

RESEARCH ARTICLE

Open Access

# Preferential regulation of miRNA targets by environmental chemicals in the human genome

Xudong Wu<sup>1,2,3\*</sup> and Yijiang Song<sup>2</sup>

## Abstract

**Background:** microRNAs (miRNAs) represent a class of small (typically 22 nucleotides in length) non-coding RNAs that can degrade their target mRNAs or block their translation. Recent disease research showed the exposure to some environmental chemicals (ECs) can regulate the expression patterns of miRNAs, which raises the intriguing question of how miRNAs and their targets cope with the exposure to ECs throughout the genome.

**Results:** In this study, we comprehensively analyzed the properties of genes regulated by ECs (EC-genes) and found miRNA targets were significantly enriched among the EC-genes. Compared with the non-miRNA-targets, miRNA targets were roughly twice as likely to be EC-genes. By investigating the collection methods and other properties of the EC-genes, we demonstrated that the enrichment of miRNA targets was not attributed to either the potential collection bias of EC-genes, the presence of paralogs, longer 3'UTRs or more conserved 3'UTRs. Finally, we identified 1,842 significant concurrent interactions between 407 miRNAs and 497 ECs. This association network of miRNAs-ECs was highly modular and could be separated into 14 interconnected modules. In each module, miRNAs and ECs were closely connected, providing a good method to design accurate miRNA markers for ECs in toxicology research.

**Conclusions:** Our analyses indicated that miRNAs and their targets played important roles in cellular responses to ECs. Association analyses of miRNAs and ECs will help to broaden the understanding of the pathogenesis of such chemical components.

## Background

miRNAs are a class of small non-coding RNAs, which act through binding in a sequence-specific manner to the 3'UTR of target genes [1]. With a very short recognition sequence (~8bp), each miRNA can potentially regulate hundreds of transcripts. At least one-third of human genes are estimated to be miRNA targets, so the regulation mediated by miRNA at the post-transcriptional level is pervasive in animals [2]. Transcriptomic studies suggest that miRNAs can regulate the expression and stability of targets [3-6]. miRNAs also provide a genetic buffer to constrain the variation of their targets' expression, playing an important role in regulating embryo development and maintaining the identity of mature tissues [7]. In many situations, miRNAs and their targets are co-expressed at intermediate levels;

miRNAs serve to buffer the fluctuation of the targets' expression through feed-forward loop architecture [8], such as the relationship between *miR-9a* and *E(spl)* in *Drosophila* [9,10] and *miR-17* and *E2F1* in human [11].

Cells change physiologically in response to signals from their external environments. To achieve this, they must activate or repress various genes and tune their products to a proper level under different situations. Many toxicological researchers, adopting RT-PCR, Northern-blotting or microarray technologies to investigate the expression of protein-coding genes, have demonstrated that exposure to ECs often has a negative effect on the normal growth of cells [12,13]. The Comparative Toxicogenomics Database (CTD, <http://ctd.mdibl.org/>) is a manually curated database, which stores high-quality chemical-gene regulatory data [14]. Its current dataset includes a large number of associations among chemicals and proteins in *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Caenorhabditis elegans* and other species; therefore, it

\* Correspondence: xudong33@yahoo.com.cn

<sup>1</sup>Key laboratory of Photosynthesis and Environmental Molecular Physiology, the Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China  
Full list of author information is available at the end of the article

can be used to determine whether a chemical binds to or regulates the expression of a protein-coding gene.

A two-tiered review system was implemented in the CTD to identify the high-quality chemicals-genes regulatory data [14,15]. First, the curators used text mining to select literature where interactions between chemicals and genes were identifiable, so that data were supported by their source references. The senior curators then proofread the entries from other curators, ensuring that the correct chemical names and gene symbols were chosen. Second, a group of prominent senior immunologists evaluated the curation guidelines and contacted the authors of papers to clarify details of the experimental procedures, to assure that the data were exactly presented as in the reference. Recently, the datasets of the CTD has been used by several independent groups and demonstrated great utility for meta-analyses of ECs [13,16,17].

miRNAs are essential for regulating many cellular processes, such as apoptosis, proliferation and metastasis [18,19]. How miRNAs function in regulating human responses to environmental chemical (EC) stimuli is an unexplored field of compound risk evaluation. In this study, we retrieved the dataset of EC-genes from the CTD and explored their propensities to be miRNA targets. By evaluating the factors that may potentially result in the enrichment, we found that miRNA targets were preferentially regulated by ECs. Through simulations and statistical analyses, we identified significantly occurring miRNA-EC pairs and reconstructed the association network. The identified miRNAs specific to EC-exposure could be used as biomarkers for determining the genotoxicity and carcinogenicity of chemicals [20,21]. The following module analysis provides us with an in-depth view of miRNA function in toxicological research.

## Results

### miRNA targets are preferentially regulated by environmental chemicals

We first collected the genes regulated by ECs (EC-genes). According to the expression regulatory information from the CTD [14], such as “chemical x results in increased expression of protein y”, “compound x results in decreased expression of protein y” or “compound x affects the expression of protein y”, we compiled the dataset of proteins regulated by ECs (see Methods), and then transformed the gene symbols to their Ensembl gene IDs using the BioMart program (<http://biomart.org>). Based on the reports of 4,162 literatures, we retrieved 42,770 regulatory relationships among compounds and human protein-coding genes (Additional file 1). Specifically, the expression profiles of 9,692 protein-encoding genes were regulated by at least one of the 1,938 ECs, including polycyclic compounds, organic

chemicals, heterocyclic compounds, inorganic chemicals, hormones and so on.

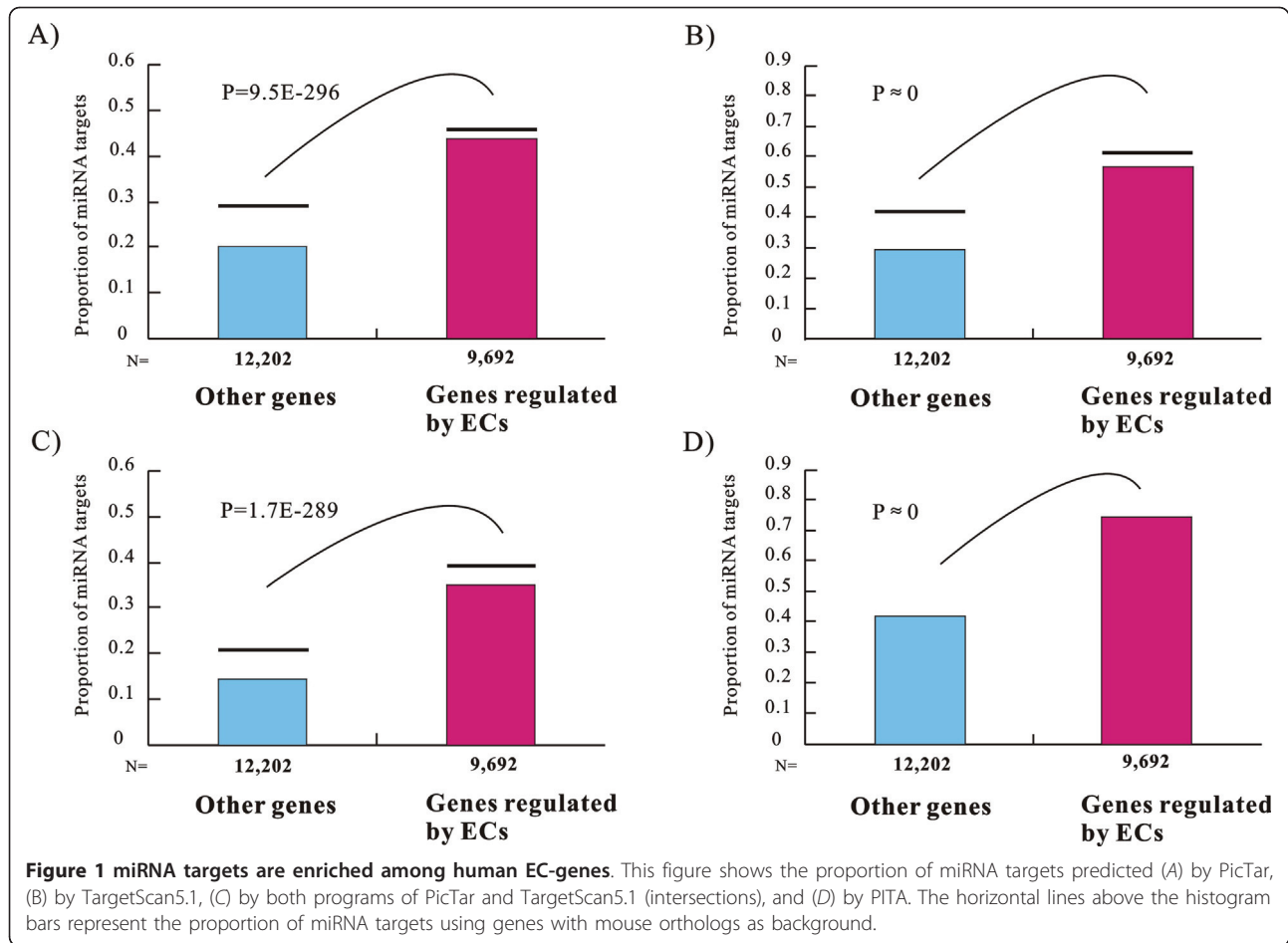
We next evaluated the probability of these EC-genes being targeted by miRNAs using TargetScan5.1 [22] and PicTar (four-way) [23], which predict miRNA targets based on sequence complementarities, sequence context information, binding energy, and were regarded by previous surveys having high confidence [24]. As the reliance of TargetScan and PicTar upon cross-species conservation might introduce potential bias, we also included a third set of predicted human miRNA targets derived from PITA [25], which only considered the sequence complementarities and site accessibility; therefore, many more genes were annotated as miRNA targets with the advantage of detecting human-specific miRNA targets.

