COMBO: A Computational Framework to Analyze RNA-seq and Methylation Data Through Heterogeneous Multi-layer Networks

Ilaria Cosentini[1], Vincenza Barresi[2], Daniele Filippo Condorelli[2], Alfredo Ferro[3], Alfredo Pulvirenti[3], and Salvatore Alaimo[3(✉)]

[1] Department of Physics and Astronomy, University of Catania, 95124 Catania, Italy
[2] Department of Biomedical and Biotechnological Sciences, Section of Medical Biochemistry, University of Catania, 95123 Catania, Italy
[3] Department of Clinical and Experimental Medicine, Bioinformatics Unit, University of Catania, 95123 Catania, Italy
salvatore.alaimo@unict.it

Abstract. Multi-layer Complex networks are commonly used for modeling and analysing biological entities. This paper presents a new computational framework called COMBO (Combining Multi Bio Omics) for generating and analyzing heterogeneous multi-layer networks. Our model uses gene expression and DNA-methylation data. The power of COMBO relies on its ability to join different omics to study the complex interplay between various components in the disease. We tested the reliability and versatility of COMBO on colon and lung adenocarcinoma cancer data obtained from the TCGA database.

Keywords: Multi-layer networks · Multi-omics analysis · Biological data integration · Next-Generation sequencing · Graph database

1 Introduction

Understanding tumor phenotype is a challenging task. Indeed, independently investigating each component is not enough to characterize its entire biology. In addition, studying the interactions and relations between the different omics (i.e., genomics, transcriptomics, and proteomics) in the cancer microenvironment is crucial to stratify case subtypes or distinguish cases from controls. For example, the study of single gene perturbations is insufficient to explain the cause of abnormal behavior in a tissue that triggers the disease. The usage of Complex Networks to describe natural phenomena such as biological systems has taken over [1]. This approach has been demonstrated to help reveal unknown mechanisms behind the different diseases. Indeed, the interdependence between multiple biological elements is responsible for disease onset and progression [2, 3]. Here, the network represents a practical approach [4, 5]. The analysis of omics functional interaction networks (e.g., genes, proteins, and ligands) is promising for unveiling their regulatory mechanisms. However, due to the vast number of elements and information provided by the biological components, using networks to highlight the main

H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2016 2022, SCI 1077, pp. 251–264, 2023.
https://doi.org/10.1007/978-3-031-21127-0_21

aspects of a condition is not straightforward. Nevertheless, it represents a significant milestone for science [6]. However, in the biological field, simple networks (where there is at most one edge between each pair of nodes) are not suitable to represent the multiple types of relationships among various constituents [7–9]. On the other hand, multi-layer networks (every node is assumed to belong to every layer, each layer represents a specific kind of relation, along with intralayer edges type for each layer) fit nicely in real-world system interactions avoiding the loss of essential properties [8, 10, 11]. There are several examples of multi-layer network usage in Biology [12, 13]. As the integration of different sources of genetic data is still challenging, modeling a multilayer network in which each layer represents specific omics can improve cancer characterization [14–17]. However, state-of-art methods focus on visualization and analysis of multi-layer networks, leaving the pre-processing phase to the user [16, 18–21]. This study proposes a bioinformatics framework to infer important information about a specific disease by integrating the transcriptional and epigenetic (methylation) features. We propose a novel pipeline, COMBO (Combining Multi Bio Omics), for multi-layer network inference and analysis to identify relevant pathways in the studied case. Taking advantage of the Boolean implication method, both transcriptomic and epigenomic data were analyzed through StepMiner and BooleanNet systems [22, 23]. The goal was to identify the implication between the different transcripts and methylated CpGs. The obtained results were used to generate heterogeneous multi-layer graphs. Subsequently, Neo4J was exploited to query the multi-layer network with properly defined Cypher queries. The model has been tested on colon cancer samples (TCGA-COAD) having the amplification of chromosome 20, also called GAIN, compared to colon tumoral samples without that chromosomal alteration, identified as DISOMIC. Only the GAIN multi-layer network exhibited interesting patterns when searching for all the paths that link the expression layer with the methylation layer, including genes located on chromosome 20. In addition, the enriched analysis of the genes composing the path shows the alteration of pathways that play an essential role in cancer. To demonstrate the plasticity of the pipeline to any human disease, we tested it in TCGA-LUAD samples. The comparison focused on samples having upregulated IWS1 transcript (upIWS1) versus those with downregulated IWS1 (downIWS1). As shown for COAD, the paths that involve IWS1 and upregulated genes are available only for the multi-layer generated for the upIWS1 condition. Furthermore, a Reactome enrichment analysis shows the alteration of crucial genes in lung adenocarcinoma. Our bioinformatics approach is a flexible and easy-to-use method in preliminary cancer studies to identify significant disease signatures.

2 Materials and Methods

2.1 Pipeline Design

The pipeline comprises five phases (Fig. 1). First, it generates multi-layer networks for each condition under investigation. Next, networks are stored in Neo4j. Then, through custom Cypher queries, the multi-layer is interrogated to generate sub-networks that can be analyzed differently (i.e., enrichment pathway analysis).

