

## Environment and bladder cancer: molecular analysis by interaction networks

Andrea Polo<sup>1</sup>, Anna Crispo<sup>1</sup>, Pellegrino Cerino<sup>2</sup>, Luca Falzone<sup>3</sup>, Saverio Candido<sup>3</sup>, Aldo Giudice<sup>1</sup>, Giuseppina De Petro<sup>4</sup>, Gennaro Ciliberto<sup>5</sup>, Maurizio Montella<sup>1</sup>, Alfredo Budillon<sup>6</sup> and Susan Costantini<sup>6</sup>

<sup>1</sup>Epidemiology Unit, Istituto Nazionale Tumori "Fondazione G. Pascale", IRCCS, Napoli, Italia

<sup>2</sup>Istituto Zooprofilattico Sperimentale del Mezzogiorno (IZSM), Napoli, Italia

<sup>3</sup>Department of Biomedical and Biotechnological Sciences, Section of General and Clinical Pathology and Oncology – Translational Oncology and Functional Genomics Laboratory, University of Catania, Catania, Italy

<sup>4</sup>Dipartimento di Medicina Molecolare e Traslazionale, Università di Brescia, Brescia, Italia

<sup>5</sup>Scientific Directorate, Istituto Nazionale Tumori "Regina Elena", IRCCS, Roma, Italia

<sup>6</sup>Experimental Pharmacology Unit, Istituto Nazionale Tumori "Fondazione G. Pascale", IRCCS, Napoli, Italia

**Correspondence to:** Alfredo Budillon, **email:** a.budillon@istitutotumori.na.it  
Susan Costantini, **email:** s.costantini@istitutotumori.na.it

**Keywords:** bladder cancer, network analysis, environmental exposure, arsenicals

**Received:** March 16, 2017

**Accepted:** April 12, 2017

**Published:** May 26, 2017

**Copyright:** Polo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

**Bladder cancer (BC) is the 9th most common cancer worldwide, and the 6th most common cancer in men. Its development is linked to chronic inflammation, genetic susceptibility, smoking, occupational exposures and environmental pollutants. Aim of this work was to identify a sub-network of genes/proteins modulated by environmental or arsenic exposure in BC by computational network approaches. Our studies evidenced the presence of HUB nodes both in "BC and environment" and "BC and arsenicals" networks. These HUB nodes resulted to be correlated to circadian genes and targeted by some miRNAs already reported as involved in BC, thus suggesting how they play an important role in BC development due to environmental or arsenic exposure. Through data-mining analysis related to putative effect of the identified HUB nodes on survival we identified genes/proteins and their mutations on which it will be useful to focus further experimental studies related to the evaluation of their expression in biological matrices and to their utility as biomarkers of BC development.**

### INTRODUCTION

Bladder cancer (BC) is the 9th most common cancer worldwide, and the 4th most common cancer in more developed world [1–3]. In the last years, BC risk has been linked to the smoking, to the occupational exposures [4], to the use of drugs like cyclophosphamide or chlornaphazine [5], to the environmental pollutants and the arsenicals [6, 7], and to the chronic inflammation with the related up-regulation of some pro-inflammatory proteins like interleukin (IL)-6, tumor necrosis factor (TNF) and C-reactive protein (CRP) [8]. Moreover, a strong correlation between altered expression of various

clock genes (PER1 and PER2) and common tumor markers (TP53, PTEN and PAI-1) evidenced that a disturbed function in the cellular clock system can clearly represent an another mechanism of BC progression [9].

In particular, many studies reported experimental evidences about the role of the environmental chemicals in BC carcinogenesis. In fact, the increased COX-2 expression and the activation of the mitogen activated protein kinase pathway in BC cells exposed to monomethylarsonous acid, a metabolite of inorganic arsenic, resulted a strong mechanism for BC carcinogenesis [10]. Moreover, it has been demonstrated in mice and rats the BC carcinogenicity of both 3-amino-1-methyl-5H-pyrido[4,3-b]indole, that is

formed in cooked meat and fish [11], and pellets containing crude tryptophan pyrrolisate and its derivatives [12]. On the other hand, the exposure to high concentrations of diesel engine emissions, belonging to the sub-category of pollutants named as vehicle emissions, as well as insecticides, fungicides and pesticides resulted to contribute to carcinogenesis [13–15].

Our research group works in Campania, a region of Italy, in which BC incidence is very high with 75.3 cases per 100000 inhabitants, in comparison to only 19 cases per 100000 inhabitants in Europe in 2012 [3]. In the last decades, large areas of the Naples and Caserta provinces have been extensively contaminated by the widespread burial and open air dumping and incineration of industrial toxic waste [16]. Recently, the serum levels of arsenic measured in living subjects of five municipalities in the Naples province resulted to be higher than expected compared to national average [16].

Since the etiology of BC development is multifactorial, only its early detection can reduce mortality. Hence, the development of new non-invasive biomarkers would benefit patients. Some reviews in literature report a summary of studies conducted through different omics approaches aimed to evaluate the transcriptome, the miRNome, the metabolome and the long non coding RNAs profiling involved in BC [16–21].

However, any detailed information is reported until now about what genes/proteins modulated by environmental exposure or by only arsenicals are involved in BC development or progression, how they are correlated, and what metabolic pathways are affected. Therefore, our aim was to select the genes/proteins modulated by environmental exposure (comprising arsenicals, pollutants, smoking, insecticides, fungicides and pesticides) or by only arsenicals in BC starting from Comparative Toxicogenomics Database (CTD), and to correlate them by a computational network approach. Through this method, we identified: i) the specific HUB nodes, that are nodes with a large degree and have connections with many other nodes, ii) the HUB-HUB interaction sub-network between HUB nodes that can be considered as specific of the correlation between BC and environment, and, hence, of the BC development due to environmental exposure, iii) the correlations between HUB nodes with circadian genes and miRNAs, iv) effect of HUB nodes on survival outcome of BC patients and their mutational status in this cancer. Therefore, further experimental studies could be focused on these HUB nodes to verify their utility as new diagnostic and/or prognostic biomarkers as well as potential targets for chemoprevention approach.

## RESULTS AND DISCUSSION

### Selection of environmental chemicals implicated in BC and the related modulated proteins

To create an interaction network between proteins modulated by environmental exposure in BC, we followed

a protocol reported in Figure 1. Firstly we extracted the list of environmental chemicals implicated in BC and of the related proteins modulated by them using Comparative Toxicogenomics Database (CTD) [22]. The selected chemicals were grouped in the four following categories: arsenicals, smoking, pollutants and others, comprising insecticides, fungicides and pesticides, by considering also the studies already reported in literature [10–15] (Supplementary Table 1). About the identified proteins, we found 51 proteins modulated by arsenicals, 40 by smoking molecules, 94 by pollutants and 73 by other chemicals (Figure 2 and Supplementary Table 2). A molecular pathway analysis was conducted on all the proteins belonging to these four categories (Supplementary Table 3), and the molecular pathways, that were common or specific between the four categories, were extracted (Supplementary Table 4). In details, we can evidence the presence of some common pathways, that are strictly correlated to the cancer, among which: i) TNF and NOD-like receptor signaling pathway were common between arsenicals and smoking sub-groups; ii) Rap1, estrogen, prolactin, TGF-beta and VEGF signaling pathways were common between arsenicals, pollutants and “others” sub-groups; iii) hematopoietic cell lineage and Cytokine-cytokine receptor interaction molecular pathway were common between smoking and “others” sub-groups; iv) chemical carcinogenesis was common between smoking, pollutants and “others” sub-groups, and v) drug metabolism - cytochrome P450 was common between smoking and pollutants sub-groups. However, it is important to underline that, although different categories of chemicals are able to regulate similar pathways even if through different proteins along the same pathway, they can generate distinct molecular/cellular outputs.

