De Novo A-to-I RNA Editing Detection in Human mRNAs by RNA-Seq Data: Towards the Inosinome?

Picardi E (1), Gallo A (2), Raho S (2), Galeano F (2), Pesole G (1, 3)

(1) Dipartimento di Biochimica e Biologia Molecolare “E. Quagliariello”
Università degli Studi di Bari, Bari, Italy
(2) Laboratorio di ricerca RNA editing, - Ospedale Pediatrico Bambino Gesù, IRCCS, Roma, Italy
(3) Istituto Biomembrane e Bioenergetica del Consiglio Nazionale delle Ricerche, Bari, Italy

Motivation
RNA editing is a post-transcriptional process occurring in a wide range of organisms including prokaryotes, plants, viruses and animals (1). In human, the A-to-I RNA editing, in which individual adenosine (A) bases in pre-mRNA are modified to yield inosine (I), is the most frequent event (1). Such conversion is due to members of ADAR (adenosine deaminase acting on RNA) family of enzymes. ADAR1 and ADAR2 have been detected in many human tissues whereas ADAR3 is expressed specifically in brain (1). Since inosine is read as guanosine during translation, A-to-I conversion in coding sequences can lead to amino acid changes and protein function alterations. RNA editing also occurs in non-coding RNAs and especially in Alu repeat elements (1). A peculiarity of editing is that both the edited and unedited versions of affected transcripts are co-expressed in the same cell and the ratio between the unedited and edited variants can be regulated by a variety of factors depending on tissue type or developmental stage. RNA editing is very frequent in human brain and it is essential for cellular homeostasis. Indeed, its deregulation has been linked to several diseases such as epilepsy, schizophrenia, amyotrophic later sclerosis and cancer (2). To date thousand sites have been identified by computational approaches employing mRNA/EST alignments onto the genome of origin. Recent technologies for massive RNA sequencing are improving the study of entire eukaryotic transcriptomes as well as post-transcriptional modifications occurring herein (3). RNA-Seq data can be extremely informative to investigate editing sites in a variety of experimental conditions. However the detection of new genuine RNA candidates can be optimally performed using genomic reads selected from the same individual in order to exclude errors due to SNPs. Although sequencing costs are going down, the concomitant sequencing of transcriptome and genome from the same individual is yet an expensive solution. However a possible suitable alternative is proposed, in which, a simple computational strategy is employed to predict de novo A-to-I RNA editing sites in human RNA-Seq experiments.
Methods
To accurately predict RNA editing sites we first map millions of Illumina short reads onto the reference human genome tolerating at maximum two mismatches for each unique alignment. Next, we explore all reads supporting each reference position and calculate the empirical probability to observe a substitution. Such probabilities are 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. Bonferroni or Benjamini-Hochberg correction is 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 a classical Sanger sequencing from RNA/DNA extracted from the same individual.

Results
We 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 methodology we found 35 highly significant A-to-I conversions in known human coding regions. Interestingly, 14 of such changes have been already described in literature and experimentally confirmed. Moreover, the A-to-I variation in 12 of remaining positions is supported by public ESTs/mRNAs. To confirm the reliability of our detections we selected several sites for independent validation by Sanger sequencing. Preliminary results show clear editing evidence for new candidates that have never been characterized. Looking at potential events falling in mRNA untranslated regions, we found 119 A-to-I conversions in 3’ UTRs and only 1 change in 5’ UTRs. Approximately 50% of our non-coding predictions are known in literature and 75% fall in Alu elements. Our results, therefore, clearly demonstrate the feasibility and effectiveness of the above-described strategy to detect de novo A-to-I RNA editing events in human.

References

Contact email
graziano.pesole@biologia.uniba.ut