

VIRES: visualization and identification of A-to-I RNA editing sites in genomic sequences

R. Distefano¹, G. Nigita¹, V. Macca¹, R. Giugno², A. Pulvirenti²✉, A. Ferro²

¹Dipartimento di Matematica e Informatica Università degli Studi di Catania, Catania, Italy

²Dipartimento di Biomedicina Clinica e Molecolare, Università degli Studi di Catania, Italy

Motivations

RNA Editing is a type of post-transcriptional modification that takes place in the eukaryotes and represents one of the last frontiers of molecular biology. It alters the sequence of primary RNA transcripts by deleting, inserting or modifying residues. Several forms of RNA editing have been discovered including A-to-I, C-to-U, U-to-C and G-to-A editing. A-to-I RNA editing (Adenosine-to-Inosine) is the most frequent and common post-transcriptional modification, where adenosine (A) deamination produces its conversion into inosine (I), which in turn is interpreted by the machinery translation and splicing as guanosine (G) and so this causes the change of the RNA sequence. This biological phenomenon is catalyzed by members of the Adenosine Deaminase Acting on RNA (ADAR) family of enzymes and occurs only on dsRNA structures. Thus, A-to-I editing changes RNA molecules in various ways including: the translation of its codons; the creation and/or destruction of splicing sites; the micro-RNA/mRNA binding. Therefore, it is not surprising that the malfunction of the editing machinery has been implicated in various human diseases. In the last years, the application of global approaches to the study of A-to-I editing, including high throughput sequencing and bioinformatics, has led to important advances. However, in spite of enormous efforts, the real biological functioning of this phenomenon remains unknown. In this work, starting from genomic sequences, given as input, we present a bioinformatics approach to discover and visualize A-to-I editing sites.

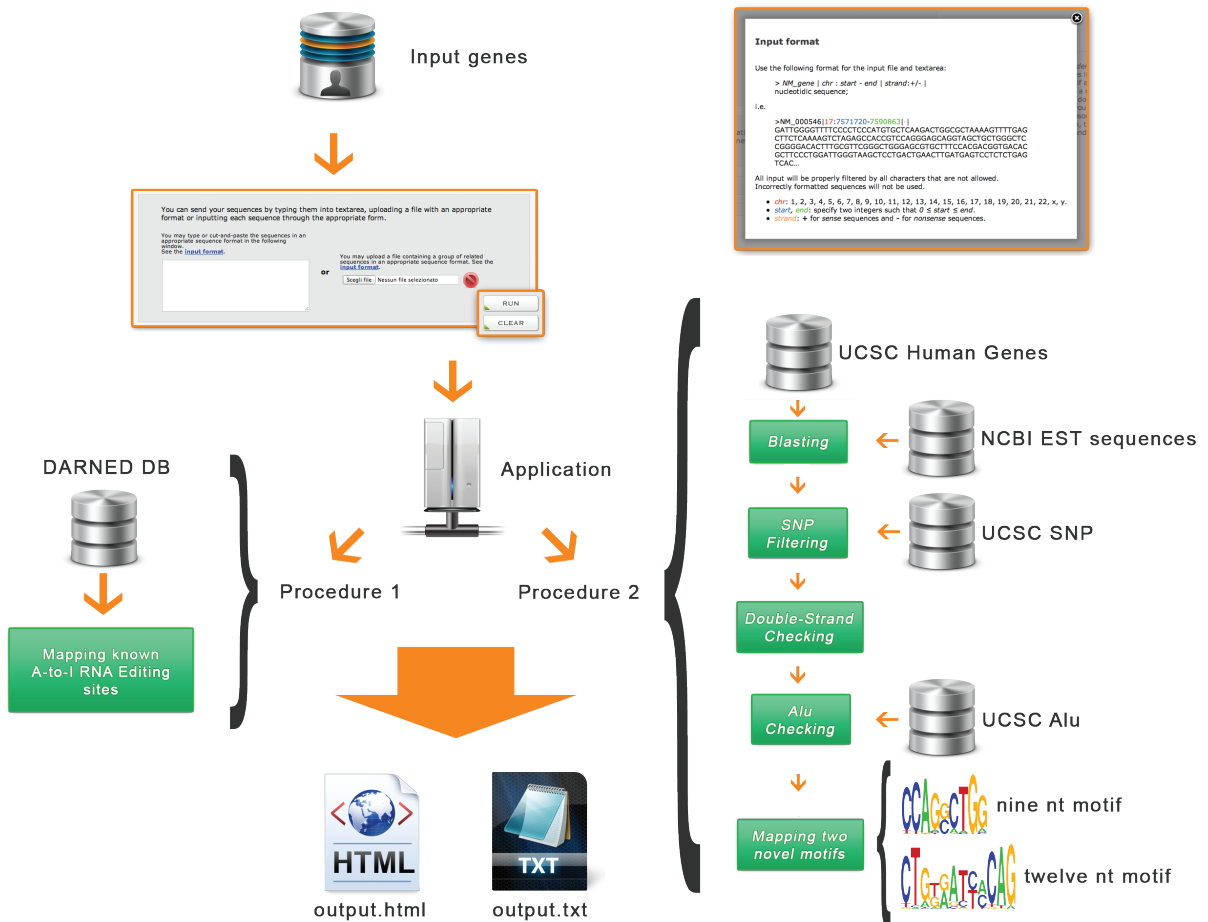
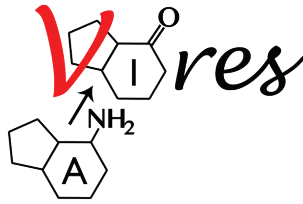
Methods