Moreover, we evaluated also the pathways that are specific for each chemical category, and found for example: i) regulation of actin cytoskeleton, Wnt and Chemokine signaling pathway for arsenicals; ii) drug metabolism - other enzymes and metabolism of xenobiotics by cytochrome P450 for pollutants; iii) AK-STAT signaling pathway for “others” sub-group. Any specific molecular pathway for smoking was evidenced (Supplementary Table 4).

Moreover, a detailed analysis of the proteins modulated by only one chemical sub-category showed that FBXW8 was modulated only by arsenicals, ANXA2R only by smoking chemicals, and nineteen proteins (ASXL2, RALGPS1, STAG2, SMC1B, SMC1A, TYMP, SLC12A7, ARID1A, TACC3, APOBEC3B, MLL2, LOXL4, CALMHM1, NRSN1, EOMES, KMT2A, CHD6, VWA3A, KDM6A) only by pollutants (Figure 2 and Supplementary Table 2).

In details, FBXW8 is a member of the F-box protein family that acts in phosphorylation-dependent ubiquitination whereas ANXA2R may act as a receptor for annexin A2 that plays a role in the regulation of cellular growth and in signal transduction pathways. Regarding the nineteen proteins modulated by the pollutants, only

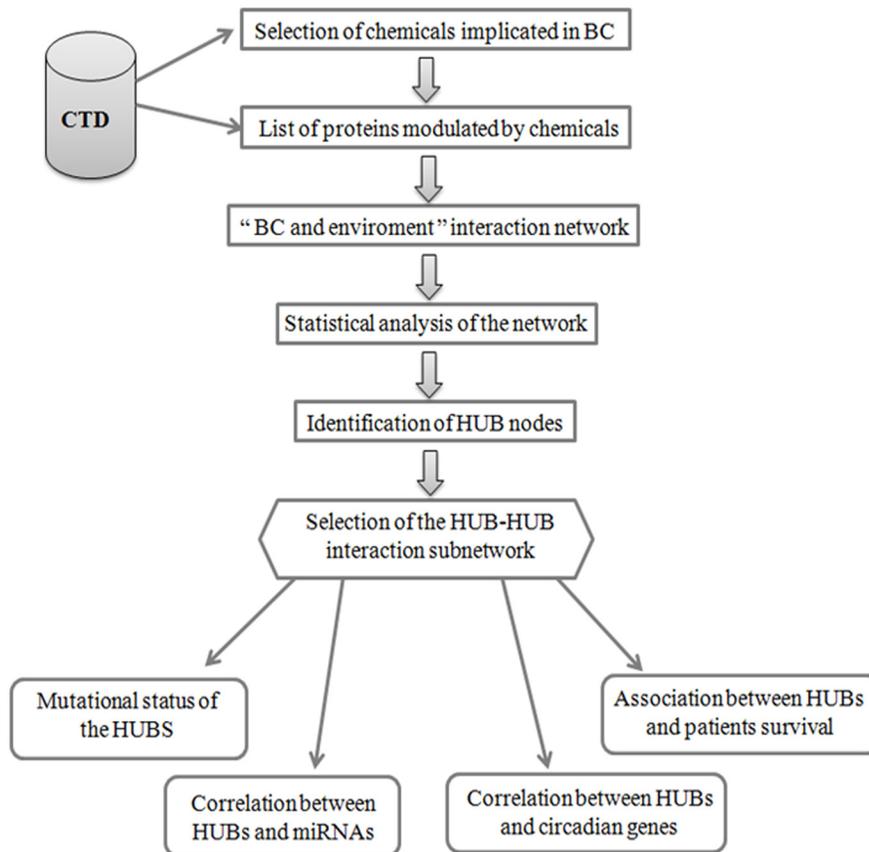


Figure 1: Flow-chart of the protocol used to identify and to analyse HUB nodes in “BC and environment” network.

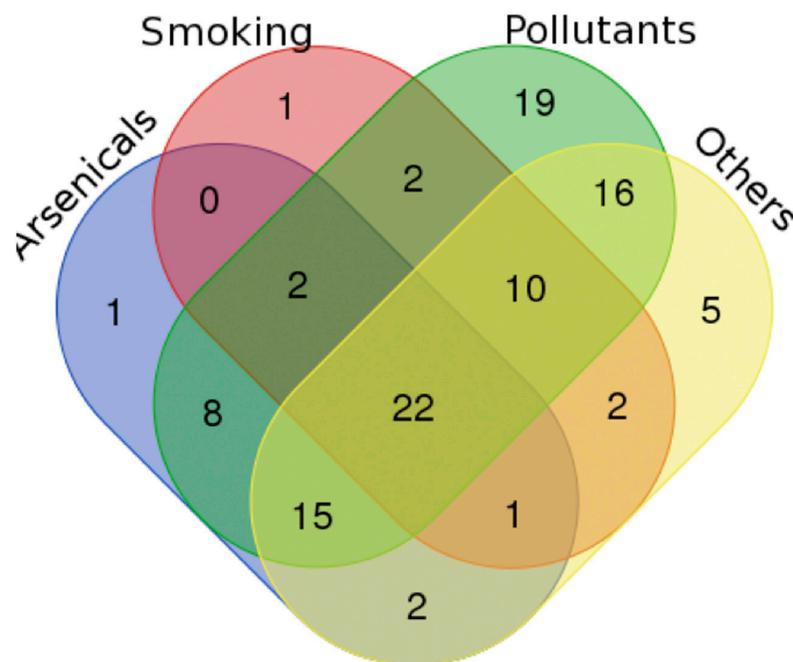


Figure 2: Venn diagram evidences the proteins that are modulated by four chemical sub-groups (arsenicals, smoking, pollutants and others). In this way it is possible to see the number of proteins modulated by only a sub-group or from two, three or four groups.

STAG2, SMC1A and SMC1B resulted to be involved in the enriched molecular pathway named cell cycle.

However, through the Venn diagram analysis, we verified also the presence of twenty-two proteins that are modulated in BC by all four chemical sub-groups: CDH1, KRAS, IGF1, IGFBP3, GSTM1, IGF1, TP53, IGFBP5, GSTP1, PTGS2, TNF, GPX1, CXCL8, CDKN2A, SOD2, EGFR, NQO1, MYC, CDKN1A, MT2A, BIRC3, FAS and ESR1 (Figure 2). These proteins are involved in the following molecular pathways: p53 signaling, MAPK signaling, Apoptosis, ErbB signaling, Cell cycle and Glutathione metabolism (Supplementary Table 5).

Since our principal aim was to identify a sub-network of genes/proteins modulated by environmental exposure in BC, we decided to focus our attention on the proteins modulated by all four chemical sub-groups and to create the general network of interaction between BC and environment.

