BITS :: Call for Abstracts 2023 - Oral communication

Type	Oral communication
Session	Systems Biology
Title	Development of a risk class prediction methodology using longitudinal data.
All Authors	Pernice S. (1,2), Sirovich R.(3), Ferrero S.(4,5), Genuardi E.(6), Ladetto M. (7), Beccuti M. (1,2), Cordero F. (1,2)

Affiliation

(1) Department of Computer Science, University of Turin, Turin, Italy
(2) Laboratorio InfoLife, Consorzio Interuniversitario Nazionale per l'Informatica (CINI)
(3) Department of Mathematics, University of Turin, Turin, Italy
(4) Department of Molecular Biotechnologies and Health Sciences, University of Torino, Torino, Italy;
(5) AOU Città della Salute e della Scienza di Turin, Turin, Italy
(6) Department of Molecular Biotechnologies and Health Sciences, University of Turin, Turin, Italy;
(7) Division of Hematology, Antonio e Biagio e Cesare Arrigo Hospital, Alessandria, Italy.

Motivation

Minimal residual disease (MRD) detection is a validated outcome predictor in mantle cell lymphoma (MCL). Recently, the clinical relevance of repeated MRD monitoring has been described in a prospective clinical trial [1]. Nonetheless, complex patterns of MRD kinetics over time and the availability of results obtained by heterogeneous tissues (bone marrow, BM, vs peripheral blood, PB) might generate substantial interpretation issues and hamper an easy-to-use application of this predictive biomarker. From the computational and mathematical point of view, any collection of observations recorded at several different time points, describing a system evolution, is defined as longitudinal data. This type of data is useful to explore the evolution of a given event, but the classical methods developed for finite dimension and independent observations cannot be directly applied. In literature, the analysis of longitudinal data is limited in fitting the data with a specific function (i.e. exponential, sigmoids, etc) and then comparing the parameter values of the function to aggregate similar evolution trends.
To overcome the limitation of empirical analysis we used CONNECTOR [2], a novel automated computational framework to facilitate the interpretation of MRD kinetics and stratify patients in risk classes based on a solid, algorithm-derived, classification.

Methods

We used CONNECTOR a computational framework for unsupervised analysis of longitudinal data [2]. It can analyze any sample consisting of measurements collected sequentially over time by a model-based approach for clustering functional data, called Functional Clustering Model (FCM) [3]. CONNECTOR is based on a functional clustering algorithm based on a mixed-effect model that is particularly effective when observations are sparse and irregularly spaced. The CONNECTOR software comprises two modules consisting of (i) an R library, and (ii) a docker image. Docker containerization is utilized to simplify the distribution, use, and maintenance of the analysis tools. The R library provides an easier user interface for which no knowledge of the docker commands is needed. The framework pipeline is defined by three necessary steps: (i) the data importing and processing to create the R object exploited through the entire analysis, (ii) the model selection by identifying the optimal values of the two free parameters of the FCM approach, (iii) the inspection and visualization of the obtained clusters.

A new CONNECTOR module for the prediction of the risk class label was implemented. This was performed using Bayes' optimal allocation rule, on the estimated probabilities of belonging to each class given the observed values.

Results

We used CONNECTOR to facilitate the interpretation of MRD kinetics and stratify patients in risk classes based on a solid, algorithm-derived, classification. Two cases of study are proposed. In the first, ASO RQ-PCR data to follow the MRD was generated from bone marrow and peripheral blood samples of 117 patients in the FIL MCL0208 trial, offering first-line high-dose chemoimmunotherapy and autologous transplantation to younger MCL patients.
CONNECTOR identified four patient risk clusters predicting patients' outcomes: median time to progression (TTP) was not reached for favorable MRD kinetics clusters while 36 and 27 months for unfavorable MRD kinetics ($p<0.0001$). 89 out of 95 patients (94%) were correctly reclassified. CONNECTOR allowed unsupervised identification of four CrC of patients with different MRD kinetics and highly significant different outcomes. Such clusterization proved effective using BM, PB, and mixed tissues. Moreover, an independent validation on a European cohort, on behalf of the MCL network, is now running.
The CONNECTOR framework and its modules are a user-friendly tool to efficient comprehended longitudinal data, providing hints to increase interpretability and molecular accuracy and to use the new model found for improving the outcome prediction.

Info

[1] Ferrero S., et al. Punctual and kinetic MRD analysis from the Fondazione Italiana Linfomi MCL0208 phase 3 trial in mantle cell lymphoma. Blood (2022) 140 (12): 1378–1389.

[2] Pernice S., et al. CONNECTOR, fitting and clustering of longitudinal data to reveal a new

risk stratification system. Bioinformatics. (2023) Apr 20

[3] Gareth MJ and Sugar AC, Clustering for sparsely sampled functional data. Journal of the American Statistical Association (2003), 98 (462), 397-408.

filename	-
Figure	
-	
Availability	-

Corresponding Author

Name, Surname	Simone, Pernice
Email	simone.pernice@unito.it
Submitted on	02.05.2023