VIRES is a web-based tool that maps newly predicted and known A-to-I editing site in genomic sequences. The tool is equipped with a user-friendly interface allowing users to analyze

genomic sequences in order to identify candidate A-to-I editing sites. VIRES action can be subdivided in two different tasks: the identification of known editing sites and the mapping of new ones. The system highlights the known editing events falling into the input sequences by searching the genomic positions of sequences in the DARNED dataset containing approximately 42,000 human genome loci corresponding to validated A-to-I RNA editing sites. In the second phase, we search for new putative editing events in these sequences. This task is performed by analyzing more than 38,000 Human genes (build GRCh37/hg19). In more details, the searching of new editing sites can be divided into the following steps. First, we execute BLAST between Human genes and EST (Expressed Sequence Tag) sequences selecting only the alignments which contains at least one A-G mismatch between genomic and EST sequence. Next, we remove all mismatches that are SNPs (Single Nucleotide Polymorphisms). Finally, we verify if this candidate edited site belongs to a double strand region. Once we identify the candidate A-to-I editing sites, we add the information for the functional enrichment including the presence of repetitive elements, the ESTs sequences with the candidate editing sites, and the location of two novel motifs characterizing A-to-I editing events (CCAG[G|C]CTGG and CTG[T|G][G|A]AT[C|T][A|C]CAG) in flanking regions of putative editing sites. The user can choose whether to download this information in a text or xml file.

Results

VIRES is an easy to use bioinformatic tool that allows the in-silico-identification of putative and validated A-to-I RNA editing sites. It is built on top of several knowledge bases such as DARNED, EST, SNPs.



Below, you can see results obtained by analyzing input sequences.

Seq #1	Gene name	Chr	Start (bp)	End (bp)	Strand
88401	1	1	124095	124094	-
88402	2	1	124095	124094	-
88403	3	1	124095	124094	-
88404	4	1	124095	124094	-
88405	5	1	124095	124094	-
88406	6	1	124095	124094	-
88407	7	1	124095	124094	-
88408	8	1	124095	124094	-
88409	9	1	124095	124094	-
88410	10	1	124095	124094	-
88411	11	1	124095	124094	-
88412	12	1	124095	124094	-
88413	13	1	124095	124094	-
88414	14	1	124095	124094	-
88415	15	1	124095	124094	-
88416	16	1	124095	124094	-
88417	17	1	124095	124094	-
88418	18	1	124095	124094	-
88419	19	1	124095	124094	-
88420	20	1	124095	124094	-
88421	21	1	124095	124094	-
88422	22	1	124095	124094	-
88423	23	1	124095	124094	-
88424	24	1	124095	124094	-
88425	25	1	124095	124094	-
88426	26	1	124095	124094	-
88427	27	1	124095	124094	-
88428	28	1	124095	124094	-
88429	29	1	124095	124094	-
88430	30	1	124095	124094	-

Chromosome	Strand	Double Strand	Absolute position	Relative position
chr1	-	993	124094	9613

Motives (the nearest motives to RNA Editing site)	Start (bp)	End (bp)	Strand	Chromosome
9motif	124094	124095	-	chr1
12motif	124094	124095	-	chr1

EST sequences	Repetitive Regions																									
<table border="1"> <thead> <tr> <th>Gene</th> <th>Range (start-end)</th> <th>Name</th> <th>Class</th> <th>Family</th> </tr> </thead> <tbody> <tr> <td>00222773</td> <td>124094</td> <td>h1</td> <td>EST</td> <td>h1</td> </tr> <tr> <td>00322266</td> <td>124094</td> <td>h1</td> <td>EST</td> <td>h1</td> </tr> </tbody> </table>	Gene	Range (start-end)	Name	Class	Family	00222773	124094	h1	EST	h1	00322266	124094	h1	EST	h1	<table border="1"> <thead> <tr> <th>Gene</th> <th>Range (start-end)</th> <th>Name</th> <th>Class</th> <th>Family</th> </tr> </thead> <tbody> <tr> <td>124094</td> <td>124094</td> <td>h1</td> <td>EST</td> <td>h1</td> </tr> </tbody> </table>	Gene	Range (start-end)	Name	Class	Family	124094	124094	h1	EST	h1
Gene	Range (start-end)	Name	Class	Family																						
00222773	124094	h1	EST	h1																						
00322266	124094	h1	EST	h1																						
Gene	Range (start-end)	Name	Class	Family																						
124094	124094	h1	EST	h1																						

A fragment of output.html page

Figure 1. Vires usage workflow and architecture.