

# New level of genetic diversity in human RNA sequences uncovered

May 19 2011

---

A detailed comparison of DNA and RNA in human cells has uncovered a surprising number of cases where the corresponding sequences are not, as has long been assumed, identical. The RNA-DNA differences generate proteins that do not precisely match the genes that encode them.

The finding, published May 19, 2011, in *Science Express*, suggests that unknown cellular processes are acting on RNA to generate a sequence that is not an exact replica of the DNA from which it is copied. Vivian Cheung, the Howard Hughes Medical Institute investigator who led the study, says the RNA-DNA differences, which were found in the 27 individuals whose genetic sequences were analyzed, are a previously unrecognized source of genetic diversity that should be taken into account in future studies.

Genes have long been considered the [genetic blueprints](#) for all of the proteins in a cell. To produce a protein, a gene's DNA sequence is copied, or transcribed, into RNA. That RNA copy specifies which amino acids will be strung together to build the corresponding protein. "The idea that RNA and protein sequences are nearly identical to the corresponding [DNA sequences](#) is strongly held and has not been questioned in the past," says Cheung, whose lab is at the University of Pennsylvania School of Medicine.

With recent advances in sequencing technology, however, it has become possible to perform the kind of analysis necessary to test that

assumption. In their study, Cheung and her colleagues compared the sequences of DNA and RNA in [B cells](#) (a type of white blood cell) from 27 individuals. The [DNA sequences](#) they analyzed came from large, ongoing genomics projects, the International HapMap Project and the 1000 Genomes Project. They used high-throughput sequencing technology to sequence the RNA of B cells from the same individuals.

Within the sequences' protein-coding segments, they found 10,210 sites where RNA sequences were not the same as the corresponding DNA. They call these sites RNA-DNA differences, or RDDs. They found at least one RDD site in about 40 percent of genes, and many of these RDDs cause the cell to produce different [protein sequences](#) than would be expected based on the DNA. In the cells they studied, the sequences of thousands of proteins may be different from their corresponding DNA, the scientists say. "It is important to note that since these RDDs were found with just 27 individuals, they are common," Cheung points out.

To test whether the phenomenon was specific to B cells, the team also searched for RDDs in DNA and RNA sequences in human skin and brain cells. They found that most of the RDD sites occurred in at least some samples of all three cell types and were present in cells from both infants and adults, indicating that the RNA-DNA differences are not due to aging or specific to certain developmental stages.

Cheung says the particular RNA-DNA discrepancies they found appear systematic. There are four bases, or letters, that make up the DNA code: A, T, G, and C. The RNA equivalents are A, U, G, and C. In individuals who had RNA-DNA differences at a specific site in the genome, the mismatched bases were always the same. In other words, if the team found a C in the RNA sequence where they expected an A, all individuals who had an RDD at this point also had a C in their RNA sequence—never a G or a U. "Such uniformity makes us believe that

there is a 'code' or 'guide' that mediates the RDDs and they are not random events," Cheung explains.

In the 1980s, scientists found the first examples of RNA sequences that did not match the corresponding DNA. Today, many genes in humans and other organisms are now known to be targets of RNA editing. The known examples of such editing are mediated by enzymes called deaminases, which chemically modify specific As and Cs in the RNA sequence, converting the As to Gs and the Cs to Us. Cheung says abnormal RNA editing of glutamate and serotonin receptors has been associated with psychiatric disorders and resistance to certain drugs—evidence that traditional RNA editing is critical for maintaining normal cellular function.

Nearly half of the RDDs uncovered in the new study cannot be explained by the activity of deaminase enzymes, however, indicating that unknown processes must be modifying the RNA sequence, either during or after transcription. Cheung says there are several possibilities. For example, the DNA might be chemically or structurally modified so that certain bases look different to the enzyme that copies DNA to RNA, causing it to insert a mismatched RNA base during transcription. Alternatively, newly synthesized RNAs might be folded in such a way to signal enzymes to convert certain bases to other ones. The biological significance of these modifications remains to be determined, but since they are widespread among individuals and cell types, Cheung and her colleagues expect they have some function.

Although all of the individuals analyzed in the study had a large number of RDDs, there was a great deal of variability in the specific RDDs found in each person's genetic material. This variability likely contributes to differences in disease susceptibility, Cheung says. Scientists have generally searched for DNA sequence differences to explain why some people are more prone to certain diseases, whereas

studies of RNA and proteins have considered levels of expression, but not sequences. But major genetic contributors for many diseases remain unknown, and Cheung says it will be valuable to begin to include RNA sequences in disease-association studies.

Cheung notes that her team's analysis would not have been possible without the large-scale genomics projects, which until now have focused on DNA. "Without these large-scale genome projects, we would not have the volume of DNA sequences for comparisons and would not have the technologies that enabled us to sequence our RNA samples," she says.

"Our study provides support for why large-scale data are important. Previously the focus was on DNA, now our results suggest that RNA [sequences](#) also need to be examined. Exploration of these data, when founded on fundamental biology, will lead to fruitful scientific discoveries."

**More information:** Mingyao Li, et al. Widespread RNA and DNA Sequence Differences in the Human Transcriptome. *Science*, May 19, 2011 [DOI: 10.1126/science.1207018](https://doi.org/10.1126/science.1207018)

Provided by Howard Hughes Medical Institute

Citation: New level of genetic diversity in human RNA sequences uncovered (2011, May 19) retrieved 19 April 2024 from <https://phys.org/news/2011-05-genetic-diversity-human-rna-sequences.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--