STEP 1: Input Pre-preprocessing. The input files for the analysis consist of the expression and methylation matrices containing pre-processed counts and B-values,

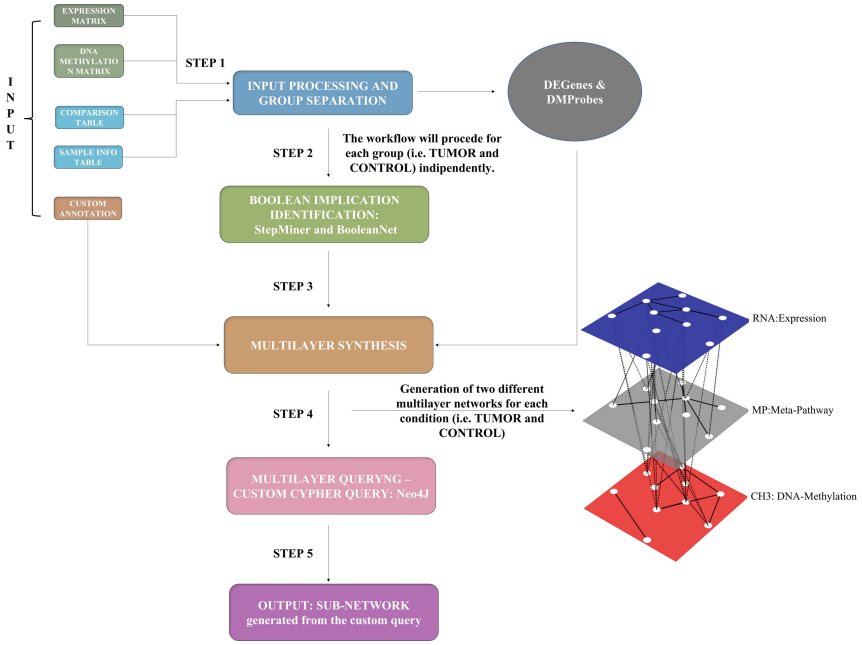


Fig. 1. The COMBO pipeline workflow.

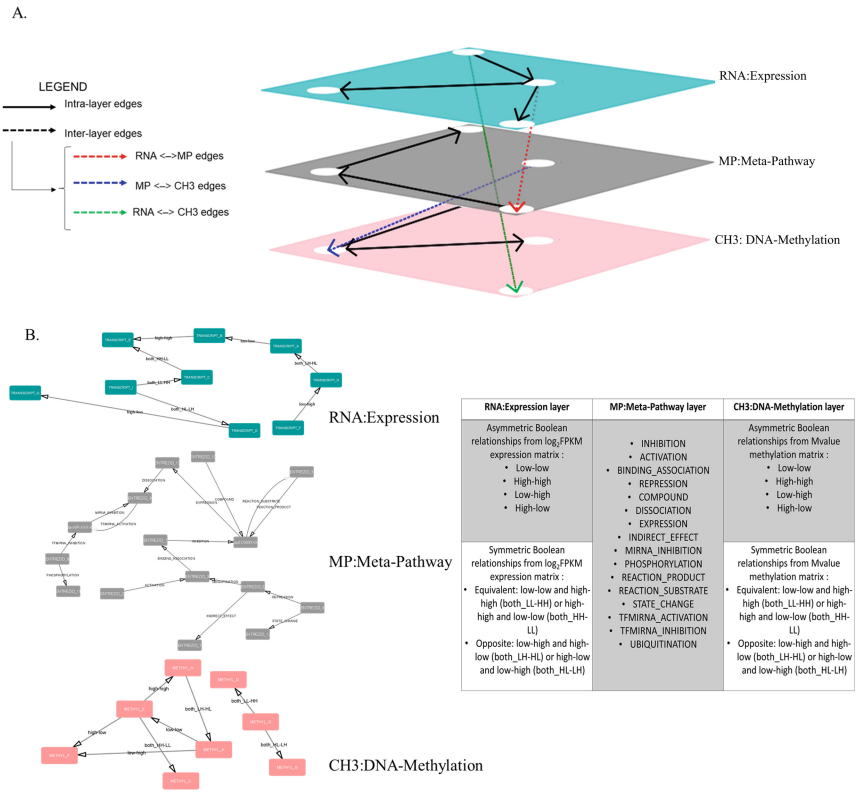
respectively. The initial matrix contains all the samples representing the conditions to compare. Two tables are needed for the differential expression and methylation analysis. The first one contains the name of each sample and its group. The second table represents the comparisons between groups. The number of columns in the comparison table equals the number of multi-layer networks. The first step is processing the input matrices (expression and methylation) to identify differentially expressed genes and differentially methylated CpGs. The differentially expressed genes and the methylated probes are generated using the R package limma [24]. The user can choose the threshold for the identification of differentially expressed genes (default $= 1.0$, corresponding to logFC < -1 and logFC > 1) and differentially methylated probes (default $= 0$, equivalent to logFC < 0 and logFC > 0) along with the adjusted p-value (default < 0.05). In addition, the count and the B-value in the expression and methylation matrices are converted in $\log_2$FPKM and M-value, respectively. These will be provided as input for the next phase.

STEP 2: Boolean Implication. The second step is the generation of boolean implications with the StepMiner [22] and BooleanNet [23] algorithms. **StepMiner Algorithm** was initially developed to extract binary patterns in time-series microarray gene expression data. After sorting values in ascending order, StepMiner fits the data by a step function to discretize the jump from low to high values. In this step, it finds the best threshold t for each gene. Then, values above t + 0.5 are classified as "High," values below t − 0.5 are classified as "low," and values between t − 0.5 and t + 0.5 are classified as "intermediate." Two different discretized matrix was generated for expression and methylation data. $\log_2$FPKM for expression data and M-value for methylation data were

used as input for StepMiner. The M-value is calculated as the $\log_2$ ratio of the intensities of the methylated probe versus the unmethylated probe, as reported in [25]. **The Boolean Values** obtained from Step Miner are converted in the gene-pairs relationship using BooleanNet. The Boolean implications between gene expression and methylation levels are extracted as "if-then" relationships. All the potential pairs of values of genes in each sample are plotted in a scatter plot which is divided into four quadrants "low-low", "low-high", "high-low", and "high-high." For each pair of genes, the algorithm checks which quadrant is significantly sparsely populated with points compared to the other quadrants. In the scatter plot, each point represents the expression value of the pair of genes for a specific sample. Based on the empty quadrant in the scatter plot, the algorithm detects six types of Boolean implications. The Boolean relationships can be symmetric or asymmetric. Symmetric relationships include the equivalent type, which corresponds to highly positively correlated genes, or the opposite type, which corresponds to the highly negatively correlated genes. The asymmetric ones correspond to every sparse quadrant and are Divided into low—low, high—low, low—high, and high—high. An implication between two genes is significant when one or more quadrants is sparsely populated according to a statistical test, and there are enough high and low values for each gene. Although both software packages are pretty old, recent papers confirm that StepMiner and BooleanNet are the state-of-art methods to infer Boolean implications between biological entities [26–29]. The user can select the delta threshold for StepMiner, the BooleanNet statistic parameter (s), and the *p*-value threshold. The statistical parameters used for Boolean analysis in COAD samples were $s = 4$ for methylation data and $s = 2$ for RNA-seq data. Whereas $s = 2$ was used for expression and methylation data in LUAD samples. In the COMBO pipeline the user can choose the *p*-value (default $= 0.01$) and the BooleanNet statistic (default $= 6$).