Using each dataset of miRNA targets (Additional file 2), we observed that miRNA targets were significantly enriched among EC-genes. As shown in Figure 1A, miRNA targets comprise 43% of the EC-genes (4,195 out of 9,692), but only make up 20% of the non-EC-genes (2,473 out of 12,202) as predicted by PicTar. Using the targets detected by TargetScan (Figure 1B), both programs of PicTar and TargetScan (Figure 1C), and PITA (Figure 1D), we obtained similar results. Therefore, miRNA targets were roughly twice as likely to be EC-genes as comparable to the other genes (*Chi-square test*,  $p$ -values  $< 1.7E-289$  for the four datasets). Because the methods of TargetScan and PicTar depend upon alignments of human-mouse orthologs, we only included the human genes with mouse orthologs and repeated the comparative analyses. In Figure 1A-C, the horizontal lines above the histogram bars represent the proportion of miRNA targets using the genes with mouse orthologs as background. Significant enrichment of miRNA targets among EC-genes were observed in this dataset.

In *Mus musculus* and *Rattus norvegicus*, we retrieved 9,552 and 5,064 genes respectively, which were regulated by ECs and found miRNA targets were over-represented among them as compared to the other genes (Additional file 3), indicating that the enrichment of miRNA targets in EC-genes seems to be common in mammal systems. In the following analysis, we only focused on the human genes because similar conclusions were drawn from the analysis results of *Mus musculus* and *Rattus norvegicus*.

### Enrichment of miRNA targets is not dependent on the collection bias of EC-genes

To evaluate whether the above-observed enrichment of miRNA targets among EC-genes was caused by some sample collection bias, we performed the following analysis. First, if the experiments in the literatures from which the EC-genes were extracted, were designed to



deal with miRNA-related scientific questions, the enrichment would be intuitively expected but with serious bias. We downloaded all of the abstracts of the 4,261 papers from the NCBI via the PubMed IDs (<http://www.ncbi.nlm.nih.gov/sites/batchentrez>) and searched keywords such as “miR”, “miRNA”, “micro-RNA” or “let-7”. As a result, only 5 papers were directly related to the study of miRNAs and about 30 genes appearing in the datasets were included by them (Additional file 4); therefore, there should be no bias in the literature collection results.

Second, some EC-genes may belong to a certain class of genes that are preferentially regulated by miRNAs; therefore, the enrichment may be only contributed by that class of genes. As cancer-related genes were extensively studied [26] and often found to be miRNAs targets [18,19], it is possible that the exposure of the cancer-related genes to ECs were more likely to be investigated and eventually made the miRNA targets over-represented among the EC-genes. To test this, we retrieved a separate set of genes over-expressed in cancer tissues [27]. Specifically, 2,362 proteins corresponding to 2,062 Ensembl

genes were at least over-expressed 4-fold in brain (*astrocytoma* and *glioblastoma*), breast, colon, endometrial, kidney, liver, lung, ovary, prostate, skin, and thyroid cancers as compared to healthy tissues of the same type. Significant enrichment of miRNA targets among EC-genes were still observed even after filtering out these cancer-related genes from the datasets (Additional file 5).

Third, for the 12,202 genes not observed to be regulated by ECs (non-EC-genes), some of them may in fact be regulated by ECs but not analyzed or reported thus far. If this is indeed the case, the potential false-negatives from the non-EC-genes may seriously challenge the enrichment conclusion. Because 45% of human genes (calculated by  $9,692/(9,692 + 12,202)$ ) were confirmed to be regulated by ECs genome-wide, we arbitrarily sampled genes with the probability of 0.45 from the non-EC-genes and assumed them to be non-annotated EC-genes. To investigate the impact of the potential false-negatives of non-EC-genes, we performed the following procedures: (a) randomly sampled genes from non-EC-genes with the probability of 0.45,  $S = 0.45$ ; (b) constructing the dataset of pseudo-EC-genes with

$n = 9,692 + 12,202 * S$  and a dataset of pseudo-non-EC-genes with  $n = 12,202 - 12,202 * S$ ; (c) comparing the proportion of miRNA targets between the pseudo-EC-genes and pseudo-non-EC-genes. We repeated this simulation several times and always obtained significant differences of miRNA targets between pseudo-EC-genes and pseudo-non-EC-genes (see Additional file 6 for the results of eight simulations). Thus, the potential false-negatives of non-EC-genes would not affect the enrichment tendencies of miRNA targets.

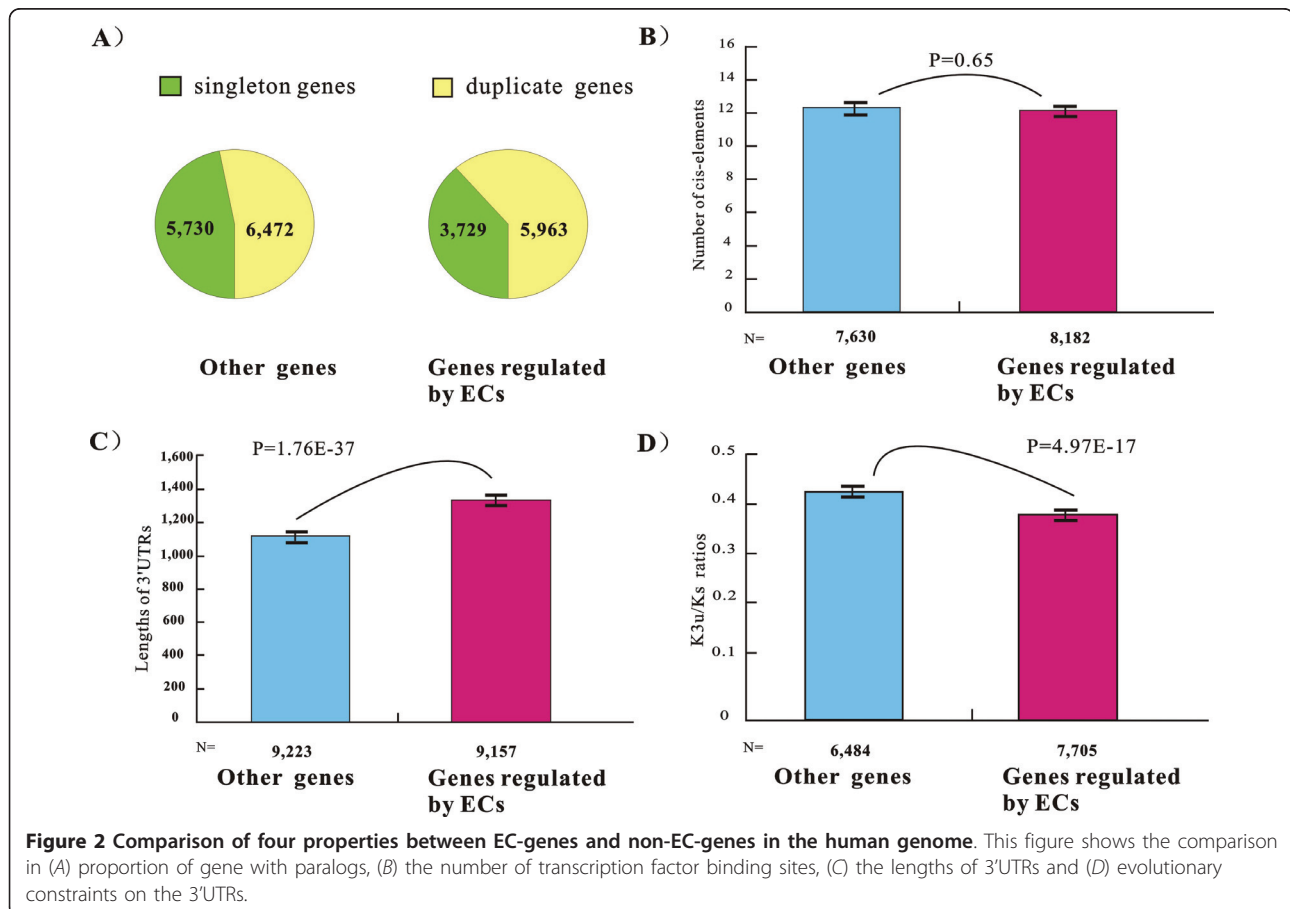
Fourth, it is interesting and important to know whether the enrichment of miRNA targets could still be observed in a single experiment. In Perl scripts, the keywords such as “microarray”, “array”, “affymetrix” and “chip” were used to search the abstracts of 4,162 papers. Many reports did not provide the raw datasets (.cel files), but rather only displayed the differentially expressed genes in tables in the main texts or supplemental materials. We read the full-text of  $\geq 50$  papers and investigated whether the raw datasets (.cel files) were available from GEO (NCBI Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>) or ArrayExpress (EBI Gene Expression Atlas, <http://www.ebi.ac.uk/gxa/>). Finally, we manually selected six

affymetrix raw datasets and used a uniform pipeline to identify the differentially expressed genes in seven cell lines treated with chemicals (see Methods). The enrichment of miRNA targets among differentially expressed genes was observed in each dataset, respectively (Additional file 7), implying the pervasive roles of miRNAs in responding to the various chemicals in different cells.

