Artificial Intelligence allows for identification of new cancer genes

The method, published in the journal *Nature Communications*, has been biologically tested in breast, prostate, lung and colon cancer cell lines, as well as retrospective analysis of survival of thousands of patients.

It pointed 63 genes and experimentation confirmed that 36 of them contribute to the irregular cell growth.

The method computationally recreates the biological interactions that take place in our cells.

Barcelona Supercomputing Center's researcher Nataša Průša has led the creation of a new artificial intelligence-based computational method that accelerates the identification of new genes related to cancer. The method and its results, which have been biologically tested, have been published in *Nature Communications*.

Prof. Průša uses Machine Learning techniques to relate large amounts of omic data and recreate them in a computational prototype, Integrated Cell or iCell. Specifically, it fuses three tissue-specific molecular interaction networks: protein-protein interaction, gene co-expression, and genetic interaction networks. The
The technique by which this fusion is performed is Non-negative Matrix Tri-Factorization, a machine learning technique originally proposed for co-clustering and dimensionality reduction that was recently used for data-integration.

The authors of the *Nature Communications* article have applied this method to reconstruct cells from four of the most common types of cancer (breast, prostate, lung and colon) and in all of them, it has proven useful in locating new genes related to these diseases. The method has indicated 63 genes, and a biological validation process has confirmed that at least 36 of them contribute to the irregular growth of the cells. The validation has been carried out through gene deactivation experiments followed by cell viability tests and analysis of patient survival data.

The experimentation revealed, for instance, that breast cancer patients with high expression of MRPL3, a mitochondrial ribosomal protein which was not related to cancer previously, have reduced survival. This is an example of how the new method may be used to uncover new biomarker genes, which may be relevant in the stratification and prediction of survival in cancer patients.

**Nataša Pr?ulj** is ICREA Professor and has just joined the BSC as the leader of the Computational Integrative Network Biology group.

**Alfonso Valencia**, ICREA Professor and Director of BSC's Life Sciences Department states that "Nataša's iCells perfectly complement our BSC cancer genome analysis portfolio, and it is only the first one of the many strong computational methods that we are expecting to see developed by her new group in the coming years ".

Prof. **Pr?ulj** highlights that this new method to analyze cells "enables the identification of perturbed genes in cancer that do not appear as perturbed in any data type alone. This discovery emphasizes the importance of integrative approaches to analyze biological data and paves the way towards comparative integrative analyzes of all cells".

Possible applications range from various other diseases to aging, with the ultimate goal being uncovering intrinsic principles of inner organization of life on Earth.

**Article**: Towards data-integrated cell - DOI: 10.1038/s41467-019-08797-8