STEP 3: Multi-layer Network Synthesis. Networks (or graphs) are mathematical tools to represent relationships between entities or objects. For example, interactions, reactions, or structural relations are depicted as the links between the interacting elements (genes, proteins, metabolites) in a biological network. Multi-layer networks have different layers containing a set of homogeneous nodes equal or not between the layers [30]. The edges between nodes in the same layer are indicated as intra-layer edges, whereas the edges linking different layers are called inter-layer edges. When inter-layer edges connect nodes in two different layers representing the same entity, the network is classified as a multiplex network [31].

The third step is the multi-layer generation. We show an example in Fig. 2. In the example, we include three layers: expression, meta-pathway, and methylation. The expression level contains nodes and directed intra-layer edges selected from BooleanNet [23] analysis conducted with RNA-seq data. Similarly, the nodes and directed intra-layer edges that participate in the methylation level are selected from BooleanNet methylation output. The meta-pathway layer was added to connect the other two layers. The meta-pathway is a network obtained by merging all KEGG pathways through their common nodes, annotated with experimentally validated miRNA-target and Transcription Factor-miRNA interactions [32–34]. Consequently, the meta-pathway



RNA:Expression layer	MP:Meta-Pathway layer	CH3:DNA-Methylation layer
Asymmetric Boolean relationships from log_2FPKM expression matrix : • Low-low • High-high • Low-high • High-low	• INHIBITION • ACTIVATION • BINDING_ASSOCIATION • REPRESSION • COMPOUND • DISSOCIATION • EXPRESSION • INDIRECT_EFFECT • MIRNA_INHIBITION • PHOSPHORYLATION • REACTION_PRODUCT • REACTION_SUBSTRATE • STATE_CHANGE • TFMIRNA_ACTIVATION • TFMIRNA_INHIBITION • UBIQUITINATION	Asymmetric Boolean relationships from Mvalue methylation matrix : • Low-low • High-high • Low-high • High-low
Symmetric Boolean relationships from log_2FPKM expression matrix : • Equivalent: low-low and high-high (both_LL-HH) or high-high and low-low (both_HH-LL) • Opposite: low-high and high-low (both_LH-HL) or high-low and low-high (both_HL-LH)		Symmetric Boolean relationships from Mvalue methylation matrix : • Equivalent: low-low and high-high (both_LL-HH) or high-high and low-low (both_HH-LL) • Opposite: low-high and high-low (both_LH-HL) or high-low and low-high (both_HL-LH)

Fig. 2. (A) A heterogeneous multi-layer network example (generated with pymnet python library http://www.mkivela.com/pymnet/). The interlayer edges are depicted with different colors. (B) Example of the structure of the single layer. The table shows the edges attribute for each layer.

layer includes genes, miRNAs, and metabolites linked with several edges, which represent a validated biological interaction (i.e., activation, repression, binding association, compound, dissociation, expression, indirect effect, inhibition, miRNA inhibition, phosphorylation, reaction product, reaction substrate, state change, transcription factor—miRNA activation, ubiquitination). We followed the procedure described in Alaimo et al. [32, 33] for the KEGG meta-pathway and the one described in Alaimo et al. [34] for the Reactome-KEGG meta-pathway, which integrates REACTOME pathways to the meta-pathway environment, yielding a richer and more comprehensive model. A unique nodeID is generated for each node in the three layers. If a node is present in two or more layers, the ID will differ because each layer reports different aspects of the same gene. In addition, directed inter-layer edges are inserted between two nodes representing the same gene to connect the other layers. Each node has an annotation that the users will employ for querying the network. The default annotations are divided into four categories: (i) nodes details: nodeID, layer, entrezId, symbol, and chromosomal localization of the gene; (ii) expression of the genes generated in the first step: using the parameters defined by the user, the upregulated genes are indicated as “UP_REG,” whereas

the downregulated genes are annotated as "DOWN_REG." The remaining genes do not have any annotation; (iii) methylation information: two columns are included in this section, one for the hypomethylated probes and the other for the hypermethylated one. In addition, for each hyper- or hypo-methylated CpG site, we also report the location (Body, Exon, 3'UTR, 5'UTR, Promoter); (iv) custom annotation: the user can choose to add custom annotations for each node, loading a table for each annotation reporting the name of the genes with the respective annotation in the different conditions.

STEP 4–5: Multi-layer Network Querying. The last two steps involve using Neo4j to store the multi-layers network and query its content. Neo4j (https://neo4j.com/) is a native graph database that implements a true graph model down to the storage level. Once the multi-layer is stored on Neo4J, the user can query the network using Cypher language to find a correlation between the nodes in the transcription and methylation layers. Cypher is Neo4j's query language to retrieve data from the graph and was inspired by SQL. Based on specific node annotation, the user can choose custom Cypher queries.