#### Enrichment of miRNA targets is not caused by other properties of EC-genes

Previous studies have reported significant differences between miRNA-targets and non-miRNA-targets. These differences may be potential sources of bias and contribute to the enrichment of miRNA targets in EC-genes. We adopted a sampling method to control these biases and then examined whether the enrichment of miRNA targets in EC-genes was still observable.

The first potential bias stems from the observation showing that genes with paralogs (gene duplication mechanisms) have a high probability of being targeted by miRNAs [28]. In Figure 2A, the EC-genes have a higher tendency of having paralogs compared with non-EC-genes (61% vs. 53%,  $p = 2.45E-36$ , *Chi-square test, two tailed*). We sampled genes with a probability of 0.71



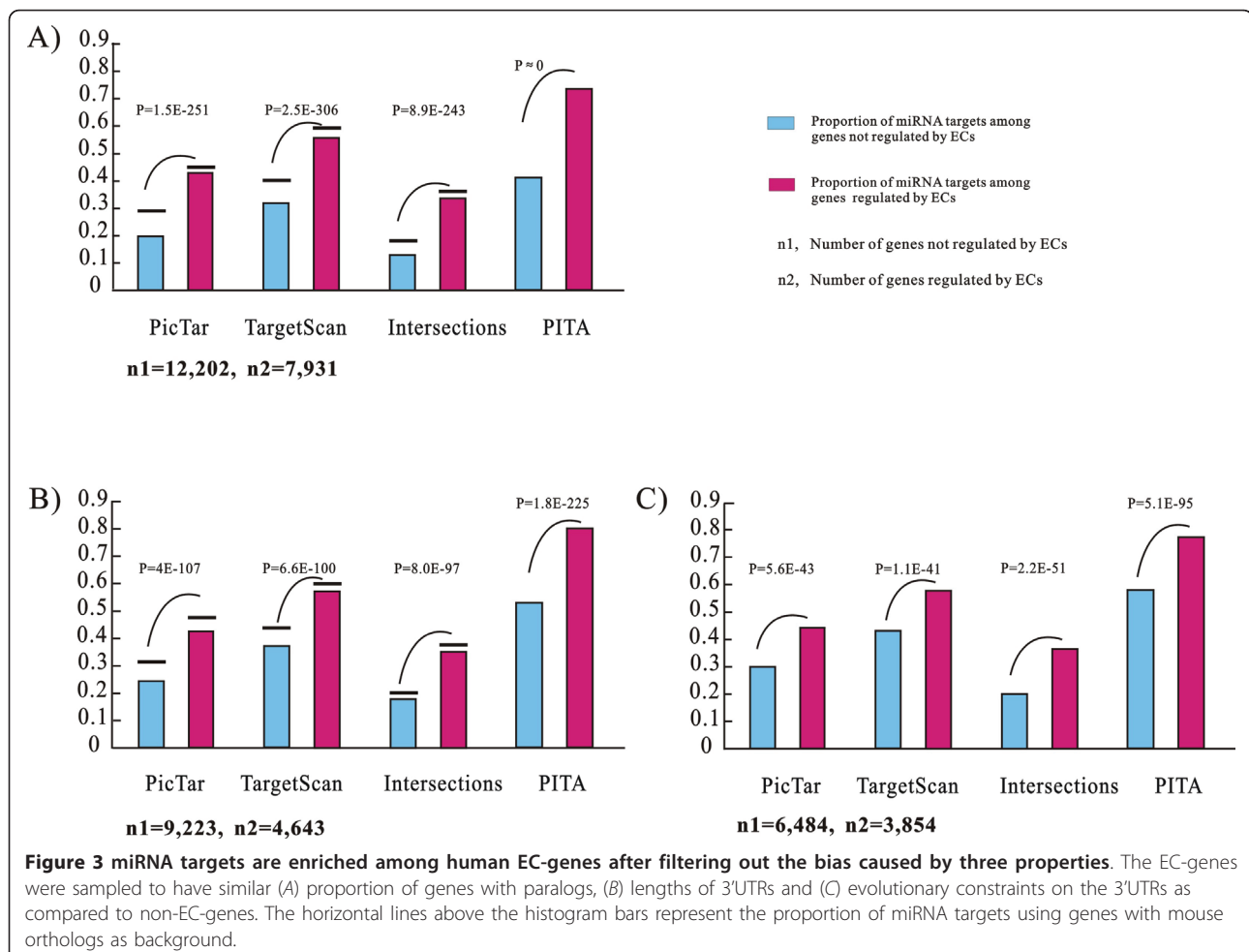
( $S = 0.71$ , calculated by  $[3,729 * (6,472/5,730)] / 5,963$ ) from the 5,963 EC-genes with paralogs. In this way, the dataset of sampled EC-genes were constructed with  $n = 3,729 + 5,963 * S$  and had the same proportion of genes with paralogs as that of the non-EC-genes. By eliminating potential bias from the higher propensity for gene duplication, we observed similar enrichment of miRNA targets among EC-genes (Figure 3A).

A second potential bias comes from the observation that genes with more TF-binding sites in the 5' upstream regions have a higher probability to become targets of miRNAs [29]; thus, an increased number of transcription factor binding sites may result in an over-representation of miRNA targets among EC-genes. We tested this hypothesis by using cis-elements that were exclusively predicted from conserved motif sequences among a set of vertebrate genome sequences (cisRED database, <http://www.cisred.org/>) [30]. A total of 94,321 and 100,112 predicted cis-elements were found to be in the proximity of 7,630 non-EC-genes and 8,182 EC-genes, respectively. Thus, EC-genes and non-EC-genes

have a similar average number of cis-elements ( $p = 0.65$ , Mann-Whitney U, two-tail test) (Figure 2B), suggesting that the enrichment of miRNA targets among EC-genes was not likely related to the over-presence of TF-binding sites in 5'-upstream regions.

A third potential bias results from the observation showing that genes with longer 3'UTRs are more likely to be regulated by distinct types of miRNA [31]. As shown in Figure 2C, the EC-genes tended to have longer 3'UTRs than non-EC-genes ( $p = 1.76E-37$ , Mann-Whitney U, two-tailed test), indicating that the enrichment may be attributed to the higher probability of EC-genes of being detected as miRNA targets. To test this, we sampled 4,643 EC-genes whose 3'UTR length falls into the 1<sup>st</sup> to 3<sup>rd</sup> quartiles of the 3'UTR lengths among the 9,223 non-EC-genes with available 3'UTR annotation. By eliminating potential bias caused by longer 3'UTRs, we further observed that miRNA targets were significantly enriched among EC-genes (Figure 3B).

Fourth, we inspected whether the 3'UTRs of the EC-genes and non-EC-genes were under the same level of



selective pressure. The substitution rates of the 3'UTR ( $K3u$ ) of each gene were normalized against the synonymous substitutions per synonymous site ( $Ks$ ) in the coding region of the same gene (see Methods). Using the ratio of  $K3u/Ks$  to estimate the evolutionary constraints on the 3'UTRs, we found the 3'UTR of the EC-genes tended to evolve more conservatively than non-EC-genes ( $p = 4.97E-17$ , Mann-Whitney U, two-tailed test) (Figure 2D). To explore the possibility that the elevated level of overall sequence conservation led to an over-representation of miRNA targets among EC-genes, we sampled 3,854 EC-genes whose  $K3u/Ks$  ratios fall into 1<sup>st</sup> to 3<sup>rd</sup> quartiles of the  $K3u/Ks$  ratios calculated from the 6,484 non-EC-genes. After getting rid of the potential bias caused by the more conserved 3'UTR, the enrichment of miRNA targets was again observable among sampled EC-genes (Figure 3C).

