

REDportal: millions of novel A-to-I RNA editing events from thousands of RNAseq experiments

Luigi Mansi^{1,†}, Marco Antonio Tangaro^{2,†}, Claudio Lo Giudice¹, Tiziano Flati³, Eli Kopel⁴, Amos Avraham Schaffer⁴, Tiziana Castrignanò⁵, Giovanni Chillemi⁶, Graziano Pesole^{1,2,7} and Ernesto Picardi^{1,2,7,*}

¹Department of Biosciences, Biotechnologies and Biopharmaceutics (DBBB), University of Bari, Via Orabona 4, 70125 Bari, Italy, ²Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), National Research Council, Via Amendola 122/O, 70126 Bari, Italy, ³SCAI-Super Computing Applications and Innovation Department, CINECA, Via dei Tizii 6B, 00185 Rome, Italy, ⁴Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, 52900 Ramat Gan, Israel, ⁵Department of Ecological and Biological Sciences (DEB), University of Tuscia, Via S. Camillo de Lellis 44, 01100 Viterbo, Italy, ⁶Department for Innovation in Biological, Agro-food and Forest systems (DIBAF), University of Tuscia, Via S. Camillo de Lellis 44, 01100 Viterbo, Italy and ⁷National Institute of Biostructures and Biosystems (INBB), 00136 Roma, Italy

Received September 14, 2020; Revised September 30, 2020; Editorial Decision October 01, 2020; Accepted October 05, 2020

ABSTRACT

RNA editing is a relevant epitranscriptome phenomenon able to increase the transcriptome and proteome diversity of eukaryotic organisms. ADAR mediated RNA editing is widespread in humans in which millions of A-to-I changes modify thousands of primary transcripts. RNA editing has pivotal roles in the regulation of gene expression or modulation of the innate immune response or functioning of several neurotransmitter receptors. Massive transcriptome sequencing has fostered the research in this field. Nonetheless, different aspects of the RNA editing biology are still unknown and need to be elucidated. To support the study of A-to-I RNA editing we have updated our REDportal catalogue raising its content to about 16 millions of events detected in 9642 human RNAseq samples from the GTEx project by using a dedicated pipeline based on the HPC version of the REDtools software. REDportal now allows searches at sample level, provides overviews of RNA editing profiles per each RNAseq experiment, implements a Gene View module to look at individual events in their genic context and hosts the CLAIRE database. Starting from this novel version, REDportal will start collecting non-human RNA editing changes for comparative genomics investigations. The database is freely available at <http://srv00.recas.ba.infn.it/atlas/index.html>.

INTRODUCTION

RNA editing refers to a group of non-transient epitranscriptome modifications altering primary RNA transcripts through the insertion/deletion of specific nucleotides or base substitutions (1). The deamination of adenosine (A) in inosine (I) is the most common type of RNA editing, affecting thousands of nuclear and cytoplasmic transcripts in a variety of eukaryotic organisms (2). In humans, as well as in other mammals, the A-to-I conversion is mediated by members of the ADAR family of enzymes acting on double stranded (ds) RNAs which comprises ADAR (also known as ADAR1) and ADARB1 (also known as ADAR2), expressed in the majority of tissues, and ADARB2 (also known as ADAR3) found mainly in the nervous central system and thought to be catalytically inactive (3). A-to-I editing events are prominent in long double-stranded RNAs (dsRNAs) located in non-coding regions and formed by repeated elements in opposite orientation (mainly Alu sequences) (2). By contrast, the list of ADAR substrates in protein coding genes is relatively small (4).

Since Inosine mimics the properties of guanosine (G), it is commonly recognised as G by transcription and translation machineries (other than sequencing enzymes). As a consequence, A-to-I RNA editing can increase the transcriptome and proteome diversity, generate or destroy splice sites, alter codon identity or base-pairing interactions within higher-order RNA structures. Several evidences indicate that ADAR mediated RNA editing plays pivotal functional roles, tuning gene expression (5,6) or modulating the innate immune response through the MDA5-MAVS axis (2,7). Additionally, its deregulation is under active investigation be-

*To whom correspondence should be addressed. Tel: +39 0805442179; Fax: +39 0805443317; Email: ernesto.picardi@uniba.it

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

ing linked to different human disorders (8) including neurological (9,10), autoimmune (11), cardiovascular diseases (12) and cancer (13,14).

RNA editing has a great and promising therapeutic potential (15). Indeed, in contrast to CRISPR gene editing in which the phenotype rescue could be associated to undesired immune system responses (16) or accidental permanent genome changes (17), RNA editing could allow temporary fixes that eliminate candidate mutations with reduced adverse collateral effects (18,19).

