

De novo detection of A-to-I RNA editing sites in human mRNAs by massive transcriptome sequencing

E. Picardi^{1,2}✉, A. Gallo³, S. Raho³, F. Galeano³, G. Pesole¹

¹Dipartimento di Bioscienze, Biotecnologie e Scienze Farmacologiche, Università di Bari A.Morō Bari, Italy

²Istituto Biomembrane e Bioenergetica del Consiglio Nazionale delle Ricerche, Bari, Italy

³Laboratorio di ricerca RNA editing, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy

Motivations

RNA editing is a widespread molecular phenomenon which modifies primary transcripts at specific positions [1]. It occurs in a variety of organisms including human and cooperates with alternative splicing in increasing both proteomic and transcriptomic complexity. RNA Editing can modulate gene expression and affect protein functionality. In human, such phenomenon is highly frequent in brain and its deregulation has been linked to a variety of neurological and neurodegenerative diseases [2]. Many editing events have been identified by next generation sequencing technologies employing massive transcriptome sequencing [3] together with whole genome or exome sequencing. Nowadays numerous RNA-Seq experiments are available through public databases and represent a relevant source of yet unexplored RNA editing sites. Hereafter we propose a simple computational strategy to identify de novo genomic positions enriched in novel potential RNA editing events through a new, two-step mapping procedure.

Methods

To accurately predict RNA editing sites we developed a double mapping procedure in which millions of Illumina short reads were independently mapped onto the human transcriptome and the reference human genome tolerating at maximum two mismatches for each unique alignment. Only concordant alignments from the double mapping procedure were used for downstream analyses. All reads supporting each reference position were explored to calculate the empirical probability to observe a substitution. Such probabilities were then used to detect statistically significant base conversions by applying the Fisher's exact test by comparing the observed and expected occurrences in the aligned reads. Benjamini-Hochberg correction was finally employed to reduce the false discovery rate. A-to-I RNA editing candidates may be then selected according to P-value, coverage

and editing extent for the experimental validation using the classical Sanger sequencing from RNA/DNA extracted from the same individual.

Results

We initially tested our computational method on the SRA study SRA012427 involving high-throughput transcriptome sequencing of human brain tissues by Illumina technology. Over 22 millions 50 nt long paired-end reads were aligned onto the human reference genome (assembly hg18). Applying the above-described double mapping methodology and stringent filters we found 19 highly significant A-to-I conversions in known human coding regions. Interestingly, 11 of such changes have been already described in literature and 6 were experimentally confirmed. To further corroborate our strategy we carried out an RNA-Seq experiment on total RNA extracted from human spinal cord. More than 20 million of directional paired-end reads were analysed using the above mentioned procedure. At 0.05 significance level (FDR corrected) we obtained 15 RNA editing candidates covered by at least 30 independent reads and showing only A-to-G changes. In this case potential editing candidates were confirmed by whole exome sequencing performed on the individual and tissue in order to optimally exclude SNPs and somatic mutations. Notably, we were able to confirm 12 predicted candidates. Our results, therefore, indicate the feasibility and effectiveness of the above-described strategy to detect de novo A-to-I RNA editing events in human.

References

1. Gott, J.M. and Emeson, R. (2000) Functions and mechanisms of RNA editing. *Annual review of genetics*, 34, 499-531.
2. Maas, S., Kawahara, Y., Tamburro, K.M. and Nishikura, K. (2006) A-to-I RNA editing and human disease. *RNA biology*, 1-9.
3. Picardi, E., Horner, D.S., Chiara, M., Schiavon, R., Valle, G. and Pesole, G. (2010) Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic acids research*, 38, 4755-4767.