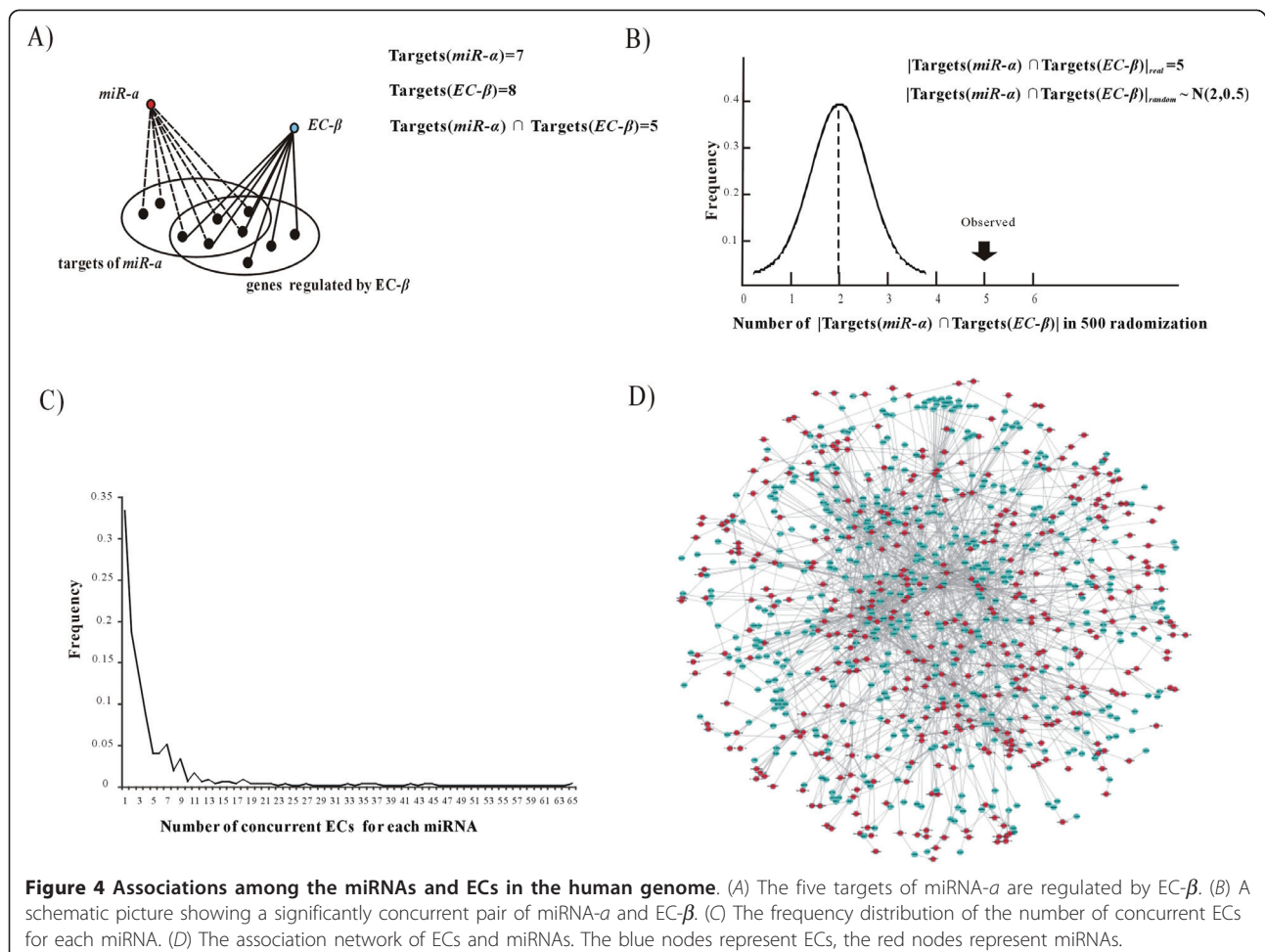
Collectively, these results demonstrate that the enrichment of miRNA targets is not a simple by-product of ancillary features of the analyzed gene set, but is a reflection of the propensity of being targeted by

miRNAs increasing the genes' probabilities of being regulated by ECs.

#### Target preference of miRNAs on genes regulated by different environmental chemicals

Based on the above statistical analyses, we have confirmed that genome-wide miRNA targets were preferentially regulated by ECs; but, whether miRNAs have different targeting preference for genes regulated by different ECs is still an open question, and vice versa. If the preference exists, we would expect a large number of concurrent miRNA-EC pairs, which tend to co-regulate the same genes.

We devised a randomization method to identify significant concurrent miRNA-ECs pairs (see a detailed description in Methods). As an example to illustrate the identifying pipelines: In Figure 4A, miRNA-*a* has 7 target genes, 5 of which are regulated by EC- $\beta$ ; therefore, the  $|\text{Targets}(\text{miR-}a)|$  is assigned to 7,  $|\text{Targets}(\text{EC-}\beta)|$  to 8 and  $|\text{Targets}(\text{miR-}a) \cap \text{Targets}(\text{EC-}\beta)|_{\text{real}}$  to 5. In each simulated run, the 7 targets for miRNA-*a* were



**Figure 4** Associations among the miRNAs and ECs in the human genome. (A) The five targets of miRNA-*a* are regulated by EC- $\beta$ . (B) A schematic picture showing a significantly concurrent pair of miRNA-*a* and EC- $\beta$ . (C) The frequency distribution of the number of concurrent ECs for each miRNA. (D) The association network of ECs and miRNAs. The blue nodes represent ECs, the red nodes represent miRNAs.

randomly replaced by the targets of other miRNAs by an edge-swapping procedure [32] (the algorithm of edge-swapping can sufficiently randomize the content of targets, while keeping the number of targets for each miRNA), then  $|\text{Targets}(miR-\alpha) \cap \text{Targets}(EC-\beta)|_{random}$  was recorded. Repeating this simulation 500 times, the  $|\text{Targets}(miR-\alpha) \cap \text{Targets}(EC-\beta)|_{random}$  followed a normal distribution as  $N(2, 0.5)$  (Figure 4B). The Z-score was adopted to assess the statistical significance of whether the miRNAs and ECs tend to regulate the same genes. Here, the Z-score was calculated by  $(5-2)/0.5 = 6$  and transformed to a *p-value* as  $9.87E-10$ ; thus, the targets of miRNA-*a* were considered to be preferentially regulated by the component- $\beta$ .

Using two lists, EC-Genes (Additional file 1) and miRNA-Target by TargetScan5.1, a miRNA-EC pair was considered to be significantly concurrent if the FDR-corrected *p-value* (the *q-value*) was less than 0.01. Finally, we identified 1,842 concurrent interactions among 407 miRNAs and 497 ECs (Additional file 8), which tend to synergistically regulate the same gene sets. Therefore, distinct miRNAs tend to be “adopted” to regulate genes in response to different ECs.

#### Association network of miRNAs and ECs

Graph theory provides paradigms to study biological networks [33]. Here, miRNAs and ECs can be represented respectively by different colored nodes, the concurrent relationship by links. We constructed the association network of miRNAs and ECs to provide a global view of how miRNAs function in concert with ECs. As shown in Figure 4C, the number of the concurrent ECs for each miRNA followed a power-law distribution, where a small proportion of miRNAs connect to many ECs; whereas, a large number of miRNAs only connected to one or two ECs. In this way, it is possible to select a single miRNA or a combination of miRNAs as biological markers in functional studies of their concurrent ECs.

