A computational framework for the identification of genes modules epigenetic regulated and targetable by drugs at the level of single patient.

Tagliazucchi G(1)†, Taccioli C(2)

(1) Azienda Ospedaliero Universitaria di Parma, Department of Research and Innovation, 43126 Parma, Italy
(2) Dipartimento di Medicina Animale, Produzione e Salute, Università di Padova, 35020 Legnaro, Padova, Italy.

† Email: guidantonio.mt@gmail.com

Motivation
The integration of “omics” data with data-fusion methods [1] from Next Generation Sequencing (NGS) technologies provides a great potential to better understand the molecular basis of tumors and stratify different cancer subtypes [2]. Moreover, the increasing amount of NGS data, introduces the possibility to further develop precision and personalized-medicine. The USNRC define the personalized medicine as "the situation in which therapeutics are synthesized for specific individuals" [3]. In this context, the integration of omics data can also provide an important step for the identification of altered biological mechanisms at the level of single patient. Diseases can arise from multiple steps that involved also epigenetic mechanisms. However, at the present moment, few tools able to integrate omics data and study the role of epigenetic components (e.g. histone marks) on groups of genes at the level of single patient. Here, we describe a framework, GMIEC, that integrates omics data (gene-expression, copy number variation, mutation, methylation) with data containing genomic coordinates of cis and/or trans epigenetic regulators (e.g. ChIP-seq). Then, at the level of the single patient, GMIEC identify groups of genes (modules) that share common genomic features. Then, using an external database, the genes in each module are associated with their own target drugs. Therefore, GMIEC allows the identification, at the level of the single patient, groups of genes that are regulated by the same epigenetic regulators (e.g. TFs), sharing common genomic features and associated with different number of drugs.

Methods
The input of GMIEC are: i) a file with the genomic regions of the regulators of interest; ii) omics data; ii) a file that includes the association between genes and drugs. The four principal steps (Figure 1A) of analysis are: i) the annotation of the epigenetic components (ECs) derived from a catalogue (ReMap [3]) with their target genes using ChIPpeakAnno[4]; ii) a step of discretization used to mark the presence of one alteration (1) or not (0) for all genes of single patient, considering all datasets (e.g. gene-expression data, copy-number alterations data, mutations data); iii) the genes with the similar genomic features grouped using Association Rules Discovery (ARD) technique. This is a data mining method used to discover associations
between subsets of items and large dataset extensively used in bioinformatics [5], [6]. In our case, (Figure 1B) if two genes, target of the ECs, share similar common genomic features (e.g. amplification, hypermethylation), then they are grouped together. Finally, a combined score (SAD) is computed to verify if a specific group of genes are altered or not and if they are targets of drugs or not. GMIEC is provided as an R-script (https://github.com/guidnt/gmiec) and depends on the R-package “arules”. To test GMIEC we used The Cancer Genome Atlas BRCA datasets.

Results

We applied GMIEC on a dataset of Breast Cancer TCGA data and considering RUNX1 as regulator. RUNX1 was identified as strongly expressed in breast cancer epithelia and dysregulated during tumorigenesis. Here, GMIEC was set to select only the group of genes with few genomic alterations. Considering only the samples of the basal-like subtype we identified 18 patients with only 1 gene in their module, 23 patients with more than 2 genes. Nine of these samples contained modules with more than 16 genes (range from 16 to 131). Our analysis indicates that some of the patient modules were also associated with anti-neoplastic agents. These results were confirmed by further analyses performed using oncoScore and DrugPattern [7]. Thus, our strategy proved its effectiveness in the identification of subtle differences along samples and patients. In summary, GMIEC is able to: i) stratify samples and patients; ii) identify patient specific genes; iii) plan new treatment strategies personalized for cancer patients.

References

3. Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. Grif-


