BMC Bioinformatics

RESEARCH ARTICLE   **Open Access**

Uncovering packaging features of co-regulated modules based on human protein interaction and transcriptional regulatory networks

Lina Chen[1*†], Hong Wang[1†], Liangcai Zhang[1†], Wan Li[1], Qian Wang[1], Yukui Shang[1], Yuehan He[1], Weiming He[2], Xu Li[1], Jingxie Tai[1], Xia Li[1*]

Abstract

Background: Network co-regulated modules are believed to have the functionality of packaging multiple biological entities, and can thus be assumed to coordinate many biological functions in their network neighbouring regions.

Results: Here, we weighted edges of a human protein interaction network and a transcriptional regulatory network to construct an integrated network, and introduce a probabilistic model and a bipartite graph framework to exploit human co-regulated modules and uncover their specific features in packaging different biological entities (genes, protein complexes or metabolic pathways). Finally, we identified 96 human co-regulated modules based on this method, and evaluate its effectiveness by comparing it with four other methods.

Conclusions: Dysfunctions in co-regulated interactions often occur in the development of cancer. Therefore, we focussed on an example co-regulated module and found that it could integrate a number of cancer-related genes. This was extended to causal dysfunctions of some complexes maintained by several physically interacting proteins, thus coordinating several metabolic pathways that directly underlie cancer.

Background

One of key challenges of the post-genomic era is to understand the complexity of molecular networks, and describe their applications to elucidate essential principles of cellular systems and disease machinery [1,2]. Spurred by advances in technology, several types of molecular networks, *e.g.* protein-protein interaction networks (PPINs), transcriptional regulatory networks (TRNs), and phenotype networks have been identified, providing us with a global landscape of how biological molecules may interact with one another. Many studies have demonstrated that PPINs and TRNs are essential for controlling the expression levels of genes and the activity of proteins, which mediates coordinated responses and adapted modifications to multifarious cellular stimuli [3,4]. Given this landscape, integrative analysis of both PPINs and TRNs is a major focus in systems biology and bioinformatics.

Many computational strategies based on integrated PPIN and TRN networks have been devised and used to decipher specific network structures [4,5] or their potential biological implications [6] that underlie disease traits.

In molecular networks, genes, proteins, and other molecules form components called 'functional modules' that are densely interconnected, but relatively isolated from other networks [7]. Recent surveys have shown that genes within a module or a cluster appear to have similar expression patterns, share common underlying regulatory mechanisms, and thus have strong associations with specific biological functions that determine the behaviour or phenotype of the cell [8,9]. Complex diseases are known to result from the loss of one or more normal essential functions. One such example is cancer. In the recent years, an increasing number of cancer studies have combined human gene expression profiling and computational-based module searching algorithms to obtain a more comprehensive view of the molecular underpinnings and regulatory relationships of cancer [10]. Segal *et al.* [11] have identified gene sets with similar behaviour across

* Correspondence: chenlina_2004@yahoo.com.cn; lixia@hrbmu.edu.cn
† Contributed equally
[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Hei Longjiang Province, China

microarrays, and constructed 'cancer module maps' to characterize a variety of clinical conditions. Whitfield *et al.* [12] have detected modules in which genes shared both similar expression profiles and similar transcription factor binding profiles. Pomeroy *et al.* [13] have explored regulatory modules using the conservation of co-expression relationships across a diverse range of organisms. The utility of microarray analysis provides more interpretable results than using gene lists alone. A study by Chuang *et al.* [14] have combined microarray analysis and the human PPIN to identify sub-network biomarkers for breast cancer, and proposed that integrated network-based approaches could help researchers acquire additional and more accurate molecular mechanisms for cancers. Another study by Cui *et al.* [15] have demonstrated that the co-regulatory mechanism of molecular networks could mediate cancer-related genes, convey their abnormal states through several functional modules, and eventually lead to uncontrolled cell growth, invasion, and metastasis in distant planes of the body. Thus, uncovering co-regulated modular structures in integrated molecular networks could provide valuable insights into the pathogenesis of cancer.

In this paper, we introduce a probabilistic model termed Co-Regulatory Analysis using Integrated Networks (CRAIN) to detect human co-regulated modules using an integrative weighted network of a PPIN and a TRN. Then the performance of our analysis is evaluated by cross-validation with biological evidence. Furthermore, we figure out biological relevance of our modules for assembling or rewiring biological entities such as genes, protein complexes, and metabolic pathways. Finally, exemplified by cancer, we investigate whether co-regulated modules are capable of assembling different biological entities with underlying mechanisms in tumorigenesis.

Results and Discussion
Overview of the identification of co-regulated modules
We scaled and merged a human PPIN and TRN, and constructed a highly quality integrated network of protein and transcription regulation interactions. Adopting a probabilistic model, we evaluated whether a cluster of co-regulated proteins was likely to form a module in the integrated network. Under this model, we formulated a log-likelihood ratio to compare the fit of a cluster to the desired structure with its likelihood, given that the interaction map was randomly constructed. Highly scoring sub-networks corresponded to likely modules. We used a heuristic strategy for module-detecting procedures consisting of: *(i)* seed initialization; *(ii)* seed expanding; and *(iii)* overlap filtering. Finally, we obtained 96 co-regulated modules (Additional file 1), each of which was co-regulated by one or more specific transcription factors (TFs). And furthermore, we used three bipartite

graphs to map our modules onto the biological entities of genes, protein complexes, and metabolic pathways to uncover the underlying biological significance of the modules. From our analysis, we concluded that in each module, co-regulated relationships might play important roles in packaging their binding genes, then extending to regulating complexes maintained by several physical interacting proteins, and thus involving in some metabolic pathways or disease traits.

