

Protein coding 'junk genes' may be linked to cancer

November 17 2013

By using a new analysis method, researchers at Karolinska Institutet and Science for Life Laboratory (SciLifeLab) in Sweden have found close to one hundred novel human gene regions that code for proteins. A number of these regions are so-called pseudogenes, which may be linked to cancer. The expectation is now that this recently developed protein analysis method, published in the scientific journal *Nature Methods*, will open up a whole new field of research.

All information about the human genome is stored in the DNA sequence in the [cell nucleus](#), and was mapped in the early 2000s. Genes are defined sections of DNA encoding different types of proteins. In recent decades, researchers have been able to define around 21,000 protein coding [human genes](#), using DNA analysis, for example. In the different cell types of the body, different protein producing [genes](#) are active or inactive, and many medical conditions also depend on altered activity of specific genes.

In humans, only about 1.5% of the human genome or DNA consists of protein-coding genes. Of the remaining DNA, some sequences are used to regulate the genes' production of proteins, but the bulk of the DNA is considered to lack any purpose and is often referred to as "junk DNA". Within this junk DNA there are so-called called pseudogenes. Pseudogenes have been considered as non-functional genes, which are believed to be gene remnants that lost their function during evolution.

In the current paper in *Nature Methods*, researchers present a new

proteogenomics method, which makes it possible to track down protein coding genes in the remaining 98.5% of the genome, something that until now has been an impossible task to pursue. Among other things, the research shows that some pseudogenes produce proteins indicating that they indeed have a function.

"To be able to do this we had to match experimental data for sequences of peptides with millions of possible locations in the whole genome", says Associate Professor and study leader Janne Lehtiö. "We had to develop both new experimental and bioinformatics methods to allow protein based gene detection, but when we had everything in place it felt like participating in a Jules Verne adventure inside the genome."

The Lehtiö team found evidence for almost one hundred new protein-coding regions in the [human genome](#). Similar findings were made in cells from mice. Many of the new proteins encoded by pseudogenes could also be traced in other cancer cell lines, and the next objective on the researchers' agenda is to investigate if these genes in the "junkyard" of the genome play a role in cancer or other diseases.

"Our study challenges the old theory that [pseudogenes](#) don't code for proteins", says Dr Lethiö. "The presented method allows for protein based genome annotation in organism with complex genomes and can lead to discovery of many novel [protein coding](#) genes, not only in humans but in any species with a known DNA sequence."

More information: *Nature Methods* [DOI: 10.1038/nmeth.2732](https://doi.org/10.1038/nmeth.2732)

Provided by Karolinska Institutet

Citation: Protein coding 'junk genes' may be linked to cancer (2013, November 17) retrieved 24

April 2024 from <https://phys.org/news/2013-11-protein-coding-junk-genes-linked.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.