

RELAZIONE FINALE PROGRAMMA DI RICERCA STM 2016

Il Fruitore: *Antonella Iuliano*

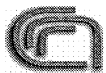
Istituto di afferenza: Istituto per le Applicazioni del Calcolo “Mauro Picone” (IAC) – Napoli
con qualifica di *Assegnista di Ricerca*

Istituzione ospitante: Computer Laboratory, Università di Cambridge - William Gates Building, 15 JJ Thomson Ave, Cambridge CB3 0FD, Regno Unito

Titolo del programma: *Statistical algorithms for Cancer Survival Analysis*

L'obiettivo del mio soggiorno presso il *Computer Laboratory* dell'Università di Cambridge, sotto la supervisione del Dr. Pietro Liò ed in collaborazione con il suo gruppo di ricerca, è stato quello di mettere a punto una strategia statistico-computazionale basata su reti per risolvere problemi ad alta dimensionalità con applicazioni all'analisi della sopravvivenza di pazienti oncologici utilizzando l'integrazione di dati *multi-omici*. Infatti, l'analisi a diversi livelli biologici (genomica, trascrittomica, epigenomica) mi ha permesso di eseguire una caratterizzazione molecolare più completa del fenomeno in esame. In particolare, è stato possibile individuare sottoclassi differenti di pazienti e importanti marcatori biologici aventi un ruolo fondamentale nell'analisi della sopravvivenza con l'obiettivo di programmare interventi farmacologici mirati e di migliorare le capacità diagnostiche e prognostiche del paziente.

Durante questo periodo è stato estremamente importante sviluppare ed implementare delle procedure computazionali basate sull'integrazione di più omiche al fine di mettere in evidenza meccanismi non osservabili quando i diversi livelli biologici sono analizzati in maniera indipendente. Infatti, uno dei principali limiti a questo sviluppo è il fatto che i dati

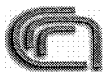


provenienti dai diversi livelli biologici sono eterogenei e hanno differenti scale e unità di misura, con la conseguenza che una comparazione diretta risulta a volte impossibile o estremamente complicata.

Nella prima fase di questo soggiorno ho svolto un'analisi accurata dei modelli di Cox basati sulla combinazione di tecniche di *screening* [1,2] e modelli di *network-penalizzati* [3]. Si tratta essenzialmente di un approccio statistico che permette prima di ridurre lo spazio delle covariate, ossia i geni, estraendo quelle che sono fortemente correlate alla variabile d'interesse, ossia il tempo di sopravvivenza, per poi procedere con una ricerca più dettagliata dei geni altamente coinvolti nella malattia attraverso l'utilizzo di tecniche di penalizzazione basate su reti. Infatti, attraverso vari studi effettuati in precedenza è chiaro che i metodi network funzionano meglio dopo aver ridotto la dimensionalità iniziale delle variabili in esame [4].

Nella seconda fase ho implementato tali metodi prendendo in considerazione dati provenienti da vari consorzi internazionali, tra cui "The Cancer Genome Atlas" (TCGA), "International Cancer Genome Consortium" (ICGC Data Portal) e "Gene Expression Omnibus" (GEO). I dati provenienti da ciascuna omica (come varianti geniche, dati di espressione genica ed epigenetici) sono stati analizzati prima separatamente e poi integrati utilizzando un metodo computazionale MANCIE [5], con l'obiettivo di individuare una serie di geni aventi profili concordanti nei diversi livelli.

Nella terza fase ho effettuato la validazione dei dati al fine di selezionare marcatori tumorali e gruppi di pazienti fortemente a rischio ottenendo così una capacità predittiva di sopravvivenza molto più precisa e significativa. Infatti, tale analisi ci ha permesso di identificare varianti geniche e pattern epigenetici che simultaneamente influenzano l'espressione genica di un sottogruppo di geni, eventualmente in un sottogruppo di pazienti. Più precisamente, l'analisi della sopravvivenza è stata condotta utilizzando il



metodo di Kaplan-Meier per rappresentare graficamente i risultati dei pazienti a rischio e l'analisi dei "pathway biologici" sulla base dei geni altamente coinvolti nella malattia.

Dunque, l'utilizzo di questo nuovo approccio (riduzione della dimensionalità e metodi network-penalizzati) per la caratterizzazione e l'analisi di diverse omiche consente di indirizzare meglio la conoscenza dei meccanismi tumorali a livello molecolare permettendo, dove possibile, di individuare dei sottogruppi di pazienti e biomarcatori che migliorano la diagnosi o addirittura pianificare delle terapie personalizzate. Questo ci consentirebbe di intervenire in modo più mirato e raffinato per migliorare la salute dei pazienti e l'utilizzo delle cure attualmente disponibili.

Inoltre, un valore aggiunto molto importante per questo tipo di studio sarà lo sviluppo di un software, al quale già stiamo lavorando, che darà la possibilità di visualizzare i risultati con grafici costruiti ad hoc rendendo l'analisi dei dati fruibile anche da chi non ha specifiche competenze di bioinformatica.

Posso, pertanto, fortemente confermare che questo breve periodo è stato di estrema importanza per me e per entrambe le parti coinvolte, infatti la possibilità di continuare la collaborazione è fondamentale per i futuri sviluppi di ricerca. Lo scopo è creare un "ponte" di comunicazione tra i due laboratori condividendo risorse comuni e complementari nelle diverse aree di ricerca, per la disseminazione dei risultati raggiunti e progetti internazionali al fine di combattere le gravi malattie con la genomica computazionale guidata dai dati omici.

[1] Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 849-911.



- [2] Fan, J., Feng, Y., & Wu, Y. (2010). High-dimensional variable selection for Cox's proportional hazards model. In *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown* (pp. 70-86). Institute of Mathematical Statistics.
- [3] Sun, H., Lin, W., Feng, R., & Li, H. (2014). Network-regularized high-dimensional cox regression for analysis of genomic data. *Statistica Sinica*, 24(3), 1433.
- [4] Iuliano, A., Occhipinti, A., Angelini, C., De Feis, I., & Liò, P. (2016) Cancer markers selection using network-based Cox Regression: a methodological and computational practice. *Front. Physiol.* doi: 10.3389/fphys.2016.00208
- [5] Zang, C., Wang, T., Deng, K., Li, B., Qin, Q., Xiao, T., ... & Liu, J. S. (2016). High-dimensional genomic data bias correction and data integration using MANCIE. *Nature communications*, 7.

Luogo e data

Napoli, 15/06/2016

Il Fruitore

Antonella Iuliano