martin[4], Eli Eisenberg [6], Erez Y. Levanon[2,3], Graziano Pesole [1,5] and Ernesto Picardi [1,5,*]

[1]Department of Biosciences, Biotechnologies and Environment, University of Bari Aldo Moro, via Orabona 4, 70125, Bari, Italy
[2]Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat Gan, 52900, Israel
[3]Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, 52900, Israel
[4]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton CB10 1SD, UK
[5]Institute of Biomembranes, Bioenergetics and Molecular Biotechnology, National Research Council, via Amendola 122/O, 70126, Bari, Italy
[6]School of Physics and Astronomy, Tel Aviv University, Tel Aviv, 699781, Israel

*To whom correspondence should be addressed. Tel: +39 0805442179; Fax: +39 0805442134; Email: ernesto.picardi@uniba.it
†The second and third authors contributed equally to this work.

Abstract

A-to-I RNA editing is the most common non-transient epitranscriptome modification. It plays several roles in human physiology and has been linked to several disorders. Large-scale deep transcriptome sequencing has fostered the characterization of A-to-I editing at the single nucleotide level and the development of dedicated computational resources. REDIportal is a unique and specialized database collecting ~16 million of putative A-to-I editing sites designed to face the current challenges of epitranscriptomics. Its running version has been enriched with sites from the TCGA project (using data from 31 studies). REDIportal provides an accurate, sustainable and accessible tool enriched with interconnections with widespread ELIXIR core resources such as Ensembl, RNAcentral, UniProt and PRIDE. Additionally, REDIportal now includes information regarding RNA editing in putative double-stranded RNAs, relevant for the immune-related roles of editing, as well as an extended catalog of recoding events. Finally, we report a reliability score per site calculated using a deep learning model trained using a huge collection of positive and negative instances. REDIportal is available at http://srv00.recas.ba.infn.it/atlas/.

Graphical abstract
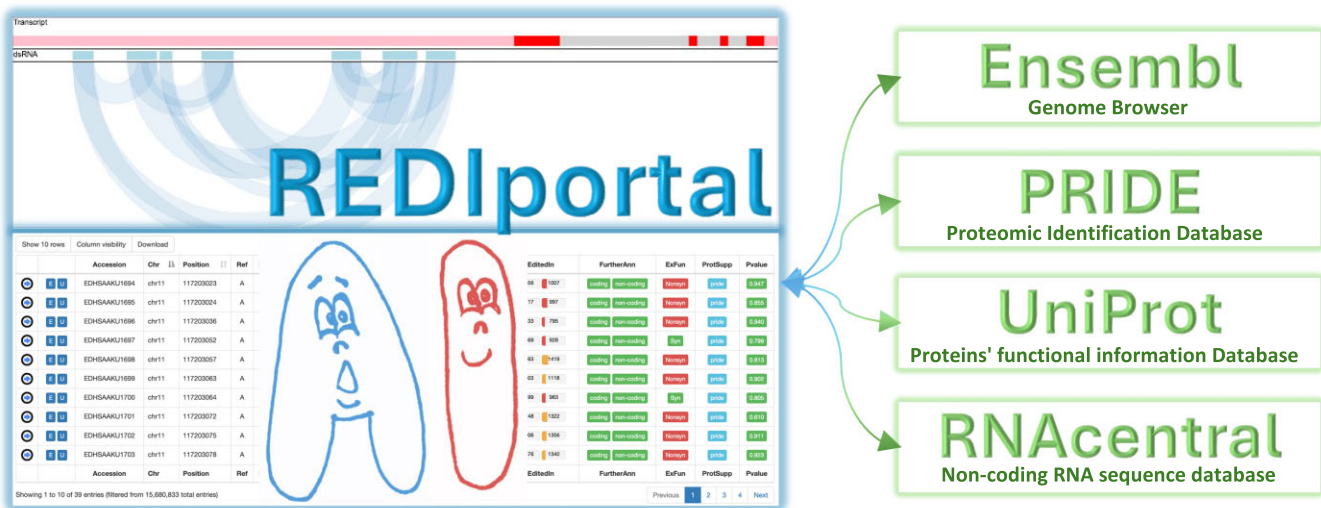
Introduction

Epitranscriptome modifications are emerging as critical factors in fine-tuning gene expression and regulation (1,2). Among them, A-to-I RNA editing by ADAR (adenosine deaminase acting on RNA) enzymes is pervasive in eukaryotic transcriptomes (3–5). In humans, where this phenomenon has been investigated for a while, it has been shown to play several key biological roles depending on its location in the target RNA (i.e. Coding Sequence (CDS), intron or Untranslated Regions (UTRs)) (6). A-to-I editing can modify the basic proper-

ties of neurotransmitter receptors (7,8), modulate alternative splicing (9,10), influence the biogenesis of circRNAs and microRNAs (11–13) or tune their effects acting on their target RNAs (14). The majority of known events occur in repetitive regions and mostly in double-stranded RNAs (dsRNAs) formed by Alu elements in opposite orientations (15–17). A-to-I editing in dsRNAs is pivotal in cytosolic innate immunity by suppressing type I interferon signaling (18,19). On the contrary, only a small fraction of A-to-I conversions has been described in protein-coding regions. These can lead to amino acid changes (recoding sites) (20). Recoding sites are extensively investigated due to the potential for downstream functional effects, and their deregulation was associated with different human disorders including but not limited to neurological, neurodegenerative, autoimmune, cardiovascular diseases and cancer (21–23).

Large-scale deep transcriptome sequencing has fostered the characterization of A-to-I editing at the single nucleotide level and several computational tools for its accurate detection have been released (24–27). Harnessing >9000 RNA sequencing (RNAseq) experiments from the GTEx project and a rigorous bioinformatics protocol based on our REDItools package (25,28), we collected ∼16 million of A-to-I events across 31 tissues and 54 body sites of 549 individuals, populating the REDIportal database (5). Our knowledgebase tool represents a unique and comprehensive A-to-I catalog and is becoming a reference resource for the scientific community and all researchers involved in the epitranscriptome field. REDIportal is queried to investigate RNA editing in personal datasets or downloaded for developing novel computational resources or serving as a benchmark. For instance, several databases such as REIA (29), TCEA (30), CAeditome (31) and GPedit (32) are based on REDIportal annotations. In the current year, its main webpage has been viewed >23 000 times with a mean of ∼2300 views per month (∼77% more than last year) (source Google Analytics—https://analytics.google.com).