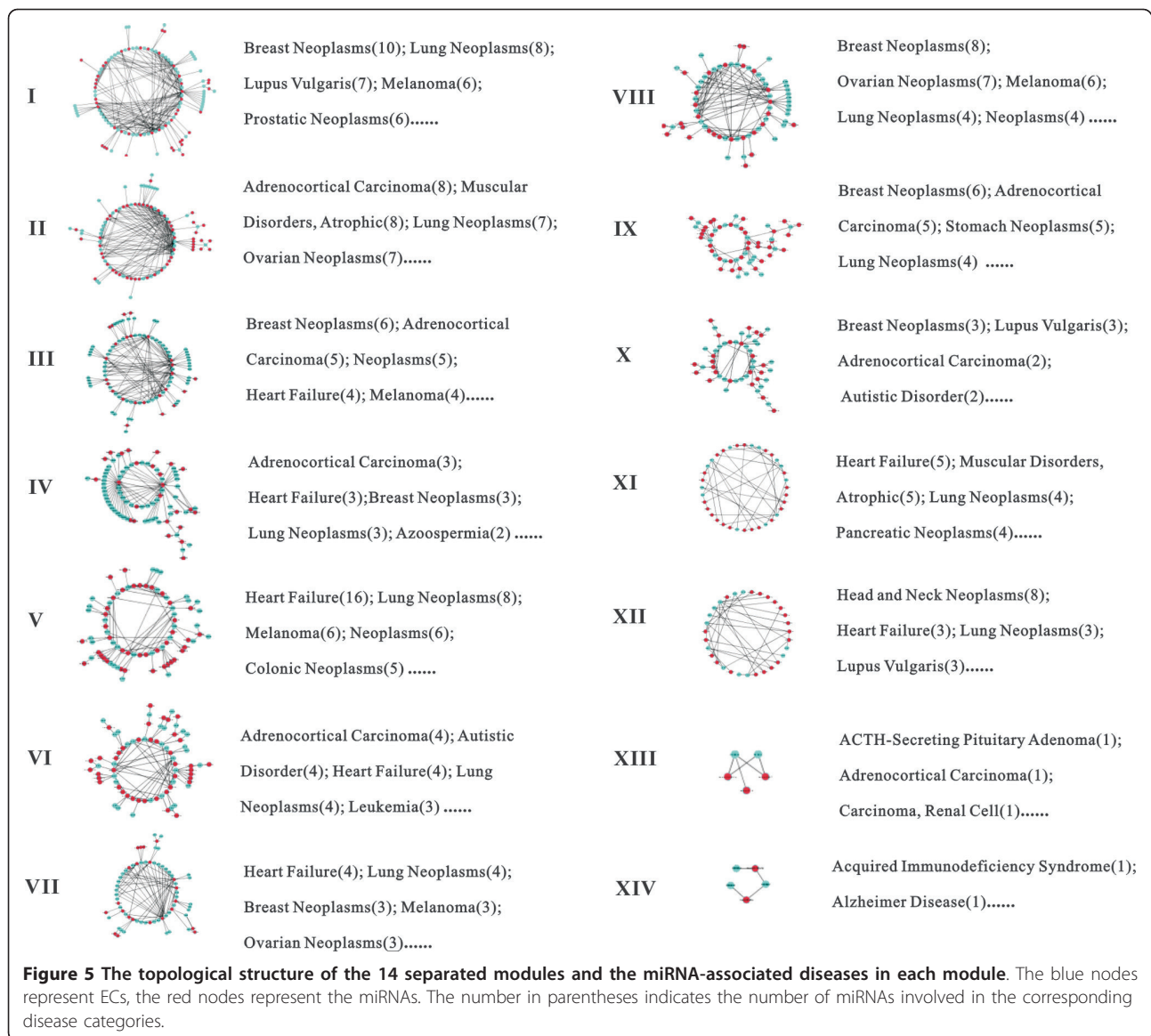
Besides a scale-free structure, the network also demonstrated a modular structure (Figure 4D); that is, a set of miRNAs and ECs were found to be densely concurrent in community-like modules, suggesting that miRNAs in some functional pathway may be co-operatively “adopted” to respond to ECs [34]. We used the algorithm of *Guimera* and *Amaral* [35] to measure the modularity (see Methods), because it performed well in making links within modules much denser than those across modules [36] and has been validated in our previous network analysis [37]. With a final value of modularity being 0.65, 14 topological modules could be separated. In each module, the miRNAs and ECs were closely connected (see Figure 5 for topological structures and Additional file 9 for description of the 14 modules).

We used available disease information of miRNAs to explore the potential function of each module. Three databases have been recently developed, HMDD [38], miR2Disease [39] and PhenomiR [40], which contain a large number of miRNA-disease associations from the literatures (*i.e.*, the abnormal regulations of miRNAs correlated with or leading to diseases). Therefore, the correlation of chemicals associated with human diseases could be interpreted by integrating the available disease information on miRNAs with the network modules. As shown in Figure 5, eight out of 19 miRNAs of module XII involved in “Head and Neck neoplasm”, indicated that the concurrent ECs of this module had a high probability to be risk factors for head and neck neoplasms. For 16 out of 40 miRNAs of module V involved in “Heart Failure”, the information on the concurrent ECs of this module could aid greater understanding of the regulatory mechanisms of heart disease.

#### Discussion

In this study, we showed that miRNA targets were preferentially regulated by ECs in *Homo sapiens*, *Mus musculus* and *Rattus norvegicus*. The enrichment of miRNA targets in genes associated with ECs was also confirmed by the STITCH database (<http://stitch.embl.de/>), another well-known free resource of associations between chemicals and proteins [41] (see Additional file 10 for the results of comparative analyses). Therefore, miRNA mediated post-transcriptional regulation may be a pervasive strategy for mammals to cope with irritation caused by ECs. The concurrent analysis revealed that distinct miRNAs regulated genes in response to different ECs. Based on these findings, we proposed that abnormal regulation of miRNAs and protein-coding genes by ECs may eventually disrupt normal signal transduction pathways or destroy the dosage balance of protein complexes. Hence, miRNAs and their targets should be given more attention in studies on environmental health.

It is known that the miRNAs identified to date are incomplete. More and more miRNAs have been recently discovered using next-generation sequencing platforms. One may suspect that the enrichment of miRNA targets may be only established using the current version of miRNAs and these tendencies might disappear with the identification of more miRNAs. However, considering the principles of miRNA-Target recognition, we speculate that findings of more miRNAs will not weaken but rather consolidate this enrichment. First, if a new miRNA belongs to a known family, its targets should have been predicted by their homologous miRNAs because of similar mature seed sequences [42]. Second, different miRNAs tend to synergistically target the same genes [43]. Although a new miRNA cannot be classified



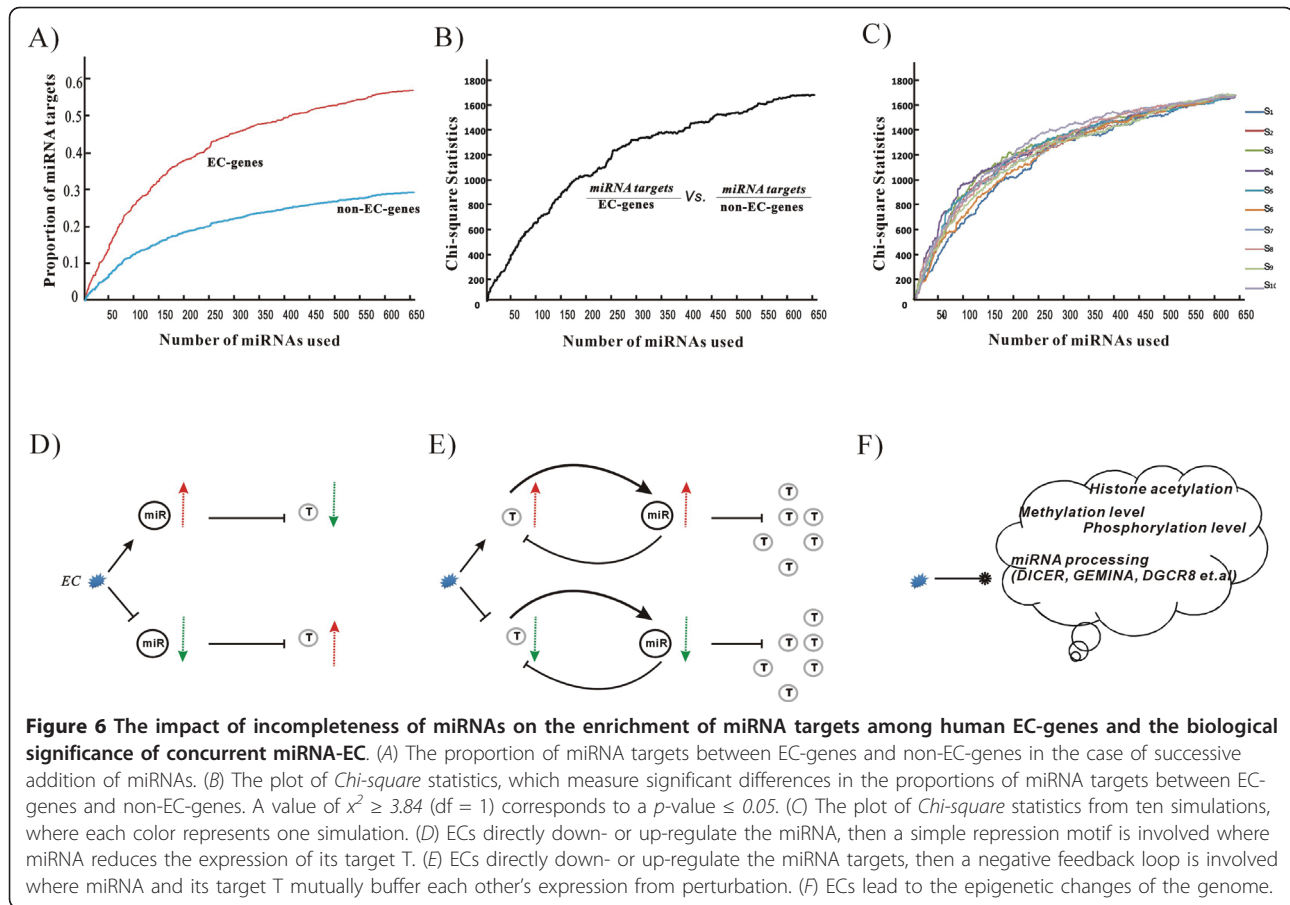
into a known family, it may not dramatically increase the number of new targets. This expectation has been further confirmed by the following simulation. Of 659 miRNAs contained in TargetScan 5.1, we successively sampled one miRNA into a new miRNA dataset, and then used these miRNAs to compare the proportion of their targets between EC-genes and non-EC-genes. As more miRNAs were placed into the new miRNA dataset ( $n = 1, 2, \dots, 658, 659$ ), the difference in proportion of miRNA targets between EC-genes and non-EC-genes became more and more pronounced (Figure 6A and 6B). Surprisingly, in the very early stage, when accumulatively transferring  $\geq 5$  miRNAs, we observed the enrichment of miRNA targets among EC-genes (see Figure 6C for the plot of Chi-square statistics from ten simulations). Therefore, miRNA-target propensity

increases the preference of being regulated by ECs and discovering new miRNAs will consolidate the tendency for enrichment.