### Identification and analysis of HUB nodes in “BC and environment” network

On the basis of the human molecular interactome and of the list of the twenty-two proteins, reported above, the “BC and environment” interaction network was created and analyzed. In detail, we mapped our twenty-two proteins on the human molecular interactome (INTACT) [23], extracted their related interaction network (named as “BC and environment”) composed by 1839 nodes and 2376 interactions and analyzed it by the related topological properties, reported in the Methods section, in order to understand the position and the role of these proteins present as nodes in the network (Supplementary Figure 1) [24]. In this way, we selected HUB nodes, which are the nodes with the strongest coordination role by considering a consensus of four out of six measures of centrality and topology (see Methods section). The following fifteen HUB nodes were identified: KRAS, IGFBP3, GSTM1, TP53, GSTP1, TNF, CXCL8, CDKN2A, SOD2, EGFR, NQO1, MYC, CDKN1A, FAS, and ESR1. They resulted to be up-regulated in BC by microarray studies with the exception of GSTP1 and ESR1 [25–31], and be involved in specific molecular functions and pathways (Supplementary Tables 5 and 6). We can speculate that the most part of identified HUB nodes is involved in molecular pathways correlated to cancer development and do not seem specific for BC. However, among these nodes, it is important to evidence the presence of GSTP1 and ESR1, about which few information is reported in BC. In general, GSTP1 catalyzes the conjugation of many hydrophobic and electrophilic compounds with reduced glutathione. It is reported that its polymorphisms are among the genetic determinants related to lead-induced inflammatory response and may modulate the response to epithelial oxidative changes caused by air pollutant exposure in lung [32]. Hence, it could be interesting and useful to study how GSTP1 is involved in BC development

after environmental chemical exposure. In regard to ESR1, it is an estrogen receptor involved in sexual development and reproductive function. Considering that some environmental chemicals such as endosulfan and dieldrin are endocrine disruptors that can cause negative effects on the endocrine functions by miming the action of steroid hormones due to their structure similar to these last ones, we can think in future to investigate in more detail how ESR1 can be affected by these chemicals and through what mechanisms it can be involved in BC carcinogenesis.

About the obtained network a detailed analysis of the statistical centrality and topological measures has permitted to evidence its effectiveness and its robustness. In fact, it is important to underline that our network has a centralization value of 0.423, a network density of 0.001, a heterogeneity value of 9.899 and characteristic path length of 3.330. Overall these data evidenced that: i) the network effectiveness is elevated with nodes that are highly correlated between them; ii) the network is of small world type characterized by short path lengths [33]. Moreover, the plot of the node degree distribution showed a decreasing trend demonstrating that our network had scale free property indicating that it follows the role that “riches get richer” (Supplementary Figure 2) [34–37]. On the other hand the clustering coefficient graph showed a decreasing trend highlighting the tendency of our network to contain HUB nodes [37]. However, considering that the betweenness centrality is a measure to obtain inferences on the importance of inter-connected proteins on the basis of load placed on the given node in the network, the increasing trend of the betweenness centrality in our network demonstrated that the following five HUB nodes had the maximum load: i) TP53 (tumor protein p53), a DNA binding tumor suppressor protein, ii) MYC (c-MYC), a multifunctional and nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation, iii) EGFR (Epidermal Growth Factor Receptor), a cell surface receptor, iv) ESR1 (estrogen receptor), a nuclear hormone receptor and v) CDKN1A (Cyclin Dependent Kinase Inhibitor 1A) which functions as a regulator of cell cycle progression at G1.

Then, to study if in the network there were clusters and/or modules characterized by groups of nodes correlated between them, a cluster analysis was performed considering as statistically significant only the clusters with p-values lower than 0.001. In this way four clusters with density values ranging from 0.004 to 0.170 were selected (Supplementary Figure 7). They comprised 639, 345, 68 and 25 proteins/nodes, respectively, and five HUB nodes. In detail, EGFR was in cluster 1, MYC in cluster 2, ESR1 in cluster 3, and GSTP1 and NQO1 in cluster 4. These clusters were named as EGFR cluster, MYC cluster, ESR1 cluster and GSTP1-NQO1 cluster, respectively. A functional analysis performed on the proteins present in the four clusters evidenced that each cluster comprised proteins involved in different metabolic pathways in comparison to

those present in the other clusters. In fact, the only common pathways between at least two clusters were: i) “cell cycle and Epstein-Barr virus infection“ in the case of EGFR and MYC clusters, ii) “RNA transport” in the case of ESR1 and MYC clusters, and iii) “proteoglycans in cancer” in the case of EGFR and ESR1 clusters (Supplementary Table 7 and Supplementary Figure 4).

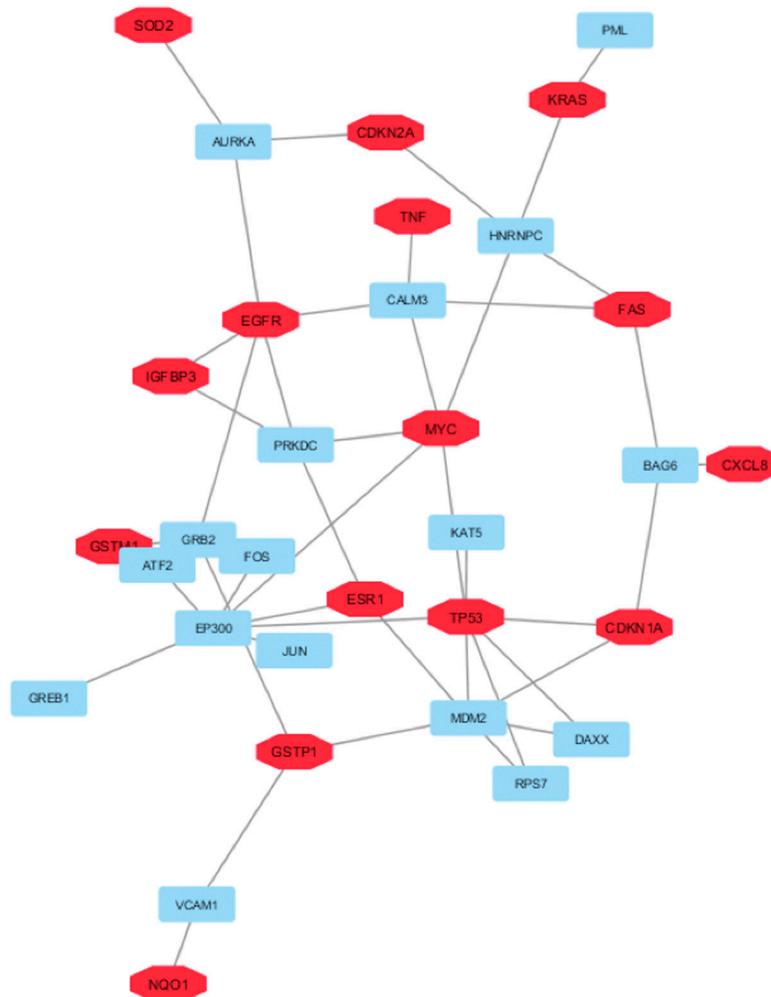
Overall the cluster analysis demonstrated that our “BC and environment” network comprised specific functional sub-networks in which some HUB nodes play crucial roles.

