The TAVNIT1.0 pipeline (Hebrew *tavnit* – pattern) is dedicated to cancer subtyping and subsequent subtype signature extracting. The pipeline compares cancer and non-cancer samples to extract quantified signatures that strictly detach cancer subtypes from all non-cancer cells. These signatures describe shifts in transcription levels in different cancers. The goal is to use these signatures as targets for cancer drugs.

The pipeline is based on two algorithms: the Constrained–Divisive algorithm for hierarchical clustering and Tjala (an Ant Colony Optimization [ACO] algorithm) for extracting signatures of hierarchical clusters. The Constrained–Divisive algorithm was developed by Kestler et al. [1]. Tjala is inspired by the ℎ𝑚AntMinerorder [2] and Ant-Miner [3] algorithms.

The clustering package consumes transcriptomic data from cancer and non-cancer samples. Notably, cancer and non-cancer samples can be prelabeled to cluster them apart. The clustering output serves as the input for signature extraction. The signatures describe the transcription levels of a number of genes to outline a number of samples. The signatures can be extracted for whole cells or for surfaceome only. An advantageous strategy is “sufaceome-only signatures from all-genes-included clustering.” The surfaceome signatures can be used as targets for surface-recognizing gated drugs (e.g., CAR-Ts).

The pipeline was tested on input whole-genome data from Wang et al., 2018 [4] and an adapted list of surfaceome genes from Bausch-Fluck et al., 2018 [5].

Future prospects include the following:

- the implementation of specific omic distance metrics [6,7]

- the implementation of more options for term selection [8]

(N.B. “Term” refers to a “brick” for building signatures.)

In addition to the people whose ideas directly served as the foundation for this pipeline, Dr. Marco Dorigo, the creator of ACO [9], and those who developed the applications of ACO and hierarchical multilabel classification in biology (for a review, see [10]), are deeply appreciated.

References

1. Kestler HA, Kraus JM, Palm G, Schwenker F. On the Effects of Constraints in Semi-supervised Hierarchical Clustering. In: Schwenker, F., Marinai, S. (eds) Artificial Neural Networks in Pattern Recognition. ANNPR 2006. Lect Notes Comput Sci. 2006; 4087: 57-66. Springer, Berlin, Heidelberg.

2. Liang Zh, Guo R, Sun J, Ming Zh, Zhu Z. Orderly Roulette Selection Based Ant Colony Algorithm for Hierarchical Multilabel Protein Function Prediction. Math Probl Eng, 2017 Jun 28.

3. Parpinelli RS, Lopes HS, Freitas AA. Data mining with an ant colony optimization algorithm. In IEEE Transactions on Evolutionary Computation. 2002 Aug;6(4):321-332.

4. Wang Q, Armenia J, Zhang C, Penson AV, Reznik E, Zhang L, Minet T, Ochoa A, Gross BE, Iacobuzio-Donahue CA, Betel D, Taylor BS, Gao J, Schultz N. Unifying cancer and normal RNA sequencing data from different sources. Sci Data. 2018 Apr 17;5:180061.

5. Bausch-Fluck D, Goldmann U, Müller S, van Oostrum M, Müller M, Schubert OT, Wollscheid B. The *in silico* human surfaceome. Proc Natl Acad Sci U S A. 2018 Nov 13;115(46):E10988-E10997.

6. Nguyen VA, Lió P. Measuring similarity between gene expression profiles: a Bayesian approach. BMC Genomics. 2009 Dec 3;10 Suppl 3:S14.

7. Wang V, Cicalese PA, Louis Sam Titus ASC, Mohan C. Polaratio: A magnitude-contingent monotonic correlation metric and its improvements to scRNA-seq clustering. bioRxiv; 2020.

8. Dulebenets MA. Evaluation of Non-Parametric Selection Mechanisms in Evolutionary Computation: A Case Study for the Machine Scheduling Problem. Nature-inspired Methods for Stochastic, Robust and Dynamic Optimization [Internet]. 2018 Jul 18.

9. Dorigo M. Optimization, learning and natural algorithms. PhD Thesis. Politecnico di Milano. 1992.

10. Otero FEB, Freitas AA, Johnson CG. A hierarchical multi-label classification ant colony algorithm for protein function prediction. Memetic Comp. 2010;2:165–181.