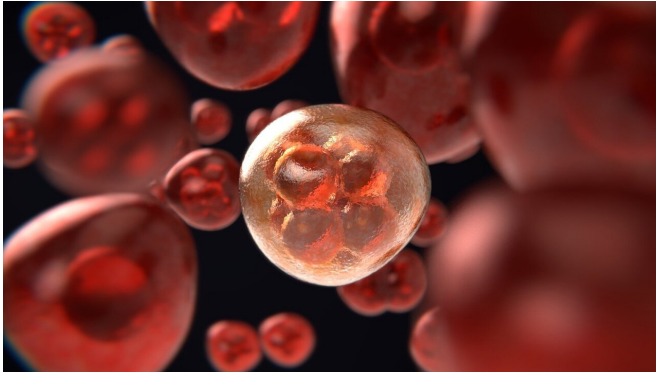# Researchers develop new method for identifying mutational signatures in cancer

30 March 2021

Credit: Pixabay/CC0 Public Domain

Researchers at the Johns Hopkins Kimmel Cancer Center used machine learning techniques to detect mutational signatures in cancer patients. Their algorithm outperformed the current standard of analysis and revealed new mutational signatures associated with obesity, which is believed by cancer prevention experts to be becoming the most significant lifestyle factor contributing to cancer in the U.S. and most of the Western world. The study was published in the Jan. 25 issue of the journal *eLife*.

"Mutational signatures are important in current cancer research as they enable you to see the signs left by underlying factors, such as aging, smoking, alcohol use, UV exposure, and BRCA inherited mutations that contribute to the development of a cancer," says study leader, Cristian Tomasetti, Ph.D., associate professor of oncology at the Johns Hopkins Kimmel Cancer Center, with a joint appointment in Biostatistics at the Johns Hopkins Bloomberg School of Public Health.

The new technique uses an application of artificial intelligence called machine learning, via a

computer algorithm that accesses and analyzes data to uncover what they call SuperSigs, which are mutational signatures that reveal the genetic effects of the underlying contributors to cancer. Their algorithm is classified as "supervised" because it is an analysis that includes known exposures during the training of the algorithm for the genetic analysis of a cancer. The most widely used mutational signatures used for assessing genomic data are classified as "unsupervised" because they do not take known exposures into consideration. Instead, it notes patterns and then goes back to correlate them with exposures. The new method also allows for a mix of supervised and unsupervised approaches, controlling or blocking out the effect of known exposures to carcinogens to explore the possible effect of potential unknown factors.

The researchers found that the new supervised technique outperformed the unsupervised methodology in terms of prediction accuracy. The supervised methodology had a median area under the curve (AUC) of 0.73 for aging and 0.90 for all other factors, while the unsupervised methodology had a median AUC of 0.57 for aging and 0.77 for all other factors.

"A 0.5 or below AUC means the method is not better than pure chance. The highest value you can get is 1," says first author Bahman Afsari, Ph.D., an instructor at the Johns Hopkins Kimmel Cancer until a few months before publication.

They also revealed what they believe are the first mutational signatures associated with cancers of obese patients, providing evidence for a mutational mechanism related to obesity and the origination of cancers.

"Obesity is arguably the most important lifestyle factor contributing to cancer, but its mechanism for causing cancer has been unknown," says Tomasetti. "As cancers of obese patients often do

not appear to have an increased number of mutations, it was thought that the mechanism through which obesity increases cancer risk was not via mutations. Our results show that it is, at least in part, mutational."

Their method also showed that an etiological, or underlying, factor does not always cause the same mutational effect on all tissues, a discovery that was contrary to assumptions of the unsupervised methodology.

"Aging yields different mutational signatures in different tissues, and so do smoking and several other environmental exposures," says co-first author Albert Kuo, Ph.D. candidate at the Johns Hopkins Bloomberg School of Public Health. "Also, in lungs, the signature for aging and the signature for smoking are very different, but in other tissues, the signature of smoking is relatively similar to the signature for aging, suggesting inflammation as the main mechanism."

Additionally, the research provided validation for the key role of random mutations—normal mistakes occurring within the DNA of cells during replication—in the development of a cancer.

"Every time a cell divides, it has to duplicate its DNA. As the duplication and repair machinery copies the billions of letters—the molecules that make up our DNA—mistakes are made. It is estimated that there are between three to six DNA mutations occurring every time a cell divides," explains Tomasetti. "A major source of the mutations that cause cancer appears to be these endogenous processes that have nothing to do with genetic defective genes or harmful exposures."

Through the algorithm, Tomasetti and team determined that 69% of the mutations found in cancer patients across all tumor types can be attributed to randomly occurring mutations, pointing to a need for a greater focus of effort and resources on early detection, he says.

"If we can't avoid cancer from occurring, then the next best thing is to find it before it is too late. If we can find a cancer at an early stage, then typically, you can save the life of the patient," he says.