

Gut microbiome composition and small RNA spectra in human stool for colorectal cancer detection

Ferrero G(1)[†], Tarallo S(2), Francavilla A(2), Gallo G(3,4), Clerico G(4), Manghi P(5), Thomas A(5), Segata N(5), Pardini B(2,6), Naccarati A(2,7), Cordero F(1,2)

(1) Department of Computer Science, University of Turin, Italy

(2) Italian Institute for Genomic Medicine (IIGM), Turin, Italy

(3) Department of Surgical and Medical Sciences, University of Catanzaro

(4) Department of Colorectal Surgery, Clinica S. Rita, Vercelli, Italy

(5) Centre for Integrative Biology (CIBIO), University of Trento, Italy.

(6) Department of Medical Sciences, University of Turin, Turin, Italy

(7) Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Prague, Czech Republic.



[†] Email: giulio.ferrero@unito.it

Motivation

Many evidence suggest a contribution of an altered gut microbiota composition in the onset and progression of ColoRectal Cancer (CRC). Human small RNAs (hs-sRNAs) were shown to be involved in both tumorigenesis as well as inter-kingdom interactions between microbiome and human cells [1]. However, little is known on how much informative are a study of gut microbiota composition and a small RNA-Seq (sRNA-Seq) experiment performed on the same stool samples from CRC patients. In this study we integrate data from sRNA-Seq with shotgun DNA sequencing data to study stool bacteria sRNAs (bsRNAs) and to evaluate the accuracy of the combinatoric use of these data in the classification of healthy individuals from patients with CRC or precancerous lesions.

Methods

We performed a DNA shotgun sequencing experiment and sRNA-Seq in 80 stool samples collected from healthy individuals and patients with adenoma or CRC. Metaphlan2 was applied to identify bacteria relative abundances from shotgun sequencing data. Adopting the strategy proposed in [2], sRNA-Seq reads where firstly used to identify the hs-miRNA and human non miRNA sncRNAs (hs-sncRNA) collections, then the unmapped reads were used to define the bsRNA profiles by estimation of bsRNAs relative abundances using Kraken algorithm. sRNA-Seq reads assigned to bacteria were also aligned against bsRNA annotations from Bacterial Small RNAs Database (BSRD). Secondary structure of candidate bsRNA was predicted using RNAFold applied on each sRNA-Seq reads. Comparison between abundances of bacterial DNA and bsRNAs was performed by correlation analysis and prediction of bsRNA transcriptional rates. Patient classification accuracy was computed using Random Forest classifier using different combination of input data.

Results

Our analysis revealed an extended and significant correlation between relative abundances of bacterial DNA and bsRNAs (median $r=0.89$) with a consistent predicted

increase of Proteobacteria from healthy to CRC patients. *Escherichia Coli* emerged as significantly more abundant bacteria at both DNA and sRNA level but also associated to a low transcriptional rate. Secondary structure analyses revealed that bsRNA reads assigned to bacteria could form more stable structures compared to reads mapped on hs-miRNAs or hs-sncRNAs consistent with the highest structured property of bacteria RNAs. Analysing the reads assigned to *Escherichia Coli* gene annotations, as well as annotations from BSRD, led us to identify a set of bsRNAs whose expression significantly increased from healthy to CRC patients.

The random forest approach provided a high classification accuracy (AUC=0.83) when both bsRNAs and bacteria DNA data were used, while bacterial DNA profiles-only had no sufficient power to discriminate samples (AUC=0.65). Noteworthy, using hs-miRNA, bsRNA, and bacterial DNA profiles together, we obtained a high classification accuracy (AUC=0.88). The features selection analysis revealed that 32 features were sufficient to achieve the classification accuracy. Our results support the hypothesis that the integration of shotgun sequencing and sRNA-Seq data can be efficiently employed to classify samples as well to extract novel insight on human and bacterial sRNAs involved in a disease.

References

1. Liu, S., da Cunha, A. P., Rezende, R. M., Cialic, R., Wei, Z., et al., (2016). The host shapes the gut microbiota via fecal microRNA. *Cell host & microbe*, 19(1), 32-43.
2. Ferrero, G., Cordero, F., Tarallo, S., Arigoni, M., Riccardo, F et al., (2018). Small non-coding RNA profiling in human biofluids and surrogate tissues from healthy individuals: description of the diverse and most represented species. *Oncotarget*, 9(3), 3097.