Then, to select the sub-network of HUB nodes that were more correlated between them and to define the related HUB – HUB interaction sub-network, we extracted the interactions between HUB nodes in “BC and environment” network. In this way, it was possible to evidence that: i) four HUB nodes exhibited direct HUB–HUB interactions in the network (EGFR-IGFBP3 and TP53-CDKN1A) and ii) the other HUB nodes were linked between them through only one node (Figure 3). This finding suggested the strict functional relationship between these HUB nodes and how the HUB – HUB sub-network

could represent a panel of proteins on which we can focus further studies in order to verify the possibility to use them as specific of the involvement of the environmental chemicals in BC initiation.

Moreover, the mutational status of the identified HUB nodes was evaluated. This analysis showed that: i) eight HUB nodes (ESR1, FAS, GSTM1, IFGBP3, KRAS, SOD2 and TNF) did not show mutations in BC, ii) CXCL8, NQO1 and MYC had one missense mutation, iii) GSTP1 had two missense mutations, iv) CDKN1A and EGFR had three missense mutations, v) CDKN2A had four missense mutations and vi) TP53 had thirty-eight missense mutations (Supplementary Table 8). Moreover, the co-occurrence of three mutations, i.e. T384S in EGFR and P151H in TP53, D136H in CDKN1A and R273S in TP53, and G146V in GSTP1 and G245S in TP53, resulted as statistically significant (p-value <0.001) suggesting their importance in BC initiation.

Then, to highlight in greater detail the involvement of these HUB nodes in BC development, we searched if they correlated with clock genes already reported to be



**Figure 3: HUB – HUB interaction sub-network in “BC and environment” network.** In details, HUB nodes are reported in red whereas the other nodes in cyan.

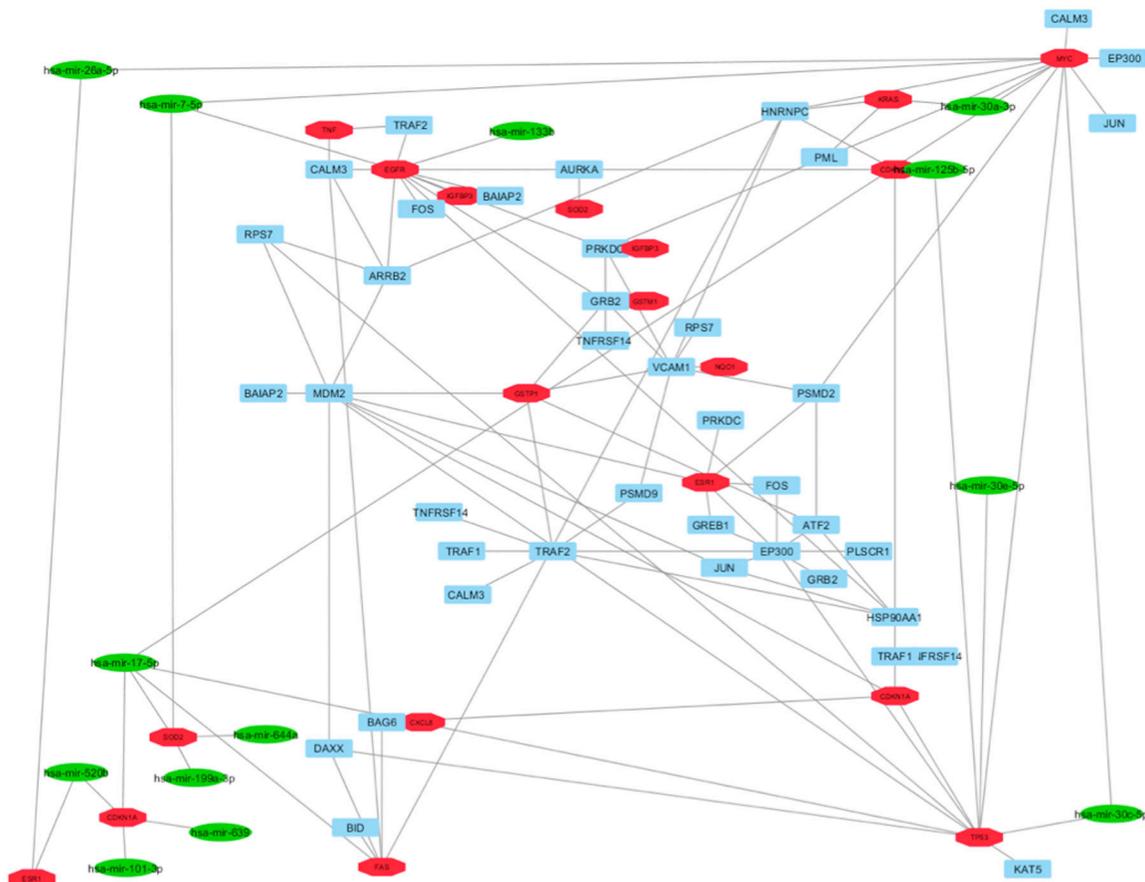
altered in BC [9], and if they can be target of miRNAs reported to be implicated in BC [38].

In general, the circadian clock system comprises both negative and positive regulators, based on an auto-regulatory transcriptional and translational feedback program. In this context, PER and CRY proteins bind to the promoter region of BMAL1 and CLOCK, that are two transcription factors, and are capable to reduce the transcription of many genes during ambient light exposure [9]. In particular, the basic helix-loop-helix (bHLH)/PAS domain transcription factor plays a crucial role in the controlling the biological clock that controls the circadian rhythms.

Also the urinary system is regulated from the circadian rhythms. In fact, during day and night both urine excretion and extrusion are actively regulated by several internal factors and hormones. Such circadian variations led us to postulate that similar to other organs, the perturbation of the clockwork may contribute to the dysregulation that develops during BC development [9]. Since clock genes are able to modify the gene regulation, they may interact with the transcription of oncogenes and/or tumour suppressor-genes. In fact, a recent paper reported the close correlation between altered expression of various clock genes and common tumor markers in BC evidencing how a disturbed function in the cellular clock

may be an important additional mechanism contributing to cancer progression [9]. Hence, we searched if there were correlations between the proteins codified from circadian genes (PER1, PER2, PER3, CRY1, CRY2, BMAL1, CLOCK, ANTI-CSNK1 $\alpha$ 1L, CSNK1 $\alpha$  and CSNK1 $\epsilon$ ) and our HUB nodes. As visible in Supplementary Figure 5, there was a direct correlation between TP53 (HUB node) and CSNK1 $\epsilon$  (circadian node) and other circadian nodes correlated with the other HUB nodes through one or two nodes. This demonstrated the strict relationship between circadian rhythms and HUB nodes and confirmed how our HUB nodes can have an important role in BC development in according to the other previous analysis.

The list of the miRNAs implicated in BC by microarray studies and verified experimentally by RT-PCR, and the list of the genes targeted from these miRNAs was extracted by mirNET database (Supplementary Table 9) [38]. In this way, it was possible to select only the miRNAs able to target our HUB nodes. In detail, our analysis evidenced that thirteen miRNAs (hsa-mir-7-5p, hsa-mir-17-5p, hsa-mir-26a-5p, hsa-mir-30a-3p, hsa-mir-30c-5p, hsa-mir-30e-5p, hsa-mir-101-3p, hsa-mir-125b-5p, hsa-mir-133b, hsa-mir-199a-3p, hsa-mir-520b, hsa-mir-639 and hsa-mir-644a) correlated with seven HUB nodes (EGFR, SOD2, MYC, KRAS, ESR1, CDKN1A and TP53) (Figure 4).