Performance evaluation
Analysis of module robustness
We assessed the internal connectivity of each co-regulated module by comparison with its control clusters. To generate a control for a given module, we conducted random replacements for 10%, 20% or 30% of the module nodes with an equal number of proteins/TFs outside the module. We repeated this replacement process 100 times, and used the average connectivity for all analytical runs. Figure 1A shows the internal connectivity of the extracted modules and their controls. Inside connections of co-regulated modules decreased significantly with an increase in the replacement size during randomization experiments. We also studied the average connecting ratio of the nodes within each module to the ones outside of it. We found that the ratio in the real dataset was higher than in the randomization experiments (Figure 1B), suggesting that each of the identified modules was indeed densely connected, and robustly formed a local sub-network.
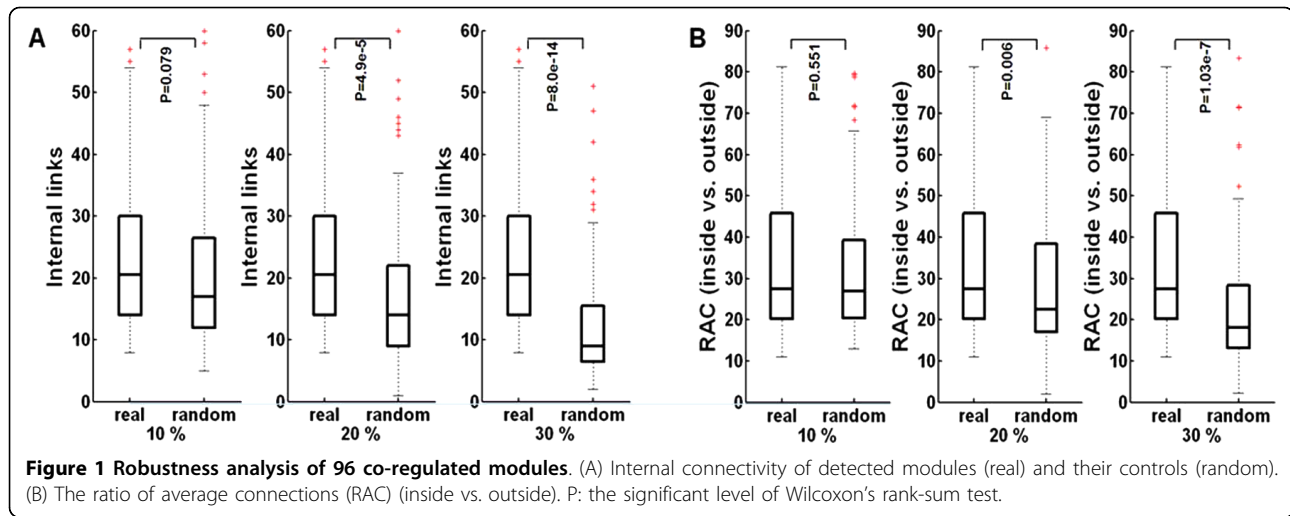
Analysis of module functional coherency
Using the TANGO toolkit [16], we performed Gene Ontology (GO) enrichment analysis for our extracted 96 modules, to identify strongly-associated functional categories. The TANGO algorithm includes all levels of GO, and computes raw enrichment p-values using a standard hyper-geometric test with a significant level of $p < 0.001$. Annotation results showed that 77 modules (80%) were significantly enriched in biological function (Additional file 2).

To quantify the functional consistency of each discovered module, we computed the Hit-rate and Miss-rate proposed by Milenkoviae *et al.* [17] for each module M (GO enrichment significant level $p < 0.001$):

$$Hit(M) = \frac{\max\{\left|N_M^1\right|, \left|N_M^2\right|, ..., \left|N_M^i\right|, ..., \left|N_M^t\right|\}}{\left|M\right|}$$

$$Miss(M) = \frac{\left|M\right| - \left|N_M^1 \cup N_M^2 \cup ... \cup N_M^i \cup ... \cup N_M^t\right|}{\left|M\right|}$$

For a given module M, N_M^i ($i = 1, 2,..., t$, where t represents the number of GO terms for which the

Figure 1 Robustness analysis of 96 co-regulated modules. (A) Internal connectivity of detected modules (real) and their controls (random). (B) The ratio of average connections (RAC) (inside vs. outside). P: the significant level of Wilcoxon's rank-sum test.

module M enriched) is the intersection gene set of module M and its enriched GO term i, and $|M|$ is the size of M. A higher Hit-rate indicated that more genes in module M convey a centralized biological function; a lower Miss-rate provided additional confirmation of our deduction. We binned the Hit-rates and Miss-rates in grades of 10%, and compared the Hit-rates and Miss-rates between our predicted modules and their controls (30% nodes replacement) (Figure 2). In the GO: biological process (BP) branch, 50 investigated modules in the real team had a Hit-rate above 90%, and 79 had a Miss-rate below 10%, while 17 modules in the control team had a Hit-rate above 90%, and 38 had a Miss-rate below 10%. The same observations for higher Hit-rate and lower Miss-rate were seen when analyzing the functional consistency of our investigated modules in the molecular function (MF) and cellular component (CC) categories. These results suggested that our method was capable of finding co-regulated modules with strong biological relevance. Similar results were found for the 10% and 20% node replacements (data not shown).