Using statistical analysis, we found miRNAs had different preferences for targeting genes in response to different ECs, (i.e., a set of miRNAs were often concurrent with a specific EC). From a biological view, three paradigms could be used to explain the strong concurrence between ECs and miRNAs, where some have been verified by recent toxicological studies.

In the first paradigm, the ECs directly down- or up-regulate the miRNAs [44] and subsequently their targets are up- or down-regulated accordingly (Figure 6D). In brain tissue, hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), a common environmental contaminant, induced the over-expression of *miR-206*, *miR-30* and *miR-195*,





which then inhibited the expression of the target *BDNF* gene and contributed to neuro-toxicity and CNS disorders. Exposure to RDX also induced aberrant expressions of other onco-miRNAs and tumor-suppressing miRNAs, such as *let-7*, *miR-10b*, *miR-15*, *miR-16*, *miR-26* and *miR-181*, which regulated tumor pathogenesis or genes related to the cell cycle (e.g., *TNKS*) [45]. In human airway epithelial cells, diesel exhaust particles (DEP), the largest source of emitted airborne particulate matter (PM), induced *miR-513b*, *miR-513c*, *miR-923*, *miR-494* and *miR-338*, and repressed *miR-31\**, *miR-26b*, *miR-96*, *miR-27a*, *miR-135* and *miR-374a*. The subsequent target genes such as *IL-8*, *IL-6*, *tumor necrosis factor- $\alpha$* , *B7-H1* and *PRMT5*, were found to be strongly associated with inflammatory-response pathways and tumorigenic disease signatures [46].

The second paradigm is that ECs do not regulate miRNAs but rather their targets directly (Figure 6E). In the expression-buffering motifs, where a miRNA and its target mutually buffer each other's expression from perturbation in a negative feedback loop [47], miRNAs may be induced to be inversely regulated to buffer the expression fluctuation of their targets when exposed to ECs [9,10]. It is difficult to distinguish the second paradigm

from the first, but the biological significance is clear. The proposal for drug design follows that, if the drug is directly designed to a miRNA target (e.g., down-regulating an onco-gene), whose expression level is maintained by miRNA, the down-regulation of targets may promote the corresponding miRNA to be down-regulated, leading to the unexpected up-regulation of other targets and even deleterious phenotypes.

The third paradigm is that exposure to ECs alters the methylation level of global DNA, histone acetylation [48], methylation [49], and phosphorylation. In this way, epigenetic changes will lead to expression variation of both the miRNAs and their targets (Figure 6F). For example, the association between the chemical black-carbon and blood pressure was mediated by the modification of nucleotide polymorphisms in miRNA processing genes, such as *DICER*, *GEMIN4*, *DGCR8*, *GEMIN3* and *GEMIN4* [50].

No matter which particular paradigm is, the interactions between miRNAs and ECs add to the potential roles of miRNAs in chemical modulation of gene expression throughout the entire genome. The regulatory mechanisms of miRNAs will help us to design more accurate biological markers of ECs or drugs in toxicology research.

## Conclusions

To the best of our knowledge, this is the first genome-wide association analysis among human miRNAs, their targets and ECs. Our analysis will pave the way for future studies for the functional characterization of miRNAs. This network study reveals more clear roles of miRNAs involvement in toxicology and is also valuable for studying the impact of ECs on human health.

## Methods

### The compilation of EC-genes

The protein-coding genes regulated by various ECs were retrieved from the publicly-available Comparative Toxicogenomics Database (CTD, <http://ctd.mdibl.org/downloads/>) [14]. Gene expression data are presented in the CTD such as “chemical x can increase, decrease, affect or not affect the expression of protein y”. Perl scripts were used to remove associations with negation such as “chemical x does not affect the expression of protein y”.

### The compilation of miRNA targets

The miRNAs and their predicted targets were taken from two previously published studies: TargetScan (<http://www.targetscan.org> version 5.1, updated April, 2009) [22] and PicTar (USSC genome browser database, <http://genome.ucsc.edu> four-way) [23]. Targets predicted by TargetScan 5.1 with a total context score of -0.3 or lower were ignored, where the score could quantitatively measure the overall target efficacy [51]. Targets with at least one conserved 7-mer or 8-mer were selected as reliable miRNA targets. The intersection dataset was constructed by the targets predicted both by the TargetScan5.1 and PicTar (four-way). The PITA targets were downloaded from the Weizman Institute website ([http://genie.weizmann.ac.il/pubs/mir07/mir07\\_data.html](http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html), updated August, 2008) [25], where a score less than -10 was used as the cutoff to select reliable miRNA targets.

### Analysis of microarray datasets

The Affymetrix raw datasets were downloaded from [ftp://ftp.ncbi.nih.gov/pub/geo/data/geo/raw\\_data/series/](ftp://ftp.ncbi.nih.gov/pub/geo/data/geo/raw_data/series/) and <http://www.ebi.ac.uk/arrayexpress/> GSE6013, the lung adenocarcinoma cells (A549) and SV40-transformed bronchial epithelial cells (Beas-2B) were treated with asbestos; GSE5679, normal dendritic cells were treated with RAR $\alpha$ -specific agonists (AM580) and the synthetic PPAR $\gamma$  ligand rosiglitazone (RSG), respectively; GSE6907, the HepG2 cells were treated with N-nitrosodimethylamine (NMN) and phenol, respectively; E-MEXP-1327, normal epithelia prostatic cells were treated with selenium and vitamin E, respectively, and the stromal cells were treated with selenium; E-MEXP-390,

colorectal carcinoma cells (HCT116) were treated with fluorouracil; E-MEXP-1171, *HCT116* cells were treated with 7-ethyl-10-hydroxy-20(S)-camptothecin (SN38).

In each test, the raw datasets were normalized using the Affymetrix detection algorithms in the MAS5 library and the background levels and PM/MM ratios were corrected according to the Affymetrix Statistical Algorithms. Based on the estimated expression values of probes (Affy library), the expression values of corresponding Ensembl genes were obtained by BioMart. Finally, each test consisted of one control (no treatment) and a series of treatments (a chemical). Afterwards, the t-test (two-tailed) was used to determine if a gene's expression intensity after treatment ( $G_a$ ) was significantly different from that before treatment ( $G_b$ ).

The null hypothesis was  $H_0 = G_a - G_b = 0$ .

A gene was considered to be differentially expressed if  $H_0$  was rejected ( $p$ -value  $\leq 0.01$ ) after treatment.

### The compilation of human genes

The human protein-coding genes annotated as ‘known genes’, human paralogs, human-mouse orthologs, the sequences of proteins, coding regions and 3'UTRs were downloaded from Ensembl using BioMart software (<http://www.biomart.org/>) [52]. For genes with multiple splice isoforms, the transcripts with the longest sequences were used for analysis.

### Measurement of the evolutionary constraints on the 3'UTRs of human genes

The Clustalw software [53] was used to globally align the protein sequences of human-mouse orthologs, and the corresponding coding sequences were realigned with the gaps in the alignment trimmed. The  $K_s$  was estimated from the codon-based nucleotide sequence alignment by using the *Yang-Nielsen* maximum-likelihood method implemented in the *yn00* program of the PAML package [54]. The Clustalw software was used to globally align the 3'UTRs of orthologs, the substitution rate per site  $K_{3u}$  with the *Kimura* two-parameter model was calculated by *distmat* program of the EMBOSS package [55]. Finally, the ratios of  $K_{3u}/K_s$  were used to estimate the evolutionary constraints on the 3'UTRs for individual genes.

### Identification of significant concurrence between miRNAs and ECs

This model tests whether a given pair of miRNA-EC co-regulates the same genes at a higher rate, while considering the distribution of the number of different miRNAs regulating each target as a background. Two steps were followed: (a) the assignment matrix of miRNAs to targets were subjected to 100,000 iterations of the edge-swapping procedure, while keeping the number of

targets for each miRNA and keeping the number of regulator miRNA for each target [32], (b) for a pair of miRNA,  $\alpha$ , and EC,  $\beta$ , with their set of targets, Targets(*miR- $\alpha$* ) and Targets(*EC- $\beta$* ), respectively, the number of  $|\text{Targets}(\text{miR-}\alpha) \cap \text{Targets}(\text{EC-}\beta)|_{\text{random}}$  was recorded if it was larger than 0. The steps from (a) to (b) were repeated 500 times to obtain the distribution of  $|\text{Targets}(\text{miR-}\alpha) \cap \text{Targets}(\text{EC-}\beta)|_{\text{random}}$ .