The application of programmable A-to-I editing as well as the study of key events linked to human disorders or the investigation of still unknown RNA editing properties and functions or the development of advanced computing methods based on artificial intelligence require large and accurate collections of RNA editing events from a variety of transcriptome experiments. A few years ago we released REDiportal (20), a specialized database for A-to-I editing comprising >4.5 millions of events in 55 body sites of 150 healthy individuals from the Genotype-Tissue Expression (GTEx) project (<https://gtexportal.org/home/>). Currently, REDiportal is the unique and comprehensive resource for human A-to-I editing. Indeed, other competing databases such as RADAR (21) and DARNED (22) are outdated or in dismission.

REDiportal has been initially developed contrasting 2660 GTEx RNAseq experiments with a large collection of known A-to-I editing sites detected in six human tissues with the addition of data from the RADAR database. Although the bioinformatic identification of A-to-I editing is easy, its profiling at single nucleotide level in a huge amount of RNAseq data is computationally intensive. Also, in a recent work dealing with the dynamic landscape and regulation of RNA editing in mammals, authors describe A-to-I editing in 8551 human GTEx RNAseq data but only 381 of these were really profiled at single nucleotide resolution (23).

To provide a comprehensive overview of RNA editing in humans, we describe here a novel release of REDiportal comprising about 16 millions of A-to-I events detected *de novo* at single nucleotide resolution in 9642 GTEx RNAseq data. Editing candidates have been identified using an *ad hoc* bioinformatics protocol based on our REDiportal algorithm optimized for High Performance Computing (HPC) systems (24–26). The current version includes also events detected in hyper-edited reads that fail to be correctly aligned to the genome (27).

The novel REDiportal database stores individual A-to-I positions as well as statistics and relevant metrics per each GTEx sample, such as the Alu Editing Index (AEI) or the Recoding Editing Index (REI) or the expression of ADAR genes, that are expected to facilitate the RNA editing investigations. Now users can browse and visualize editing sites through our embedded and updated JBrowse or explore them in their genic context by our novel Gene View functionality.

With this release, REDiportal officially starts collecting RNA editing in non-human organisms, providing annotations for 107 094 A-to-I events from mouse nascent RNAseq data (5).

The REDiportal resource also includes the Cell Line A-to-I Rna Editing (CLAIRE) database (<http://srv00.recas.ba.infn.it/atlas/claire.html>) (28), making our portal a reference point for the scientific community and all researchers involved in the RNA editing field.

DATA COLLECTION AND PROCESSING

Data collection

We downloaded 9642 GTEx RNAseq experiments from the database of Genotypes and Phenotypes (dbGaP) (<https://www.ncbi.nlm.nih.gov/gap/>) (29) under the accession number phs000424.v7.p2 in sra format and converted them in standard fastq using the fastq-dump program of the SRA toolkit (<http://ncbi.github.io/sra-tools/>). In all cases in which WGS data (DNAseq) were available, they were downloaded from dbGaP in sra format and converted in raw fastq by fastq-dump. RNAseq reads were aligned onto the human genome (hg19/GRCh37) using STAR (version 2.5) (30) providing known gene annotations from Gencode (version 31) (31) (Figure 1A). DNAseq reads, instead, were aligned onto the human genome (hg19/GRCh37) using BWA (version 0.7) (32) (Figure 1A). All alignments were saved in sorted and indexed BAM files using SAMtools (version 1.9) (33).

RNA editing detection

The *de novo* detection of RNA editing events was performed at the CINECA HPC Data Center (Italy) consuming ~30 millions of CPU hours running an optimized version of our REDiportal package whose algorithm scales almost linearly with the number of available cores (25). Indeed, the editing identification in massive transcriptome sequencing data is computationally intensive and time consuming, requiring the screening of the entire human genome, position by position, in order to look at nucleotide differences between RNA reads and the corresponding reference genomic site. In addition, depending on the sequencing throughput, individual genomic positions are supported by a very different number of RNA reads (sequencing depth), sometimes higher than 8000 counts. To speed up the browsing and traversing of aligned RNAseq reads, the editing detection was distributed over multiple computing nodes, each working on a given genomic interval. Moreover, the editing identification at nucleotide level was improved by a novel routine developed to increase the data loading efficiency, raising the algorithm performances of 8–10 times (25).

