Reviewer's report

Title: Dark Matter RNA Illuminates the Puzzle of Genome-Wide Association Studies

Version:2**Date:**2 May 2014

Reviewer: Eduardo Reis

Reviewer's report:

In this well-written and timely short review, the authors elaborate the argument that disease-associate sequence variants identified by GWAS studies that lie in unannotated genomic regions may exert their effects by targeting regulatory long noncoding RNAs.

The starting point is the observation that most GWAS variants lie outside the coordinates of known genes. While only a modest fraction of these are occupied by DNA regulatory elements, it is conceivable that a significant fraction may overlap with ncRNAs pervasively transcribed across the human genome that have been detected at increasing number in the last decade. In spite of the ongoing debate regarding the extent to which the noncoding transcriptome is functionally relevant, the few lncRNAs that have been characterized in detail have revealed novel mechanisms of gene regulation that affect basic cellular processes.

To demonstrate the possible association between disease causing GWAS variants and lncRNAs the authors provide an original and interesting analysis based on published data showing that GWAS variants associated to specific disease types (mental health, cellular proliferation) have a bias toward genomic locations occupied by introns or long ncRNAs.

The few known examples implicating disease associated GWAS SNPs with the perturbation of the function of lncRNAs are discussed, highlighting the importance to develop more sensitive and scalable methods to probe the functionality of lncRNAs by genetic approaches in order to establish functional associations between the two.

However, I would disagree that there is consistent evidence to support the claim that complex diseases are better explained by mutations in ncRNAs than in protein-coding mRNAs (pg. 7). The ability of GWAS studies to reveal the genetic basis of complex diseases (e.g. cancer, neurological) has been reduced by limited sample sizes, which is likely to affect the identification of both coding and noncoding variants.

Minor Compulsory Revisions

To allow reproducibility, the analysis shown in Figure 1 and discussed in the text should be accompanied by more detailed information regarding the primary data

used (what was the GWAS catalog?) and the analytic method (What is the impact of counting the same variant multiple times if it appears in more than one publication? Could it introduce biases?).

Discretionary Revisions

Pg. 2: It would be nice to mention what kind of errors are potentially involved in GWAS studies and/or provide appropriate references to this issue.

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests