Scientists identify seventh and eighth bases of DNA

July 21 2011

For decades, scientists have known that DNA consists of four basic units -- adenine, guanine, thymine and cytosine. Those four bases have been taught in science textbooks and have formed the basis of the growing knowledge regarding how genes code for life. Yet in recent history, scientists have expanded that list from four to six.

Now, with a finding published online in the July 21, 2011, issue of the journal *Science*, researchers from the UNC School of Medicine have discovered the seventh and eighth bases of DNA.

These last two bases – called 5-formylcytosine and 5 carboxylcytosine – are actually versions of cytosine that have been modified by Tet proteins, molecular entities thought to play a role in DNA demethylation and stem cell reprogramming.

Thus, the discovery could advance stem cell research by giving a glimpse into the DNA changes – such as the removal of chemical groups through demethylation – that could reprogram adult cells to make them act like stem cells.

"Before we can grasp the magnitude of this discovery, we have to figure out the function of these new bases," said senior study author Yi Zhang, Ph.D., Kenan Distinguished Professor of biochemistry and biophysics at UNC and an Investigator of the Howard Hughes Medical Institute. "Because these bases represent an intermediate state in the demethylation process, they could be important for cell fate

reprogramming and cancer, both of which involve DNA demethylation."

Much is known about the "fifth base," 5-methylcytosine, which arises when a chemical tag or methyl group is tacked onto a cytosine. This methylation is associated with gene silencing, as it causes the DNA's double helix to fold even tighter upon itself.

Last year, Zhang's group reported that Tet proteins can convert 5 methylC (the fifth base) to 5 hydroxymethylC (the sixth base) in the first of a four step reaction leading back to bare-boned cytosine. But try as they might, the researchers could not continue the reaction on to the seventh and eighth bases, called 5 formylC and 5 carboxyC.

The problem, they eventually found, was not that Tet wasn't taking that second and third step, it was that their experimental assay wasn't sensitive enough to detect it. Once they realized the limitations of the assay, they redesigned it and were in fact able to detect the two newest bases of DNA. The researchers then examined embryonic stem cells as well as mouse organs and found that both bases can be detected in genomic DNA.

The finding could have important implications for stem cell research, as it could provide researchers with new tools to erase previous methylation patterns to reprogram adult cells.

It could also inform cancer research, as it could give scientists the opportunity to reactivate tumor suppressor genes that had been silenced by DNA methylation.

Provided by University of North Carolina