2.2 Colon Cancer Case Study

The Gene expression and methylation data of tumor samples and normal mucosa were obtained from The Cancer Genome Atlas (TCGA; https://www.cancer.gov/tcga) database. The samples were grouped into three classes, the "GAIN-chr20" with the amplification of chromosome 20, the "DIS-chr20" without amplification in that chromosome, and the mucosae group ("MU"). The corresponding clinical annotations and chromosomal assets were obtained from the cBioPortal database. The samples having both the RNA-seq and methylation data were selected: DIS-chr20 (n = 68), GAIN-chr20 (n = 89), MU (n = 19). Samples presenting microsatellite instability were inserted in this analysis to have a significant comparison.

2.3 Lung Cancer Case Study

TCGA-LUAD project samples were downloaded from The Cancer Genome Atlas (TCGA; https://www.cancer.gov/tcga) databases. We divided the samples into "upIWS1" and "downIWS1," identifying the distribution of IWS1 transcript across the samples. Since the distribution is normal (Kolmogorov-Smirnov test, p-value = 0.1309), we selected the samples using the threshold calculated as follows: $averageFPKM \pm \sigma(standarddeviation)$ (mean = 0.8, σ = 0.2). Gene expression and methylation data of tumor samples and normal tissue were downloaded. The samples having both the RNA-seq and methylation data were selected: upIWS1 (n = 95), downIWS1 (n = 47), normal tissues (n = 21).

2.4 Data Analysis

The raw counts and B-values were downloaded and pre-processed using the TCGAbiolinks [35] package. The statistical parameters used for the Boolean analysis in COAD samples were s = 4 for methylation data and s = 2 for RNA-seq data. Whereas s = 2 was used for expression and methylation data in LUAD samples. We used the default

parameters for the analysis. For each node, we added some custom attributes to query the network. For example, we added the results of the DEG analysis of GAIN-chr20vsDIS-chr20 and upIWS1vsDownIWS1 in the two case studies, respectively. The genes are indicated as "OVER_T" or "DOWN_T." In addition, the protein expression, mutation status, and the transcription factor annotation for each node were also inserted. The protein expression data were retrieved from The Cancer Protein Atlas—TCPA https://tcpaportal.org/tcpa/) for the selected TCGA-COAD samples. The expression was determined by calculating the unpaired two-sample t-test and the mean between the sample in the same group. The mutation data were collected in cBioPortal (https://www.cbioportal.org/) for each condition. The complete list of transcription factors was retrieved from the Human Transcription Factor database (http://humantfs.ccbr.utoronto.ca/) [36]. The insertion of custom information allowed us to query the multi-layer networks more comprehensively. The multi-layer networks were generated using Neo4J (https://neo4j.com/). The query results from the two conditions were compared to maintain only the exclusive nodes and edges. Then, the remaining nodes were subjected to Reactome enrichment analysis using the ReactomePA R package [37].

3 Results and Discussion

Our multi-layer approach is an easy-to-use bioinformatics pipeline to generate and analyze multi-layer networks combining multi-omics data. To show the capability of this method, we performed two case studies with colon and lung adenocarcinoma cancers data retrieved from TCGA. The user can use this method for a preliminary study to identify critical genes in a specific disease and evaluate the interplay between DNA methylation and transcription. Finally, the multi-layer approach is also helpful for the biomedical and research lab to filter the vast amount of RNA-seq and methylation data.

3.1 TCGA-COAD Results

The chromosome-20 amplification is a hallmark of colorectal cancer, along with the amplification of chromosome 13, chromosome 8, and chromosome 7 [38, 39]. In recent years, scientists have identified the importance of chr-20 amplification in colon cancer, mainly since it is the location of multiple oncogenes such as BCL2L1, AURKA, TPX2, and SRC [40–44]. However, despite numerous functional studies of various genes located in chromosome 20, the amplification of this chromosome in colon cancer remains under-investigated, among other critical molecular features [45]. By constructing a multi-layer network, we could identify remarkable pathways in COAD having the chr20 amplification.

Initially, we selected the COAD samples deposited on TCGA, distinguishing the samples having the amplification of chromosome 20 "GAIN-chr20" (Fig. 3A) and "DIS-chr20" (Fig. 3B), which do not show aberration in that chromosome. In the GAIN-chr20 group, chromosome 13 amplification appears in 72% of samples because of the correlation between the two chromosomal abnormalities [38, 39]. In addition, normal tissues were also analyzed as a control. To demonstrate how the usage of Boolean implication to generate the layer and, in general, how our approach is representative of disease

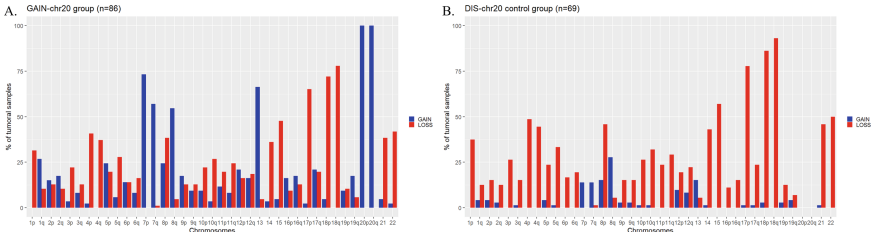


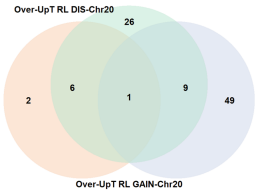
Fig. 3. (A) Chromosomal aberration distribution on TCGA GAIN-chr20 COAD samples (n = 89). 100% of samples show the amplification of chromosome 20. A small group also includes the amplification of chromosome 13. (B) Chromosomal aberration distribution on TCGA DIS-chr20 COAD samples (n = 69). The samples were selected without the amplification of chromosome 20.

profile, we compared the Over-UpT genes located on chromosome 20 listed in Condorelli et al. [38] with the nodes included in the RNAlayer of both multi-layers (GAIN-chr20 and DIS-chr20). A gene located on chromosome 20 is defined as Over-UpT if it is upregulated in cancer and further overexpressed by gene dosage ("positive caricature transcriptomic effect") [38, 39]. Firstly, all the nodes on chromosome 20 selected using the Boolean implication were compared with the Over-UpT genes reported in Condorelli et al.. 11 nodes are shared between the GAIN-chr20 RNAlayer and the Over-UpT list. These genes are absent in the DIS-chr20 RNAlayer. Then, more accurate research was conducted, selecting all the nodes with annotations "UP_REG" and "OVER_T." Although Over-UpT genes were detected using a different methodology (microarray in Condorelli et al. and RNA-seq in our approach) and a different logFC cutoff, 9 Over-UpT genes are in common (Fig. 4A). The list of shared genes, reported in Fig. 4B, shows bona fide cancer driver genes, already implicated in cancer pathogenesis [46–54]. The enrichment analysis of the common Over-UpT genes is reported in Fig. 4C. These genes are involved in cancer-critical pathways.