Multiple methods comparison

We validated the performance of CRAIN by comparison with four other module identification algorithms [18-20]: connected components (Connected), biconnected components (Biconnected), clique percolation method (CPM), and Markov cluster algorithm (MCL). For this process, we predicted modules using these four methods. Enrichment was computed using the standard hyper-geometric test by TANGO toolkit (significance level p < 0.001). For each method, we defined sensitivity as the proportion of annotations enriched in at least one module at p < 10^{-4}, and specificity as the proportion of modules enriched with at least one annotation at p < 10^{-4} [21]:

$$\text{Sensitivity} = \frac{\text{No. of annotations enriched in at least one module at P<}10^{-4}}{\text{No. of annotations}}$$

$$\text{Specificity} = \frac{\text{No. of modules enriched at least one annotation at P<}10^{-4}}{\text{No. of modules}}$$

The F-score summarizes the two measures, and is defined as follows:

$$\text{F} = \frac{2 \times \text{Sensitivity} \times \text{Specificity}}{\text{Sensitivity} + \text{Specificity}}$$
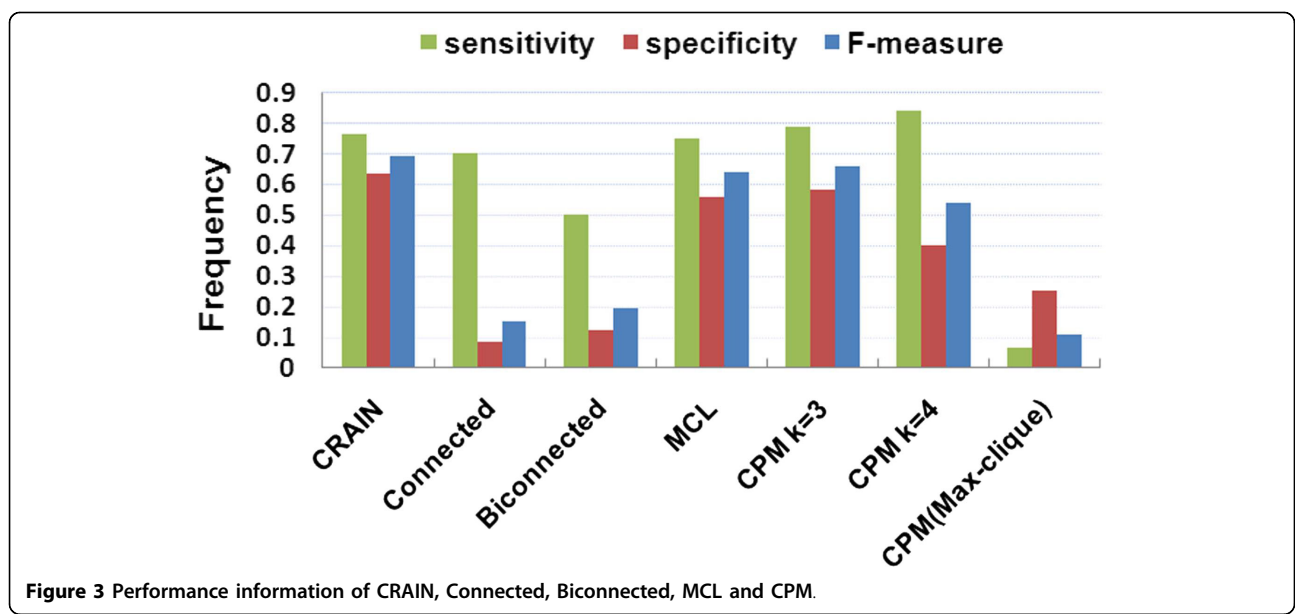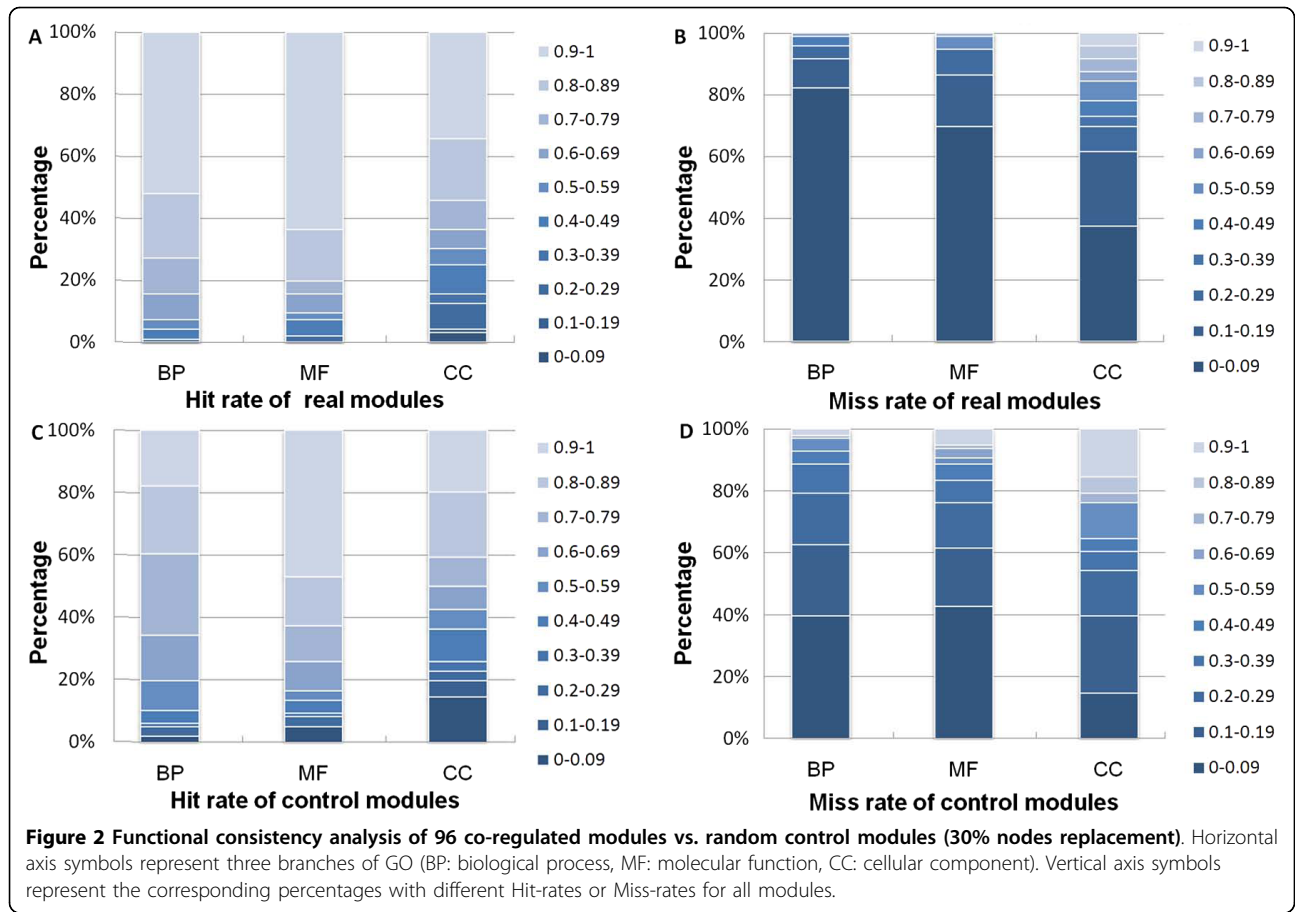
Figure 3 is a histogram of three measures: sensitivity, specificity and the summary measurement F-measure, for each algorithm. The results indicated that the F-score of our method was superior to the other methods. This suggested that CRAIN could return co-regulated modules with more affluent biological meanings.

Biological association of co-regulated modules with cancer

Cancer-related genes are often assumed to mediate each other through the co-regulatory mechanisms of molecular networks, causing abnormal states through several functional modules, and eventually leading to uncontrolled cell growth, invasion, and metastasis to distant planes of the body [15]. To investigate this, we used Fisher's exact test to check biological associations between cancer-mutated genes and each of the 96 identified modules. We found that 42 (43%) of modules were associated with cancer (p < 0.05, Additional file 3).

Packaging features of co-regulated modules

Furthermore, to determine the biological importance of the co-regulated modules, we investigated the role of transcription regulation in assembling or rewiring genes,

Figure 2 Functional consistency analysis of 96 co-regulated modules vs. random control modules (30% nodes replacement). Horizontal axis symbols represent three branches of GO (BP: biological process, MF: molecular function, CC: cellular component). Vertical axis symbols represent the corresponding percentages with different Hit-rates or Miss-rates for all modules.

Figure 3 Performance information of CRAIN, Connected, Biconnected, MCL and CPM.

protein complexes, and metabolic pathways within modules.

Gene packaging

For all 96 co-regulated modules, we labelled TFs and proteins with their associated biological functions. We found that each module could work as an 'assembler' to assemble a set of genes with similar biological functions that were regulated by one or more TFs. For example, Figure 4 illustrates one module associated with a 'biopolymer metabolic process' (module 27). In this module, two groups of regulated subsets were identified: one group consisted of *JUN* and three tumour-mutated genes (*CCND1*, *MSH2* and *BRCA1*). Recent studies have reported that *JUN*, a key cancer-related regulator, is important in carcinogenesis: inappropriate gene activation or numerous different genetic defects of *JUN* or its target genes could lead to cell growth inhibition, DNA damage or cell cycle delay, and these series of unexpected variations could finally have effects on tumour emergence, promotion and metastasis [22,23]. Another group contained five TFs (*RPA1*, *RPA2*, *TP53BP1*, *FUBP1*, and *JUN*) and their target genes (*BRCA1* and *BRCA2*). *BRCA1* and *BRCA2* are important tumour suppressor genes, whose loss of function is closely associated with tumorigenesis [24,25]. Several studies have reported that these two genes are involved in DNA recombination and DNA repair [26-28]. A mutation in *BRCA1* or *BRCA2* compromises interaction with replication protein A (*RPA1* and *RPA2*), and these two proteins are essential for DNA replication, repair, and recombination [29,30]. Lack of interaction first inhibits the recruitment of double-strand break repair proteins, then leads to an accumulation of carcinogenic DNA abnormalities, eventually causing predisposition to early onset

Figure 4 A sample module associated with 'biopolymer metabolic process'. Blue nodes represent TFs, red nodes represent cancer-mutated genes, and green nodes represent non-mutated genes. Interactions between TFs and proteins are blue; interactions between protein pairs are gray.

cancer. These findings demonstrated that one or more TFs in co-regulated modules could package different genes with specific functions. Cancer-related modules could assemble a set of cancer-mutated genes and regulate specific biological functions associated with cancer, thus contributing to the pathogenesis of disease traits.

To address whether genes that link to genes mutated in cancer in co-regulated modules are more likely to be cancer-associated, we interrogated non-mutated genes within modules associated with 'biopolymer metabolic process' (module 27), using manual literature validation. We found that all non-mutated genes were implicated in tumorigenesis (Additional file 4). These results suggested that genes in cancer-related co-regulated modules had a high disease risk for tumours, and might be tumour candidate biomarkers. Additional analysis found that similar results could be obtained for all other cancer related co-regulated modules (data not shown).

Complex packaging

To access the association of co-regulated modules with protein complexes, we acquired 1347 human protein complexes from the MIPS database as a reference set, and analyzed the packaging characteristic of our modules [31,32]. A hyper-geometric test was used to evaluate the significance of overlap between our modules and the MIPS functional categories. The results showed 90 (94%) modules that could organize numerous protein complexes ($p < 0.05$, Additional file 5). As an example of these significant results, a sample module that is involved in 'biopolymer metabolic process' (module 27), packages 98 protein complexes involved in eight functional classes (Figure 5, Additional file 6). The complexes and this module share a set of cancer-related functions such as DNA repair, cell cycle regulation, and transcription from RNA polymerase II promoters. Many studies have shown that gene alterations in cancer patients, such as malignant changes in DNA sequence and chromosomal fragment amplifications, cause subtle divergence of the DNA sequence with subsequent mistakes in replication during 'DNA repair' and 'DNA replication', altering 'transcription activity' and 'cell cycle', resulting in the evolution of mutinous cells, and resulting in the ability to invade and metastasise [33-37]. Similar packaging results for the other co-regulated modules are in additional file 5. These results suggested that our co-regulated modules had the functionality of rewiring different protein complexes, and that cancer-related modules could package complexes that underlie carcinogenesis.

