HHS Public Access

Author manuscript

*Proceedings (IEEE Int Conf Bioinformatics Biomed)*. Author manuscript; available in PMC 2017 December 01.

Published in final edited form as:

*Proceedings (IEEE Int Conf Bioinformatics Biomed)*. 2016 December ; 2016: 1302–1306. doi:10.1109/BIBM.2016.7822706.

# Transcriptional Responses to Ultraviolet and Ionizing Radiation: An Approach Based on Graph Curvature

**Yongxin Chen**[1,†], **Jung Hun Oh**[1,†], **Romeil Sandhu**[2], **Sangkyu Lee**[1], **Joseph O. Deasy**[1], and **Allen Tannenbaum**[3,*]

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY, USA

[2]Department of Biomedical Informatics, Stony Brook University, NY, USA

[3]Department of Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY, USA

## Abstract

More than half of all cancer patients receive radiotherapy in their treatment process. However, our understanding of abnormal transcriptional responses to radiation remains poor. In this study, we employ an extended definition of Ollivier-Ricci curvature based on LI-Wasserstein distance to investigate genes and biological processes associated with ionizing radiation (IR) and ultraviolet radiation (UV) exposure using a microarray dataset. Gene expression levels were modeled on a gene interaction topology downloaded from the Human Protein Reference Database (HPRD). This was performed for IR, UV, and mock datasets, separately. The difference curvature value between IR and mock graphs (also between UV and mock) for each gene was used as a metric to estimate the extent to which the gene responds to radiation. We found that in comparison of the top 200 genes identified from IR and UV graphs, about 20~30% genes were overlapping. Through gene ontology enrichment analysis, we found that the metabolic-related biological process was highly associated with both IR and UV radiation exposure.

## I. Introduction

Approximately 60% cancer patients receive radiotherapy in the treatment process [1]. Radiation-induced toxicity is a common side effect for patients treated with radiotherapy. Therefore, it is important to find the biological processes implicated in radiation to develop personalized treatment for those who are predicted as being at high risk of developing radiation-induced side effects. Oh *et al.* surveyed many published studies that had shown associations between genes and radiation exposure, and summarized a list of 221 radiosensitive genes and the corresponding biological processes [2]. Eschrich et al. proposed a linear regression predictive model of cellular radiosensitivity using the 10 hub genes from the top 500 genes identified by a univariate linear regression [3]. Jen and Cheung assessed transcriptional response of lymphoblastoid cells to ionizing radiation (IR) at various time points with 3 Gy and 10 Gy of ex vivo IR exposure [4]. They found that the higher radiation

*Correspondence: allen.tannenbaum@stonybrook.edu.

†These authors contributed equally to this work

dose exposure induced transcriptional changes in a more number of genes. Popanda et al. performed a literature review to identify radioresponsive single nucleotide polymorphisms (SNPs) associated with irradiation [5]. Andreassen and Alsner summarized studies that had reported associations between genetic variants and normal tissue complications in various cancers and proposed an allelic architecture model that shows relative risk for genetic variants associated with normal tissue radiosensitivity [6]. Rieger and Chu investigated radiosensitive genes and biological processes associated with IR and ultraviolet radiation (UV) using cell lines from 15 individuals [7].

In this study, we reanalyze this dataset by employing an extension of the notion of Ricci curvature to the case of weighted graphs known as *Ollivier-Ricci curvature*, and apply this to study a gene interaction topology. We rank genes based on the curvature difference relative to mock after IR and UV treatment and perform gene ontology enrichment analysis to identify significant radiosensitive biological processes.

Our motivation for employing curvature to study cancer comes from the recent work of modeling various biological systems as complex networks. In general, the growing importance of studying complex networks has been documented in a huge and growing literature, and has even been referred to as the field of *network science* [8]. In particular, as argued in various works (see [9] and the references therein), the onset and proliferation of cancer stems from dynamic changes that result from a series of changes in cellular interactions governing a complex network. As described in [10], there is a strong relationship between network functionality in terms of robustness and topological and geometric properties of networks such as curvature. One of the key ideas underpinning the present study, is based on the positive correlation between an increase of curvature and network functional robustness. Since a fundamental hurdle to cancer therapy is to acquire tumor robustness, it is essential to quantify the robustness of cancer networks in some easily computable manner. The notion of curvature as described below, turns out to be a powerful technique for accomplishing this purpose.

