Gaining power from multi-omic data integration and pathway topology: improvement of prognostic gene module identification

Martini P(1)† , Calura E(1), Chiogna M(2) & Romualdi C(1)

(1) Department of Biology, University of Padova, Padova
(2) Department of Statistical Sciences, University of Padova, Padova

† Email: paolo.martini@unipd.it

Motivation
The molecular medicine has rapidly grown thanks to advances in genome-wide assay that can produce expression or methylation profiles for many patients. Each assay has several analysis methods and the results of these methods are used to guide diagnosis, prognosis or treatment of diseases. In our lab, we devised tools for the topological pathway analysis of expression data: clipper for 2-class comparison, timeClip for time series and survClip for survival analysis. Using these tools in a cohort of patients with EOC stage I, we demonstrated that the patients with the worst prognosis have a higher activity of a molecular circuit around the lncRNA PVT1. The new challenge is to integrate many omics dimensions. A naive approach would rely on the analysis of each omics data singularly and then should merge results trying to draw a mechanistic framework that explain the relation between the phenotype and the molecular defects reported by the assay. Here, we propose a method to integrate different omics and test them together.
Methods
We devised a new method, called MOSClip, to aggregate multiple omics dimensions and predict patients' survival. We proposed to shift from the gene-centric to the module view. Using pathway topology (Reactome pathway from Graphite), we define survival modules as a group of genes closely connected in a pathway. In the survival modules, we integrate the expression, methylation and somatic mutations of the genes and we test the predictive power of each module. Along with the test, we implemented several visualizing tools to explore the results. This inspection instruments make easier the results interpretation (Fig1).
Results
MOSClip is an R package that allows the simultaneous integration of different omics. Watching the problem from three perspective (mRNA, SNP and methylation) enhances the correct stratification of the patients. Furthermore, thinking in the frame of survival modules makes easier the formulation of hypothesis of pathological mechanisms because the connection between the function of survival module and the phenotype is straightforward. MOSClip has been tested both in simulated and TCGA – Ovarian Cancer dataset. Our results on the Ovarian Cancer agree with the results recently published by the PanCancer Consortium (Sanchez-Vega et al. Cell 2018). Moreover, the results tell us that different omics complement each other in

biological processes that can better explaining patients' survival.

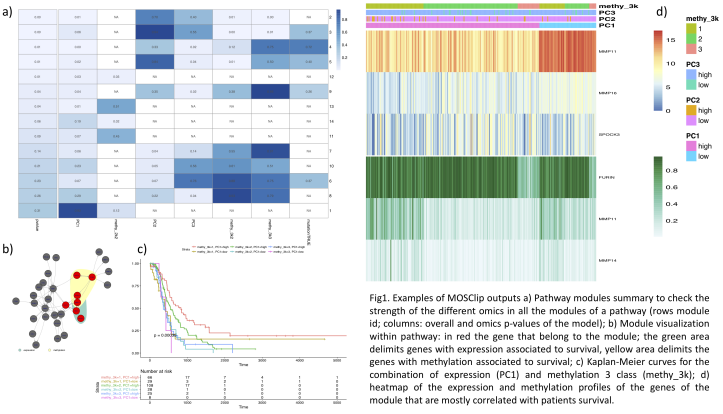

Fig1. Examples of MOSClip outputs a) Pathway modules summary to check the strength of the different omics in all the modules of a pathway (rows module id; columns: overall and omics p-values of the model); b) Module visualization within pathway: in red the gene that belong to the module; the green area delimits genes with expression associated to survival, yellow area delimits the genes with methylation associated to survival; c) Kaplan-Meier curves for the combination of expression (PC1) and methylation 3 class (methy_3k); d) heatmap of the expression and methylation profiles of the genes of the module that are mostly correlated with patients survival.