Pathway packaging

To further investigate the assembling power of co-regulated modules on metabolic pathways, we performed KEGG annotation analysis for each module using DAVID (Count > = 2; EASE < = 0.05) [38,39],
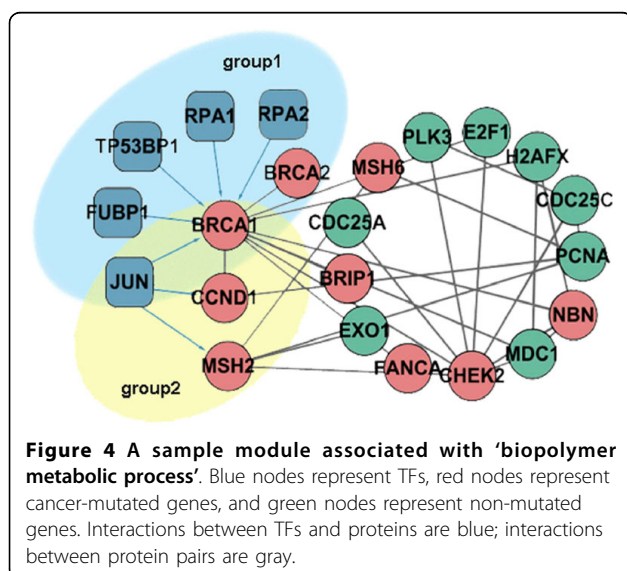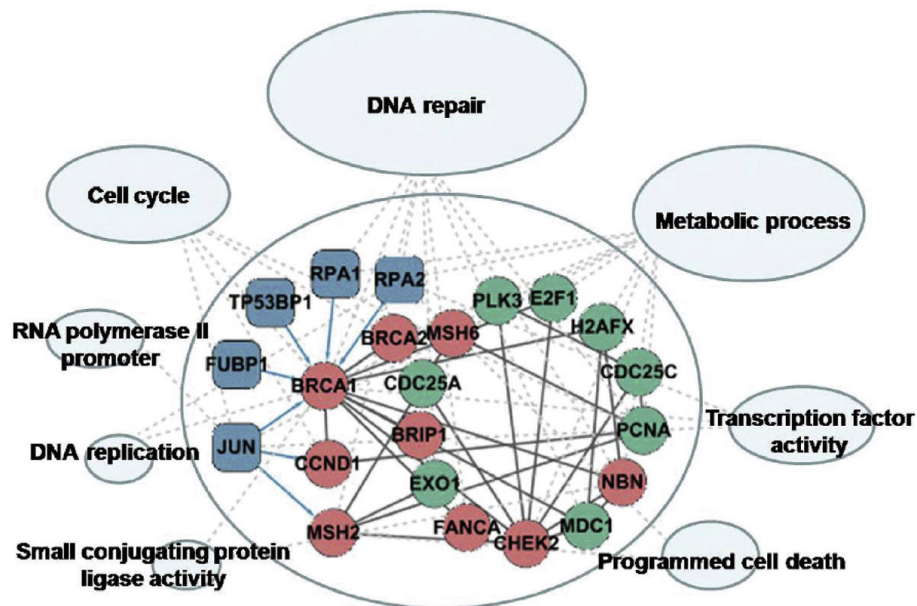
Figure 5 Complex packaging of the 'biopolymer metabolic process' module. As an assembler, this co-regulated module organized eight groups of protein complexes. The ellipses mark groups of complexes belonging to similar functional classes. Node size depends on the number of complexes in each function group.

a useful tool that integrates different sources of biological information to obtain biological annotations, and ranks them by statistical significance. We found that 79 (82%) modules had significant annotated pathway information (Additional file 7). A sample module ('biopolymer metabolic process') assembled eight divergent metabolic pathways (Figure 6). We discovered two

cancer-related TFs (*RPA1* and *RPA2*) that function as hub TFs, forming focal nodes in information exchange between eight metabolic pathways. These two TFs and their binding proteins in the module work in a complementary manner to rewire the mismatch repair, cell cycle, and homologous recombination pathways leading to the dysfunction of different cancer pathways [40-42].
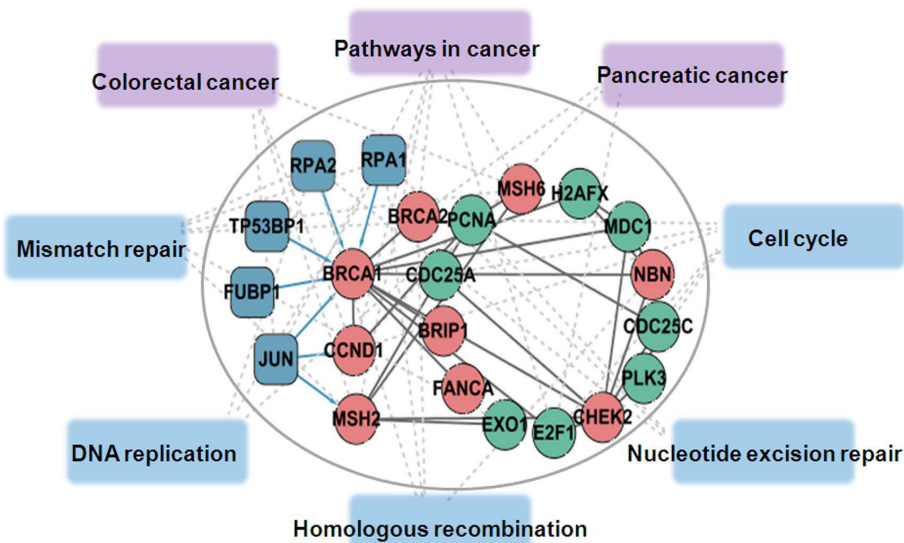

Figure 6 Pathway packaging of the 'biopolymer metabolic process' module. As an assembler, this co-regulated module organized eight divergent metabolic pathways. Cancer pathways are in purple, and cancer-related pathways are in blue.