Investigating RNA editing may facilitate a better understanding of yet unknown molecular aspects of this non-transient epitranscriptome modification, unveil associations with human disorders, or discover biomarkers with a prognostic or diagnostic value (33). Understanding editability properties is also crucial for designing therapeutic interventions based on site-directed RNA editing approaches, such as correcting disease-causing mutations (34). From a computational point of view, large collections of A-to-I events can be exploited to feed machine and deep learning algorithms for comprehensive editing profiling without the need for specific genomic data or tedious filtering steps implemented in cumbersome bioinformatics pipelines. To improve the investigation of RNA editing and face the current challenges of epitranscriptomics, we have updated REDIportal adding new data and functionalities, maintaining the same layout to leave unchanged the user experience. This third major REDIportal release adds A-to-I changes from >9000 TCGA RNAseqs of primary tumors (covering 31 studies) (35). Available cancer editing databases, such as TCEA (30), CAeditome (31), GPedit (32) or REIA (29), are all built using the first or second REDIportal release. Our new release improves on these resources and includes the AEI (Alu editing index) (36) and REI (recoding index) (21) indices, two well-established metrics to quantify RNA editing activity globally or at recoding sites only, as well as expression values for the ADAR genes (37). Given the relevance of A-to-I editing in innate immunity (6,15,38,39),

REDIportal implements a novel module to study ADAR-mediated editing in putative dsRNAs by providing dsRNA indices per transcript calculated likewise the AEI index. To provide an accurate, sustainable and accessible database, this new release of REDIportal is interconnected with Ensembl (40), UniProt (41), RNAcentral (42) and PRIDE (43), which are widespread and internationally recognized ELIXIR core data resources, part of the European infrastructure for life sciences.

This new REDIportal release is also enriched with an increased portfolio of recoding sites detected in GTEx data by a recent ad hoc computational procedure (20). Finally, REDIportal events have been associated with a reliability score calculated using our REDInet (https://github.com/BioinfoUNIBA/REDInet) algorithm implementing a deep learning model trained using a huge collection of positive and negative examples (44).

All the updates and database improvements described in this manuscript are mainly related to human data. REDIportal is still freely available at http://srv00.recas.ba.infn.it/atlas/index.html.

Data collection and processing

Data collection

TCGA RNAseq data comprising 9683 experiments from primary tumors across 31 studies (and 24 disease types) were downloaded from the GDC Data Portal (https://portal.gdc.cancer.gov/) (dbGaP accession number phs000178) in BAM format using the GDC Data Transfer Tool Client (https://gdc.cancer.gov/access-data/gdc-data-transfer-tool). Since BAM files were processed by aligning reads onto the hg38/GRCh38 human reference genome by STAR (45) and using Gencode annotations, only a preliminary filtering step was performed to retain uniquely mapped reads. All alignments were saved in sorted and indexed BAM files using SAMtools (version 1.9) (46).

Recoding editing sites in GTEx samples, detected using the computational strategy described in (20) implementing stringent filters for protein-coding regions, were downloaded from the publisher website and merged into REDIportal. Resulting A-to-I editing sites were checked to rule out duplicates and positions showing strand biases. The non-redundant list of 15 680 833 bona fide A-to-I editing events was annotated using ANNOVAR and the following updated databases: (i) RepeatMasker containing known repetitive elements; (ii) db-SNP (version 155) collecting genomic single nucleotide polymorphisms (35); (iii) Gencode (v45) (31), RefSeq (36) and UCSC (Genome Browser database) (37) storing gene and transcript annotations; and (iv) PhastCons providing conservation scores across 100 species.

For each site, a unique alphanumeric accession number was generated, allowing the recovery of editing sites independently from the genome assembly and database version, promoting interoperability. To cover annotated sites and those that may originate from future releases, each accession number is composed of eight alphanumeric characters prefixed with ED (for 'EDiting') and an organism code (of at least two letters such as HS for human or MM for mouse), and followed by four upper-case letters and four digits, allowing >4.5 billion combinations per organism.

A-to-I sites collected in REDIportal have been detected using an empirical computational protocol based on our RED-

Itools software (25,28). It identifies editing events according to specific filters that are tuned depending on the RNAseq features (i.e. base quality, mean depth and so on) (28). To improve the reliability of sites collected in REDIportal, we have now introduced an aggregated *P*-value calculated by our REDInet algorithm (https://github.com/BioinfoUNIBA/REDInet). It is based on a deep learning model trained using a balanced dataset comprising >9 million positives and negatives. For each site of a given RNAseq, REDInet explores its surrounding bases and assigns an editing *P*-value. REDInet was applied to all GTEx RNAseq data and resulting *P*-values, aggregated per tissue and body site, were included in the main REDIportal table. REDInet *P*-values have been calculated for human data only since no large A-to-I collections for training are yet available for other organisms.

A-to-I editing detection in TCGA data and editing metrics

RNA variants per sample were called using an HPC-optimized version of our REDItools package (47) and stored in tabular format. REDItools was launched on individual BAM files with non-stringent parameters to provide a comprehensive catalog of all putative and non-putative RNA variants. Output tables were gzipped and indexed by tabix. Known A-to-I editing sites from the latest REDIportal list were extracted from indexed REDItools tables by a custom Python script. *De novo* events were not included in the current release due to computational restrictions. The detection of novel editing sites, indeed, requires stringent and time-consuming filtering steps to take into account somatic changes and regions with altered copy numbers, other than extensive quality checks.

On average 2 166 001 known editing sites appeared edited in TCGA studies with sufficient coverage (>10 reads). The highest number of events was found in TCGA-DLBC with samples from the lymphoid neoplasm diffuse large B-cell lymphoma (Figure 1A). In contrast, TCGA-KIRC with samples from kidney renal clear cell carcinoma contained the lowest number of editing sites (Figure 1A).

In contrast with the available cancer editing databases, REDIportal includes the AEI and REI metrics to quantify the RNA editing activity globally or at recoding sites, respectively. The AEI, defined as the weighted average of editing events occurring at all adenosines within Alu elements, was calculated using the RNAEditingIndexer program (https://github.com/a2iEditing/RNAEditingIndexer) (36). The REI index, instead, was computed as the weighted average over all known recoding sites, as described in (21) and implemented in (37) (https://github.com/BioinfoUNIBA/QEdit). The AEI index is particularly relevant, allowing users to select samples and tumor types according to the RNA editing activity (Figure 1B) and, thus, focus only on events characterized in a given cancer study. The REI index is also as important as AEI in cancer genomics because it allows the identification of tumor studies with an altered recoding activity that could lead to cancer-specific biomarkers. In GBM, for instance, the study of the recoding activity allowed the finding of the COG3 I/V as an eligible site for personalized targeted gene therapy (21).

RNA editing in dsRNAs

ADAR-based editing is pivotal in suppressing innate immune interferon responses triggered by cellular dsRNAs through the MDA5–MAVS axis (18,19). Consequently, investigating RNA editing in dsRNAs can offer valuable insights into the study of immunogenic dsRNAs (6,15) and deciphering the roles of dsRNA editing in inflammatory and autoimmune diseases and cancer, fostering the design and development of novel therapeutic approaches.