To identify all potential editing candidates the HPC version of REDiportal was initially launched on individual BAM files of aligned RNAseq reads with non-stringent parameters. DNAseq support was subsequently added if available. Each REDiportal table was then filtered according to our protocol described in Lo Giudice *et al.* (26) (Figure 1A). Briefly, all detected positions were annotated using known SNP sites, repeated elements in RepeatMasker and known editing events stored in the first release of the REDiportal database. SNPs and sites not supported by at least 10 DNAseq reads (if available) were removed, and the remaining positions were grouped in ALU (residing in

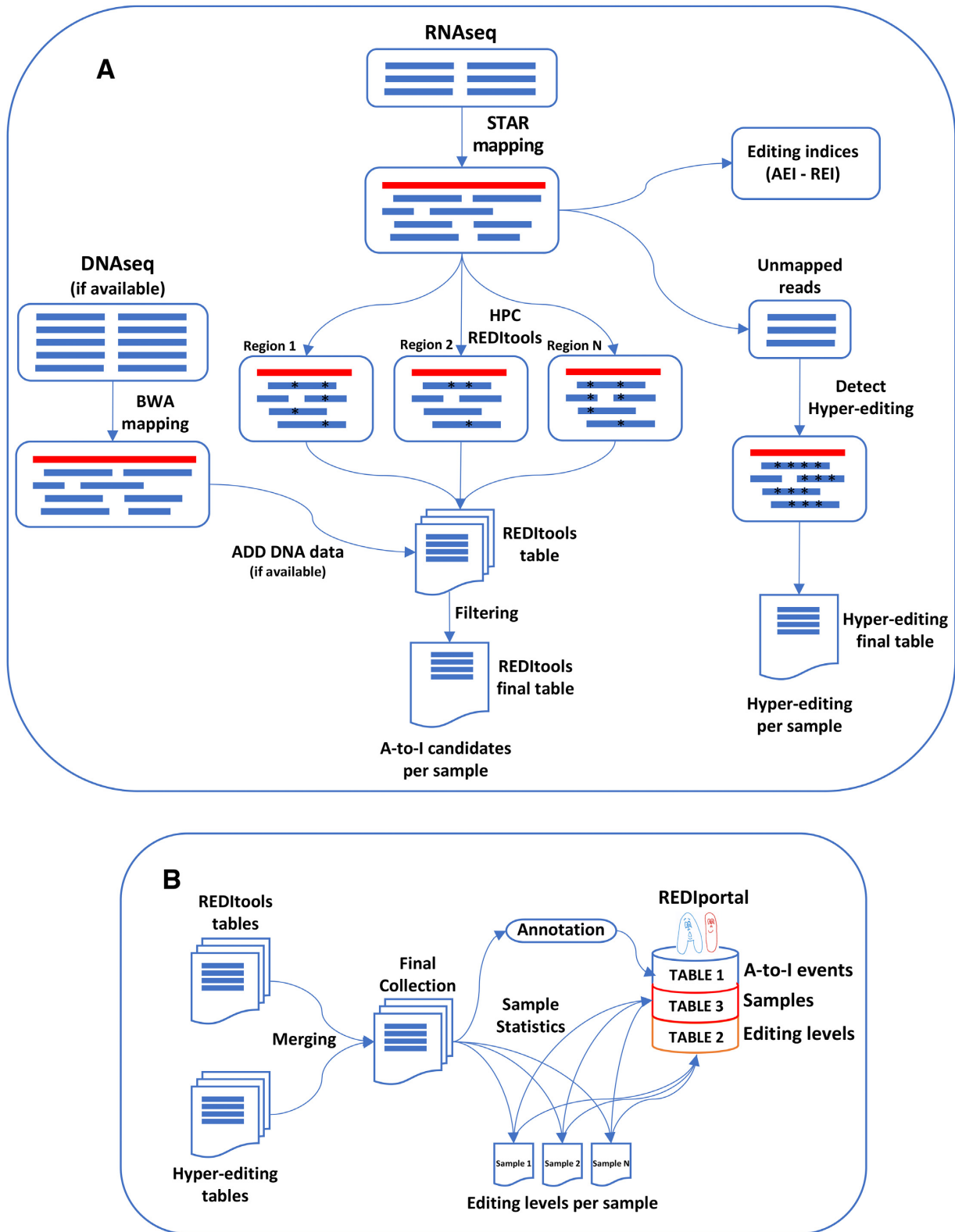


Figure 1. Data processing and database construction. (A) RNAseq data in fastq format are aligned on the human genome by STAR and converted in BAM files. In parallel, if DNaseq reads are available, are aligned on the same genome version by BWA. RNAseq BAM files are analyzed by HPC REDIttools and the editing calling is distribute across different computing nodes, each working on a given genomic region. Resulting REDIttools tables undergo to further filtering steps before the generation of the final table of A-to-I candidates. RNAseq unmapped reads are re-analyzed to detect hyper-edited reads and provide a list of hyper-editing sites per sample. RNAseq BAM files are further used to calculate the AEI and REI indices. (B) REDIttools table and hyper-editing tables are merged in the final REDiportal collection. All events are annotated and stored in the MySQL TABLE1. They are also used to interrogate all RNAseq data to recover RNA editing levels and populate the MySQL TABLE2. Main RNA editing statistics per sample are also computed and collected in the MySQL TABLE3. Blue rectangles are reads, red rectangles are genomic regions, while black stars are A-to-I candidates.