## II. Curvature

In the theory of differential geometry, curvature is the amount by which a geometric object deviates from being flat. Ricci curvature is a notion of curvature that captures this change along some specific direction. In [11], Ollivier extended the concept of Ricci curvature to general metric measure spaces, in particular, weighted graphs. This is of great interest to us since the graph may represent gene regulatory networks or other biological networks. Next we describe the Ollivier notion of curvature on graphs.

Ollivier curvature relies on the Wasserstein distance, which is a metric based on the theory of optimal mass transport problems [12], [13]. Let $X$ be a metric measure space equipped with distance $d$. On a graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$ with nodes $\mathcal{V}$ and edges $\mathcal{E}$, one can simply choose the distance $d$ as the hop distance. That is, the distance between two nodes $x,y\in\mathcal{V}$ is given by the minimum of number of hops to go from $x$ to $y$ or vice versa. Given two probability measures $\mu$, $\nu$ on $X$ with finite $p$-th ($1\leq p<\infty$) moments, one can define the $L^p$ Wasserstein distance between $\mu$ and $\nu$ as

$$W_p(\mu,\nu)=\left(\inf_{\pi\in\Pi(\mu,\nu)}\int_{X\times X}d(x,y)^p d\pi(x,y)\right)^{1/p},$$

where $\Pi(\mu, \nu)$ denotes a set of all joint probability measures on $X \times X$ whose marginals are $\mu$ and $\nu$. This means that the Wasserstein distance is associated with the minimum cost of moving mass distributed according to $\mu$, to mass distributed according to $\nu$, when the cost of moving unit mass from $x$ to $y$ is $d(x, y)^p$. In most applications, $p = 1, 2$ were used, and following [11], we will take $p = 1$. This is also known as the Earth Mover's Distance (EMD) [14]. We note that when the initial and terminal mass concentrate on $x$ and $y$ respectively, i.e., $\mu = \delta_x$, $\nu = \delta_y$, then $W_p(\mu, \nu) = d(x, y)$. The Ollivier-Ricci curvature captures the change of transportation cost after adding small diffusions to the mass distributions. If we attach to each point $x \in X$ a probability measure $\mu_x$, which corresponds to the diffusion at $x$ or "fuzzifying" the point $x$, then the Ollivier-Ricci curvature is defined as

$$\kappa(x,y)=1-\frac{W_1(\mu_x,\mu_y)}{d(x,y)}.$$

When the curvature is positive, $W_1$ is less than $d$, which implies that a small diffusion would help to reduce the transportation cost. On the other hand, if the curvature is negative, then $W_1$ is greater than $d$, which implies that diffusion would increase the transportation cost. Let $\mathcal{G}=(\mathcal{V},\mathcal{E})$ denote a weighted undirected graph with vertices $\mathcal{V}$ and edges $\mathcal{E}$. Let $w_{xy}$ denote the weight of the edge $(x, y) \in \mathcal{E}$. We assume that $w_{xy}$ is positive. For any pair of points $x$ and $y$ that are not directly connected on the graph, we define $w_{xy} = 0$. We define a probability measure $\mu_x$ for a given node $x \in \mathcal{V}$ by

$$\mu_x(y)=\frac{w_{xy}}{d_x},$$
$$d_x=\sum_y w_{xy}.$$

Namely, the mass at $x$ spreads to the neighbors of $x$, and the amount of mass is proportional to the weight between them. The Ollivier-Ricci curvature $Ric$ and the entropy $S$ are closely related. It was established in [10] that they are actually positively correlated, namely,

$$\Delta S\times\Delta Ric\geq 0.$$

Furthermore, via the Fluctuation Theorem [15], [16], [17] entropy and robustness, the latter denoted by $R$, are positively correlated, namely,

$$\Delta S\times\Delta R\geq 0.$$

Therefore, one deduces the positive correlation

$$\Delta Ric \times \Delta R \geq 0$$

between the curvature *Ric* and the robustness *R* follows (see [10] for more details). The intuition behind this is that graphs with higher curvature will have more alternative ways to transport mass or information from one node to another, meaning that the possible damage caused by a random perturbation will be smaller, and therefore the network will be more robust. Further, from the above discussion, higher curvature is an indication of greater heterogeniety, another hallmark of cancer.

