

Future of Exosome Bioinformatics

Dimitrios P. Vlachakis^{1,2,3}✉, Kalliopi Io Diakou¹, George P. Chrousos^{1,2,3}

¹Laboratory of Genetics, Department of Biotechnology, School of Applied Biology & Biotechnology, Agricultural University of Athens, Athens, Greece

²University Research Institute of Maternal and Child Health & Precision Medicine, and UNESCO Chair on Adolescent Health Care, National and Kapodistrian University of Athens, "Aghia Sophia" Children's Hospital, Athens, Greece

³Division of Endocrinology and Metabolism, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

Competing interests: DPV none; KID none; GPC none

Exosomes are under intense study as a promising means for drug or biomarker discovery, primarily due to their implication in intercellular communication and the emergence of disease states and their potential as a cutting-edge, natural delivery system at a nanoscale level. The proteins and nucleic acid cargo of exosomes has been at the center of exosome bioinformatic analysis in the context of health and disease, towards the hunt for novel biomarkers and diagnostics. However, the exosomal lipid composition has been emerging as an interesting target of study as well. Exosomes derived from different sources exhibit enrichment of specific lipid classes and various lipid compositions under different physiological conditions. Therefore, there is a mounting need for exosome lipidomic studies to build the foundation for novel therapeutic studies that use exosomal components. Bioinformatic pipelines are under development to efficiently identify, quantify and elucidate the exosomal lipids and their roles in disease. Cutting-edge bioinformatic tools, such as LipidXplorer, LUX Score, and LipidHome allow the execution of essential analyses such as shotgun lipidomics, and the detection of systematic differences in lipid composition and metadata processing. In the case of pancreatic cancer, an admittedly prevalent and life-threatening disease, these tools have yielded novel exosome lipid biomarkers. Furthermore, bioinformatic platforms such as "Lipidomics Informatics for Life-Science" enable fast and integrated access to these pipelines in a user-friendly manner.

Modern bioinformatic methods facilitate the processing of exosome lipidomics data stemming from mass spectrometry. A set of bioinformatic tools, developed and provided by the [LIPID MAPS consortium](https://lipidmaps.org/)¹, allow the prediction of structural components from an input of mass spectrometry data. In addition to the chemical structures, information on the exosomal

lipid ontology is readily available. Meta-analysis of the exosomal lipidomics data is also made possible through the platform, which filters the data for non-lipid artifacts and corrects potential errors in the processing of MS spectra before storing the datasets.

Proteins that are secreted by exosomes can have a multitude of effects on the development of pathological conditions. In contrast to eukaryotes, where a signal peptide marks proteins that are to be secreted to allow passage through the ER/Golgi-dependent pathway, the proteins to be secreted by exosomes lack such a signal peptide. Bioinformatic methods can thus bridge the gap toward the effective prediction of protein secretion mediated by exosomes. Intuitive algorithms, such as random forest, can be trained on amino acid sequences of proteins secreted and not secreted by exosomes as a relevant feature, approaching the prediction problem as a classification problem. ExoPred, which implements this method, has been developed to identify proteins secreted by exosomes in vertebrates, marking the potential of cutting-edge algorithmic approaches for predicting and annotating distinct, exosome-secreted proteins and opening new pathways for the subsequent study of their potential role in cell communication.

The study of exosomal miRNAs can lead to discovering valuable information concerning the target mRNAs for these miRNAs. Especially in the context of disease and its progression, the differential expression of exosomal miRNAs represents a source of valuable data that can be analysed with bioinformatic tools. With the increase in computational power, machine learning algorithms are gaining popularity in medicine and exosomal analysis. The strength and adaptability of these methods allow the extraction of relevant and helpful information from the vast array of available databases, which are becoming increasingly open to the public and contain large and diverse biomedical data. The

¹<https://lipidmaps.org/>

Article history

Received: 25 February 2022

Accepted: 28 February 2022

Published: 07 March 2023

© 2023 Vlachakis *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at <http://journal.embnet.org>.

development of powerful algorithms, such as **LASSO**² enables the extraction of relevant features from a large number of unrelated features. Hence, this modern algorithm can help in the identification of key exosomal miRNAs and mRNAs involved in complex diseases. When combined with adequate cross-validation methods for the strict evaluation of features and parameters, LASSO – and similar feature extraction algorithms – can effectively generate novel information concerning the exosomal RNAs implicated in pathologies.

In cancer screening and precision nanotherapeutics, extracellular vehicles, such as exosomes, from patients are prime candidates for liquid biopsies. Machine intelligence-driven classification methods are rapidly emerging as a solid support system that, when paired with tried-and-true analytical methods, such as time-dependent spectroscopy, can aid the early and accurate

detection of malignancies. An assortment of machine learning approaches such as multilayer perceptrons, support vector machines, and **AdaBoost random forest classifiers**³, have been coupled with fluorescence correlation spectroscopy (FCS) conducted on samples of blood-derived vesicles from cancer patients, with the end goal being the accurate classification of tissue-specific extracellular vesicles. In parallel, convolutional neural networks (CNNs), quantum CNNs, and networks such as **ResNet**⁴, show promise as supplementary validation tools when trained on the spectral images generated by the FCS. Along with continuously evolving AI algorithms and refined experimental techniques for exosome characterisation, we believe that the future of exosome bioinformatics is very prominent and its role in precision and personalized medicine will prove invaluable in the years to come.

²<https://corporatefinanceinstitute.com/resources/knowledge/other/lasso/>

³<https://medium.com/machine-learning-101/https-medium-com-savanpatel-chapter-6-adaboost-classifier-b945f330af06>

⁴<https://www.mygreatlearning.com/blog/resnet/>