**Figure 4: Correlation between HUB – HUB interaction sub-network in “BC and environment” network and miRNAs reported as implicated in BC in literature.** In details, HUB nodes are reported in red, miRNAs in green and other nodes in cyan.

**Table 1: Overall survival related to high expression/high co-expression of HUB nodes in the “BC and environment” and “BC and arsenicals” networks by SynTarget online tool using public TCGA\_BLCA dataset (Bladder Urothelial Carcinoma)<sup>31</sup>**

Nodes	Survival effect	p-Value
<b>BC and environment network</b>		
EGFR	Negative	0.00124
TP53	Negative	0.00898
MYC	Negative	0.0276
GSTP1	Negative	0.0478
NQO1	Negative	0.0379
KRAS	Negative	0.017
TP53-EGFR	Negative	0.000301
MYC-EGFR	Negative	5.23e-06
KRAS-EGFR	Negative	0.000591
NQO1-EGFR	Negative	0.000134
TP53-MYC	Negative	0.00835
<b>BC and arsenicals network</b>		
KRAS	Negative	0.017
ERCC4	Negative	0.00424
ERCC4-KRAS	Negative	0.0385

The p-values lower than 0.05 were considered as statistically significant.

In addition, starting from these data, to understand if our HUB nodes can have a clinical validity and utility, we decided to perform a set of bioinformatics analyses on available gene expression datasets BC (TCGA-BLCA). The purpose was to assess if there was an association between the expression in BC of our HUB nodes and patients survival. Our analysis evidenced that high expression of EGFR, TP53, MYC, GSTP1, NQO1 and KRAS as well as the association of high co-expression of TP53-EGFR, MYC-EGFR, KRAS-EGFR, NQO1-EGFR, TP53-MYC had a negative effect on the survival (Table 1 and Supplementary Figure 6).

#### **Identification and analysis of HUB nodes in “BC and arsenicals” network**

Since BC incidence is very high in the Campania region in Italy [3] and the SEBIOREC study demonstrated a higher level of arsenic in the serum of subjects living in the province of Napoli compared to national average [16], a particular attention was directed to establish an interaction between arsenicals and fifty-one proteins known to be modulated by them (Figure 2). In this study we followed the same protocol used in the case of “BC and environment” network (Figure 1). As reported in the first Results paragraph, the functional analysis on these

proteins evidenced that they were involved in a set of molecular pathways (Supplementary Table 3). These fifty-one proteins were mapped on the human molecular interactome and the related interaction network named “BC and arsenicals” network was extracted and analyzed by different topological properties, as reported above for the “BC and environment” interaction network (Supplementary Figure 7) [24].

This network comprised 353 nodes (proteins) and 378 interactions (edges), and the following nine HUB nodes: PSMB2, TNF, BIRC3, FANCA, KRAS, CCNE1, ERCC4, PABPC1, and PRSS3. It is important to evidence that two HUB nodes, TNF and KRAS, were in common with the “BC and environment” interaction network whereas the others are uniquely regulated by arsenicals. Notably, all the HUB nodes, at exception of BIRC3 and FANCA, were already resulted to be up-regulated in BC by microarray studies [25–31, 39–43], and be involved in specific molecular functions (Supplementary Table 10). However, among the proteins already studied in BC, PSMB2 is a very interesting node; in fact, it is hypermethylated in BC and was identified as indicator of adverse health effects associated with arsenic exposure [44]. On the other hand, very few information is reported about BIRC3 and FANCA in BC. In particular, BIRC3 inhibits apoptosis by binding to tumor necrosis

factor receptor-associated factors, and is dysregulated in some cancers, and FANCA is a DNA repair protein that may operate in a post-replication repair or a cell cycle checkpoint function. Notably, it could be interesting to study in future how these two proteins can be affected by arsenicals and through what mechanisms they can be involved in BC carcinogenesis.

A detailed analysis of the obtained network showed that it had a good centralization equal to 0.402 and a network density value of 0.006. The characteristic path length of 3.752 confirmed that also this network followed the small-world rule [32] as we evidenced for “BC and environment” network. Moreover, we can underline also that: i) the decreasing trend of the node degree distribution plot indicated that our network had scale free property with the occurrences of modules (Supplementary Figure 8A) [26–29], ii) the decreasing trend of the clustering coefficient graph showed the tendency of our network to contain HUB nodes (Supplementary Figure 8B) [43], and iii) the increasing trend of the betweenness centrality demonstrated the presence of a HUB node with maximum load like PABPC1 (Polyadenylate-Binding Protein 1) that is involved in cytoplasmic regulatory processes of mRNA metabolism such as pre-mRNA (Supplementary Figure 8C).

Then, to study if in the “BC and arsenicals” network there were modules characterized by groups of nodes correlated between them, a cluster analysis was performed as in the case of “BC and environment” network. This showed the presence of two clusters (with significant P-values lower than 0.005) comprising 15 and 10 proteins/nodes with density values ranging from 0.608 to 0.644 and three HUB nodes (Supplementary Figure 9). In detail, CCNE1 was in cluster 1 whereas TNF and BIRC3 were in cluster 2. However, cluster 1 comprised also CDKN1A, a HUB node related to “BC and environmental” network. The proteins present in cluster 1 resulted to be involved in specific metabolic pathways such as PI3K-Akt signaling pathway, p53 signaling pathway and cell cycle, whereas those in cluster 2 in Apoptosis, NF-kappa B signaling pathway, TNF signaling pathway and Adipocytokine signaling pathway. However, the only common pathway between two clusters is “Pathways in cancer” (Supplementary Table 11 and Supplementary Figure 10).

As in the case of “BC and environment” network, cluster analysis highlighted that the “BC and arsenicals” network comprised two sub-networks in which the identified HUB nodes played important functional roles.

Moreover, to select the HUB – HUB interaction sub-network in “BC and arsenicals” network, we focused on the interactions between HUB nodes, and evidenced that two HUB nodes (TNF-BIRC3) exhibited direct HUB–HUB interactions in network whereas the other HUB nodes were linked among them through one or two nodes (Figure 5). Moreover, three HUB nodes identified

for the arsenicals (ERCC4, PABPC1 and CCNE1) resulted to interact directly with three HUB nodes evidenced in the “BC and environment” network (ERCC4-EGFR, PABPC1-ESR1, and CCNE1-CDKN1A) suggesting the strict correlation between “BC and arsenicals” and “BC and environment” networks.

However the analysis of HUB-HUB interaction sub-network in “BC and arsenicals” network evidenced how, with the exception of TNF and KRAS being in common between two networks, the remaining seven HUB nodes (PSMB2, BIRC3, FANCA, CCNE1, ERCC4, PABPC1, and PRSS3) can be considered as specific of the arsenicals involvement in BC development.

The analysis of the mutational status of all the HUB nodes, showed that: i) six HUB nodes (CCNE1, KRAS, PABPC1, PRSS3, PSMB2 and TNF) had not mutations in BC, ii) BIRC3, ERCC4 and FANCA had one, two and four missense mutations, respectively (Supplementary Table 12). No co-occurrence mutations resulted statistically significant.

Also in the case of “BC and arsenicals network”, we evaluated if there was a correlation between our HUB nodes and the clock genes resulted to be altered in BC [9]. As visible in Supplementary Figure 11, there was no direct correlation between HUB nodes and circadian nodes but the circadian nodes correlated with the HUB nodes through one or two nodes. This suggested that the relationship between circadian rhythms and HUB nodes in “BC and arsenicals” network was weaker in comparison to what observed for the “BC and environment” network.

