Bioinformatics and Biosystems 2007, Vol. 2, No. 1, pp. 12-16

Revealing Regulatory Networks of DNA Repair Genes in S. Cerevisiae

Min-sung Kim[1], Doheon Lee[2], Gwan-Su Yi[1]

[1]Computational Systems Biology Laboratory, Information and Communication University, South Korea
[2]Dept. of Biosystems, KAIST, South Korea

Abstract

DNA repair means a collection of processes that a cell identifies and corrects damage to genome sequence. The DNA repair processes are important because a genome would not be able to maintain its essential cellular functions without the processes. In this research, we make some gene regulatory networks of DNA repair in S. cerevisiae to know how each gene interacts with others. Two approaches are adapted to make the networks; Bayesian Network and ARACNE. After construction of gene regulatory networks based on the two approaches, the two networks are compared to each other to predict which genes have important roles in the DNA repair processes by finding conserved interactions and looking for hubs. In addition, each interaction between genes in the networks is validated with interaction information in S. cerevisiae genome database to support the meaning of predicted interactions in the networks.

Keywords: Bayesian Network, DNA Repair, S. Cerevisiae, ARACNE

Introduction

In view of thousands of damage events that genome suffer every day and errors that occur when the genome replicates, it is essential that a cell needs to have efficient repair systems. The genome would not be able to maintain its essential functions without the systems. Because of its importance, many people have been researched into them to know the detailed mechanisms of DNA repair systems. Therefore we become to know about the mechanisms of DNA repair and what kinds of gene are involved in the processes. However, how each gene interacts with others to repair DNA is not revealed clearly because there are not only genes which have direct relationships to others but also genes which have indirect relationships to others.

Therefore, gene regulatory networks, or simply gene networks, has been made to predict such kinds of interaction information. The gene regulatory networks are built from genome-wide high-throughput gene expression data that are often available from DNA microarray experiments. There have been various approaches to construct the regulatory networks. The representative approaches are Boolean network (Liang et al., 1998), Correlation Metric Construction (Arkin et al, 1997), Bayesian network, ARACNE (Basso et al, 2005), etc.

In this research, we make gene regulatory networks of DNA repair to show how each gene interacts with other genes during the processes. The target species is selected as Saccharomyces Cerevisiae, because lots of biological functions and gene information are revealed and the number of genes is smaller than other model organisms. In addition, Bayesian network and ARACNE are selected for network construction because the usefulness of the two approaches is widely accepted in current researches. Based on the two approaches, we make gene regulatory networks which are related to DNA repair and try to infer which genes and interactions are important by finding conserved interactions and looking for hubs. In addition, the annotated interaction information of each gene in

Corresponding Author: Gwan-Su Yi (Tel: +82-42-866-6815, Fax: +82-42-866-6814, Email: gsyi@icu.ac.kr)
MK and GSY were supported by Korea Research Foundation Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund) (KRF-2005-003-C00157) and grant No. R01-2005-000-10824-0 from the Basic Research Program of the Korea Science & Engineering Foundation. DL was supported by the Korea Science and Engineering Foundation(KOSEF) through the National Research Lab. Program (No. 2005-01450), and the Korean Systems Biology Research Grant (2005-00343).

S.cerevisiae genome database is used to support the meaning of predicted interactions in the networks.

Methods

Construction of Bayesian Networks

Gene regulatory networks are made by using BNArray (Chen et al, 2006) which is an R package to construct Bayesian networks from microarray experiment data. It consists of 4 modules. The first module is Least Local Square algorithm which can be used to impute missing data in microarray experiments. It has Bayesian network construction module based on Gaussian-inverse Gamma distribution for continuous variables and greedy search algorithm with random restart for network construction. Also, it can re-sample the microarray dataset to produce more reliable data using Efron's Bootstrap to construct a collection of 1st-order Bayesian networks with high scores. Finally, it has extended version of CONDENSE algorithm to reconstruct coherent regulatory modules. Because of the modules, it has almost whole required procedures in its package to construct Bayesian networks from experiment data.

S. cerevisiae cell-cycle expression data set (Spellman et al, 1998) is selected just like written in demonstration part in supplementary data of BNArray. From the 6178 genes, 799 differentially expressed genes are chosen for analysis. In addition, 17 genes which are involved in DNA repair are selected for further analysis from the 799 gene pool.

Construction of ARACNE Networks

We use ARACNE (algorithm for the reconstruction of accurate cellular networks) (Margolin et al, 2006) to make the gene regulatory networks. The approach is consists of 2 steps. It first identifies statistically significant gene-gene coregulation by using mutual information. It then eliminates indirect relationships by using data processing inequality (DPI). Whole this algorithm is implemented in geWorkbench including preprocess step such as normalization, imputation process, etc.

As a data set, the 799 differentially expressed genes are selected from the 6178 genes just like BNArray's approach. The missing values in the microarray data are imputed by averaging the values.