From a computational point of view, the detection of dsRNA editing is challenging because real RNA secondary structures remain elusive. Nonetheless, putative dsRNAs could be identified following an approach proposed by Barak *et al.* (48). Using the latest RNAseq curated annotations (downloaded from UCSC in May 2024), we selected non-overlapping gene loci and, for each locus, we extracted the longest mature transcript. Each transcript was aligned with itself by BLAST and plus/minus hits were extracted as putative dsRNAs. For each GTEx sample, we used REDItools to call A-to-I editing falling in the collection of dsRNAs (using those with length >70 bases and identity >70%), filtering out sites annotated in dbSNP (version 155). Individual dsRNAs were annotated using RepeatMasker (downloaded from UCSC in May 2024). To quantify the editing activity per dsRNA or the dsRNA editing activity per transcript, we calculated a dsRNA index likewise the AEI index. In particular, the dsRNA index is defined as the weighted average of editing events occurring in all adenosines within dsRNAs. Since most dsRNAs are due to Alu elements in opposite orientations, the dsRNA index is a proxy of the AEI index but provides insights at the transcript level.

In REDIportal, dsRNA annotations and dsRNA indices for all GTEx samples are stored in a dedicated MySQL table, queryable by gene name.

REDIportal in the ELIXIR ecosystem

ELIXIR is the European life sciences research infrastructure aimed to enable researchers to access and analyze life science data, to improve the value and impact of life science research on public health, the environment and the economy (49) (https://elixir-europe.org/). It involves scientists from 23 European countries and over 250 research institutes. ELIXIR coordinates and develops a large number of bioinformatics resources across Europe including well-established tools of fundamental importance to the wider lifescience community (49). Since epitranscriptomics is yet an under-represented field in the ELIXIR community, and REDIportal is in the ELIXIR portfolio of services (https://elixir-europe.org/services/list), our knowledge base tool has been strengthened through interconnections to existing ELIXIR core data resources such as Ensembl, UniProt, RNAcentral and PRIDE (Figure 2). Through links to Ensembl (40), users can browse editing sites in their genomic context, complementing the JBrowse module already running in REDIportal (Figure 2). On the other end, Ensembl users can now visualize RNA editing tracks and bind to REDIportal for further details. Noncoding RNAs (ncRNAs) in humans as well as other eukaryotes are critical components of cellular machinery (50,51), and many of them are subjected to RNA editing that could modulate their structure and function (52,53). To facilitate the study of A-to-I editing in ncRNAs, REDIportal annotations have been included in RNAcentral (42), a database of ncRNA sequences that aggregates data from specialized ncRNA resources and provides a single entry point for accessing ncRNA sequences of all ncRNA types from all organisms. In RNAcen-
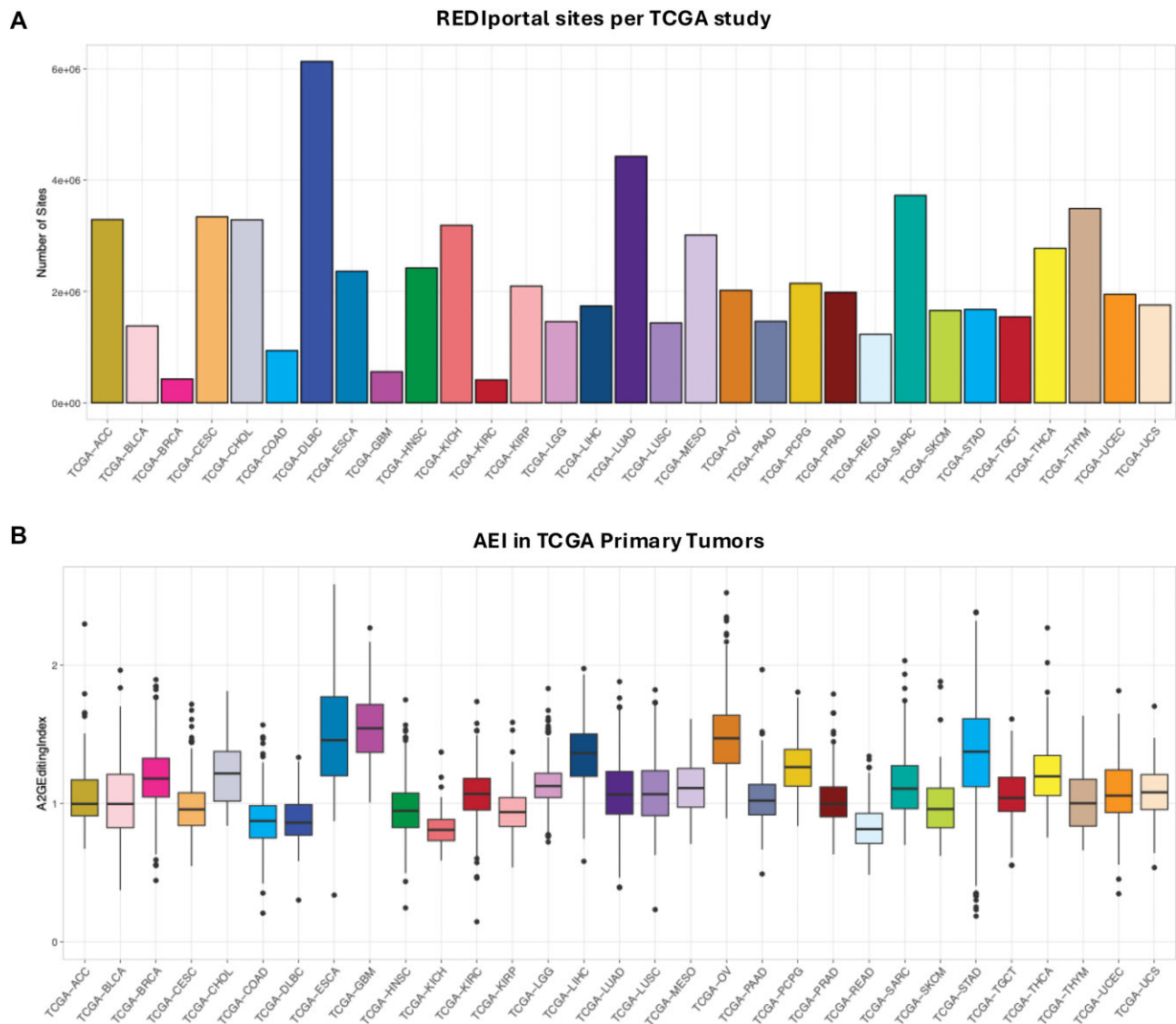
A

REDIportal sites per TCGA study

B

AEI in TCGA Primary Tumors

Figure 1. (**A**) Distribution of REDIportal sites across 31 TCGA studies. (**B**) Box plot of AEI per TCGA study.

tral (https://rnacentral.org/), users can search for all ncRNAs with an editing event and visualize editing sites in a specific non-coding context. In REDIportal, instead, users interested in ncRNAs can point to RNAcentral through dedicated hyperlinks (Figure 2). To connect REDIportal data to RNAcentral sequences, RNAcentral intersected the genomic coordinates of all editing events in REDIportal with known coordinates of all ncRNAs. They then mapped the genomic coordinates into coordinates within a sequence and used this to visualize the editing events in a non-coding context. Future work will focus on showing the editing events in our secondary structure viewer as well as within the sequence.