Using the proper Neo4j queries, it is possible to extrapolate detailed information about the phenotype studied.

Since the molecular signature that distinguishes the GAIN-chr20 from the DIS-chr20 samples is the amplification of chromosome 20, we searched in the two multi-layers for all the paths that involve nodes located on chromosome 20 (Fig. 5A). The DIS-chr20 network did not show any path. This result demonstrates how the multi-layer network method reproduces perfectly the condition studied. Figure 5 shows the sub-network structure resulting from the query (B) and the enriched pathways heat plot (C). For each pathway, the gene involved is indicated colored based on logFC generated from the comparison between GAIN-chr20 and DIS-chr20. One of the main pathways influenced is the "Signal Regulatory family interaction," in which the SRC transcript seems upregulated while the SIRP protein family results downregulated. The Src family kinase has been studied as a therapeutic target in colon cancer since it is usually overexpressed and associated with a worse prognosis [55, 56]. The multi-layer nodes annotation shows that the SRC protein is overexpressed in the analyzed data. In contrast, the downregulation of the SIRP family has been demonstrated fundamental for inhibiting cell apoptosis in prostate cancer [57].

A.



B.

Genes Over-UpT	logFC GAINvsMU	logFC GAINvsDIS
MYBL2	1,88	0,65
NELFCD	1,34	0,79
FERMT1	1,80	1,05
PLCB4	1,67	1,89
AHCY	1,48	0,78
CSEL1	1,60	0,64
DDX27	1,28	1,07
RPS21	1,30	0,96
CPNE1	1,56	1,20

C.

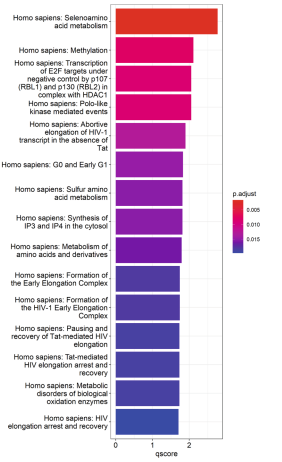
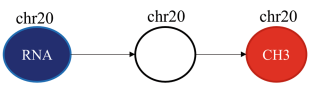
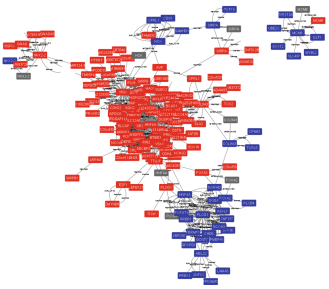


Fig. 4. (A)Venn diagram showing the common Over-UpT genes located on chromosome 20 between DIS-chr20 RNAlayer (RL) (orange circle),GAIN-chr20 RL (green circle) and Condorelli et al. list (purple circle). (B) List of Over-UpT common genes with the logFC comparing GAIN and Mucosae samples (logFC GAINvsMu) and GAIN and DIS samples (logFC GAINvsDIS). (C) Bar plot of enriched terms.

A.



B.



C.

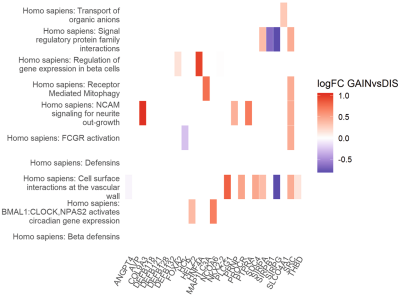


Fig. 5. (A) Cypher query pattern includes all the paths connecting nodes located on chromosome 20. (B) Schematic Cytoscape representation of paths selected by a custom query. The node colors depict the belonging layers: blue for the RNA layer, red for the CH_3 layer, and grey for the MP layer (C). The Heatmap reports the enrichment analysis starting from the nodes included in the sub-network. The logFC is generated from the difference between GAIN-chr20 and DIS-chr20.

3.2 TCGA-LUNG Results

IWS1 is one of the critical elements in lung adenocarcinoma since its phosphorylation influences the splicing and, consequently, the transcription of genes involved in cell proliferation [58]. Usually, the tumoral cells regulate the gene's transcription, promoting the alternative splicing to generate multiple transcripts and control cancer pathobiology

[59, 60]. Therefore, studying the multi-layer, we compared samples having upregulated IWS1 versus downregulated samples.

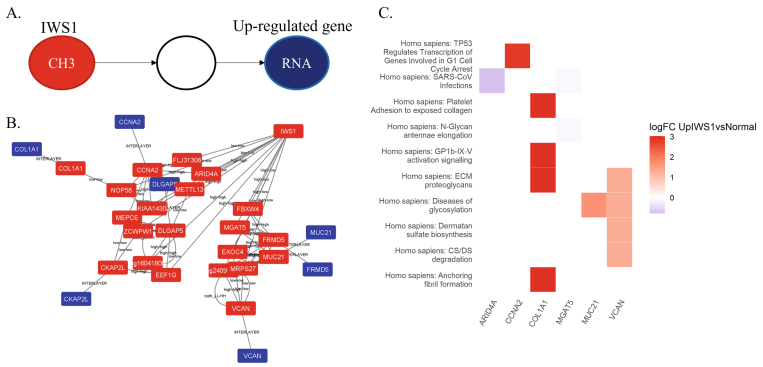


Fig. 6. (A) Custom query representation which retrieves the connection between the source node IWS1 in the CH_3 layer (red) and any target node annotated as "UP_REG" in the RNA layer (blue). Sub-network UpIWS1 obtained by Cypher query is depicted using Cytoscape in (B) No nodes lay in MP layer. (C) Heatmap reports the enrichment analysis starting from the nodes included in the sub-network. The logFC is generated from the difference between UpIWS1 and Normal tissue.

