Bioinformatics, 32(14), 2016, 2213–2215
doi: 10.1093/bioinformatics/btw204
Advance Access Publication Date: 19 April 2016
Applications Note

OXFORD

Structural bioinformatics

e23D: database and visualization of A-to-I RNA editing sites mapped to 3D protein structures

Oz Solomon[1,2,*,†], Eran Eyal[1,*,†], Ninette Amariglio[1,2], Ron Unger[2] and Gidi Rechavi[1,3]

[1]Cancer Research Center, Sheba Medical Center, Ramat-Gan, Israel, [2]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel and [3]Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Associate Editor: Anna Tramontano

Received on December 11, 2015; revised on March 21, 2016; accepted on April 9, 2016

Abstract

Summary: e23D, a database of A-to-I RNA editing sites from human, mouse and fly mapped to evolutionary related protein 3D structures, is presented. Genomic coordinates of A-to-I RNA editing sites are converted to protein coordinates and mapped onto 3D structures from PDB or theoretical models from ModBase. e23D allows visualization of the protein structure, modeling of recoding events and orientation of the editing with respect to nearby genomic functional sites from databases of disease causing mutations and genomic polymorphism.
Availability and Implementation: http://www.sheba-cancer.org.il/e23D
Contact: oz.solomon@live.biu.ac.il or Eran.Eyal@sheba.health.gov.il

1 Introduction

A-to-I RNA editing, the deamination of adenosine to inosine by ADAR family of enzymes, is the most widespread modification in the human transcriptome (Levanon *et al.*, 2004). Most of the RNA editing sites are located in non-coding regions such as introns and 3'UTRs. Still, several RNA editing sites are found in coding regions and likely change amino-acid sequence, as inosine is identified as guanosine by the translation machinery (Nishikura, 2009). These recoding sites were proved to be important for normal development. Adar2 Knock-out mice suffered from epileptic seizures and died at a very early age as a result of lack in editing, specifically at the Q-to-R site in Gria2. Changes in editing level at recoding sites in the serotonin receptor 2C (HTR2C) were found to be related to depression, schizophrenia and suicide (Iwamoto *et al.*, 2009). Recoding event from S-to-G at AZIN1 was found to result in hepatocellular carcinoma (Chen *et al.*, 2013). Recently, it was found that most of the non-synonymous editing sites in human seem to have a deleterious effect (Xu and Zhang, 2014). Most edited sites tend to have strong destabilization effect on the protein structure (Solomon *et al.*, 2014). Still, some coding RNA editing sites have a functional rule, are edited in a significant level and located at critical sites of the protein structure.

Several comprehensive databases which gather genomic data, evolutionary context, expression data and editing level for editing sites were recently developed. These include DARNED (Kiran *et al.*, 2013), RADAR (Ramaswami and Li, 2014) and Inosinome atlas (Picardi *et al.*, 2015). Still, there is a need for computational tools for convenient intersection of RNA editing information with protein structural and functional information. The current gap impedes most biologists working in RNA editing field from inferring the functional significance of editing event in the context of protein 3D structural data. Here, we present e23D, a tool which is intended to bridge this gap, and provide the community with user-friendly database and web interface to spot and visualize protein modifications resulted from A-to-I RNA editing events.

2 e23D database

e23D contains information regarding millions of editing sites collected from RADAR database. Currently we store RADAR data

regarding editing sites in human, mouse and fly (2,576,459 human sites, 8823 mouse sites and 5025 fly sites). Human RNA editing sites were also downloaded from the inosinome atlas (2,999,833 human sites). The entire information is stored as a rational database on MySQL server. For each organism (human, mouse and fly) a table with all editing sites, their read coverage, editing level and additional information is maintained. This data architecture allows relatively quick searches and is used for initial query of editing site information in the e23D front page. As shown in Figure 1A, UCSC gene annotations are used to connect the stored genomic coordinates with transcripts and protein coordinates. Protein sequences are aligned to PDB sequences using Blast. S2C (http://dunbrack.fccc.edu/Guoli/s2c/) is used to convert SEQRES with ATOM coordinates in the PDB files. e23D will be updated routinely following updates of the source editing sites databases.