ADAR-mediated editing can occur in protein-coding genes and lead to recoding events, i.e. non-synonymous substitutions that generate novel protein isoforms (6,52). Although small in number, recoding sites are over-represented in transcripts encoding proteins linked to the nervous system function and may have crucial physiological roles (20,22,54). For example, the recoding R/G site in the GRIA2 gene for the glutamate receptor subunit B regulates the desensitization kinetics of the receptor (55), while the Q/R site in the same gene

regulates the Ca^{2+} permeability of the ion channel and its alteration is related to neuronal death (56).

To better investigate recoding events and their potential functional roles, REDIportal recoding events have been interconnected to UniProt (41) (https://www.uniprot.org/). In particular, all recoding events supported by three annotation databases such as Gencode, RefSeq and UCSC, and showing editing evidence in at least five GTEx individuals and a pooled edited level >1%, were extracted from REDIportal. The resulting sites were mapped on UniProt annotations by the ID mapping tool (https://www.uniprot.org/id-mapping) to detect UniProt accession numbers. Such unique identifiers were included in REDIportal while recoding sites falling in UniProt entries were added in UniProt in the Feature Viewer and handled as a variation track (Figure 2).

Detecting recording sites in large transcriptome datasets is still a challenging task (20). Proteomics data such as those from mass spectrometry experiments can help in identifying peptides containing editing-modified amino acids. The PRIDE database (43) (https://www.ebi.ac.uk/pride/) is the world-leading database for mass spectrometry-based pro-
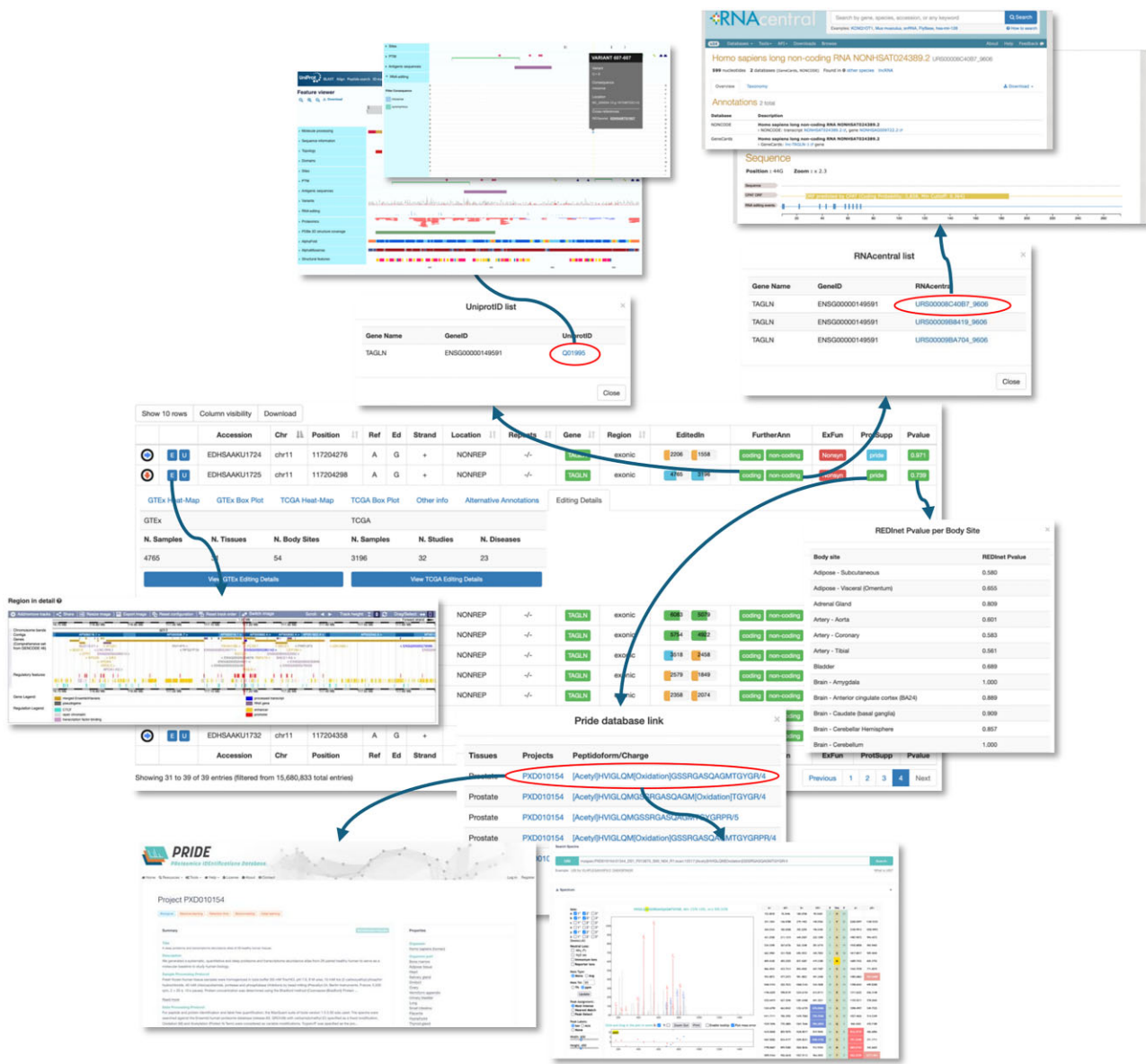
Figure 2. An overview of data integrated in REDIportal. Searched RNA editing sites are shown in sortable tables and for each genomic position several details are provided. In REDIportal v3, editing sites are interconnected to ELIXIR core data resources. If a given site falls in a protein-coding region, a link to UniProt database is provided. Proteomic support (if any) by the PRIDE database is also included by dedicated links. For sites in ncRNAs, specific links to RNAcentral have been added. Users can also browse editing sites in their genomic context by Ensembl. On the contrary, REDIportal sites can also be reached from UniProt, RNAcentral, Ensembl and PRIDE. Circles underscore REDIportal links to ELIXIR core resources. Arrows point to REDIportal popups and ELIXIR core resources.

teomics data. Public proteomics datasets can be harnessed to look at edited positions at the proteome level using a proteogenomics approach. For this purpose, public proteomics datasets were selected and reanalyzed, considering the most reliable edited sites that have been observed at the transcriptome level and annotated at REDIportal. In particular, REDIportal recoding sites used for UniProt mappings were further filtered to remove sites in repeat elements and observed in <10 GTEx samples, yielding a list of 16 184 A-to-I edited sites. Overall, 11 public proteomics datasets from the PRIDE database (see Supplementary Table S1 for the list of datasets) were reanalyzed comprising 2001 human samples. Samples were searched using MaxQuant (version 2.1.0.0) (57) against a database of peptide sequences generated using the UniProt

human proteome (downloaded in May 2023) as the basis. In brief, the protein sequences were *in silico* digested with trypsin to generate >1.2 million peptides. Combinations of 'edited' peptide sequences were generated by substituting only one potentially edited amino acid residue onto a native peptide sequence. Finally, native peptide sequences and all combinations of 'edited' peptides were pooled together into a target peptide search database. A detailed description of the methodology for generating the target peptide search database can be found in the Supplementary File.