Then, we evaluated if our HUB nodes can be targets of miRNAs, resulted already as implicated in BC using the same protocol reported above. This analysis evidenced that seven miRNAs (hsa-mir-7-5p, hsa-mir-17-5p, hsa-mir-26a-5p, hsa-mir-30a-3p, hsa-mir-125b-5p, hsa-mir-520b and hsa-mir-646) correlated with four HUB nodes (FANCA, CCNE1, KRAS and PABPC1) (Figure 6 and Supplementary Table 13).

Among these miRNAs, only hsa-mir-646 did not correlate with the HUB nodes of “BC and environmental” network. Hence, it could be specific of the involvement of arsenicals in BC development.

Finally, we evaluated if there was an association between the expression of our HUB nodes in BC and patients survival as in the case of “BC and environmental” network. Our analysis showed that only the high expression of ERCC4 and KRAS and their high co-expression (ERCC4-KRAS) resulted to have a negative effect on the survival in BC patients (Supplementary Figure 12 and Table 1).

## CONCLUSIONS

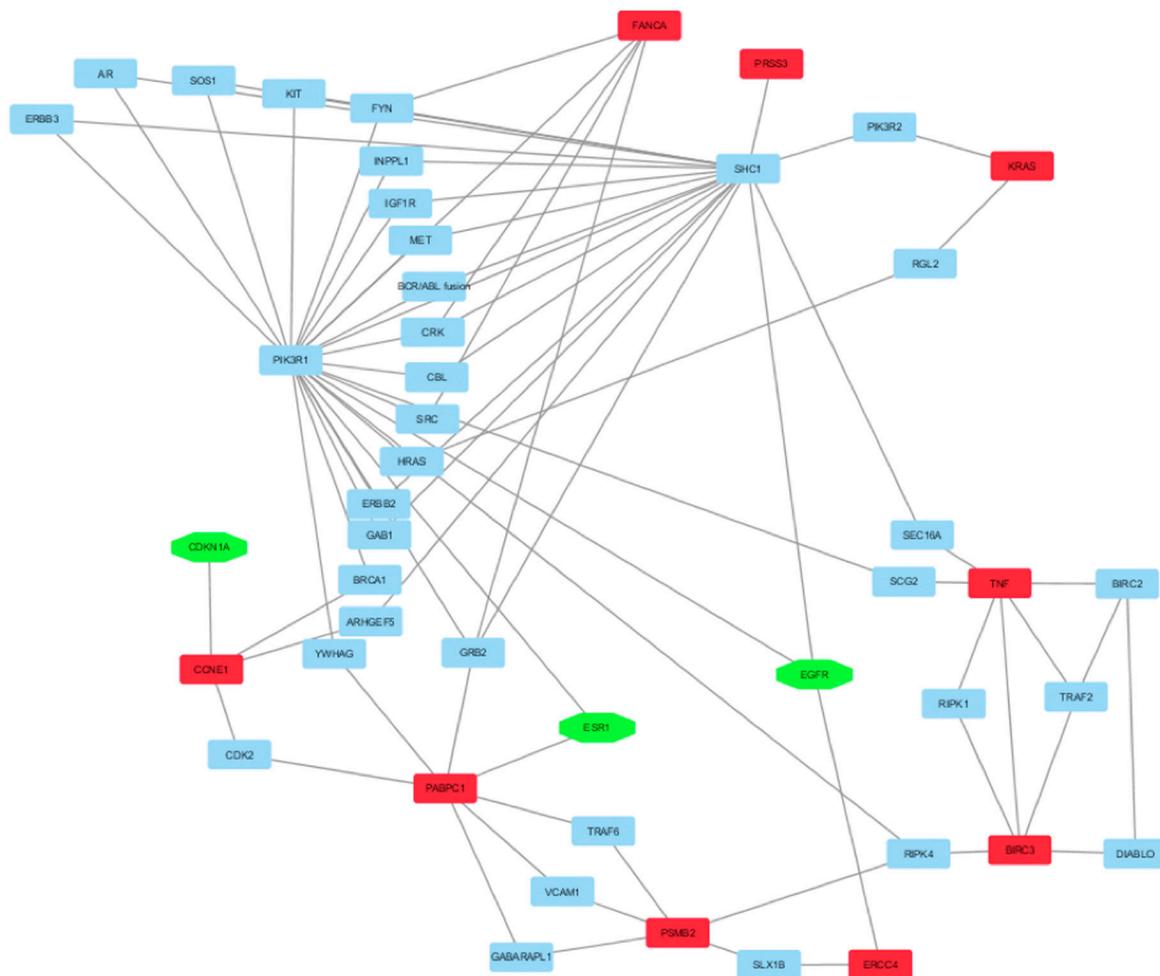
No many papers were already published about the direct linking between the environmental or arsenic

exposure and the related modulated proteins in BC. For example, it has been shown that arsenic exposure is positively associated with PRSS3 promoter methylation levels in BC [45] whereas the effect of NQO1 Pro187Ser polymorphism resulted to be more important in never smokers because no consistent results were obtained about tobacco-related BC risk [46]. Therefore, aim of this work was to highlight the correlation between genes and proteins that are modulated by environmental exposure or only by the arsenicals by network approaches, and to identify the related HUB nodes that can be considered as specific of the BC development due to environmental or arsenicals exposure, and on which can be useful to focus further experimental studies in order to verify their utility as new diagnostic and/or prognostic biomarkers or targets.

In summary, in our study we reported the creation of two networks named as “BC and environment” and “BC and arsenicals”, defined a set of HUB nodes and the related HUB-HUB interactions, and evidenced that these HUB nodes showed significant mutations implicated in BC and correlated with genes involved in circadian rhythms. In addition, we identified: i) a sub-network of interactions

between miRNAs and genes that is specific of the correlation between BC and environment or arsenicals, and ii) the genes correlated to negative effect on the survival in BC patients.

On the basis of our results, we can underline that many identified HUB nodes in “BC and environment” and “BC and arsenicals” networks are proteins already known as to be involved in molecular pathways correlated to the development of cancer and BC. However, we evidenced also the presence of other proteins for which molecular interactions in BC were unknown. For example in “BC and environment” network we found the estrogen receptor, ESR1, that can be object of further investigations because some environmental chemicals are endocrine disruptors that can mimic its structure. Hence, this finding can cause problems in the endocrine system and could represent an initiation point for BC. Moreover, in the case of “BC and arsenicals” network, we identified two HUB nodes, BIRC3 and FANCA for which the role in BC is unknown. Therefore, it can be useful to study if and how these two proteins can be affected by arsenicals and if there are specific mechanisms through which they can contribute to BC carcinogenesis. Certainly these hypothesis may be verified



**Figure 5: HUB – HUB interaction sub-network in “BC and arsenicals” network.** In details, HUB nodes are reported in red, HUB nodes that were present also in “BC and environmental network” in green and the other nodes in cyan.