3 e23D interface

At the e23D front page (http://www.sheba-cancer.org.il/e23D) the user can select between different organisms: human, mouse or fly (Fig. 1B). Editing sites can be searched according to gene symbol, genomic interval and number of reads that cover the editing sites or editing level. Alternatively, the user can choose to view all the coding editing sites. In the following page a table with editing sites matching the search criteria, is retrieved (Fig. 1C). By selecting one site, all PDB structures which satisfactory aligned to its protein are retrieved. The user can next filter the alignments between the protein sequence in Uniprot and sequences from PDB which are listed, by different options. One of the most important features of e23D is the ability to interactively explore the structural context of the amino acid change. The user can visualize each aligned structure in a JSmol session (Fig. 1D), in which all editing sites in this protein are colored. For human editing site, variants from dbSNP, ClinVar and Cosmic can be identified (Fig. 1D). It was recently suggested that RNA editing contributes to tumor heterogeneity in addition to somatic mutations (Paz-Yaacov *et al.*, 2015). Specifically, COPA I-to-V site was found to be hypo-edited in multiple cancers (Chan *et al.*, 2014; Paz-Yaacov *et al.*, 2015). Interestingly, this same I-to-V change in COPA is included in the Catalogue of Somatic Mutations in Cancer (Cosmic) (Fig. 1C) and additional somatic mutations are found in close proximity to this editing site (Fig. 1D). These examples highlight the value of showing both disease causing mutations and RNA editing sites together. Model of the new amino acid introduced by the editing can be shown. A list of contacts with the original or the modified amino-acid can also be accessed from this page. The user can also view the alignment between the query sequence and the actual sequence of the 3D structure (Fig. 1E). This alignment allows close examination of the sequence-structure relations alongside additional layers of information. Residues can be selected on the sequence and viewed on the structure allowing convenient exploration of the protein. Conservation scores of each amino-acid, disorder profiles and prosite motifs are shown below