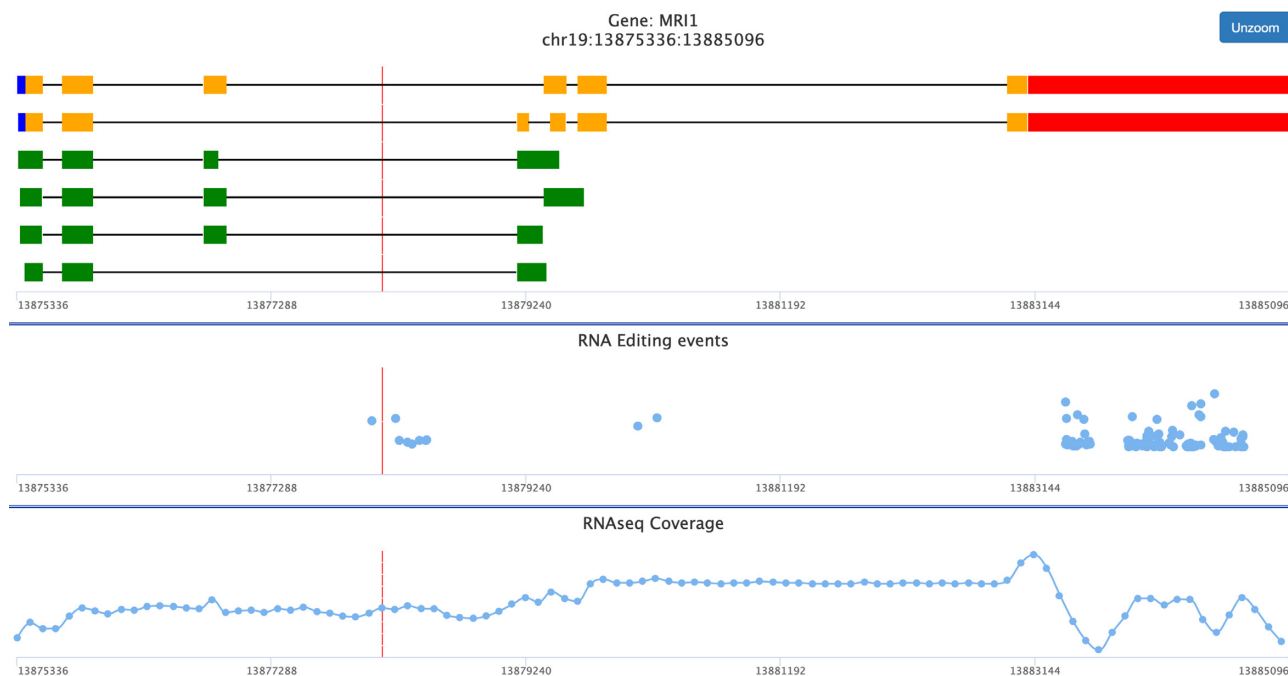


Figure 2. Example of Gene View page for gene MRI1. Gene View is a novel REDIttools functionality in which users can visualize RNA editing events in their genic context. The web page shows at most three panels: (i) the gene structure with details about individual transcripts and specific features such as 5'UTR for 5' UTR in blue, 3'UTR for 3' UTR in red, CDS for the protein coding region in orange, Intron for intervening sequences in black and Exon for a non-coding exon in green; (ii) the list of RNA editing events in blue circles with related levels (the mean value is included in case of searches from the Gene View search page); (iii) the RNAseq coverage for the specific genomic region and related to the specific RNAseq experiment. This panel is visible only at sample level.

Alu elements), REP NON ALU (residing in repetitive non Alu elements) and NON REP (residing in non-repetitive regions) groups, according to RepeatMasker annotations. While NON REP and REP NON ALU variants underwent a second round of REDIttools using stringent call criteria to exclude multimapping reads and PCR duplicates, changes in ALU regions were filtered only by coverage (at least 5 reads) and base quality (phred score of at least 30). At the end, all filtered positions were collected returning the final list of RNA editing candidates per RNAseq (Figure 1A).

In parallel, unmapped reads per sample were analysed using the pipeline by Porath *et al.* (27) in order to identify A-to-I events in hyper-edited reads (Figure 1A).