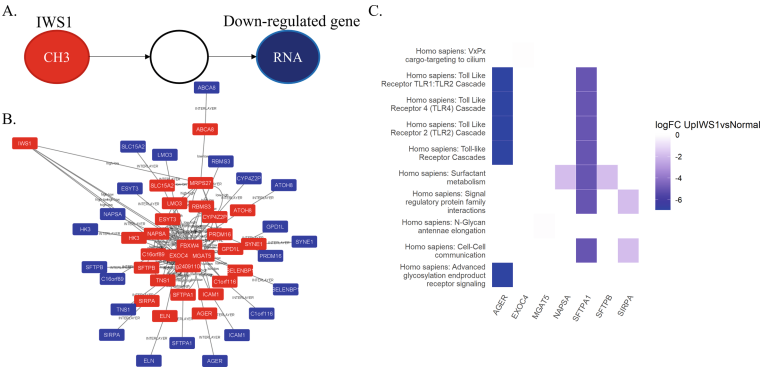


Fig. 7. (A) Designed Cypher query, which combines IWS1 node on CH_3 layer (red) with down-regulated RNA layer (blue) genes. (B) Schematic representation of the subnetwork resulting from the before cited query. (C) The enrichment analysis result for the selected nodes is reported as Heatmap. The logFC is generated from the difference between UpIWS1 and Normal tissue.

To identify the influence of IWS1 over other genes, the multi-layers were queried, looking for all the paths that connect IWS1 in the methylation layer with the upregulated genes (Fig. 6A) or the downregulated genes (Fig. 7A) in the expression layer. The query retrieved a result only for the upIWS1 multi-layer and not for the downIWS1 condition. The enrichment analysis (Fig. 6C) of the genes involved in the paths that connect IWS1 with upregulated genes shows an alteration of *CCNA2*, *COL1A1* and *MUC21* identified as key molecules in the lung adenocarcinoma progression [61–63]. On the other hand,

evaluating the enrichment analysis (Fig. 7C) on the genes obtained by Cypher query looking for the connection between IWS1 and downregulated genes, *AGER* and *SFTPA1* are highlighted. Both genes are usually downregulated in lung cancer and are validated as risk factors in LUAD [64–66].

4 Conclusions

In conclusion, the COMBO multi-layer approach is a new pipeline to query epigenomic and transcriptomic data to identify relevant signatures in a specific disease. Furthermore, the method's versatility allows the user to analyze several conditions focusing on different aspects of the pathology.

References

1. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.: Complex networks: structure and dynamics. Phys. Rep., 175–308 (2006)
2. De Domenico, M.: Multilayer network modeling of integrated biological systems: comment on "network science of biological systems at different scales: a review" by Gosak et al. Phys. Life Rev., 149–52 (2018)
3. Rai, A., Pradhan, P., Nagraj, J., Lohitesh, K., Chowdhury, R., Jalan, S.: Understanding cancer complexome using networks, spectral graph theory and multilayer framework. Sci. Rep. **7**, 41676 (2017)
4. Barabási, A.-L., Gulbahce, N., Loscalzo, J.: Network medicine: a network-based approach to human disease. Nature Rev. Genet., 56–68 (2011)
5. Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabási, A.-L.: The human disease network [Internet]. In: Proceedings of the National Academy of Sciences, pp. 8685–90 (2007)
6. Lv, Y., Huang, S., Zhang, T., Gao, B.: Application of multilayer network models in bioinformatics. Front. Genet. **12**, 664860 (2021)
7. Zheng, W., Wang, D., Zou, X.: Control of multilayer biological networks and applied to target identification of complex diseases. BMC Bioinform. **20**, 271 (2019)
8. Boccaletti, S., Bianconi, G., Criado, R., del Genio, C.I., Gómez-Gardeñes, J., Romance, M., et al.: The structure and dynamics of multilayer networks [Internet]. Phys. Rep., 1–122 (2014)
9. Mangioni, G., Jurman, G., De Domenico, M.: Multilayer flows in molecular networks identify biological modules in the human proteome [Internet]. IEEE Trans. Netw. Sci. Eng., 411–20 (2020)
10. Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer networks [Internet]. J. Compl. Netw., 203–71 (2014)
11. Kenett, D.Y., Perc, M., Boccaletti, S.: Networks of Networks—An Introduction [Internet]. Chaos, Solitons and Fractals, pp. 1–6 (2015)
12. Hammoud, Z., Kramer, F.: Multilayer networks: aspects, implementations, and application in biomedicine. Big Data Anal. (2020)
13. McGee, F., Ghoniem, M., Melançon, G., Otjacques, B., Pinaud, B.: The state of the art in multilayer network visualization. Comp. Graph. Forum. 125–49 (2019)
14. Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., et al.: Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinform. (2016)