The datasets were reanalyzed considering a statistical threshold of 1% false discovery rate at a peptide level. A total of 24 edited amino acid sites were identified in 22 proteins corresponding to 92 peptide spectrum matches

Figure 3. Searching the REDIportal database. In REDIportal v3, RNA editing data can be queried by four dedicated search pages (indicated by arrows) depending on the required details. As a novelty, users can now explore RNA editing in dsRNAs.

(PSMs). Additionally, after the manual inspection of PSMs of the detected sites, low-quality PSMs were discarded. Overall, 11 peptides with an edited amino acid substitution were kept, which were supported by 48 PSMs. These peptides were included in REDIportal. Universal Spectrum Identifiers (USI) (58) can be used to visualize the source mass spectra corresponding to these peptides, linked from REDIPortal into PRIDE (Supplementary Table S2). Examples of proteins in which edited peptides were identified include Transgelin (TAGLN, UniProt accession: H0YCU9) at amino acid position 132 N>S (chr 11: 117204298, A>G) in prostate and smooth muscle samples (USI: https://www.ebi.ac.uk/pride/archive/usi?usi=mzspec% 3APXD010154%3A01323_D01_P013562_S00_N04_R1% 3Ascan%3A15878%3A%5BAcetyl%5DHVIGLQM% 5BOxidation%5DGSSRGASQAGMTGYGR%2F4) (Figure 2) and Filamin A (FLNA, UniProt accession: H0Y5C6) at amino acid position 281 Q>R (chr X: 154351582, T>C) in vermiform appendix (USI: https: //www.ebi.ac.uk/pride/archive/usi?usi=mzspec:PXD015079: 190319-Labrie-3454-APP:scan:10364:LTVSSLR/2).

Detecting peptides changed by RNA editing is quite challenging and computationally intensive. As more datasets are reanalyzed, we hope it will be possible to provide additional results in future REDIportal releases.

Database content and web interface

Update of the database content

REDIportal collects 15 680 833 sites from 9642 human GTEx RNAseq samples (across 31 tissues and 54 body sites) and 9683 human TCGA RNAseq samples (across 31 studies and 24 disease types). Although human-oriented, REDIportal collects also 107 095 sites from mouse RNAseq samples, covering three different tissues.

REDIportal is organized in four MySQL tables. MySQL TABLE1 includes individual sites, their annotations and cross-links to external resources (Ensembl, UniProt, RNA-central and PRIDE). MySQL TABLE2 stores RNA editing levels per RNAseq, while statistics and RNA editing metrics (AEI and REI indices) per GTEx and TCGA sample are allocated in MySQL TABLE3. RNA editing data for dsR-NAs are instead included in the novel MySQL TABLE4. It comprises genomic coordinates of dsRNAs and dsRNA indices.

Update of the web interface

REDIportal inherits the layout from the previous release to leave the user experience unchanged. All web pages are developed in Bootstrap, CSS and JavaScript. Server-side operations to query MySQL tables and retrieve data are performed
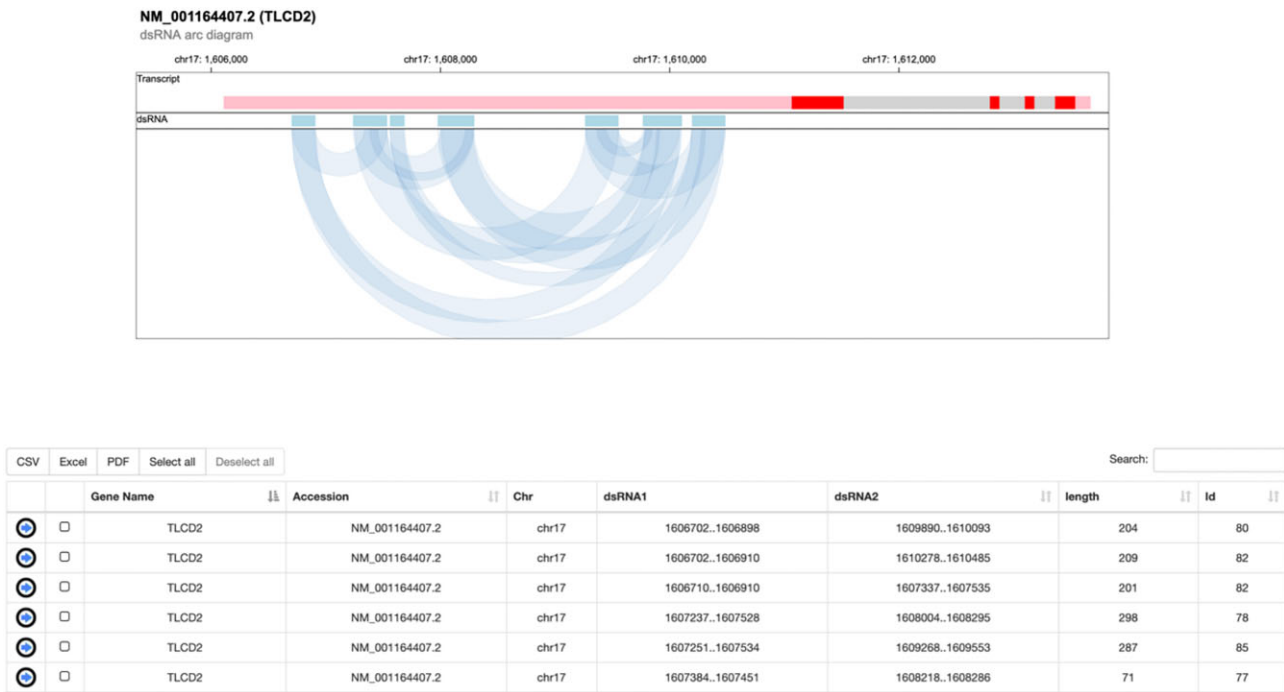
Figure 4. DsRNAs in REDIportal. The new REDIportal version allow the investigation of dsRNAs reported along the transcript structure by arc diagrams using gosling. A table comprising dsRNA coordinates and dsRNA indices is also showed.

in Python and require MySQLdb (https://pypi.python.org/pypi/MySQL-python/1.2.5) and mxTextTools (http://www.egenix.com/products/python/mxBase/mxTextTools/) as external modules for MySQL connections and high-performance text manipulation, respectively.

The REDIportal search dropdown menu now includes four pages: (i) 'Search Positions' to perform queries at the position level by providing a genomic region (in the format chr:start-end) or a gene symbol. A specific GTEx tissue, GTEx body site, TCGA study or a combination of them can also be selected. (ii) 'Search Sample' to interrogate one or more GTEx or TCGA samples by providing a string of run accessions (for GTEx) or aliquots IDs (for TCGA). Samples can also be selected according to AEI or ADAR expression values. (iii) 'Search dsRNA' to retrieve editing in dsRNAs by providing a gene name. (iv) 'Gene View' to visualize RNA editing events in their genomic context (Figure 3).