## MATERIALS AND METHODS

### Network analysis

Comparative Toxicogenomics Database (CTD) was used to extract the list of environmental chemicals implicated in BC and the related chemical–protein interactions [22]. Through the Cytoscape software platform for the visualization of complex networks and their integration (<http://www.cytoscape.org/>), a network related to the interactions between the proteins modulated by the selected environmental chemicals was constructed using as reference the human molecular interactome (INTACT) [23] where all interactions are derived from literature curation or direct user submissions and are freely available. Some statistical analyses were performed the following three measures of centrality: i) the degree that indicates the number of interactions of a particular node with other nodes in the network; ii) the betweenness centrality that evaluates the importance of a node in the network and how the other interactions in the network are controlled by this node [47]; and iii) the closeness centrality of a node that is calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph and ranges from 0 to 1 [48]. Then, we evaluated also other topological analyses like average characteristic path length, network density, and centralization [48–50]. The characteristic path length is calculated by finding the shortest path between all pairs of nodes, adding them up, and then dividing by the total number of pairs. This indicates the number of steps that takes to get from one member of the network to another. The density of a network is defined as a ratio of the number of edges to the number of possible edges [48] whereas the centralization produces rankings, which seek to identify the most important nodes in a network model ranging from 1 to 0 [49].

Finally, we performed a cluster analysis by means of Cluster-One [50] that is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters) [51].

Functional and Pathway Analyses were performed by DAVID program [52].

### miRNA evaluations

The miRNAs able to target the HUB nodes were selected by MirNet tool [53]. In details, we performed the following protocol: firstly, we extracted the list of miRNAs involved in BC; secondly, starting from this miRNA list we extracted all their targets; finally we selected only the miRNAs that targeted our HUB nodes. Then, using the list of our HUB nodes and the related miRNAs, an interaction network between selected miRNAs and HUB nodes was constructed by the Cytoscape package.

### Survival gene effect analysis

Bionformatic analyses were performed by SynTarget online tool able to test the synergetic effect of genes on survival outcome in cancer (<http://www.bioprofiling.de>) using public TCGA\_BLCA dataset (Bladder Urothelial Carcinoma) [53]. In details, PPISURV tool was used to test if our genes can be used as biomarker of cancer survival.

### Gene mutation analysis

Mutation analysis was conducted by cBioPortal for Cancer Genomics tool able to analyze cancer genomics data in order to test the presence of mutations on genes involved in cancer (<http://cbiportal.org>) [54, 55].

### Author contributions

Alfredo Budillon, Gennaro Ciliberto, Susan Costantini and Maurizio Montella conceived this study; Andrea Polo and Susan Costantini performed the experiments; Anna Crispo, Pellegrino Cerino, Luca Falzone, Saverio Candido, Aldo Giudice, Giuseppina De Petro collaborated to the discussion of the results; All the Authors approved the final version of the manuscript.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## FUNDING

This study was partially supported by Italian Ministry of Health to Istituto Nazionale Tumori di Napoli (progetto 5x1000 2016-2017). Andrea Polo is supported by a fellowship from the SPES project (<http://spes.campaniatrasparente.it/>)

## REFERENCES

1. Malats N, Real FX. Epidemiology of bladder cancer. *Hematol Oncol Clin North Am.* 2015; 29:177-189.
2. Mahdavifar N, Ghoncheh M, Pakzad R, Momenimovahed Z, Salehiniya H. Epidemiology, incidence and mortality of bladder cancer and their relationship with the development index in the world. *Asian Pac J Cancer Prev.* 2016; 17:381-386.
3. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer.* 2012; 6:1374-1403.
4. Rota M, Bosetti C, Boccia S, Boffetta P, La Vecchia C. Occupational exposures to polycyclic aromatic hydrocarbons and respiratory and urinary tract cancers: an

- updated systematic review and a meta-analysis to 2014. *Arch Toxicol.* 2014; 88:1479-1490.
5. Murta-Nascimento C, Schmitz-Dräger BJ, Zeegers MP, Steineck G, Kogevinas M, Real FX, Malats N. Epidemiology of urinary bladder cancer: from tumor development to patient's death. *World J Urol.* 2007; 25:285-295.
  6. Christoforidou EP, Riza E, Kales SN, Hadjistavrou K, Stoltidi M, Kastania AN, Linos A. Bladder cancer and arsenic through drinking water: a systematic review of epidemiologic evidence. *J Environ Sci Health A Tox Hazard Subst Environ Eng.* 2014; 48:1764-1775.
  7. Saint-Jacques N, Parker L, Brown P, Dummer TJ. Arsenic in drinking water and urinary tract cancers: a systematic review of 30 years of epidemiological evidence. *Environ Health.* 2014; 2:13-44.
  8. Shivappa N, Hébert JR, Rosato V, Rossi M, Libra M, Montella M, Serraino D, La Vecchia C. Dietary inflammatory index and risk of bladder cancer in a large Italian case-control study. *Urology.* 2017; 100:84-89.
  9. Littlekalsoy J, Rostad K, Kalland KH, Hostmark JG, Laerum OD. Expression of circadian clock genes and proteins in urothelial cancer is related to cancer-associated genes. *BMC Cancer.* 2016; 16:549.
  10. Eblin KE, Bredfeldt TG, Buffington S, Gandolfi AJ. Mitogenic signal transduction caused by monomethylarsonous acid in human bladder cells: role in arsenic-induced carcinogenesis. *Toxicol Sci.* 2007; 95:321-30.
  11. Takahashi M, Toyoda K, Aze Y, Furuta K, Mitsumori K, Hayashi Y. The rat urinary bladder as a new target of heterocyclic amine carcinogenicity: tumor induction by 3-amino-1-methyl-5H-pyrido[4,3-b]indole acetate. *Jpn J Cancer Res.* 1993; 84:852-858.
  12. Hashida C, Nagayama K, Takemura N. Induction of bladder cancer in mice by implanting pellets containing tryptophan pyrolysis products. *Cancer Lett.* 1982; 17:101-105.
  13. Latifovic L, Villeneuve PJ, Parent MÉ, Johnson KC, Kachuri L; Canadian Cancer Registries Epidemiology Group, Harris S. Bladder cancer and occupational exposure to diesel and gasoline engine emissions among Canadian men. *Cancer Med.* 2015; 4:1948-1962.
  14. Fujii T, Mikuriya H, Kamiya N, Hiraga K. Enhancing effect of thiabendazole on urinary bladder carcinogenesis induced by sodium o-phenylphenate in F344 rats. *Food Chem Toxicol.* 1986; 24:207-211.
  15. Ito N, Fukushima S, Shirai T, Hagiwara A, Imaida K. Drugs, food additives and natural products as promoters in rat urinary bladder carcinogenesis. *IARC Sci Publ.* 1984; 56:399-407.
  16. Di Lorenzo G, Federico P, De Placido S, Buonerba C. Increased risk of bladder cancer in critical areas at high pressure of pollution of the Campania region in Italy: a systematic review. *Crit Rev Oncol Hematol.* 2015; 96:534-541.
  17. Wang H, Niu L, Jiang S, Zhai J, Wang P, Kong F, Jin X. Comprehensive analysis of aberrantly expressed profiles of lncRNAs and miRNAs with associated ceRNA network in muscle-invasive bladder cancer. *Oncotarget.* 2016; 52:86174-86185. doi: 10.18632/oncotarget.13363.
  18. Bao Z, Zhang W, Dong D. A potential prognostic lncRNA signature for predicting survival in patients with bladder urothelial carcinoma. *Oncotarget.* 2017; 8:10485-10497. doi: 10.18632/oncotarget.14441.
  19. Wei L, Chintala S, Ciamporcero E, Ramakrishnan S, Elbanna M, Wang J, Hu Q, Glenn ST, Murakami M, Liu L, Gomez EC, Sun Y, Conroy J, et al. Genomic profiling is predictive of response to cisplatin treatment but not to PI3K inhibition in bladder cancer patient-derived xenografts. *Oncotarget.* 2016; 7:76374-76389. doi: 10.18632/oncotarget.13062.
  20. Enokida H, Yoshino H, Matsushita R, Nakagawa M. The role of microRNAs in bladder cancer. *Investig Clin Urol.* 2016; 57:S60-S76.
  21. Cheng Y, Yang X, Deng X, Zhang X, Li P, Tao J, Qin C, Wei J, Lu Q. Metabolomics in bladder cancer: a systematic review. *Int J Clin Exp Med.* 2015; 8:11052-11063.
  22. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegiers J, Wiegiers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res.* 2017; 45:D972-D978.
  23. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E. et al. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 2014; 42:D358-D363.
  24. Sharan R, Ideker T. Modeling cellular machinery through biological network comparison. *Nat Biotechnol.* 2006; 24:427-433.
  25. Fernández-Medarde A, Santos E. Ras in cancer and developmental diseases. *Genes Cancer.* 2011; 2:344-358. doi: 10.1177/1947601911411084.
  26. Korkolopoulou P, Konstantinidou AE, Thomas-Tsagli E, Christodoulou P, Kapralos P, Davaris P. WAF1/p21 protein expression is an independent prognostic indicator in superficial and invasive bladder cancer. *Appl Immunohistochem Mol Morphol.* 2000; 8:285-292.
  27. Sarkis AS, Bajorin DF, Reuter VE, Herr HW, Netto G, Zhang ZF, Schultz PK, Cordon-Cardo C, Scher HI. Prognostic value of p53 nuclear overexpression in patients with invasive bladder cancer treated with neoadjuvant MVAC. *J Clin Oncol.* 1995; 13:1384-1390.
  28. Schmitz-Dräger BJ, Schulz WA, Jürgens B, Gerharz CD, van Roeyen CR, Bültel H, Ebert T, Ackermann R. c-myc in bladder cancer. Clinical findings and analysis of mechanism. *Urol Res.* 1997; 25:S45-S49.