Based on the Ollivier-Ricci curvature, we can define a scalar curvature for each node $x \in \mathcal{V}$ as the sum over all the Ricci curvature values between *x* and its neighbors, namely,

$$\eta(x) := \sum_{y \in N_x} \kappa(x, y),$$

where $N_x$ is a set of nodes that are directly connected to *x*. To avoid the bias induced by the topology, one can also incorporate the weights and define a scalar curvature as $\eta_w(x) := \sum_{y \in N_x} w_{xy} \kappa(x, y)$. Note that unlike $\kappa$ which gives a value to each pair of interaction, the scalar curvature $\eta$ assigns a value to each node. Therefore, it provides a possible way to compare different genes on a network since each gene corresponds to a node.

## III. Materials and Methods

To investigate transcriptional responses to radiation, we analyzed a gene microarray dataset (GSE1977) downloaded from the Gene Expression Omnibus (GEO) repository, which consisted of gene expression levels of ~10000 genes for 45 samples: lymphoblastoid cell lines collected from 15 healthy individuals (mock group), 15 IR-treated samples (IR group), and 15 UV-treated samples (UV group). For detailed information, refer to [7]. The network topology was constructed using gene interaction information derived from the Human Protein Reference Database (HPRD) [18]. After incorporating the gene expression data, and discarding redundant genes, a final graph consisted of 5568 nodes (genes) and 23689 edges (interactions). For each of the three treatments (mock, IR, and UV), we computed the Pearson correlation $c_{xy}$ between two connected genes (*x* and *y*) and assigned a non-negative weight $w_{xy} = (1 + c_{xy})/2$ on the edge $(x, y) \in \mathcal{E}$. With this weighting strategy, we built 3 different graphs ( $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$) on the same topology, but with different weights for the mock, IR, and UV groups, respectively. We computed the scalar curvatures for all genes on the network topology. After that, for each gene, we computed the curvature difference between IR and mock graphs, and between UV and mock graphs. The absolute difference value was used as a metric to assess the extent to which each gene responds to radiation. Based on the curvature difference, we ranked genes and performed gene ontology enrichment analysis using a curated database (MetaCore, Thomson Reuters).

## IV. Results

We computed the scalar curvatures on the 3 graphs and assessed the curvature change for each gene after IR and UV radiation, i.e, mock–IR and mock–UV. As shown in Figure 1, overall the differences were small for most of the genes in response to IR and UV treatment. Only a small portion of genes showed significant changes in curvature. We selected the top 200 genes with the most significant changes: the top 100 genes with a curvature increase (negative difference) relative to mock results and another top 100 genes with a curvature decrease (positive difference). Increasing the number of the top ranked genes from 10 to 100 in steps of 10, we compared these genes to investigate how many genes responded to both UV and IR treatment. As can be seen in Figure 2, about 20~30% percent genes were overlapping with a slightly higher percentage for genes with the negative curvature difference. Table I shows the top 10 genes with the most significant curvature changes after IR and UV treatment for the positive difference and negative difference, separately. For IR treatment, TP53 and MYC were top-ranked, whereas for UV treatment CALM1 and AR were top-ranked with respect to the positive and negative difference, respectively. It should be noted that in our previous research [2], MYC gene was found to be the most significant radiosensitive gene. The confirmation using the Ollivier-Ricci curvature approach in this study implies that MYC appears to play a crucial role in radiation-induced biological processes. We performed gene ontology enrichment analysis with the top 200 genes using the MetaCore. Figures 3 and 4 show the top 10 biological processes for IR and UV treatment, respectively, resulting from the MetaCore analysis. Interestingly, the metabolic-related process was top-ranked in both treatments, implying that there may be common biological processes in response to both IR and UV radiation.

## V. Conclusion

We employed a graph-based curvature concept to identify radiosensitive genes and biological processes using a microarray dataset. Using the idea that the change of robustness of a biological network after radiation treatment is positively correlated with the network curvature, we computed the curvature differences between mock and IR or UV treatment for all genes and used the difference value as a metric to assess the degree to which each gene responds to radiation. It was found that about 20~30% genes among the top-ranked genes responded to both IR and UV radiation. Despite the relatively low overlapping fraction, the metabolic-related process was top-ranked in both IR and UV treatment, suggesting that the key biological processes associated with IR and UV radiation exposure appear to be similar. However, a further evaluation on larger datasets is needed to elucidate these observations.