In our prior studies, we found that genes in cancer development and progression are distributed sparsely among different metabolic pathways. According to pathway analysis, we concluded that our modules had the functionality of organizing multiple biological pathways and controlling numerous cell behaviours, which eventually contribute to cancer pathogenesis.

Conclusions

We devised and implemented a probabilistic model and a bipartite graph framework to infer human co-regulated modules. We analyzed their specific features in packaging different biological entities from an integrated molecular network with high confidence. Through robustness analysis, we demonstrated that our algorithm identified probable co-regulated modules for *Homo sapiens*. The performance of our approach was evaluated by comparison with other four module identification approaches. Further analysis using the bipartite graph framework uncovered packaging features for co-regulated modules, and showed that modules appeared to act as 'assemblers' dominated by several transcriptional regulations, and tended to coordinate complexes maintained by several physical interacting proteins, and indicating involvement in metabolic pathway cross-talk within neighbouring regions.

The success of our method can be attributed to the following factors. PPINs and TRNs are based on the curated literature and experimentally-determined interactions, so an integrated molecular network can be used to identify co-regulatory modules. In addition, we introduced a bipartite graph framework to evaluate packaging features of co-regulated modules with different biological entities, which easily divided biological entities into piles according to each module. As shown by various examples, our method appears to be effective in the identification of human co-regulated modules, and in searching for their packaging features in biological entities.

However, our proposed method has some limitations. We introduced a greedy algorithm aimed to make the locally optimal choice at each expanding step. Greedy algorithms are known to generally fail in finding globally optimal solutions, because they usually do not operate exhaustively on all the data. However, from our analysis results, we believe that the greedy algorithm was effective for module identification. The limitations of the proposed method for packaging (overlap) analysis are that two-thirds of human genes are annotated by at least one functional annotation, but the remaining one-third has yet to be annotated [43]. In addition, the incompleteness of information about complexes and biological pathways might miss some significant overlaps or packing relationships. Although our proposed method has these limitations, the packaging features of co-regulated modules could still be deciphered in integrated molecular networks. With the accumulation of human data, we expect that our framework may facilitate the identification of additional modules and their packaging features.

Methods
Human interaction data sources
Human protein-protein interaction data was extracted from the HPRD databases (Release7) [44]. The derived network contained 34,083 interactions between 9014 proteins. We determined edge reliability weights for these interactions with supporting evidence information including experimental validation, computational methods, and public literature mining for a number of proteins [45].

Transcriptional regulatory data was acquired from the Transfac Database (Release11.4) [46]. The resulting regulatory network consisted of 281 TFs and 624 genes with 1603 interactions. For further analysis, we assigned an empirical weight to be 0.99 (a balanced confidence level of each edge in a TRN) for each transcriptional regulatory interaction.

Cancer mutated genes
Cancer mutated genes (384) were obtained from the Cancer Gene Census [47], a well-known online database cataloguing genes in which mutations have been causally implicated in a wide variety of tumour types.

Human co-regulated module identification
Integrative weighted network construction
The human integrated network was represented as a weighted graph. The vertices of the graph were proteins or TFs, and the edges were protein-protein interactions or transcription regulation interactions. All edges are set confidence scores, as described above.
Probabilistic statistical model
We constructed a probabilistic model to evaluate whether a cluster of co-regulated proteins is likely to form a module in an integrated network. An underlying assumption was that a module corresponds to a sub-network that is typically dense. Under the probabilistic model, we formulated a log-likelihood ratio used to compare the fit of our model of a module against the likelihood that it arose at random. Highly scoring sub-networks corresponded to likely modules.

This approach requires constructing a sub-network model and a background model for interactions [48]. We defined two models: the sub-network model, Ms, assumed that interactions between proteins have a high probability α (set to 0.8), and that interactions between transcription factors and their target genes have high probability β (set to 0.9), according to the average level of

the interaction's confidence weight in the real PPIN or TRN. In contrast, the background model, Mb, was obtained from a long series of random edge crosses by Monte Carlo simulations [49]. In this process, we chose two links *(a, b)*, *(c, d)* uniformly at random from the integrated network, and rewired them by exchanging their partners. Note that this procedure preserved their degrees distribution. We estimated the probabilities of interactions in the random network based on the percentage of the observed edges. We defined the likelihood model as:

$$L(P,T) = L_{PPI} + L_{TRI} \qquad (I)$$
$$where,$$
$$L_{PPI} = \sum_{u,v \in P} \log \frac{\alpha P(u,v) + (1-\alpha)(1-P(u,v))}{R(u,v)P(u,v) + (1-R(u,v))(1-P(u,v))}$$
$$and$$
$$L_{TRI} = \sum_{u \in P, t \in T} \log \frac{\beta P(u,t) + (1-\beta)(1-P(u,t))}{R(u,t)P(u,t) + (1-R(u,t))(1-P(u,t))}$$

Here, the ratio score of each candidate cluster is calculated by adding the log likelihood ratio score of the PPIN to that of the TRN. $P(u, v)$ represents the confidence weight between two proteins u and v, and $P(u, t)$ represents the confidence weight between protein u and transcription factor t. The probabilities $R(u, v)$ and $R(u, t)$ of the random network were estimated based on the percentage of the observed edge.

Module searching

Each candidate cluster was generated from our searching algorithm. The searching process consisted of three basic processes: *(i)* seed initialization; *(ii)* seed expanding; and *(iii)* overlap filtering.

We defined candidate seeds as a set with a TF and two of its binding genes, and restricted to include two protein-protein interactions. A greedy approach was used to filter the candidate seeds, retaining those with the highest L-score as the staring seed subunits.

In the second step, we expanded the starting seed subunits using a local search. In each seed expanding, we iteratively added a node (a protein or TF) to modify the current cluster, ensuring that each newly built candidate cluster had the highest ratio score. This procedure was repeated until the contribution gains passed a predefined threshold, which we defined as 4. Finally, after all expansion rounds, we checked overlaps between our resulting modules via a simple overlap ratio (OR):

$$OR = \frac{NO_i}{NO_u}$$

Where NO_i is the size of the overlaps between any two modules, and NO_u is the union size of any two modules. If the OR score of two modules was larger than 0.8, we merge the module with lower L-score into larger one.