29. Colquhoun AJ, Mellon JK. Epidermal growth factor receptor and bladder cancer. *Postgrad Med J.* 2002; 78:584-589.
30. Shen SS, Smith CL, Hsieh JT, Yu J, Kim IY, Jian W, Sonpavde G, Ayala GE, Younes M, Lerner SP. Expression of estrogen receptors-alpha and -beta in bladder cancer cell lines and human bladder tumor tissue. *Cancer.* 2006; 106:2610-2616.
31. Gao L, Fang YQ, Zhang TY, Ge B, Xu B, Huang JF, Zhang ZF, Tan N. GSTP1 arrests bladder cancer T24 cells in G0/G1 phase and up-regulates p21 expression. *Int J Clin Exp Med.* 2014; 7:2984-2991.
32. Reddy P, Naidoo RN, Robins TG, Mentz G, Li H, London SJ, Batterman S. GSTM1 and GSTP1 gene variants and the effect of air pollutants on lung function measures in South African children. *Am J Ind Med.* 2012; 55:1078-86.
33. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011; 12:56-68.
34. Wu J, Tan YJ, Deng HZ, Zhu DZ. A new measure of heterogeneity of complex networks based on degree sequence. *Unifying Themes in Complex Systems.* Berlin, Heidelberg: Springer, 2010; 66-73.
35. Dong J, Horvath S. Understanding network concepts in modules. *BMC Syst Biol.* 2007; 1:24.
36. Sharma A, Costantini S, Colonna G. The protein-protein interaction network of the human Sirtuin family. *Biochim Biophys Acta.* 2013; 1834:1998-2009.
37. Radrich K, Tsuruoka Y, Dobson P, Gevorgyan A, Swainston N, Baart G, Schwartz JM. Integration of metabolic databases for the reconstruction of genome-scale metabolic networks. *BMC Syst Biol.* 2010; 4:114.
38. Fan Y, Siklenka K, Arora SK, Ribeiro P, Kimmins S, Xia J. miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic Acids Res.* 2016; 44:W135-W141.
39. Pollard C, Nitz M, Baras A, Williams P, Moskaluk C, Theodorescu D. Genoproteomic mining of urothelial cancer suggests  $\gamma$ -glutamyl hydrolase and diazepam-binding inhibitor as putative urinary markers of outcome after chemotherapy. *Am J Pathol.* 2009; 175:1824-1830.
40. Kim SH, Ho JN, Jin H, Lee SC, Lee SE, Hong SK, Lee JW, Lee ES, Byun SS. Upregulated expression of BCL2, MCM7, and CCNE1 indicate cisplatin-resistance in the set of two human bladder cancer cell lines: T24 cisplatin sensitive and T24R2 cisplatin resistant bladder cancer cell lines. *Investig Clin Urol.* 2016; 57:63-72.
41. Chen R, Feng C, Xu Y. Cyclin-dependent kinase-associated protein Cks2 is associated with bladder cancer progression. *J Int Med Res.* 2011; 39:533-540.
42. García-Closas M, Malats N, Real FX, Welch R, Kogevinas M, Chatterjee N, Pfeiffer R, Silverman D, Dosemeci M, Tardón A, Serra C, Carrato A, García-Closas R, et al. Genetic variation in the nucleotide excision repair pathway and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2006; 15:536-542.
43. Enokida H, Nakagawa M. Epigenetics in bladder cancer. *Int J Clin Oncol.* 2008; 13:298-307.
44. Rager JE, Tilley SK, Tulenko SE, Smeester L, Ray PD, Yosim A, Currier JM, Ishida MC, González-Horta Mdel C, Sánchez-Ramírez B, Ballinas-Casarrubias L, Gutiérrez-Torres DS, Drobná Z, et al. Identification of novel gene targets and putative regulators of arsenic-associated DNA methylation in human urothelial cells and bladder cancer. *Chem Res Toxicol.* 2015; 28:1144-1155.
45. Marsit CJ, Karagas MR, Danaee H, Liu M, Andrew A, Schned A, Nelson HH, Kelsey KT. Carcinogen exposure and gene promoter hypermethylation in bladder cancer. *Carcinogenesis.* 2006; 27:112-116.
46. Hung RJ, Boffetta P, Brennan P, Malaveille C, Gelatti U, Placidi D, Carta A, Hautefeuille A, Porru S. Genetic polymorphisms of MPO, COMT, MnSOD, NQO1, interactions with environmental exposures and bladder cancer risk. *Carcinogenesis.* 2004; 25:973-978.
47. Yoon J, Blumer A, Lee K. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics.* 2006; 22:3106-3108.
48. Newman M. A measure of betweenness centrality based on random walks. *Soc Networks.* 2005; 27:39-54.
49. Freeman LC. Centrality in social networks conceptual clarification. *Soc Networks.* 1978; 1:215-239.
50. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003; 4:2.
51. Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods.* 2012; 9:471-472.
52. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37:1-13.
53. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature.* 2014; 507:315-322.
54. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013; 6:p11.
55. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012; 2:401-404.