## References

1. Rieger KE, Hong WJ, Tusher VG, Tang J, Tibshirani R, Chu G. Toxicity from radiation therapy associated with abnormal transcriptional responses to dna damage. Proceedings of the National Academy of Sciences of the USA. 2004; 101(17):6635–6640. [PubMed: 15096622]
2. Oh JH, Wong HP, Wang X, Deasy JO. A bioinformatics filtering strategy for identifying radiation response biomarker candidates. PLoS One. 2012; 7(6):e38870. [PubMed: 22768051]

3. Eschrich S, Zhang H, Zhao H, Boulware D, Lee JH, Bloom G, Torres-Roca JF. Systems biology modeling of the radiation sensitivity network: a biomarker discovery platform. International journal of radiation oncology, biology, physics. 2009; 75(2):497–505.
4. Jen KY, Cheung VG. Transcriptional response of lymphoblastoid cells to ionizing radiation. Genome Research. 2003; 13(9):2092–2100. [PubMed: 12915489]
5. Popanda O, Marquardt JU, Chang-Claude J, Schmezer P. Genetic variation in normal tissue toxicity induced by ionizing radiation. Mutation Research. 2009; 667:58–69. [PubMed: 19022265]
6. Andreassen CN, Alsner J. Genetic variants and normal tissue toxicity after radiotherapy: a systematic review. Radiotherapy and oncology. 2009; 92(3):299–309. [PubMed: 19683821]
7. Rieger KE, Chu G. Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells. Nucleic Acids Research. 2004; 32(16):4786–4803. [PubMed: 15356296]
8. Barabasi A. The network takeover. Nature Physics. 2012; 8:14–16.
9. West J, Bianconi G, Severini S, Teschendorff A. Differential network entropy reveals cancer system hallmarks. Scientific Reports. 2012; 2(802)
10. Sandhu R, Georgiou T, Reznik E, Zhu L, Kolesov I, Senbabaoglu Y, Tannenbaum A. Graph curvature for differentiating cancer networks. Scientific Reports. 2015; 5
11. Ollivier Y. Ricci curvature of markov chains on metric spaces. Journal of Functional Analysis. 2009; 256(3):810–864.
12. Rachev, ST., Rüschendorf, L. Mass Transportation Problems: Volume I: Theory. Vol. 1. Springer Science & Business Media; 1998.
13. Villani, C. Topics in optimal transportation. American Mathematical Soc; 2003.
14. Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision. 2000; 40(2):99–121.
15. Demetrius LA. Boltzmann, darwin and directionality theory. Physics Reports. 2013; 530(1):1–85.
16. West J, Bianconi G, Severini S, Teschendorff AE. Differential network entropy reveals cancer system hallmarks. Scientific Reports. 2012; 2
17. Demetrius L, Manke T. Robustness and network evolutionan entropic principle. Physica A. 2004; 346:682–696.
18. Prasad, TS Keshava, Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, DS., Sebastian, A., Rani, S., Ray, S., Kishore, CJ Harrys, Kanth, S., Ahmed, M., Kashyap, MK., Mohmood, R., Ramachandra, YL., Krishna, V., Rahiman, BA., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R., Pandey, A. Human protein reference database-2009 update. Nucleic Acids Research. 2009; 37:D767–D772. [PubMed: 18988627]