Annotation and downstream analyses

Filtered REDIttools tables from 9642 samples were merged yielding 10 089 202 editing positions, while the union of hyper-edited tables returned 9 982 214 editing sites (Figure 1B). The merging between both collections yielded a comprehensive and non-redundant list of 15 683 855 *bona fide* A-to-I editing events. All positions were then annotated using ANNOVAR (34) (a standalone software to functionally annotate genetic variants) and the following updated databases: (i) RepeatMasker containing known repetitive elements; (ii) dbSNP (version 151) collecting genomic single nucleotide polymorphisms (35); (iii) Gencode (v34) (31), RefSeq (36) and UCSC (Genome Browser database) (37) storing gene and transcript annotations; (iv) Phast-Cons providing conservation scores across 100 species. Al-

though RADAR and DARNED databases are not fully functional, known A-to-I changes from both repositories were added for continuity with the previous REDItportal database. The complete collection was also lifted to human genome hg38/GRCh38 by liftover (a free software to convert genome coordinates and genome annotation files between assemblies) and annotated using the corresponding hg38 databases. RADAR and DARNED positions were converted to hg38/GRCh38 accordingly.

Hg19 and hg38 annotated positions were finally loaded in the MySQL TABLE1 of REDItportal (Figure 1B).

All positions were also used to interrogate 9642 RNAseq samples to extract RNA editing levels as well as the number of reads supporting As and Gs per site. These values were collected in the MySQL TABLE2 of REDItportal (Figure 1B).

For each aligned RNAseq experiment, FeatureCounts (38) (a software for counting reads to genomic features such as genes, exons, promoters and genomic bins) was applied to count the number of reads per gene and a custom script was used to normalized these values in TPM (transcripts per million). Same aligned reads were also used as inputs to calculate the AEI and REI indices that are relevant metrics to measure the RNA editing activity, globally or at recoding sites, respectively (Figure 1A). The AEI was calculated using the RNAEditingIndexer program (<https://github.com/a2iEditing/RNAEditingIndexer>) (39), while the REI was computed as described in Silvestris *et al.* (13) and implemented in Lo Giudice *et al.* (40) (<https://github.com/BioinfoUNIBA/QEdit>).



Figure 3. Example of sample search for run SRR1069188. The output of the sample search provides basic RNA editing info per RNAseq in a tabular format and enables the browsing of five panels with further details. (A) ‘Genomics Facts’ includes the location and distribution of detected sites in different genomic and genic regions; (B) ‘Base Distribution’ shows the graphical distribution of all detected variants by the REDIttools based pipeline; (C) ‘RNA Editing Indices’ depicts box plots of AEI and REI indices calculated on the body site group of the retrieved sample. Specific index values are indicated in each plot by an empty circle; (D) ‘RNA Editing Levels’ shows the distribution of RNA editing levels; (E) ‘Transcriptome Coverage’ displays statistics about the fraction of edited and unedited genes, and the distribution of detected events in mRNAs or ncRNAs. Panels C and E include two additional buttons for further details about recoding events (Figure 3) and edited genes.

The AEI was defined as the weighted average of editing events occurring in all adenosines within Alu elements, while the REI was the weighted average over all known recoding sites (i.e. residing in coding protein genes).

Finally, a custom script was used to collect main statistics for each GTEx sample and generate the MySQL TABLE3 of REDiportal (Figure 1B).

DATABASE CONTENT AND WEB INTERFACE

Database construction and content

As in the previous release, REDiportal allocates all 15 683 855 sites in two main MySQL tables. TABLE1 includes individual sites and their annotations, while TABLE2 stores RNA editing levels per RNAseq, tissue and body site. Statistics and RNA editing metrics per sample are instead stored in the novel TABLE3 MySQL table. It comprises the RNAseq run accession (according to dbGAP), the DNaseq run accession (if available), the organism name, the data source (GTEx or SRA), the body site and its status (healthy or diseased), the tissue type (bulk or single cell), the number

of detected A-to-I events as well as the number of hyper-editing sites, the AEI and REI indices, and the expression of ADAR genes (ADAR, ADARB1 and ADARB2). TABLE3 contains also further statistics such as the distribution of editing sites over the three ALU, REP NON ALU and NON REP groups, the fraction of synonymous or non-synonymous events in protein coding regions, the distribution of editing sites across the gene structure, the fraction of edited and hyper-edited positions, and the distribution of RNA editing levels. In addition, it includes the fraction of edited and unedited genes, and the distribution of all variants detected per sample as a sort of quality check of the prediction.