The 'Search Positions' allows now the retrieval of A-to-I sites in TCGA studies while 'Search Sample' enables also the browsing of TCGA samples and includes five panels per sample: (i) 'Genomics Facts'; (ii) 'Base Distribution'; (iii) 'RNA Editing Indices'; (iv) 'RNA Editing Levels'; and (v) 'Transcriptome Coverage'. The button 'REI details' now allows the browsing of >2000 known recoding events.

RNA editing sites retrieved by the 'Search Positions' page are shown in dynamic and sortable tables. For each human A-to-I site, REDIportal now shows the following info: (i) the unique accession number; (ii) two links to Ensembl and UCSC genome browsers; (iii) the genomic position; (iv) the reference and edited nucleotide; (v) the strand; (vi) the editing location; (vii) the repeated element (if any); (viii) the gene symbol according to Gencode v45 linked to GeneCards; (ix) the genic region; (x) the number of edited samples in GTEx and TCGA projects; (xi) links to UniProt (if in protein-coding gene) or RNAcentral (if in ncRNA); (xii) the potential amino

acid change; (xiii) the PRIDE proteomic support with links to the original dataset and the mass spectrum of the detected edited peptide; and (xiv) the aggregated REDInet *P*-value calculated as the mean of *P*-values across all body sites (Figure 2).

Extra info per site is located in child rows and now consists of seven panels: (i–ii) an interactive heat map and box plot to look at RNA editing levels across GTEx body sites; (ii–iii) an interactive heat map and box plot for editing levels in TCGA studies; (iv) alternative gene/transcript annotations according to RefSeq and UCSC databases; (v) editing details with the number of edited samples in GTEx and TCGA, and links to the corresponding RNA editing levels; and (vi) other info containing the Ensembl gene ID, the gene type (according to Gencode), the phastCons score across 100 species, the db-SNP accession (if any) and a flag indicating whether the site was already known in previous editing databases [including RADAR (59) and DARNED (60)] (Figure 2).

Double-stranded RNAs can be retrieved through the 'Search dsRNA' page and be visualized as arc diagrams in their transcript context by gosling (61) (http://gosling-lang.org/). A table reporting dsRNAs genomic coordinates as well as dsRNA editing indices is also displayed (Figure 4).

Conclusion and future plans

A-to-I RNA editing is a functionally relevant epitranscriptome modification that plays crucial roles in humans and other organisms (6,62). To foster its investigation, we developed REDIportal (4,5), a specialized database comprising the largest collection of A-to-I changes. Its third release, mainly human-oriented, comprises ~16 million events in >18 000 RNAseqs from GTEx and TCGA projects. REDIportal has now been included in the ELIXIR ecosystem by interconnections with widespread core resources such as Ensembl, RNA-

central, UniProt and PRIDE, and enables the study of edited dsRNAs.

To face the current challenges of epitranscriptomics, we plan to extend the repertoire of editing sites with *de novo* events from TCGA samples [facilitating the prediction of cancer neo-antigens (63,64)] and large transcriptome data produced by several consortia. Although REDIportal is mainly human-oriented, in the next releases we plan to increase the number of organisms and adapt the current output layout also to sites from other organisms whose data are already collected in REDIportal (such as mouse).

New sequencing technologies such as direct RNA sequencing by Oxford Nanopore Technologies (ONT) are now available to profile RNA editing and RNA modifications in general (65–67). They hold the promise to detect inosine and modified adenosines at the transcript level (66,68). We plan to update REDIportal with a dedicated module to deal with editing annotations from ONT data. Our future ambition is to collect also modified adenosines (m^6A or m^1A) to investigate the interplay between A-to-I editing and RNA modifications and provide an overview of epitranscriptome changes at the single and transcript level. Since ONT also allows the accurate detection of C-to-U editing events (69), we plan to include them in a future REDIportal release.

Data availability

REDIportal is available through the following web page: http://srv00.recas.ba.infn.it/atlas/. The full list of sites can be downloaded from the REDIportal download page (http://srv00.recas.ba.infn.it/atlas/download.html). REDItools, used to profile RNA variants in RNAseq data, is available at the following GitHub repository: https://github.com/BioinfoUNIBA/REDItools and https://doi.org/10.5281/zenodo.13981272.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank the ReCaS computing center at the University of Bari for hosting the database and providing the needed computational resources. We also acknowledge the European ELIXIR infrastructure (https://www.elixir-europe.org/) and its Italian node (ELIXIR-IT) for supporting the development and maintenance of REDIportal. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP under the accession number phs000424. The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga, available also in dbGaP under the accession number phs000178. TCGA is a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), which are both part of the National Institutes of Health, the US Department of Health and Human Services.

Conflict of interest statement

The authors declare that there aren't interest conflicts.

References

1. Cerneckis,J., Ming,G.-L., Song,H., He,C. and Shi,Y. (2024) The rise of epitranscriptomics: recent developments and future directions. *Trends Pharmacol. Sci.*, **45**, 24–38.
2. Roundtree,I.A., Evans,M.E., Pan,T. and He,C. (2017) Dynamic RNA modifications in gene expression regulation. *Cell*, **169**, 1187–1200.
3. 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.
4. Picardi,E., D'Erchia,A.M., Lo Giudice,C. and Pesole,G. (2016) REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.*, **45**,D750–D757.
5. Mansi,L., Tangaro,M.A., Lo Giudice,C., Flati,T., Kopel,E., Schaffer,A.A., Castrignanò,T., Chillemi,G., Pesole,G. and Picardi,E. (2021) REDIportal: millions of novel A-to-I RNA editing events from thousands of RNAseq experiments. *Nucleic Acids Res.*, **49**, D1012–D1019.
6. Eisenberg,E. and Levanon,E.Y. (2018) A-to-I RNA editing—immune protector and transcriptome diversifier. *Nat. Rev. Genet.*, **19**, 473–490.
7. Hood,J.L. and Emeson,R.B. (2012) Editing of neurotransmitter receptor and ion channel RNAs in the nervous system. *Curr. Top. Microbiol. Immunol.*, **353**, 61–90.
8. Lundin,E., Wu,C., Widmark,A., Behm,M., Hjerling-Leffler,J., Daniel,C., Öhman,M. and Nilsson,M. (2020) Spatiotemporal mapping of RNA editing in the developing mouse brain using *in situ* sequencing reveals regional and cell-type-specific regulation. *BMC Biol.*, **18**, 6.
9. Rueter,S.M., Dawson,T.R. and Emeson,R.B. (1999) Regulation of alternative splicing by RNA editing. *Nature*, **399**, 75–80.
10. Licht,K., Kapoor,U., Amman,F., Picardi,E., Martin,D., Bajad,P. and Jantsch,M.F. (2019) A high resolution A-to-I editing map in the mouse identifies editing events controlled by pre-mRNA splicing. *Genome Res.*, **29**, 1453–1463.
11. Ivanov,A., Memczak,S., Wyler,E., Torti,F., Porath,H.T., Orejuela,M.R., Piechotta,M., Levanon,E.Y., Landthaler,M., Dieterich,C., *et al.* (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.*, **10**, 170–177.
12. Blow,M.J., Grocock,R.J., van Dongen,S., Enright,A.J., Dicks,E., Futreal,P.A., Wooster,R. and Stratton,M.R. (2006) RNA editing of human microRNAs. *Genome Biol.*, **7**, R27.
13. Heale,B.S., Keegan,L.P. and O'Connell,M.A. (2011) The effect of RNA editing and ADARs on miRNA biogenesis and function. *Adv. Exp. Med. Biol.*, **700**, 76–84.
14. Zhang,L., Yang,C.-S., Varelas,X. and Monti,S. (2016) Altered RNA editing in 3′ UTR perturbs microRNA-mediated regulation of oncogenes and tumor-suppressors. *Sci. Rep.*, **6**, 23226.

