

# RNA editing study under intense scrutiny

Debate highlights pitfalls in interpreting genomic data.

Erika Check Hayden

15 March 2012

Three groups<sup>1–3</sup> today throw fuel on the debate surrounding a researcher’s claim to have discovered a twist in the mechanism whereby genes are translated into proteins. The paper, published in *Science* last May<sup>4</sup>, suggested a revision in the central dogma, which holds that the RNA transcripts used as templates for protein assembly are generally faithful matches to the original DNA. In technical commentaries published today in *Science*, the groups suggest that errors in multiple aspects of the study, led by Vivian Cheung of the University of Pennsylvania in Philadelphia, seriously undermine its claim to have found a new mechanism of genetic regulation.

Cheung and her group reported that they had found 10,210 sites in the human genome where an RNA sequence did not match the sequence of the DNA from which it had presumably been transcribed. The researchers said that their results provided evidence for the existence of a new mechanism of RNA editing, in which transcribed RNA is altered to yield proteins that do not exactly match the underlying DNA sequence. Scientists already know that this occurs in some cases, but the Cheung paper proposed that much more RNA editing was occurring than was previously known (see [Evidence of altered RNA stirs debate](#)).

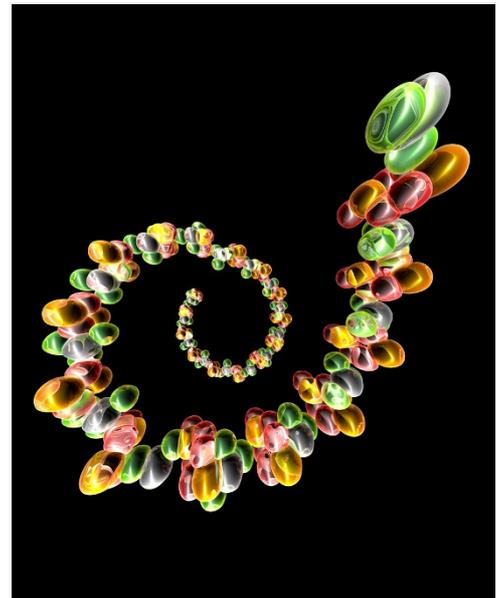
Today, the three groups estimate that up to 94% of the putative RNA-editing sites identified in Cheung’s paper are wrong. The groups, which worked independently, say that multiple sources of error contributed to the original paper’s overestimate of ‘RNA–DNA differences’ (RDDs). Other researchers had previously criticized Cheung’s findings<sup>5</sup> (see [‘Estimates of errors’](#)).

In their comment<sup>3</sup>, genomicists Claudia Kleinman and Jacek Majewski of McGill University in Montreal, Canada, write: “We provide evidence that the authors overestimate the frequency of RDDs by an order of magnitude. In view of the above shortcomings, we question some of the boldest findings of this work.”

Cheung is defending her findings and has conducted new analyses to support the work.

“We are certain that there are many more mismatches between RNA and their underlying DNA sequences than previously known,” Cheung’s group writes in its response, which is also published today in *Science*<sup>6</sup>.

The potential errors identified by the critics of Cheung’s work come from multiple sources.



PASIEKA/SCIENCE PHOTO LIBRARY

The idea that RNA is often edited after being transcribed from DNA is a controversial one.

## Estimates of errors

### Keeping track

Four groups, including three publishing today in *Science*, have now estimated that the vast majority of RNA–DNA differences reported by Cheung’s group were the result of error.

Study	Estimated rate of false positives	Possible sources of error
Lin <i>et al.</i> <sup>1</sup>	89%	Mapping errors, genetic variation, gene duplications
Pickrell <i>et al.</i> <sup>2</sup>	88–94%	Sequencing errors, mapping errors, genetic variation
Kleinman & Majewski <sup>3</sup>	68–90%	Sequencing errors, mapping errors, genotyping errors
Schrider <i>et al.</i> <sup>5</sup>	>90%	Gene duplications

All three groups say that the work is riddled with errors that arose during the process of sequencing RNA fragments and mapping the reads back to their corresponding stretches of DNA. Cheung's group sequenced RNA using machines, sold by Illumina of San Diego, California, that read the sequences of short fragments of genetic material.

Kleinman and Majewski and a group led by Jonathan Pritchard at the University of Chicago<sup>2</sup> found that the putative RNA–DNA differences often occurred in the ends of a sequencing read — sites where Illumina machines are most likely to make errors. Pritchard's group estimates that 87% of the RNA–DNA differences could be caused by such errors.

A third group, led by Jin Billy Li of Stanford University in California<sup>1</sup>, reports that 82% of the supposed mismatches could be caused by primers, short fragments of genetic material that attach to the ends of RNA to enable the sequencing process. Because the primers may not precisely match the RNA that they stick to, the resulting sequence could appear to represent RNA that differs from its corresponding DNA.

Other possible sources of error identified include natural human genetic variation and gene duplications. Many genes in the human genome exist in multiple, slightly modified versions, and RNA read out from one of these may be wrongly identified as having originated from a separate, slightly different copy of the gene, as has been argued in a previous paper<sup>5</sup>

Cheung's group addressed these concerns by using two different programs to align the RNA and DNA sequences and by sequencing two individuals from the previous study in more depth. They claim that both the new mapping programs and extra sequences identified many of the original RNA–DNA mismatches, and that there are reasons why the mismatches tend to cluster near the ends of sequence reads. In response, Cheung's group also cites other studies that have found a higher than expected rate of mismatches.

But Cheung's response does not address her critics' concerns — ones that they say also affect the subsequent studies. "There is a lot of room for improvement in the field," says Li.

In the meantime, critics of Cheung's work say that the ongoing argument is a lesson for anyone attempting to make use of large sets of sequencing data.

"These sequencing techniques are great, but you've really got to get to know and understand the biases and potential errors that accompany them," says Joseph Pickrell, an evolutionary genomicist at Harvard Medical School in Boston, Mass., who co-authored the comment with Pritchard.

In a post today on the blog [Genomes Unzipped](#), Pickrell opines that Cheung's study "should have been outright retracted", and elaborates: "By selecting for the most 'odd-looking' regions of the genome in an analysis, one enriches for strange and unexpected technical artefacts."

In other words: geneticists, proceed with caution.

*Nature* | doi:10.1038/nature.2012.10217

## References

---

1. Lin, W., Piskol, R., Tan, M. H. & Li, J. B. *Science* <http://dx.doi.org/science.1210624> (2012).
2. Pickrell, J. K., Gilad, Y. & Pritchard, J. K. *Science* <http://dx.doi.org/science.1210484> (2012).
3. Kleinman, C. L. & Majewski, J. *Science* <http://dx.doi.org/science.1209658> (2012).
4. Li, M. *et al. Science* **33**, 53–58 (2011).
5. Schrider, D. L., Gout, J.-F. & Hahn, M. W. *PLoS ONE* **6**, e25842 (2011).
6. Li, M., Wang, I. X. & Cheung, V. G. *Science* <http://dx.doi.org/science.1210419> (2012).