

Supercomputing reveals the genetic code of cancer

February 2 2015, by Yngve Vogt



Something goes wrong with the transcription from DNA to RNA in cancer cells. There is either too much transcription, which means that far too many RNA, and subsequently proteins of a specific type are formed, or the composition of base pairs in RNA is wrong. Researchers can gain a better understanding of the



mechanisms that lead to cancer by comparing RNA in diseased and healthy tissue. Credit: Colourbox

Cancer researchers must use one of the world's fastest computers to detect which versions of genes are only found in cancer cells. Every form of cancer, even every tumour, has its own distinct variants.

"This charting may help tailor the treatment to each patient," says Associate Professor Rolf Skotheim, who is affiliated with the Centre for Cancer Biomedicine and the Research Group for Biomedical Informatics at the University of Oslo, as well as the Department of Molecular Oncology at Radiumhospitalet, Oslo University Hospital.

His research group is working to identify the genes that cause bowel and prostate cancer, which are both common diseases. There are 4,000 new cases of bowel cancer in Norway every year. Only six out of ten patients survive the first five years. Prostate cancer affects 5,000 Norwegians every year. Nine out of ten survive.

Comparisons between healthy and diseased cells

In order to identify the genes that lead to cancer, Skotheim and his research group are comparing the <u>genetic material</u> in tumours with the genetic material in healthy cells. In order to understand this process, a fast introduction to our genetic material is needed.

Our genetic material consists of just over 20,000 genes. Each gene consists of thousands of <u>base pairs</u>, represented by a specific sequence of the four <u>building blocks</u> adenine, thymine, guanine, and cytosine, popularly abbreviated to A, T, G, and C. The sequence of these building blocks is the very recipe for the gene. Our whole DNA consists of some



six billion base pairs.

The DNA strand carries the molecular instructions for activity in the cells. In other words, DNA contains the recipe for proteins, which perform the tasks in the cells. DNA, nevertheless, does not actually produce proteins. First a copy of DNA is made. This transcript is called RNA, and it is this molecule that is read when proteins are produced.

RNA is only a small component of DNA, and is made up of its active constituents. Most of DNA is inactive. Only 1-2% of the DNA strand is active.

In cancer cells, something goes wrong with the RNA transcription. There is either too much RNA, which means that far too many proteins of a specific type are formed, or the composition of base pairs in RNA is wrong. The latter is precisely the area being studied by the UiO researchers.

Wrong combinatorics

All genes can be divided into active and inactive parts. A single gene may consist of tens of active stretches of nucleotides (exons).

"RNA is a copy of a specific combination of the exons from a specific gene in DNA."

There are many possible combinations, and it is precisely this search for all of the possible combinations that is new in cancer research.

Different cells can combine the nucleotides in a single gene in different ways. A cancer cell can create a combination that should not exist in healthy cells. And as if that didn't make things complicated enough, sometimes RNA can be made up of stretches of nucleotides from



different genes in DNA. These special, complex genes are called <u>fusion</u> <u>genes</u>.



"Charting the versions of the genes that are only found in cancer cells may help tailor the treatment offered to each patient," says Rolf Skotheim. Credit: Yngve Vogt

In other words, researchers must look for errors both inside genes and between the different genes.

"Fusion genes are usually found in cancer cells, but some of them are also found in healthy cells."

In patients with prostate cancer, researchers have found some fusion genes that are only created in <u>diseased cells</u>. These fusion genes may then be used as a starting-point in the detection of and fight against cancer.

The researchers have also found fusion genes in bowel cells, but they were not cancer-specific.



"For some reason, these fusion genes can also be found in healthy cells. This discovery was a let-down."

Can improve treatment

There are different RNA errors in the various cancer diseases. The researchers must therefore analyse the RNA errors of each disease.

Among other things, the researchers are comparing RNA in diseased and healthy tissue from 550 patients with prostate cancer. The patients that make up the study do not receive any direct benefits from the results themselves. However, the research is important in order to be able to help future patients.

"We want to find the typical defects associated with <u>prostate cancer</u>. This will make it easier to understand what goes wrong with <u>healthy cells</u>, and to understand the mechanisms that develop cancer. Once we have found the cancer-specific molecules, they can be used as biomarkers. In some cases, the biomarkers can be used to find cancer, determine the level of severity of the cancer, the risk of spreading, and whether the patient should be given a more aggressive treatment.

Even though the researchers find deviations in the RNA, there is no guarantee that there is appropriate, targeted medicine available.

"The point of our research is to figure out more of the big picture. If we identify a fusion gene that is only found in cancer cells, the discovery will be so important in itself that other research groups around the world will want to begin working on this straight away. If a cure is found that counteracts the fusion genes, this may have enormous consequences for the cancer treatment."



Laborious work

Recreating RNA is laborious work. The set of RNA molecules consists of about 100 million bases, divided into a few thousand bases from each gene.

The laboratory machine reads millions of small nucleotides. Each one is only one hundred base pairs long. In order for the researchers to be able to place them in the right location, they must run large statistical analyses. The RNA analysis of a single patient can take a few days.

All of the nucleotides must be matched with the DNA strand. Unfortunately the researchers do not have the DNA strands of each patient. In order to learn where the base pairs come from in the DNA strand, they must therefore use the reference genome of the human species.

"This is not ideal, because there are individual differences."

The future potentially lies in fully sequencing the DNA of each patient when conducting medical experiments.

Supercomputing

There is no way the research can be carried out using pen and paper.

"We need powerful computers to crunch the enormous amounts of raw data. Even if you spent your whole life on this task, you would not be able to find the location of a single nucleotide. This is a matter of millions of nucleotides that must be mapped correctly in the system of coordinates of the genetic material. Once we have managed to find the RNA versions that are only found in <u>cancer cells</u>, we will have made



significant progress. However, the work to get that far requires advanced statistical analyses and supercomputing," says Rolf Skotheim.

The analyses are so demanding that the researchers must use the University's supercomputer, which was ranked as one of the world's fastest computers a few years ago. It is 10,000 times faster than a regular computer.

"With the ability to run heavy analyses on such large amounts of data, we have an enormous advantage not available to other cancer researchers. Many medical researchers would definitely benefit from this possibility. This is why they should spend more time with biostatisticians and informaticians. RNA samples are taken from the patients only once. The types of analyses that can be run are only limited by the imagination."

"We need to be smart in order to analyse the raw data. There are enormous amounts of data here that can be interpreted in many different ways. We have just got started. There is lots of useful information that we have not seen yet. Asking the right questions is the key. Most <u>cancer</u> researchers are not used to working with enormous amounts of data, and how to best analyse vast data sets. Once researchers have found a possible answer, they must determine whether the answer is chance or if it is a real finding. The solution is to find out whether they get the same answers from independent data sets from other parts of the world."

Provided by University of Oslo

Citation: Supercomputing reveals the genetic code of cancer (2015, February 2) retrieved 4 May 2024 from <u>https://medicalxpress.com/news/2015-02-supercomputing-reveals-genetic-code-cancer.html</u>

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.