15. Levanon,E.Y., Cohen-Fultheim,R. and Eisenberg,E. (2024) In search of critical dsRNA targets of ADAR1. *Trends Genet.*, **40**, 250–259.
16. Levanon,E.Y., Eisenberg,E., Yelin,R., Nemzer,S., Hallegger,M., Shemesh,R., Fligelman,Z.Y., Shoshan,A., Pollock,S.R., Sztybel,D., *et al.* (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.*, **22**, 1001–1005.
17. Kim,D.D., Kim,T.T., Walsh,T., Kobayashi,Y., Matise,T.C., Buyske,S. and Gabriel,A. (2004) Widespread RNA editing of embedded Alu elements in the human transcriptome. *Genome Res.*, **14**, 1719–1725.
18. Pestal,K., Funk,C.C., Snyder,J.M., Price,N.D., Treuting,P.M. and Stetson,D.B. (2015) Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity*, **43**, 933–944.
19. Liddicoat,B.J., Piskol,R., Chalk,A.M., Ramaswami,G., Higuchi,M., Hartner,J.C., Li,J.B., Seeburg,P.H. and Walkley,C.R. (2015) RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science*, **349**, 1115–1120.
20. Gabay,O., Shoshan,Y., Kopel,E., Ben-Zvi,U., Mann,T.D., Bressler,N., Cohen-Fultheim,R., Schaffer,A.A., Roth,S.H., Tzur,Z., *et al.* (2022) Landscape of adenosine-to-inosine RNA recoding across human tissues. *Nat. Commun.*, **13**, 1184.
21. Silvestris,D.A., Picardi,E., Cesarini,V., Fosso,B., Mangraviti,N., Massimi,L., Martini,M., Pesole,G., Locatelli,F. and Gallo,A. (2019) Dynamic inosinome profiles reveal novel patient stratification and gender-specific differences in glioblastoma. *Genome Biol.*, **20**, 33.
22. Khermesh,K., D'Erchia,A.M., Barak,M., Annese,A., Wachtel,C., Levanon,E.Y., Picardi,E. and Eisenberg,E. (2016) Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA*, **22**, 290–302.
23. Gallo,A., Vukic,D., Michalik,D., O'Connell,M.A. and Keegan,L.P. (2017) ADAR RNA editing in human disease; more to it than meets the I. *Hum. Genet.*, **136**, 1265–1278.
24. Diroma,M.A., Ciaccia,L., Pesole,G. and Picardi,E. (2019) Elucidating the editome: bioinformatics approaches for RNA editing detection. *Brief. Bioinform.*, **20**, 436–447.
25. Lo Giudice,C., Tangaro,M.A., Pesole,G. and Picardi,E. (2020) Investigating RNA editing in deep transcriptome datasets with REDItools and REDIportal. *Nat. Protoc.*, **15**, 1098–1131.
26. Ramaswami,G., Lin,W., Piskol,R., Tan,M.H., Davis,C. and Li,J.B. (2012) Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods*, **9**, 579–581.
27. Ramaswami,G., Zhang,R., Piskol,R., Keegan,L.P., Deng,P., O'Connell,M.A. and Li,J.B. (2013) Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods*, **10**, 128–132.
28. Picardi,E. and Pesole,G. (2013) REDItools: high-throughput RNA editing detection made easy. *Bioinformatics*, **29**, 1813–4.
29. Zhu,H., Huang,L., Liu,S., Dai,Z., Songyang,Z., Weng,Z. and Xiong,Y. (2022) REIA: a database for cancer A-to-I RNA editing with interactive analysis. *Int. J. Biol. Sci.*, **18**, 2472–2483.
30. Lin,C.-H. and Chen,S.C.-C. (2019) The Cancer Editome Atlas: a resource for exploratory analysis of the adenosine-to-inosine RNA editome in cancer. *Cancer Res.*, **79**, 3001–3006.
31. Wu,S., Fan,Z., Kim,P., Huang,L. and Zhou,X. (2023) The integrative studies on the functional A-to-I RNA editing events in human cancers. *Genomics Proteomics Bioinformatics*, **21**, 619–631.
32. Ruan,H., Li,Q., Liu,Y., Liu,Y., Lussier,C., Diao,L. and Han,L. (2022) GPEdit: the genetic and pharmacogenomic landscape of A-to-I RNA editing in cancers. *Nucleic Acids Res.*, **50**, D1231–D1237.
33. Chen,L., Li,Y., Lin,C.H., Chan,T.H., Chow,R.K., Song,Y., Liu,M., Yuan,Y.F., Fu,L., Kong,K.L., *et al.* (2013) Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat. Med.*, **19**, 209–216.
34. Pfeiffer,L.S. and Stafforst,T. (2023) Precision RNA base editing with engineered and endogenous effectors. *Nat. Biotechnol.*, **41**, 1526–1542.
35. Cancer Genome Atlas Research Network, Weinstein,J.N., Collisson,E.A., Mills,G.B., Shaw,K.R.M., Ozenberger,B.A., Ellrott,K., Shmulevich,I., Sander,C. and Stuart,J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
36. Roth,S.H., Levanon,E.Y. and Eisenberg,E. (2019) Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat. Methods*, **16**, 1131–1138.
37. Lo Giudice,C., Silvestris,D.A., Roth,S.H., Eisenberg,E., Pesole,G., Gallo,A. and Picardi,E. (2020) Quantifying RNA editing in deep transcriptome datasets. *Front. Genet.*, **11**, 194.
38. Szymczak,F., Cohen-Fultheim,R., Thomaidou,S., de Brachène,A.C., Castela,A., Colli,M., Marchetti,P., Levanon,E., Eizirik,D. and Zaldumbide,A. (2022) ADAR1-dependent editing regulates human β cell transcriptome diversity during inflammation. *Front. Endocrinol. (Lausanne)*, **13**, 1058345.
39. Knebel,U.E., Peleg,S., Dai,C., Cohen-Fultheim,R., Jonsson,S., Poznyak,K., Israeli,M., Zamashanski,L., Glaser,B., Levanon,E.Y., *et al.* (2024) Disrupted RNA editing in beta cells mimics early-stage type 1 diabetes. *Cell Metab.*, **36**, 48–61.
40. Harrison,P.W., Amode,M.R., Austine-Orimoloye,O., Azov,A.G., Barba,M., Barnes,I., Becker,A., Bennett,R., Berry,A., Bhai,J., *et al.* (2024) Ensembl 2024. *Nucleic Acids Res.*, **52**, D891–D899.
41. UniProt Consortium (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
42. RNAcentral Consortium (2021) RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.*, **49**, D212–D220.
43. Perez-Riverol,Y., Bai,J., Bandla,C., García-Seisdedos,D., Hewapathirana,S., Kamatchinathan,S., Kundu,D.J., Prakash,A., Frericks-Zipper,A., Eisenacher,M., *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.*, **50**, D543–D552.
44. Fonzino,A., Mazzacuva,P.L., Handen,A., Silvestris,D.A., Arnold,A., Pecori,R., Pesole,G. and Picardi,E. (2024) REDInet: a TCN-based classifier for A-to-I RNA editing detection harnessing million known events. https://www.researchsquare.com/article/rs-4900829/v1.
45. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
46. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–9.
47. Flati,T., Gioiosa,S., Spallanzani,N., Tagliaferri,I., Diroma,M.A., Pesole,G., Chillemi,G., Picardi,E. and Castrignanò,T. (2020) HPC-REDItools: a novel HPC-aware tool for improved large scale RNA-editing analysis. *BMC Bioinformatics*, **21**, 353.
48. Barak,M., Porath,H.T., Finkelstein,G., Knisbacher,B.A., Buchumenski,I., Roth,S.H., Levanon,E.Y. and Eisenberg,E. (2020) Purifying selection of long dsRNA is the first line of defense against false activation of innate immunity. *Genome Biol.*, **21**, 26.
49. Harrow,J., Hancock,J., ELIXIR-EXCELERATE Community and Blomberg,N. (2021) ELIXIR-EXCELERATE: establishing Europe's data infrastructure for the life science research of the future. *EMBO J.*, **40**, e107409.
50. Mattick,J.S., Amaral,P.P., Carninci,P., Carpenter,S., Chang,H.Y., Chen,L.-L., Chen,R., Dean,C., Dinger,M.E., Fitzgerald,K.A., *et al.* (2023) Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.*, **24**, 430–447.
51. Mattick,J.S. and Makunin,I.V. (2006) Non-coding RNA. *Hum. Mol. Genet.*, **15**, R17–R29.
52. Nishikura,K. (2016) A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.*, **17**, 83–96.