15. Tordini, F., Aldinucci, M., Milanesi, L., Liò, P., Merelli, I.: The Genome Conformation As an Integrator of Multi-Omic Data: The Example of Damage Spreading in Cancer. Frontiers in Genetics (2016)
16. Zitnik, M., Leskovec, J.: Predicting multicellular function through multi-layer tissue networks. Bioinformatics., i190–8 (2017)
17. Gligorijević, V., Pržulj, N.: Methods for biological data integration: perspectives and challenges. J. Roy. Soc. Interface., 20150571 (2015)
18. Domenico, M.D., De Domenico, M., Porter, M.A., Arenas, A.: MuxViz: a tool for multilayer analysis and visualization of networks. J. Compl. Netw., 159–76 (2015)
19. De Bacco, C., Power, E.A., Larremore, D.B., Moore, C.: Community detection, link prediction, and layer interdependence in multilayer networks. Phys. Rev. E. **95**, 042317 (2017)
20. Škrlj, B., Kralj, J., Lavrač, N.: Py3plex toolkit for visualization and analysis of multilayer networks. Appl. Netw. Sci. (2019)
21. Hammoud, Z., Kramer, F.M.: An R Package to Create, Modify and Visualize Multilayered Graph. Genes, p. 519 (2018)
22. Sahoo, D., Dill, D.L., Tibshirani, R., Plevritis, S.K.: Extracting binary signals from microarray time-course data. Nucl. Acids Res., 3705–12 (2007)
23. Sahoo, D., Dill, D.L., Gentles, A.J., Tibshirani, R., Plevritis, S.K.: Boolean implication networks derived from large scale, whole genome microarray datasets. Genome Biol. **9**, R157 (2008)
24. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., et al.: limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucl. Acids Res. **43**, e47 (2015)
25. Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., et al.: Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinform. BioMed. Central **11**, 1–9 (2010)
26. Sgariglia, D., Conforte, A.J., Pedreira, C.E., de Carvalho, L.A.V., Carneiro, F.R.G., Carels, N., et al.: Data-Driven Modeling of Breast Cancer Tumors Using Boolean Networks. Frontiers in Big Data [Internet]. Frontiers Media SA (2021)
27. Xu, X., Zhu, L., Yang, Y., Pan, Y., Feng, Z., Li, Y., et al.: Low tumour PPM1H indicates poor prognosis in colorectal cancer via activation of cancer-associated fibroblasts. Br J Cancer. Nature Publ. Group **120**, 987–995 (2019)
28. Dabydeen, S.A., Desai, A., Sahoo, D.: Unbiased Boolean Analysis of Public Gene Expression Data for Cell Cycle Gene Identification. The American Society for Cell Biology, Mol Biol Cell (2019)
29. Sahoo, D., Wei, W., Auman, H., Hurtado-Coll, A., Carroll, P.R., Fazli, L., et al.: Boolean analysis identifies CD38 as a biomarker of aggressive localized prostate cancer. Oncotarget. Impact J. **9**, 6550–6561 (2018)
30. da Mata, A.S., da Mata, A.S.: Complex networks: a mini-review [Internet]. Brazilian J. Phys. 658–72 (2020)
31. Kinsley, A.C., Rossi, G., Silk, M.J., VanderWaal, K.: Multilayer and multiplex networks: an introduction to their use in veterinary epidemiology. Front. Vet. Sci. **7**, 596 (2020)
32. Alaimo, S., Giugno, R., Acunzo, M., Veneziano, D., Ferro, A., Pulvirenti, A.: Post-transcriptional knowledge in pathway analysis increases the accuracy of phenotypes classification. Oncotarget **7**, 54572–54582 (2016)
33. Alaimo, S., Marceca, G.P., Ferro, A., Pulvirenti, A.: Detecting disease specific pathway substructures through an integrated systems biology approach. Noncoding RNA. **3** (2017)
34. Alaimo, S., Rapicavoli, R.V., Marceca, G.P., La Ferlita, A., Serebrennikova, O.B., Tsichlis, P.N., et al.: PHENSIM: phenotype simulator. PLoS Comput. Biol. **17**, e1009069 (2021)