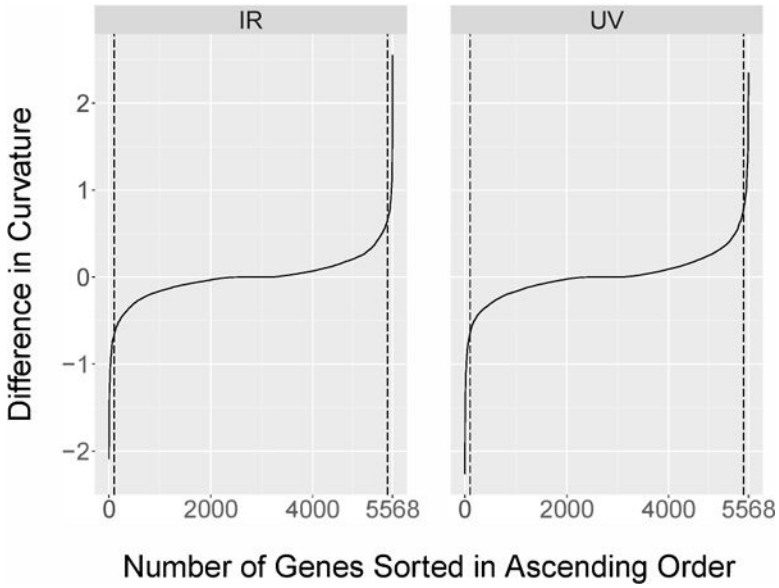


**Fig. 1.**

Sorted curvature values in ascending order. Left: curvature differences between mock and IR graphs (mock–IR) and right: curvature differences between mock and UV graphs (mock–UV) for all genes.

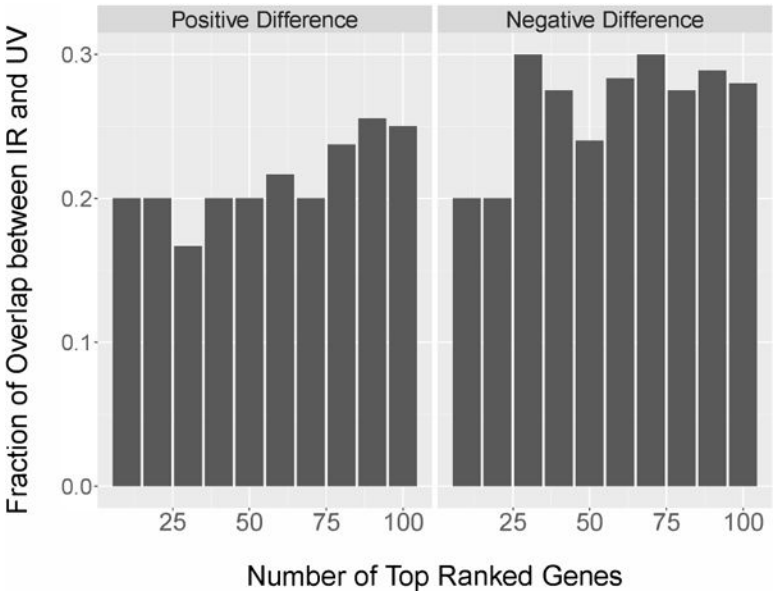


**Fig. 2.**

Faction of overlapping genes between IR and UV treatment. Left: the top 100 genes with a curvature decrease (positive curvature difference) and right: the top 100 genes with a curvature increase (negative curvature difference).

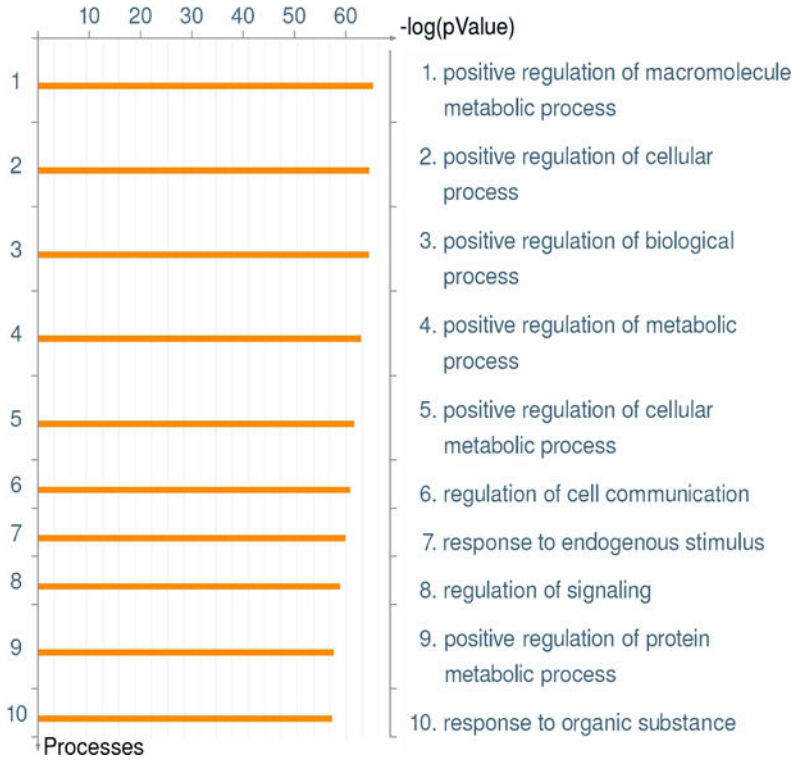


**Fig. 3.**

Gene ontology enrichment analysis using the top 200 genes associated with IR treatment.