REDiportal provides RNA editing details for 9642 RNAseq samples from 549 individuals across 31 tissues and 54 body sites. By means of the HPC REDIttools based pipeline, each sample contains 60 873 edited events on average. The mean number of A-to-I events in hyper-edited reads per sample, instead, is 26 942. The highest number of events was detected in the cerebellar hemisphere and cerebellum (Supplementary Figure S1). By contrast, skeletal

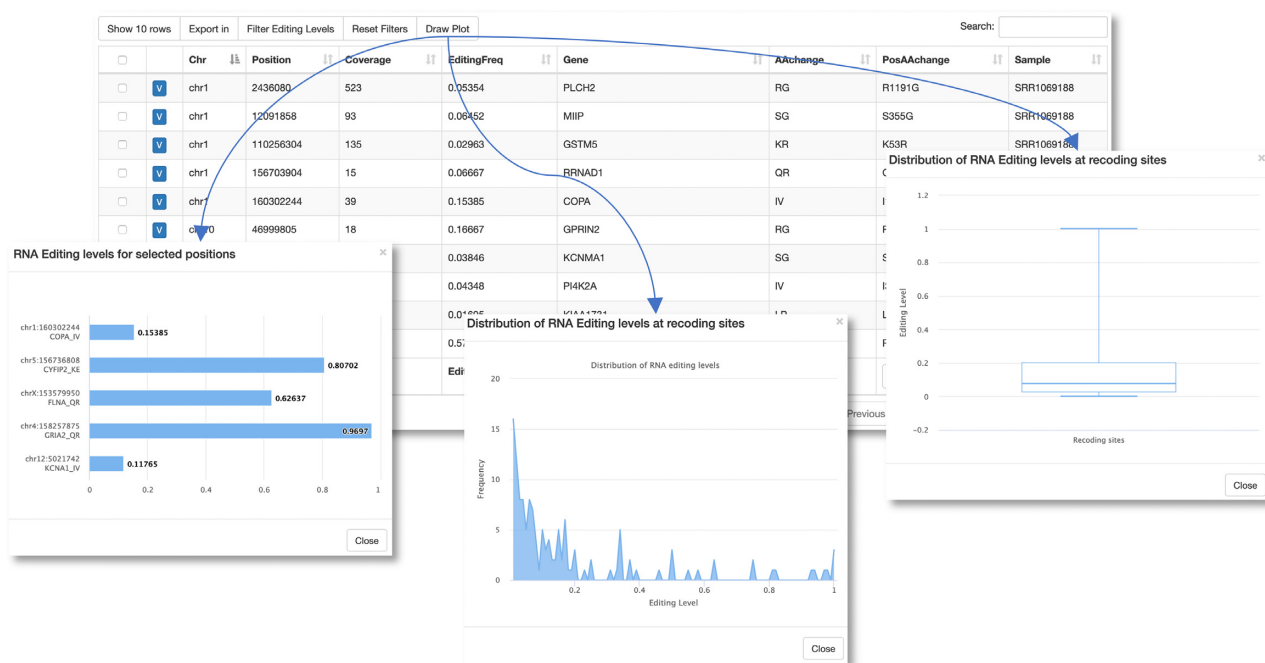


Figure 4. Details about recoding events. The ‘RNA Editing Indices’ panel comprises a button enabling the browsing of individual recoding events. Selected positions can be compared by a bar graph or all sites can be shown in dedicated plots.

muscle and heart contained the lowest number of candidates. As already known by the literature, the majority of A-to-I events resides in Alu elements located mainly in intronic regions (Supplementary Figure S2) (41). A consistent fraction of sites was identified in intergenic and 3' UTR regions (Supplementary Figure S3). A-to-I events in exonic regions, instead, were very limited (Supplementary Figure S3). The RNA editing activity, measured by the AEI index (39), was largely body site specific as well as the activity at recoding sites (Supplementary Figures S4 and S5).

Web interface

The novel REDIportal inherits the layout from the previous version and all web pages are developed in Bootstrap, CSS and JavaScript. Server side operations to query MySQL tables and retrieve data are performed in Python (v2.7) 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.

REDIportal now allows two main searches, at position level and sample level. Users can interrogate the database in the conventional way by providing a genomic region, in the format chr:start-end, or a specific gene symbol, or query the collection of samples by providing one or more run accessions and limiting results according to tunable AEI or ADAR expression values. The search dropdown menu now includes the Gene View page by which users can visualize RNA editing events in their genic context, zooming on specific gene regions if needed (Figure 2). Retrieved sites and samples are shown in dynamic and sortable tables automat-

ically generated by DataTables in server-side mode to easily handle millions of rows.

The sample search allows the browsing of main RNA editing statistics per sample and includes five panels (Figure 3): (i) ‘Genomics Facts’ with the location and distribution of detected sites in different genomic and genic regions; (ii) ‘Base Distribution’ with the graphical distribution of all detected variants (not limited to A-to-G or T-to-C); (iii) ‘RNA Editing Indices’ with box plots of AEI and REI indices calculated on the body site group of the retrieved sample; (iv) ‘RNA Editing Levels’ with the distribution of RNA editing levels; (v) ‘Transcriptome Coverage’ with statistics about the fraction of edited and unedited genes, and the distribution of detected events in mRNAs or ncRNAs.