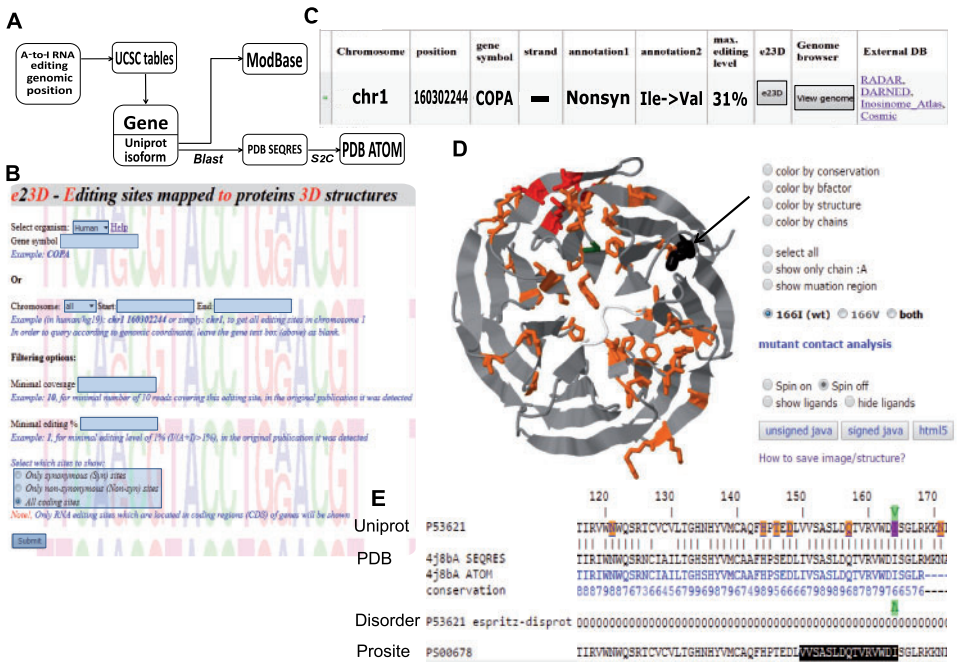

Fig. 1. (**A**) e23D pipeline. The UCSC gene annotation is used in order to convert between genomic coordinates to transcript and protein coordinates. Alignment of the protein sequence to all sequences in the PDB (PDB SEQRES) is done using BLAST. SEQRES coordinates are converted into PDB ATOM coordinates using S2C (http://dunbrack.fccc.edu/Guoli/s2c/). (**B**) e23D front page. The user can select which organism to use. Editing sites can be searched within requested genes or by providing genomic interval. Editing sites can be filtered by various criteria. (**C**) Editing sites information table. Editing sites chosen in previous step (B) are detailed in a table. Links to alignments to PDB ('e23D') or to genome browser ('View genome'), are provided. Links to external DB are also included. COPA I-to-V site is used here as an example. (**D**) Visualization of COPA 3D structure (PDB id: 4j8b, chain A). The position of the recoded residue within the structure is colored in black and marked with an arrow. Cancer mutations from Cosmic database are colored in orange. Variants from ClinVar are colored in red. Multiple options are available on the right for visualization and modeling of the edited structure. (**E**) Alignment between COPA protein sequence (Uniprot id: P53621) to sequence in PDB (PDB id: 4j8b, chain A). Cosmic variants are marked with orange background, SNPs are marked with green background, and editing sites are marked with purple background. The input editing site (here COPA I-to-V site) is marked also with green arrow above and below the alignment. Conservation scores, Prosite motifs and disordered regions are also shown below the alignment

the alignment. Detailed documentation pages are provided in the e23D website and are available for download.

To the best of our knowledge, e23D is the first database which integrates A-to-I RNA editing sites with available protein 3D structure information.

We hope that the availability of this tool will progress the use of structural information by the RNA editing community and make coding RNA editing research more effective.

Acknowledgements

The work of O.S. was done in partial fulfillment with the requirements of the Faculty of Life-Sciences, Bar-Ilan University, Israel. We thank the ModBase team for the helping us integrating ModBase models. We thank Ernesto Picardi for the help integrating inosinome atlas. G.R. is a member of the Sagol Neuroscience Network and holds the Djerassi Chair for Oncology at the Sackler Faculty of Medicine, Tel-Aviv University, Israel.

Funding

The work was supported by grants from the Varda and Boaz Dotan Research Center in Hemato-oncology affiliated with the CBRC of Tel Aviv University, Flight Attendant Medical Research Institute (FAMRI), Israeli Centers of Excellence (I-CORE) Program (ISF grants no. 41/11 and no. 1796/12).

Conflict of Interest: none declared.

References

Chan,T.H. *et al.* (2014) A disrupted RNA editing balance mediated by ADARs (Adenosine DeAminases that act on RNA) in human hepatocellular carcinoma. *Gut*, **63**, 832–843.

Chen,L. *et al.* (2013) Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat. Med.*, **19**, 209–216.

Iwamoto,K. *et al.* (2009) Serotonin receptor 2C and mental disorders: genetic, expression and RNA editing studies. *RNA Biol.*, **6**, 248–253.

Kiran,A.M. *et al.* (2013) Darned in 2013: inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res.*, **41**, D258–D261.

Levanon,E.Y. *et al.* (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.*, **22**, 1001–1005.

Nishikura,K. (2009) Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.*, **79**, 321–349.

Paz-Yaacov,N. *et al.* (2015) Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Rep.*, **13**, 267–276.

Picardi,E. *et al.* (2015) Profiling RNA editing in human tissues: towards the inosinome Atlas. *Sci Rep*, **5**, 14941.

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

Solomon,O. *et al.* (2014) Characterizing of functional human coding RNA editing from evolutionary, structural, and dynamic perspectives. *Proteins*, **82**, 3117–3131.

Xu,G. and Zhang,J. (2014) Human coding RNA editing is generally nonadaptive. *Proc. Natl. Acad. Sci. USA*, **111**, 3769–3774.