53. Picardi,E., D'Erchia,A.M., Gallo,A., Montalvo,A. and Pesole,G. (2014) Uncovering RNA editing sites in long non-coding RNAs. *Front. Bioeng. Biotechnol.*, **2**, 64.

54. Behm,M. and Ohman,M. (2016) RNA editing: a contributor to neuronal dynamics in the mammalian brain. *Trends Genet.*, **32**, 165–75.

55. Wen,W., Lin,C.-Y. and Niu,L. (2017) R/G editing in GluA2Rflop modulates the functional difference between GluA1 flip and flop variants in GluA1/2R heteromeric channels. *Sci. Rep.*, **7**, 13654.

56. Wright,A.L., Konen,L.M., Mockett,B.G., Morris,G.P., Singh,A., Burbano,L.E., Milham,L., Hoang,M., Zinn,R., Chesworth,R., *et al.* (2023) The Q/R editing site of AMPA receptor GluA2 subunit acts as an epigenetic switch regulating dendritic spines, neurodegeneration and cognitive deficits in Alzheimer's disease. *Mol. Neurodegener.*, **18**, 65.

57. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.

58. Deutsch,E.W., Perez-Riverol,Y., Carver,J., Kawano,S., Mendoza,L., Van Den Bossche,T., Gabriels,R., Binz,P.-A., Pullman,B., Sun,Z., *et al.* (2021) Universal Spectrum Identifier for mass spectra. *Nat. Methods*, **18**, 768–770.

59. Ramaswami,G. and Li,J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, **42**, D109–D113.

60. Kiran,A.M., O'Mahony,J.J., Sanjeev,K. and Baranov,P.V. (2013) Darned in 2013: inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res.*, **41**, D258–D261.

61. L'Yi,S., Wang,Q., Lekschas,F. and Gehlenborg,N. (2021) Gosling: a grammar-based toolkit for scalable and interactive genomics data visualization. *IEEE Trans. Vis. Comput. Graph.*, **28**, 140–150.

62. Danecek,P., Nellaker,C., McIntyre,R.E., Buendia-Buendia,J.E., Bumpstead,S., Ponting,C.P., Flint,J., Durbin,R., Keane,T.M. and Adams,D.J. (2012) High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.*, **13**, 26.

63. Komatsu,Y., Shigeyasu,K., Yano,S., Takeda,S., Takahashi,K., Hata,N., Umeda,H., Yoshida,K., Mori,Y., Yasui,K., *et al.* (2022) RNA editing facilitates the enhanced production of neoantigens during the simultaneous administration of oxaliplatin and radiotherapy in colorectal cancer. *Sci. Rep.*, **12**, 13540.

64. Zhou,C., Wei,Z., Zhang,L., Yang,Z. and Liu,Q. (2020) Systematically characterizing A-to-I RNA editing neoantigens in cancer. *Front. Oncol.*, **10**, 593989.

65. Furlan,M., Delgado-Tejedor,A., Mulroney,L., Pelizzola,M., Novoa,E.M. and Leonardi,T. (2021) Computational methods for RNA modification detection from nanopore direct RNA sequencing data. *RNA Biol.*, **18**, 31–40.

66. Nguyen,T.A., Heng,J.W.J., Kaewsapsak,P., Kok,E.P.L., Stanojević,D., Liu,H., Cardilla,A., Praditya,A., Yi,Z., Lin,M., *et al.* (2022) Direct identification of A-to-I editing sites with nanopore native RNA sequencing. *Nat. Methods*, **19**, 833–844.

67. Jain,M., Abu-Shumays,R., Olsen,H.E. and Akeson,M. (2022) Advances in nanopore direct RNA sequencing. *Nat. Methods*, **19**, 1160–1164.

68. Maestri,S., Furlan,M., Mulroney,L., Coscujuela Tarrero,L., Ugolini,C., Dalla Pozza,F., Leonardi,T., Birney,E., Nicassio,F. and Pelizzola,M. (2024) Benchmarking of computational methods for m^{6}A profiling with Nanopore direct RNA sequencing. *Brief. Bioinform.*, **25**, bbae001.

69. Fonzino,A., Manzari,C., Spadavecchia,P., Munagala,U., Torrini,S., Conticello,S., Pesole,G. and Picardi,E. (2024) Unraveling C-to-U RNA editing events from direct RNA sequencing. *RNA Biol.*, **21**, 1–14.