The fragments of cancer DNA analysed by the authors of this new study originate from the human genome, the sequence of which results from millions of years of evolution, and has been shaped by “copy-paste-edit” processes and co-evolution with parasitic elements. For example, 8% of our DNA comes from past viral infections.

The tortuous mutational processes that have shaped our genomes intensify and become life-threatening in the genomes of cancer cells, leading to anarchic cell mutation and proliferation.

The repeated sequences of DNA in our genomes are not only a fossil of our past evolution, but also hold a track record of how a cancer has evolved, which helps scientists understand and study cancer development and progression.

Current technologies allow scientists to read and piece together billions of short DNA sequences to study cancer genomes and identify mutations within them. But this exploration in repeated DNA has been hindered by a fundamental characteristic of the human genome: how to replace short quasi-identical sequences, often pasted from the same ancestral copy, back to their original genome location? And how to recognise mutations in those sequences?

The recent article, published in Nature Biotechnology, leverages the power of artificial intelligence to solve this problem. Applying this novel tool to the largest collection of primary cancer genomes to date has led to interesting discoveries. For example, mutations that were not detectable with common tools are found even in the coding sequence of well-known cancer genes. This means patients with cancers carrying those mutations might benefit from therapy targeting those genes. Other mutations were found in families of genes duplicated many times along the human genome. Some of these families were already associated with cancers but their mutations could not be observed. The authors have made this rich resource accessible to the scientific community, which further enriches a gold-mine in cancer genomics.

The algorithm developed by Maxime Tarabichi and his collaborators is not limited to cancer, nor to the human genome. It is a universal tool for data generated with current sequencing technologies, which is accessible to all scientists worldwide studying the evolution of life.

Contact:

Maxime Tarabichi
IRIBHM
Université libre de Bruxelles
Maxime.Tarabichi@ulb.be
Reference:

“A pan-cancer landscape of somatic mutations in non-unique regions of the human genome”
DOI number 10.1038/s41587-021-00971-y