24/03/2017

Computer program developed to diagnose and locate cancer from a blood sample

Researchers in the United States have developed a computer program that can simultaneously detect cancer and identify where in the body the cancer is located, from a patient’s blood sample. The program is described in research published this week in the open access journal *Genome Biology*.

Professor Jasmine Zhou, co-lead author from the University of California at Los Angeles, said: “Non-invasive diagnosis of cancer is important, as it allows the early diagnosis of cancer, and the earlier the cancer is caught, the higher chance a patient has of beating the disease. We have developed a computer-driven test that can detect cancer, and also identify the type of cancer, from a single blood sample. The technology is in its infancy and requires further validation, but the potential benefits to patients are huge.”

The program works by looking for specific molecular patterns in cancer DNA that is free flowing in the patients’ blood and comparing the patterns against a database of tumour epigenetics, from different cancer types, collated by the authors. DNA from tumour cells is known to end up in the bloodstream in the earliest stages of cancer so offers a unique target for early detection of the disease.

Professor Zhou explained: “We built a database of epigenetic markers, specifically methylation patterns, which are common across many types of cancer and also specific to cancers originating from specific tissue, such as the lung or liver. We also compiled the same ‘molecular footprint’ for non-cancerous samples so we had a baseline footprint to compare the cancer samples against. These markers can be used to deconvolute the DNA found freely in the blood into tumor DNA and non-tumor DNA.”

In this study, the new computer program and two other methods (called Random Forest and Support Vector Machine) were tested with blood samples from 29 liver cancer patients, 12 lung cancer patients and 5 breast cancer patients. Tests were run 10 times on each sample to validate the results. The Random Forest and Support Vector Machine methods had an overall error rate (the chance that the test produces a false positive) of 0.646 and 0.604 respectively, while the new program obtained a lower error rate of 0.265.

Twenty-five out of the 29 liver cancer patients and 5 out of 12 lung cancer patients tested in this study had early stage cancers, which the program was able to detect in 80% of cases. Although the level of tumour DNA present in the blood is much lower during the early stages of these cancers, the program was still able to make a diagnosis demonstrating the potential of this method for the early detection of cancer, according to the researchers.

Professor Zhou added: “Owing to the limited number of blood samples, the results of this study are evaluated only on three cancer types (breast, liver and lung). In general, the higher the fraction of tumor DNAs in blood, the more accurate the program was at producing a diagnostic result. Therefore, tumors in well-circulated organs, such as the liver or lungs are easier to diagnose early using this approach, than in less-circulated organs such as the breast.”

1.    Research article:

CancerLocator: Non-Invasive Cancer Diagnosis and Tissue-of-Origin Prediction Using Methylation Profiles of Cell-Free DNA

Kang et al.

*Genome Biology*March 2017

DOI: 10.1186/s13059-017-1191-5

After the embargo lifts, the article will be available at the journal website here: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1191-5>

Please name the journal in any story you write. If you are writing for the web, please link to the article. All articles are available free of charge, according to BioMed Central's open access policy.

2.    *Genome Biology*publishes outstanding research in all areas of biology and biomedicine studied from a genomic and post-genomic perspective. The current impact factor is 11.313 and the journal is ranked 4th among research journals in the Genetics and Heredity category by Thomson Reuters. Genome Biology is the highest ranked Open Access journal in the category*.*

3.    BioMed Central is an STM (Science, Technology and Medicine) publisher which has pioneered the open access publishing model. All peer-reviewed research articles published by BioMed Central are made immediately and freely accessible online, and are licensed to allow redistribution and reuse. BioMed Central is part of Springer Nature, a major new force in scientific, scholarly, professional and educational publishing, created in May 2015 through the combination of Nature Publishing Group, Palgrave Macmillan, Macmillan Education and Springer Science+Business Media. [www.biomedcentral.com](http://www.biomedcentral.com/)