Additionally, the ‘RNA Editing Indices’ panel as well as the ‘Transcriptome Coverage’ panel provide external buttons that enable the recovery of further details. The button ‘REI details’ allows the browsing and filtering of known RNA editing recoding events (Figure 4) and some visualization facilities for all or selected sites only. The button ‘Gene details’, instead, retrieves the list of edited genes per sample and enable the gene view for each gene.

RNA editing sites can also be retrieved by a dedicated Application Programming Interface (API). Differently from the web-browsing approach, API does not leverage as much bandwidth and can also be interfaced with third-party programs able to manage outputs in JSON format. API outputs can be displayed in the web-browser or in the commands shell.

CONCLUSION AND FUTURE PLANS

RNA editing is emerging as a relevant epitranscriptome phenomenon involved in a variety of cellular functions. Al-

though the massive transcriptome sequencing has accelerated the research in this field, different aspects of the RNA editing biology need to be elucidated. To facilitate the investigation of A-to-I RNA editing we have updated our just rich REDIPortal catalogue distributing to the scientific community >15 millions of A-to-I changes in 9642 human RNAseq samples from the GTEx project. In contrast with the previous release in which RNA editing events were called from a known list of A-to-I changes from only six tissues (18 RNAseq experiments) (41), the version presented here includes editing candidates identified in all GTEx RNAseq data by applying a dedicated pipeline based on our HPC version of the REDIttools software and consuming ~30 million CPU hours through PRACE (Partnership for Advanced Computing in Europe) projects.

The current REDIPortal release allows searches at sample level, provides overviews of RNA editing profiles per each RNAseq experiment, implements a Gene View module to look at individual events in their genic context, hosts the CLAIRE resource (28) and collects non-human RNA editing changes.

As a comprehensive catalogue of A-to-I RNA editing, REDIPortal is in active development and thanks to two PRACE European projects for computing resources, its short-term goal will be the inclusion of A-to-I sites from all TCGA samples as well as events from RNAseq data produced in consortia for neurological and neurodegenerative disorders (CommonMind and PsychEncode) or RNAseq experiments from the Nonhuman Primate Reference Transcriptome Resource (NHPTR) (42). Finally, REDIPortal will be expanded with A-to-I changes detected in single cells.

DATA AVAILABILITY

REDIPortal is an open source database available through the web page <http://srv00.recas.ba.infn.it/>. REDIttools is an open source software available in the GitHub repository (<https://github.com/BioinfoUNIBA/REDIttools>).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We kindly 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 the Italian node for supporting the development and maintenance of REDIPortal. We additionally thank PRACE (Partnership for Advanced Computing in Europe) for providing computing resources for running REDIttools and Shalom Hillel Roth for advises on the use of the RNAEditingIndexer software.

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.v7.p2.

We acknowledge PRACE for awarding us access to Marconi at CINECA, Italy.

Finally we kindly thank Barbara De Marzo for technical support.

FUNDING

Elixir ITA, Elixir Converge, PRACE preparatory access [2010PA3441]; PRACE call 15 [2016163924]; PRACE call 18 [2018194670]. Funding for open access charge: Elixir-ITA.

Conflict of interest statement. None declared.

