

Learning the alphabet of gene control

January 17 2013

Scientists at Karolinska Institutet in Sweden have made a large step towards the understanding of how human genes are regulated. In a new study, published in the journal *Cell*, they identified the DNA sequences that bind to over four hundred proteins that control expression of genes. This knowledge is required for understanding of how differences in genomes of individuals affect their risk to develop disease.

After the [human genome](#) was sequenced in 2000, it was hoped that the knowledge of the entire sequence of [human DNA](#) could rapidly be translated to medical benefits such as [novel drugs](#), and predictive tools that would identify individuals at risk of disease. This however turned out to be harder than anticipated, one of the reasons being that only 1 percent of the genome that code for proteins was in fact possible to read. The remaining part, much of which describes how these proteins should be expressed in different cells and tissues, could not be understood. This, in turn, because the scientists did not know which [DNA sequences](#) are functional, and bind to the specific proteins called transcription factors that regulate gene expression.

"The genome is like a book written in a foreign language, we know the letters but cannot understand why a human genome makes a human or the [mouse genome](#) a mouse", says Professor Jussi Taipale, who led the study at the Department of Biosciences and Nutrition. "Why some individuals have higher risk to develop [common diseases](#) such as heart disease or cancer has been even less understood."

The human genome encodes approximately 1000 transcription factors,

and they bind specifically to short sequences of DNA, and control the production of other proteins. In the work published in *Cell*, the scientists at Karolinska Institutet describe DNA sequences that bind to over 400 such proteins, representing approximately half of all human transcription factors. Data was generated with a new method that uses a modern DNA sequencer that produces hundreds of millions of sequences, giving the results unprecedented accuracy and reliability.

In addition, binding specificities of human [transcription factors](#) were compared to those of the mouse. Surprisingly, no differences were found. According to the scientists, these results suggest that the basic machinery of gene expression is similar in humans and mice, and that the differences in size and shape are caused not by differences in transcription factor proteins, but by presence or absence of the specific sequences that bind to them.

"Taken together, the work represents a large step towards deciphering the code that controls gene expression, and provides an invaluable resource to scientists all over the world to further understand the function of the whole human genome", says Professor Taipale. "The resulting increase in our ability to read the genome will also improve our ability to translate the rapidly accumulating genomic information to medical benefits.

More information: 'DNA-Binding Specificities of Human Transcription Factors', Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J.M., Vincentelli, R., Luscombe, N.M., Hughes, T.R., Lemaire, P., Ukkonen, E., Kivioja, T. and Taipale, J., *Cell*, online 17 January 2013. Embargoed until 12:00 (noon) US Eastern Time, on Thursday 17 January 2013.

Provided by Karolinska Institutet

Citation: Learning the alphabet of gene control (2013, January 17) retrieved 25 April 2024 from <https://phys.org/news/2013-01-alphabet-gene.html>

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.