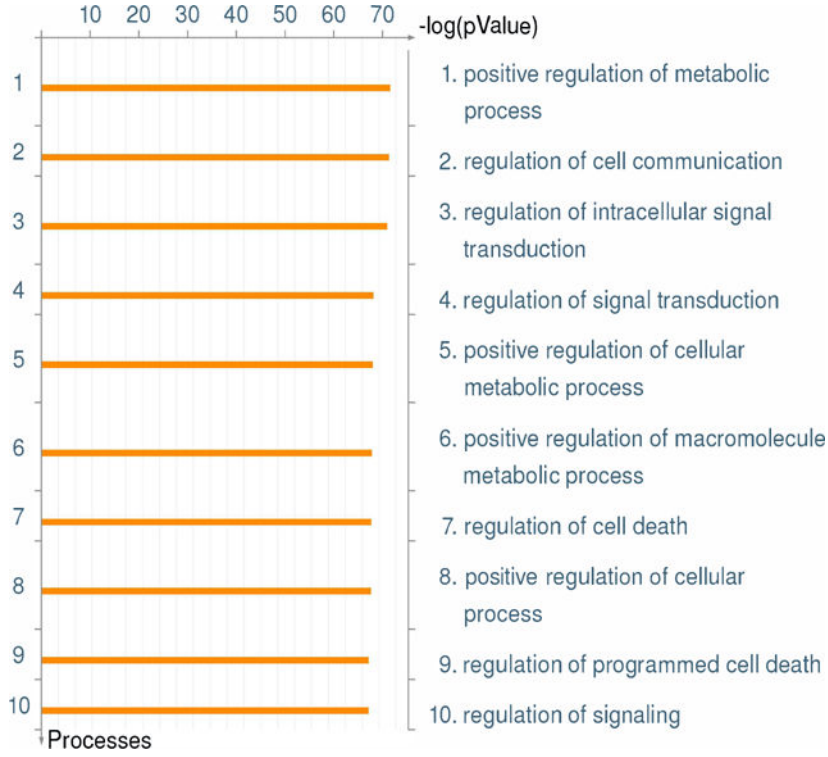


**Fig. 4.**

Gene ontology enrichment analysis using the top 200 genes associated with UV treatment.

**TABLE I**

Top ranked genes based on the curvature changes after IR and UV treatment. +: positive difference; −: negative difference.

| | IR | | | | UV | | | |
|---|---|---|---|---|---|---|---|---|
| **Ranking** | **Gene** | **+** | **Gene** | **−** | **Gene** | **+** | **Gene** | **−** |
| 1 | TP53 | 2.550 | MYC | −2.091 | CALM1 | 2.348 | AR | −2.261 |
| 2 | UNC119 | 1.898 | PRKACA | −1.886 | TSNAX | 2.232 | MET | −1.672 |
| 3 | RPS6KB1 | 1.825 | PIK3R1 | −1.725 | JUN | 2.109 | YWHAQ | −1.662 |
| 4 | SMAD3 | 1.556 | POU2F1 | −1.567 | EP300 | 2.078 | MAPK3 | −1.503 |
| 5 | GNB2L1 | 1.464 | RAF1 | −1.564 | TP53 | 2.005 | HRAS | −1.448 |
| 6 | NCOA6 | 1.320 | CDK5 | −1.540 | ERBB2 | 1.831 | PPP3CA | −1.403 |
| 7 | CDKI | 1.300 | MET | −1.477 | EEF1A1 | 1.611 | IRS1 | −1.380 |
| 8 | NFKBIB | 1.141 | PTK2 | −1.420 | SMAD4 | 1.571 | MYC | −1.365 |
| 9 | CHD3 | 1.138 | PTPN11 | −1.330 | SMAD1 | 1.550 | CRMP1 | −1.319 |
| 10 | HTT | 1.138 | VAMP2 | −1.312 | SMAD3 | 1.505 | NR2F1 | −1.272 |

*Proceedings (IEEE Int Conf Bioinformatics Biomed)*. Author manuscript; available in PMC 2017 December 01.