35. Silva, T.C., Colaprico, A., Olsen, C., D'Angelo, F., Bontempi, G., Ceccarelli, M., et al.: TCGA workflow: analyze cancer genomics and epigenomics data using bioconductor packages. F1000 Res., 1542 (2016)
36. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., et al.: The human transcription factors. Cell **175**, 598–599 (2018)
37. Yu, G., He, Q.-Y.: ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol. Biosyst. 477–9 (2016)
38. Condorelli, D.F., Spampinato, G., Valenti, G., Musso, N., Castorina, S., Barresi, V.: Positive Caricature Transcriptomic Effects Associated with Broad Genomic Aberrations in Colorectal Cancer. Scientific Reports (2018)
39. Condorelli, D.F., Privitera, A.P., Barresi, V.: Chromosomal density of cancer up-regulated genes, aberrant enhancer activity and cancer fitness genes are associated with transcriptional cis-effects of broad copy number GAINs in colorectal cancer. Int. J. Mol. Sci. **20** (2019)
40. Sillars-Hardebol, A.H., Carvalho, B., Beliën, J.A.M., de Wit, M., Delis-van Diemen, P.M., Tijssen, M., et al.: *BCL2L1*has a functional role in colorectal cancer and its protein expression is associated with chromosome 20q GAIN. J. Pathol. 442–50 (2012)
41. Carvalho, B., Postma, C., Mongera, S., Hopmans, E., Diskin, S., van de Wiel, M.A., et al.: Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. Gut **58**, 79–89 (2009)
42. Sillars-Hardebol, A.H., Carvalho, B., Tijssen, M., Beliën, J.A.M., de Wit, M., Delis-van Diemen, P.M., et al.: TPX2 and AURKA promote 20q amplicon-driven colorectal adenoma to carcinoma progression. Gut **61**, 1568–1575 (2012)
43. Ptashkin, R.N., Pagan, C., Yaeger, R., Middha, S., Shia, J., O'Rourke, K.P., et al.: Chromosome 20q amplification defines a subtype of microsatellite stable, left-sided colon cancers with wild-type RAS/RAF and better overall survival. Mol. Cancer Res. (2017)
44. Voutsadakis, I.A.: Chromosome 20q11.21 amplifications in colorectal cancer. Cancer Genom. Proteom. **18**, 487–96 (2021)
45. Bui, V.M.H., Mettling, C., Jou, J., Sun, H.S.: Genomic amplification of chromosome 20q13.33 is the early biomarker for the development of sporadic colorectal carcinoma. BMC Med. Genom. **13**, 149 (2020)
46. Liu, Q., Guo, L., Qi, H., Lou, M., Wang, R., Hai, B., et al.: A MYBL2 complex for RRM2 transactivation and the synthetic effect of MYBL2 knockdown with WEE1 inhibition aGAINst colorectal cancer. Cell Death Dis. Nature Publ. Group **12**, 1–11 (2021)
47. Song, S., Li, D., Yang, C., Yan, P., Bai, Y., Zhang, Y., et al.: Overexpression of NELFCD promotes colorectal cancer cells proliferation, migration, and invasion. Oncol. Targ. Ther. Dove Press **11**, 8741 (2018)
48. Li, L., Li, P., Zhang, W., Zhou, H., Guo, E., Hu, G., et al.: FERMT1 contributes to the migration and invasion of nasopharyngeal carcinoma through epithelial–mesenchymal transition and cell cycle arrest. Cancer Cell Int. BioMed. Central **22**, 1–14 (2022)
49. Yang, C., Li, D., Bai, Y., Song, S., Yan, P., Wu, R., et al.: DEAD-box helicase 27 plays a tumor-promoter role by regulating the stem cell-like activity of human colorectal cancer cells. Oncol. Targ. Ther. Dove Press **12**, 233 (2019)
50. Wu, S., Zhang, W., Shen, D., Lu, J., Zhao, L.: PLCB4 upregulation is associated with unfavorable prognosis in pediatric acute myeloid leukemia. Oncol. Lett. Spandidos Publ. **18**, 6057 (2019)
51. Belužić, L., Grbeša, I., Belužić, R., Park, J.H., Kong, H.K., Kopjar, N., et al.: Knock-down of AHCY and depletion of adenosine induces DNA damage and cell cycle arrest. Sci. Rep. Nature Publ. Group **8**, 1–16 (2018)
52. Pimiento, J.M., Neill, K.G., Henderson-Jackson, E., Eschrich, S.A., Chen, D.T., Husain, K., et al.: Knockdown of CSE1L gene in colorectal cancer reduces tumorigenesis in vitro. Am. J. Pathol. (2016)

53. El Khoury, W., Nasr, Z.: Deregulation of ribosomal proteins in human cancers. Biosci Rep. **41** (2021)
54. Wang, Y., Pan, S., He, X., Wang, Y., Huang, H., Chen, J., et al.: CPNE1 Enhances Colorectal Cancer Cell Growth, Glycolysis, and Drug Resistance Through Regulating the AKT-GLUT1/HK2 Pathway, Vol. 14. Onco Targets Ther. Dove Press, p. 699 (2021)
55. Chen, J., Elfiky, A., Han, M., Chen, C., Saif, M.W.: The role of Src in colon cancer and its therapeutic implications. Clin. Colorectal Cancer. **13**, 5–13 (2014)
56. Jin, W.: Regulation of Src Family Kinases during Colorectal Cancer Development and Its Clinical Implications. Cancers, pp. 12 (2020)
57. Yao, C., Li, G., Cai, M., Qian, Y., Wang, L., Xiao, L., et al.: Prostate cancer downregulated SIRP-α modulates apoptosis and proliferation through p38-MAPK/NF-κB/COX-2 signaling. Oncol. Lett. **13**, 4995–5001 (2017)
58. Sanidas, I., Polytarchou, C., Hatziapostolou, M., Ezell, S.A., Kottakis, F., Hu, L., et al.: Phosphoproteomics screen reveals akt isoform-specific signals linking RNA processing to lung cancer. Mol. Cell. **53**, 577–590 (2014)
59. Paronetto, M.P., Passacantilli, I., Sette, C.: Alternative splicing and cell survival: from tissue homeostasis to disease. Cell Death Differ. **23**, 1919–1929 (2016)
60. Coomer, A.O., Black, F., Greystoke, A., Munkley, J., Elliott, D.J.: Alternative splicing in lung cancer. Biochim. Biophys. Acta Gene Regul. Mech. **1862**, 194388 (2019)
61. Yoshimoto, T., Matsubara, D., Soda, M., Ueno, T., Amano, Y., Kihara, A., et al.: Mucin 21 is a key molecule involved in the incohesive growth pattern in lung adenocarcinoma. Cancer Sci. **110**, 3006–3011 (2019)
62. Hou, L., Lin, T., Wang, Y., Liu, B., Wang, M.: Collagen type 1 alpha 1 chain is a novel predictive biomarker of poor progression-free survival and chemoresistance in metastatic lung cancer. J. Cancer. **12**, 5723–5731 (2021)
63. Ruan, J.S., Zhou, H., Yang, L., Wang, L., Jiang, Z.S., Wang, S.M.: CCNA2 facilitates epithelial-to-mesenchymal transition via the integrin αvβ3 signaling in NSCLC. Int. J. Clin. Exp. Pathol. **10**, 8324–8333 (2017)
64. Serveaux-Dancer, M., Jabaudon, M., Creveaux, I., Belville, C., Blondonnet, R., Gross, C., et al.: Pathological implications of receptor for advanced glycation end-product gene polymorphism. Dis. Markers. **2019**, 2067353 (2019)
65. Zhang, W., Fan, J., Chen, Q., Lei, C., Qiao, B., Liu, Q.: SPP1 and AGER as potential prognostic biomarkers for lung adenocarcinoma. Oncol. Lett. **15**, 7028–7036 (2018)
66. Yuan, L., et al.: SFTPA1 is a potential prognostic biomarker correlated with immune cell infiltration and response to immunotherapy in lung adenocarcinoma. Cancer Immunol. Immunother. **71**(2), 399–415 (2021). https://doi.org/10.1007/s00262-021-02995-4