Z-scores and *p*-values were employed to determine whether the value of  $|\text{Targets}(\text{miR-}\alpha) \cap \text{Targets}(\text{EC-}\beta)|_{\text{real}}$  significantly deviated from the distribution of  $|\text{Targets}(\text{miR-}\alpha) \cap \text{Targets}(\text{EC-}\beta)|_{\text{random}}$  generated from 500 random simulations.

$$Z = \frac{N_{\text{real}} - \mu(N_{\text{random}})}{\sigma(N_{\text{random}})}$$

Where  $N_{\text{real}}$  was the number of  $|\text{Targets}(\text{miR-}\alpha) \cap \text{Targets}(\text{EC-}\beta)|_{\text{real}}$ ,  $N_{\text{random}}$  was the number of  $|\text{Targets}(\text{miR-}\alpha) \cap \text{Targets}(\text{EC-}\beta)|_{\text{random}}$ ,  $\mu$  and  $\sigma$  denoted the mean and the standard deviation of the  $|\text{Targets}(\text{miR-}\alpha) \cap \text{Targets}(\text{EC-}\beta)|_{\text{random}}$  respectively.

The Z-score was then transformed to the *p*-value  $p = 1 - \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$  (Calculated by the *NORMDIST* function in *Microsoft Excel*)

Because the above statistical significance analyses involve the simultaneous testing of thousands of hypotheses, multiple hypotheses testing is important to control the overall Type I error rate. The *p*-values in EC-miRNA concurrent analysis were FDR corrected using the *Q-value* program from R package [56].

Finally, the obtained *q*-values were used to assess the statistical significance of the concurrence between a miRNA and an EC.

### The network and modules analysis

The association network of miRNAs and ECs was displayed by the Cytoscape software <http://www.cytoscape.org/>[57]. The least-squares method was used to estimate power-law exponent of  $p(K) \propto K^{-t}$  for log-transformed data (*t*, power exponent; *K*, degree). Since the estimated power-law exponent was 2.0, the method for study of scale-free structure was applied in analysis of miRNA-EC network. The algorithm of *Guimera* and *Amaral* [35], with parameter settings as iteration factor = 1.0, cooling factor = 0.95 and number of randomization = 100, was used to measure the extent of modularity of network and separate the network into topological modules.

### The disease information of miRNAs

The disease categories associated with miRNAs were integrated based on the following three published

studies: <http://202.38.126.151/hmdd/mirna/md/>[38], <http://www.mir2disease.org/>[39] and <http://mips.helmholtz-muenchen.de/phenomir/>[40].

### Statistical analysis and computational methods

Comparison of proportions from miRNA-targets between EC-genes and non-EC-genes were performed using the *Chi-square* two-tailed test.

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where  $O_{ij}$  was the observed number in row *i* of column *j*,  $E_{ij}$  was the expected number in row *i* of column *j*. For the condition of *df* = 1, the value of  $\chi^2 \geq 3.84$  corresponded to *p* ≤ 0.05, indicating a significant difference.

The comparison of the number of cis-elements, *K3u/Ks* and the lengths of 3'UTRs between EC-genes and non-EC-genes were performed using the Mann-Whitney U with two-tailed test.

A preliminary analysis of datasets and computations were performed on a Linux cluster with 16 nodes (Intel 5130, 2.0 GHz CPU, 4G memory, Research Center for Systematic and Evolutionary Botany, Institute of Botany, CAS). The updating of datasets was performed on the Linux clusters provided by the Institute of Genomic Medicine (Wenzhou Medical College) and Dalian Institute of Chemical Physics (CAS), respectively. Perl (<http://perl.org>) and R (<http://www.r-project.org/>) scripts were used for analyses, and can be obtained on request.

### Additional material

**Additional file 1: Table S1: The 42,770 regulatory relationships among 1938 chemicals and 9,692 human protein-coding genes.**

**Additional file 2: Table S2: The list of miRNA targets predicted by PicTar, TargetScan5.1, both programs of PicTar and TargetScan5.1 (intersections), and by PITA.**

**Additional file 3: Figure S1: miRNA targets are enriched among EC-genes in *Mus musculus* and *Rattus norvegicus*.** This figure shows the proportion of miRNA targets predicted (A) by *Mus musculus* TargetScan5.1, (B) by *Mus musculus* PITA and (C) by *Rattus norvegicus* TargetScan5.1. As there are no miRNA targets from the PicTar prediction for both species and PITA prediction for *Rattus norvegicus*, this figure shows the proportion of miRNA targets predicted by TargetScan5.1 and the *Rattus norvegicus* PITA. The horizontal lines above histogram bars represent the proportion of miRNA targets using genes with human orthologs as background.

**Additional file 4: Table S3: The 4,126 papers investigating the expression profiles of human genes in response to ECs.**

**Additional file 5: Figure S2: miRNA targets are enriched among human EC-genes after filtering out the cancer-related genes in simulations.** This figure shows the proportion of miRNA targets predicted by PicTar, TargetScan5.1, both programs of PicTar and TargetScan5.1 (intersections), and PITA. The horizontal lines above the histogram bars represent the proportion of miRNA targets using genes with mouse orthologs as background.

**Additional file 6: Figure S3: miRNA targets are enriched among human EC-genes after controlling for potential false-negatives of non-EC-genes.** This figure shows the proportion of miRNA targets predicted by PicTar, TargetScan5.1, both programs of PicTar and TargetScan5.1 (intersections), and PITA. The horizontal lines above the histogram bars represent the proportion of miRNA targets using genes with mouse orthologs as background.

**Additional file 7: Figure S4: miRNA targets are enriched among human EC-genes in 11 raw-data-based microarray datasets.** This figure shows the proportion of miRNA targets between differentially vs. non-differentially regulated genes where (A) A549 cells treated with asbestos, (B) Beas-2B cells treated with asbestos, (C) Dendritic cells treated with AM580, (D) Dendritic cells treated with rosiglitazone, (E) Epithelia cells treated with selenium, (F) Epithelia cells treated with Vitamin E, (G) Stromal cells treated with selenium, (H) HCT116 cells treated with fluorouracil, (I) HCT116 cells treated with SN38, (J) HepG2 cells treated with NMN and (K) HepG2 cells treated with phenol. The horizontal lines above the histogram bars represent the proportion of miRNA targets using genes with mouse orthologs as background.

**Additional file 8: Table S4: The 1,842 pairs of significantly concurrent EC-miRNA based on the TargetScan5.1.**

**Additional file 9: Table S5: The 14 topological modules with their miRNAs and ECs.** The number in the bracket indicates the number of miRNAs associated with the corresponding disease categories.

**Additional file 10: Figure S5: miRNA targets are enriched among human EC-associated genes by STITCH2.0.** This figure shows the proportion of miRNA targets predicted (A) by PicTar, (B) by TargetScan5.1, (C) by both programs of PicTar and TargetScan5.1 (intersections), and (D) by PITA. The horizontal lines above the histogram bars represent the proportion of miRNA targets using genes with mouse orthologs as background.

#### Abbreviations

CTD: comparative toxicogenomics database; EC: environmental chemicals; EC-genes: genes observed to be regulated by environmental chemicals; Non-EC-genes: genes not observed to be regulated by the environmental chemicals; UTR: untranslated region; *Ks*: synonymous substitution rates of coding region; *K3u*: substitution rate of 3'UTR;

#### Acknowledgements

We thanked Jingjing Li for useful information and critical discussion, this work was supported by the National Basic Research Program of China (2007CB108800) and the National High Technology Research and Development Program of China (2009AA01A137).

#### Author details

<sup>1</sup>Key laboratory of Photosynthesis and Environmental Molecular Physiology, the Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China. <sup>2</sup>Institute of Genomic Medicine, Wenzhou Medical College, Wenzhou, 325035, China. <sup>3</sup>State key laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, 116023, China.