Bridging co-regulated modules with biological entities using bipartite graphs

To access the packaging features of our resulting modules, we mapped them onto biological entities of genes, protein complexes, or metabolic pathways. For each module M, we constructed three 'Module-biological Entity' bipartite graphs: *(i)* $G_{M-g} = (M,g,E_{M-g})$ as a bipartite graph of module M-gene associations, where $E_{M-g} \subseteq M \times g$; *(ii)* $G_{M-c} = (M,c,E_{M-c})$ as a bipartite graph of module M-complex associations, where $E_{M-c} \subseteq M \times c$; and *(iii)* $G_{M-p} = (M,p,E_{M-p})$ as a bipartite graph of module M-pathway associations, where $E_{M-p} \subseteq M \times p$;. Finally, we collected the biological relevance of our modules for rewiring different biological entities. As exemplified by cancer, we investigated whether cancer related co-regulated modules could assemble different cancer-related biological entities, and identified underlying biological associations of our co-regulated modules with cancer.

Additional material

> **Additional file 1: Information about the 96 co-regulated modules**
>
> **Additional file 2: Results of GO functional enrichment of co-regulated modules**
>
> **Additional file 3: Associations between co-regulated modules and cancer genes**
>
> **Additional file 4: Literature evidence for non-mutated genes in the 'biopolymer metabolic process' module**
>
> **Additional file 5: Complex packaging results for all co-regulated modules**
>
> **Additional file 6: Complex packaging results for 'biopolymer metabolic process' module**
>
> **Additional file 7: Pathway packaging results for all co-regulated modules**.

Author details

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Hei Longjiang Province, China. [2]Institute of Optoelectronics, Harbin Institute of Technology, Harbin, Hei Longjiang Province, China.

Authors' contributions

LC, HW and LZ participated in the design of the study. WL, QW and YS carried out the integrated network construction. HW carried out co-regulated modules identification and analysis of the packaging feature. YH, WH, XL and JT participated in performance evaluation of the results. XL participated in the design and coordination of the study. All authors read and approved the final manuscript.