REFERENCES

- Gott, J.M. and Emeson, R.B. (2000) Functions and mechanisms of RNA editing. *Annu. Rev. Genet.*, **34**, 499–531.
- Eisenberg, E. and Levanon, E.Y. (2018) A-to-I RNA editing—immune protector and transcriptome diversifier. *Nat. Rev. Genet.*, **19**, 473–490.
- Savva, Y.A., Rieder, L.E. and Reenan, R.A. (2012) The ADAR protein family. *Genome Biol.*, **13**, 252.
- Pinto, Y., Cohen, H.Y. and Levanon, E.Y. (2014) Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. *Genome Biol.*, **15**, R5.
- Licht, K., Kapoor, U., Amman, F., Jakobi, T., 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.
- Kapoor, U., Licht, K., Amman, F., Jakobi, T., Martin, D., Dieterich, C. and Jantsch, M.F. (2020) ADAR-deficiency perturbs the global splicing landscape in mouse tissues. *Genome Res.*, **30**, 1107–1118.
- Mannion, N.M., Greenwood, S.M., Young, R., Cox, S., Brindle, J., Read, D., Nellaker, C., Vesely, C., Ponting, C.P., McLaughlin, P.J. *et al.* (2014) The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.*, **9**, 1482–1494.
- 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.
- 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.
- Hwang, T., Park, C.K., Leung, A.K., Gao, Y., Hyde, T.M., Kleinman, J.E., Rajpurohit, A., Tao, R., Shin, J.H. and Weinberger, D.R. (2016) Dynamic regulation of RNA editing in human brain development and disease. *Nat. Neurosci.*, **19**, 1093–1099.
- Roth, S.H., Danan-Gotthold, M., Ben-Izhak, M., Rechavi, G., Cohen, C.J., Louzoun, Y. and Levanon, E.Y. (2018) Increased RNA editing may provide a source for autoantigens in systemic lupus erythematosus. *Cell Rep.*, **23**, 50–57.
- Uchida, S. and Jones, S.P. (2018) RNA editing: unexplored opportunities in the cardiovascular system. *Circ. Res.*, **122**, 399–401.
- 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.
- Paz-Yaacov, N., Bazak, L., Buchumenski, I., Porath, H.T., Danan-Gotthold, M., Knisbacher, B.A., Eisenberg, E. and Levanon, E.Y. (2015) Elevated RNA editing activity is a major contributor to transcriptomic diversity in tumors. *Cell Rep.*, **13**, 267–276.
- Reardon, S. (2020) Step aside CRISPR, RNA editing is taking off. *Nature*, **578**, 24–27.
- Chew, W.L. (2018) Immunity to CRISPR Cas9 and Cas12a therapeutics. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **10**, e1408.
- Kosicki, M., Tomberg, K. and Bradley, A. (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*, **36**, 765–771.
- Bhakta, S. and Tsukahara, T. (2020) Artificial RNA editing with ADAR for gene therapy. *Curr. Gene Ther.*, **20**, 44–54.
- Qu, L., Yi, Z., Zhu, S., Wang, C., Cao, Z., Zhou, Z., Yuan, P., Yu, Y., Tian, F., Liu, Z. *et al.* (2019) Programmable RNA editing by recruiting

- endogenous ADAR using engineered RNAs. *Nat. Biotechnol.*, **37**, 1059–1069.
20. 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.
 21. Ramaswami,G. and Li,J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, **42**, D109–D113.
 22. 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.
 23. Tan,M.H., Li,Q., Shanmugam,R., Piskol,R., Kohler,J., Young,A.N., Liu,K.I., Zhang,R., Ramaswami,G., Ariyoshi,K. *et al.* (2017) Dynamic landscape and regulation of RNA editing in mammals. *Nature*, **550**, 249–254.
 24. Picardi,E. and Pesole,G. (2013) REDiTools: high-throughput RNA editing detection made easy. *Bioinformatics*, **29**, 1813–1814.
 25. 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.
 26. 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.
 27. Porath,H.T., Carmi,S. and Levanon,E.Y. (2014) A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.*, **5**, 4726.
 28. Schaffer,A.A., Kopel,E., Hendel,A., Picardi,E., Levanon,E.Y. and Eisenberg,E. (2020) The cell line A-to-I RNA editing catalogue. *Nucleic Acids Res.*, **48**, 5849–5858.
 29. Mailman,M.D., Feolo,M., Jin,Y., Kimura,M., Tryka,K., Bagoutdinov,R., Hao,L., Kiang,A., Paschall,J., Phan,L. *et al.* (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.
 30. 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.
 31. Frankish,A., Diekhans,M., Ferreira,A.-M., Johnson,R., Jungreis,I., Loveland,J., Mudge,J.M., Sisu,C., Wright,J., Armstrong,J. *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, **47**, D766–D773.
 32. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
 33. 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**, 2078–2079.
 34. Wang,K., Li,M. and Hakonarson,H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
 35. Sherry,S.T., Ward,M.-H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
 36. O'Leary,N.A., Wright,M.W., Brister,J.R., Ciuffo,S., Haddad,D., McVeigh,R., Rajput,B., Robbertse,B., Smith-White,B., Ako-Adjei,D. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
 37. Lee,C.M., Barber,G.P., Casper,J., Clawson,H., Diekhans,M., Gonzalez,J.N., Hinrichs,A.S., Lee,B.T., Nassar,L.R., Powell,C.C. *et al.* (2020) UCSC Genome Browser enters 20th year. *Nucleic Acids Res.*, **48**, D756–D761.
 38. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
 39. 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.
 40. 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.
 41. 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.
 42. Peng,X., Thierry-Mieg,J., Thierry-Mieg,D., Nishida,A., Pipes,L., Bozinovski,M., Thomas,M.J., Kelly,S., Weiss,J.M., Raveendran,M. *et al.* (2015) Tissue-specific transcriptome sequencing analysis expands the non-human primate reference transcriptome resource (NHPRTR). *Nucleic Acids Res.*, **43**, D737–D742.