A glance at the DNA of colorectal cancer patients allows identification and validation of potential novel colorectal cancer genes

Verónica Rendo

Cancer is a genetic disease characterized by uncontrolled cell growth and proliferation. It occurs when specific genes become altered and consequently critical cell processes such as division, differentiation, programmed death, and DNA repair become disrupted. The alteration of these genes can be so significant that affected cells may be able to form tumors in local tissues and eventually invade new organs. Cancer is responsible for half a million deaths worldwide every year and it is estimated that one in every third person will be diagnosed with this disease in their lifetime. This project was based in the analysis of patients with colorectal cancer, which is one of the most common types of the disease.

One of the most effective techniques for analyzing cancer cells is 'sequencing' the DNA belonging to affected patients or, in other words, determining the order of the nucleotide bases present in the DNA molecule. This way, a selected group of genes can be analyzed and alterations in these can be detected in order to understand what is happening inside the cancer cells. This technique also allows the identification of new genes that become affected in this disease, which may lead to the discovery of new mechanisms behind cancer.

This study was performed on 96 patients with colorectal cancer. For each individual, DNA was isolated from both tumor and normal cells present in the colon tissue. These DNA samples were then sequenced on an Illumina sequencing platform, focusing only on a subset of 600 previously identified cancer genes (each person contains around 20,000 genes) that could possibly be involved in causing colorectal cancer when altered. The sequencing results showed an initial list of alterations found by Illumina present in the DNA of each patient. All these mutations were filtered by bioinformatic analysis to a group of novel mutations present only in tumor samples. In order to make sure that these alterations were actually present in cancer, validation by Sanger sequencing, another sequencing method, was performed. For this, a first group of genes were chosen. They included those genes with high mutation density, some others with alterations that seemed to have an interesting effect on the translated proteins and another set of genes that act as epigenetic modulators in the cell. Genes with epigenetic functions are involved in the modification of the DNA molecule or the histone proteins in which it is wrapped in order to regulate the expression of specific genes.

In order to start the validation, all samples were initially 'whole genome amplified', i.e. small initial amounts of DNA were copied several times in order to obtain many molecules that can be used in future validation steps. Next, all the parts of the genes containing the selected mutations were also amplified by a method called PCR, which is able to clone a specific region of the DNA many times. The obtained fragments were purified and sequenced by the Sanger method. The results showed that two epigenetic modulators named *ARID1A* and *MLL2* presented novel mutations in the tumor samples of the analyzed patients. Both genes are known to modify the level of condensation of the chromatin and regulate the expression of certain genes. In consequence, alterations in *ARID1A* and *MLL2* can lead to inactivation of genes that are normally expressed in the cell or can promote the expression of other genes that might promote cancer. Interestingly, the alteration of both genes is novel in colorectal cancer, which might contribute to the understanding of the disease and possibly the design of new drugs for anti-cancer therapy.

Degree project in biology, Master of science (1 year), 2012
Examensarbete I biologi 30 hp till magisterexamen, 2012
Biology Education Centre and Department of Immunology, Genetics and Pathology, Uppsala University
Supervisors: Tobias Sjöblom and Lucy Mathot