References

1. Albert R: **Scale-free networks in cell biology.** *J Cell Sci* 2005, **118**(Pt 21):4947-4957.
2. Lee I, Date SV, Adai AT, Marcotte EM: **A probabilistic functional network of yeast genes.** *Science* 2004, **306**(5701):1555-1558.
3. Guelzim N, Bottani S, Bourgine P, Kepes F: **Topological and causal structure of the yeast transcriptional regulatory network.** *Nat Genet* 2002, **31**(1):60-63.
4. Yeger-Lotem E, Sattath S, Kashtan N, Itzkovitz S, Milo R, Pinter RY, Alon U, Margalit H: **Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction.** *Proc Natl Acad Sci USA* 2004, **101**(16):5934-5939.
5. Greenbaum D, Jansen R, Gerstein M: **Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts.** *Bioinformatics* 2002, **18**(4):585-596.
6. Celis JE, Gromov P, Gromova I, Moreira JM, Cabezon T, Ambartsumian N, Grigorian M, Lukanidin E, Thor Straten P, Guldberg P, *et al*: **Integrating proteomic and functional genomic technologies in discovery-driven translational breast cancer research.** *Mol Cell Proteomics* 2003, **2**(6):369-377.
7. Zhang S, Jin G, Zhang XS, Chen L: **Discovering functions and revealing mechanisms at molecular level from biological networks.** *Proteomics* 2007, **7**(16):2856-2869.
8. Purmann A, Toedling J, Schueler M, Carninci P, Lehrach H, Hayashizaki Y, Huber W, Sperling S: **Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality.** *Genomics* 2007, **89**(5):580-587.
9. Michalak P: **Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes.** *Genomics* 2008, **91**(3):243-248.
10. Segal E, Friedman N, Kaminski N, Regev A, Koller D: **From signatures to models: understanding cancer using microarrays.** *Nat Genet* 2005, **37**(Suppl):S38-45.
11. Segal E, Friedman N, Koller D, Regev A: **A module map showing conditional activity of expression modules in cancer.** *Nat Genet* 2004, **36**(10):1090-1098.
12. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, *et al*: **Identification of genes periodically expressed in the human cell cycle and their expression in tumors.** *Mol Biol Cell* 2002, **13**(6):1977-2000.
13. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, *et al*: **Prediction of central nervous system embryonal tumour outcome based on gene expression.** *Nature* 2002, **415**(6870):436-442.
14. Chuang HY, Lee E, Liu YT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol* 2007, **3**:140.
15. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, *et al*: **A map of human cancer signaling.** *Mol Syst Biol* 2007, **3**:152.
16. Shamir R, Maron-Katz A, Tanay A, Linhart C, Steinfeld I, Sharan R, Shiloh Y, Elkon R: **EXPANDER–an integrative program suite for microarray data analysis.** *BMC Bioinformatics* 2005, **6**:232.
17. Milenkoviae T, Przulj N: **Uncovering Biological Network Function via Graphlet Degree Signatures.** *Cancer Inform* 2008, **6**:257-273.
18. Reimand J, Tooming L, Peterson H, Adler P, Vilo J: **GraphWeb: mining heterogeneous biological networks for gene modules with functional significance.** *Nucleic Acids Res* 2008, , **36** Web Server:W452-459.
19. Palla G, Derenyi I, Farkas I, Vicsek T: **Uncovering the overlapping community structure of complex networks in nature and society.** *Nature* 2005, **435**(7043):814-818.
20. Enright AJ, Van Dongen S, Ouzounis CA: **An efficient algorithm for large-scale detection of protein families.** *Nucleic Acids Res* 2002, **30**(7):1575-1584.
21. Ulitsky I, Shamir R: **Identifying functional modules using expression profiles and confidence-scored protein interactions.** *Bioinformatics* 2009, **25**(9):1158-1164.
22. Polakis P: **Wnt signaling and cancer.** *Genes Dev* 2000, **14**(15):1837-1851.
23. Tront JS, Hoffman B, Liebermann DA: **Gadd45a suppresses Ras-driven mammary tumorigenesis by activation of c-Jun NH2-terminal kinase and p38 stress signaling resulting in apoptosis and senescence.** *Cancer Res* 2006, **66**(17):8448-8454.
24. Schayek H, Haugk K, Sun S, True LD, Plymate SR, Werner H: **Tumor suppressor BRCA1 is expressed in prostate cancer and controls insulin-like growth factor I receptor (IGF-IR) gene transcription in an androgen receptor-dependent manner.** *Clin Cancer Res* 2009, **15**(5):1558-1565.
25. Gray SE, Kay E, Leader M, Mabruk M: **Molecular genetic analysis of the BRCA2 tumor suppressor gene region in cutaneous squamous cell carcinomas.** *J Cutan Pathol* 2008, **35**(1):1-9.
26. Li S, Ting NS, Zheng L, Chen PL, Ziv Y, Shiloh Y, Lee EY, Lee WH: **Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response.** *Nature* 2000, **406**(6792):210-215.
27. Cortez D, Wang Y, Qin J, Elledge SJ: **Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks.** *Science* 1999, **286**(5442):1162-1166.
28. Scully R, Chen J, Ochs RL, Keegan K, Hoekstra M, Feunteun J, Livingston DM: **Dynamic changes of BRCA1 subnuclear location and phosphorylation state are initiated by DNA damage.** *Cell* 1997, **90**(3):425-435.
29. Choudhary SK, Li R: **BRCA1 modulates ionizing radiation-induced nuclear focus formation by the replication protein A p34 subunit.** *J Cell Biochem* 2002, **84**(4):666-674.
30. Wong JM, Ionescu D, Ingles CJ: **Interaction between BRCA2 and replication protein A is compromised by a cancer-predisposing mutation in BRCA2.** *Oncogene* 2003, **22**(1):28-33.
31. Mewes HW, Amid C, Arnold R, Frishman D, Guldener U, Mannhaupt G, Munsterkotter M, Pagel P, Strack N, Stumpflen V, *et al*: **MIPS: analysis and annotation of proteins from whole genomes.** *Nucleic Acids Res* 2004, , **32** Database:D41-44.
32. Mewes HW, Frishman D, Mayer KF, Munsterkotter M, Noubibou O, Pagel P, Rattei T, Oesterheld M, Ruepp A, Stumpflen V: **MIPS: analysis and annotation of proteins from whole genomes in 2005.** *Nucleic Acids Res* 2006, , **34** Database:D169-172.
33. Blons H, Cote JF, Le Corre D, Riquet M, Fabre-Guilevin E, Laurent-Puig P, Danel C: **Epidermal growth factor receptor mutation in lung cancer are linked to bronchioloalveolar differentiation.** *Am J Surg Pathol* 2006, **30**(10):1309-1315.
34. Lievre A, Bachet JB, Le Corre D, Boige V, Landi B, Emile JF, Cote JF, Tomasic G, Penna C, Ducreux M, *et al*: **KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer.** *Cancer Res* 2006, **66**(8):3992-3995.
35. Zhang XY, Hu Y, Cui YP, Miao XP, Tian F, Xia YJ, Wu YQ, Liu X: **Integrated genome-wide gene expression map and high-resolution analysis of aberrant chromosomal regions in squamous cell lung cancer.** *FEBS Lett* 2006, **580**(11):2774-2778.
36. Hughes S, Yoshimoto M, Beheshti B, Houlston RS, Squire JA, Evans A: **The use of whole genome amplification to study chromosomal changes in prostate cancer: insights into genome-wide signature of preneoplasia associated with cancer progression.** *BMC Genomics* 2006, **7**:65.
37. Tsafrir D, Bacolod M, Selvanayagam Z, Tsafrir I, Shia J, Zeng Z, Liu H, Krier C, Stengel RF, Barany F, *et al*: **Relationship of gene expression and chromosomal abnormalities in colorectal cancer.** *Cancer Res* 2006, **66**(4):2129-2137.
38. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4**(5):P3.
39. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44-57.
40. Vassileva V, Millar A, Briollais L, Chapman W, Bapat B: **Genes involved in DNA repair are mutational targets in endometrial cancers with microsatellite instability.** *Cancer Res* 2002, **62**(14):4095-4099.
41. McCabe N, Turner NC, Lord CJ, Kluzek K, Bialkowska A, Swift S, Giavara S, O'Connor MJ, Tutt AN, Zdzienicka MZ, *et al*: **Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly (ADP-ribose) polymerase inhibition.** *Cancer Res* 2006, **66**(16):8109-8115.
42. Zheng YL, Kosti O, Loffredo CA, Bowman E, Mechanic L, Perlmutter D, Jones R, Shields PG, Harris CC: **Elevated lung cancer risk is associated with deficiencies in cell cycle checkpoints: genotype and phenotype analyses from a case-control study.** *Int J Cancer* 126(9):2199-2210.

43. Chen J, Aronow BJ, Jegga AG: Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics* 2009, **10**:73.

44. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N, Suresh S, *et al*: Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res* 2004, , **32** Database: D497-501.

45. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, *et al*: STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 2009, , **37** Database: D412-416.

46. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, *et al*: TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 2003, **31**(1):374-378.

47. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: A census of human cancer genes. *Nat Rev Cancer* 2004, **4**(3):177-183.

48. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T: Conserved patterns of protein interaction in multiple species. *Proc Natl Acad Sci USA* 2005, **102**(6):1974-1979.

49. Maslov S, Sneppen K: Specificity and stability in topology of protein networks. *Science* 2002, **296**(5569):910-913.