#### Authors' contributions

XW designed and performed the experiments. XW analyzed the datasets and wrote the paper. YS participated in the analysis of datasets. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 1 November 2010 Accepted: 18 May 2011  
Published: 18 May 2011

#### References

1. Bartel DP: MicroRNAs: genomics, biogenesis, mechanism and function. *Cell* 2004, **116**:281-297.

2. He L, Hannon GJ: MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 2004, **5**:522-531.
3. Bagga S, Bracht J, Hunter S, Massirer K, Holtz J, Eachus R, Pasquinelli AE: Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* 2005, **122**:553-63.
4. Ding XC, Grosshans H: Repression of *C. elegans* microRNA targets at the initiation level of translation requires GW182 proteins. *EMBO J* 2009, **28**:213-22.
5. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM: Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005, **433**:769-73.
6. Giraldez AJ, Mishima Y, Rihel J, Gockoc RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF: Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 2006, **312**:75-9.
7. Yu Z, Jian Z, Shen SH, Purisima E, Wang E: Global analysis of microRNA target gene expression reveals that miRNA targets are lower expressed in mature mouse and drosophila tissues than in the embryos. *Nucleic Acids Res* 2007, **35**:152-64.
8. Hornstein E, Shomron N: Canalization of development by microRNAs. *Nat. Genet* 2006, **38**:S20-S24.
9. Yan Li, Wang Fay, Jin-A Lee, Fen-Biao Gao: MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes Dev* 2006, **20**:2793-2805.
10. Cohen SM, Brennecke J, Stark A: Denoising feedback loops by thresholding - a new role for microRNAs. *Genes Dev* 2006, **20**:2769-2772.
11. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT: c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 2005, **435**:839-843.
12. Edwards TM, Myers JP: Environmental exposures and gene regulation in disease etiology. *Cien saude Colet* 2008, **13**:269-281.
13. Audouze K, Juncker AS, Roque FJ, Krysiak-Baltyn K, Weinhold N, Taboureaux O, Jensen TS, Brunak S: Deciphering diseases and biological targets for environmental chemicals using toxicogenomics networks. *PLoS Comput Biol* 2010, **6**:e1000788.
14. Davis AP, Murphy CG, Saraceni-Richards CA, Rosenstein MC, Wiegiers TC, Mattingly CJ: Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Res* 2009, **37** Database: D786-92.
15. Thomas C Wiegiers, Allan Peter Davis, Bretonnel Cohen K, Lynette Hirschman, Carolyn J Mattingly: Text mining and manual curation of chemical-gene-disease networks for the Comparative Toxicogenomics Database (CTD). *BMC Bioinformatics* 2009, **10**:326.
16. Gohlke JM, Thomas R, Zhang Y, Rosenstein MC, Davis AP, Murphy C, Becker KG, Mattingly CJ, Portier CJ: Genetic and environmental pathways to complex diseases. *BMC Syst Biol* 2009, **3**:46.
17. Patel CJ, Butte AJ: Predicting environmental chemical factors associated with disease-related gene expression data. *BMC Med Genomics* 2010, **3**:17.
18. Garofalo M, Croce CM: microRNAs: Master Regulators as Potential Therapeutics in Cancer. *Annu Rev Pharmacol Toxicol* 2010, **18**.
19. Bader AG, Brown D, Winkler M: The promise of microRNA replacement therapy. *Cancer Res* 2010, **70**:7027-30.
20. Chen T: The role of MicroRNA in chemical carcinogenesis. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* 2010, **28**:89-124.
21. Pogribny I, Igor P: MicroRNA dysregulation during chemical carcinogenesis. *Epigenomics* 2009, **1**:281-290.
22. Lewis BP, Burge CB, Bartel DP: Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005, **120**:15-20.
23. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: Combinatorial microRNA target predictions. *Nat Genet* 2005, **37**:495-500.
24. Chen K, Rajewsky N: Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* 2006, **38**:1452-1456.
25. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: The role of site accessibility in microRNA target recognition. *Nat Genet* 2007, **39**:1278-1284.
26. Sun J, Zhao Z: A comparative study of cancer proteins in the human protein-protein interaction network. *BMC Genomics* 2010, **11**(Suppl 3):S5.
27. Axelsen JB, Lotem J, Sachs L, Domany E: Genes overexpressed in different human solid cancers exhibit different tissue-specific expression profiles. *Proc Natl Acad Sci USA* 2007, **104**:13122-13127.

28. Li J, Musso G, Zhang Z: **Preferential regulation of duplicated genes by microRNAs in mammals.** *Genome Biol* 2008, **9**(8):R132.
29. Cui Q, Yu Z, Pan Y, Purisima EO, Wang E: **MicroRNAs preferentially target the genes with high transcriptional regulation complexity.** *Biochem Biophys Res Commun* 2007, **352**:733-8.
30. Robertson G, Bilenky M, Lin K, He A, Yuen W, Dagpinar M, Varhol R, Teague K, Griffith OL, Zhang X, Pan Y, Hassel M, Sleumer MC, Pan W, Pleasance ED, Chuang M, Hao H, Li YY, Robertson N, Fjell C, Li B, Montgomery SB, Astakhova T, Zhou J, Sander J, Siddiqui AS, Jones SJ: **cisRED: a database system for genome-scale computational discovery of regulatory elements.** *Nucleic Acids Res* 2006, **34**:D68-73.
31. Cheng C, Bhardwaj N, Gerstein M: **The relationship between the evolution of microRNA targets and the length of their UTRs.** *BMC Genomics* 2009, **10**:431.
32. Maslov S, Sneppen M, Zaliznyak A: **Detection of topological patterns in complex networks: Correlation profile of the internet.** *Physica A* 2004, **333**:529-540.
33. Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet* 2004, **5**:101-113.
34. Newman MEJ: **The structure and function of complex networks.** *SIAM Rev* 2003, **45**:167-256.
35. Girvan M, Newman ME: **Community structure in social and biological networks.** *Proc Natl Acad Sci USA* 2002, **99**:7821-26.
36. Danon L, Diaz-Guilera A, Duch J, Arenas A: **Comparing community structure identification.** *J Stat Mech* 2005, **P09008**:1-10.
37. Wu X, Qi X: **Genes encoding hub and bottleneck enzymes of the Arabidopsis metabolic network preferentially retain homeologs through whole genome duplication.** *BMC Evolutionary biology* 2010, **10**:145.
38. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q: **An analysis of human microRNA and disease associations.** *PLoS One* 2008, **3**(10):e3420.
39. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y: **miR2Disease: a manually curated database for microRNA deregulation in human disease.** *Nucleic Acids Res* 2009, **37**:D98-104.
40. Ruepp A, Kowarsch A, Schmid D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ: **PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes.** *Genome Biol* 2010, **11**:R6.
41. Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P: **STITCH: interaction networks of chemicals and proteins.** *Nucleic Acids Res* 2008, **36**:D684-8.
42. Yong Huang, Xun Gu: **A bootstrap based analysis pipeline for efficient classification of phylogenetically related animal miRNAs.** *BMC Genomics* 2007, **8**:66.
43. Shalgi Reut, Lieber Daniel, Oren Moshe, Pilpel Yitzhak: **Global and Local Architecture of the Mammalian microRNA-Transcription Factor Regulatory Network.** *PLoS Comput Biol* 2007, **3**(7):e131.
44. Zhiguang Li, William S Branham, Stacey L Dial, Wang Yexun, Lei Guo, Leming Shi, Tao Chen: **Genomic analysis of microRNA time-course expression in liver of mice treated with genotoxic carcinogen N-ethyl-N-nitrosourea.** *BMC Genomics* 2010, **11**:609.
45. Zhang B, Pan X: **RDX induces aberrant expression of microRNAs in mouse brain and liver.** *Environ Health Perspect* 2009, **117**:231-40.
46. Jardim MJ, Fry RC, Jaspers I, Dailey L, Diaz-Sanchez D: **Disruption of microRNA expression in human airway cells by diesel exhaust particles is linked to tumorigenesis-associated pathways.** *Environ Health Perspect* 2009, **117**:1745-51.
47. Chung-I Wu, Shen Yang, Tang Tian: **Evolution under canalization and the dual roles of microRNAs—A hypothesis.** *Genome Res* 2009, **19**(5):734-43.
48. Rekowski MW, Giannis A: **Histone acetylation modulation by small molecules: A chemical approach.** *Biochim Biophys Acta* 2010, **1799**(10-12):760-7.
49. Su Z, Han L, Zhao Z: **Conservation and divergence of DNA methylation in eukaryotes: new insights from single base-resolution DNA methylomes.** *Epigenetics* 2011, **6**(2):134-140.
50. Wilker EH, Baccarelli A, Suh H, Vokonas P, Wright RO, Schwartz J: **Black carbon exposures, blood pressure, and interactions with single nucleotide polymorphisms in MicroRNA processing genes.** *Environ Health Perspect* 2010, **118**:943-8.
51. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP: **The impact of microRNAs on protein output.** *Nature* 2008, **455**:64-71.
52. Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A: **BioMart-biological queries made easy.** *BMC Genomics* 2009, **14**:10-22.
53. Thompson JD, Gibson TJ, Higgins DG: **Multiple sequence alignment using ClustalW and ClustalX.** *Curr Protoc Bioinformatics* 2002, **Chapter 2**: Unit 2.3.
54. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13**(5):555-556.
55. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**(6):276-277.
56. Storey JD, Tibshirani R: **Statistical significance for genome-wide studies.** *Proc Natl Acad Sci USA* 2003, **100**(16):9440-9445.
57. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD: **Cytoscape Web: an interactive web-based network browser.** *Bioinformatics* 2010, **26**(18):2347-8.

doi:10.1186/1471-2164-12-244

**Cite this article as:** Wu and Song: Preferential regulation of miRNA targets by environmental chemicals in the human genome. *BMC Genomics* 2011 **12**:244.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