Result

The result of Bayesian Networks

Under the condition of bootstrap time=45 and confidence level=0.75, the relationship between the 17 genes are generated as shown in figure 1. As illustrated in the figure, we can find an interesting feature of the networks. The two genes, or YDR097C and YKL113C, show that have more interactions than other genes. Therefore, we can predict the two genes may play central roles during the repair processes.

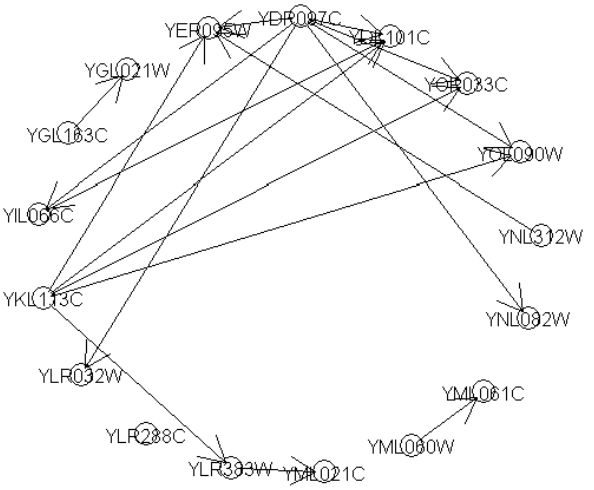

Figure 1. The Networks of the selected 17 genes Two genes (YDR097C and YKL113C) have more interactions than other genes.

To know roles of the two genes, we searched Saccharomyces genome database (Michael et al, 1998). The YDR097C (MSH6) is a protein required for mismatch repair in mitosis and meiosis. It forms a complex with YOL090W (MSH2) to repair single base and insertion-deletion mispairs (Marsischky et al, 1996; Bowers et al, 1999). The relationship between YDR097C and YOL090W is predicted in our networks, also. The other gene, or YKL113C (RAD27), is a 5' to 3' exonuclease. It is required for Okazaki fragment processing and maturation and for long-patch base-excision repair (Kao et al, 2002; Ayyagari et al, 2003).

The result of ARACNE Networks

After construction of the gene regulatory networks with varying parameter values, we choose 0.1 as MI threshold and 0.1 as DPI tolerance. From whole networks of 799 genes, we select 17 genes which are related to DNA repair. The networks of 17 genes are shown in figure 2.

To know highly interconnected genes (hubs) among these 17 genes, we count the number of first neighbors of each gene. As a result, the YDR097C (33) and YKL113C (30) are selected as hubs.

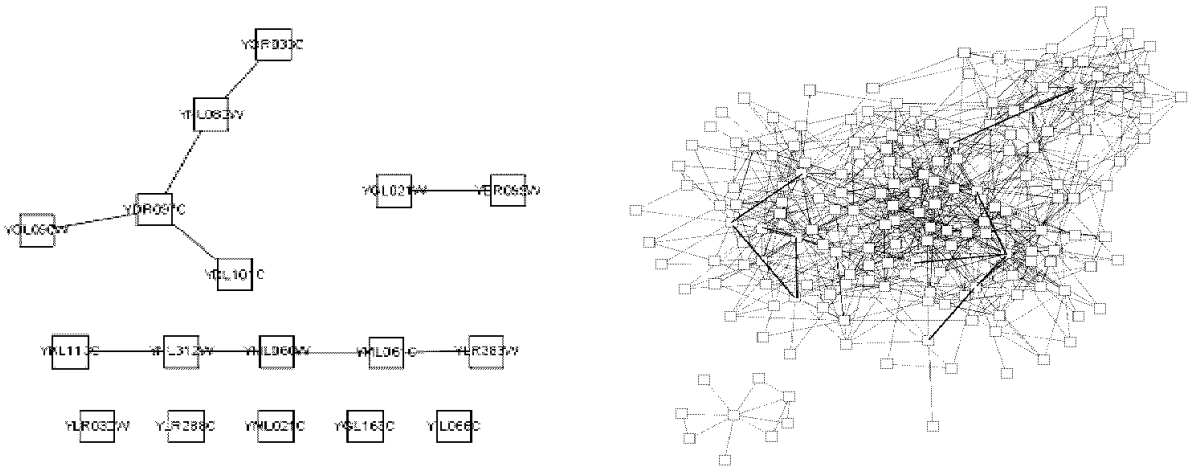

Figure 2. The generated networks of 17 genes Left figure shows relationship among 17 genes. Right figure shows networks of selected 17 genes with their first neighbors.

Comparing the result of both approaches

We compare the two networks to find central genes in DNA repair and key features of the processes. The result of comparison is illustrated in figure 3. The two left graph is generated based on ARACNE approach and the right graph is generated based on Bayesian network. The networks indicate that two genes may play central roles in DNA repair processes in common; YDR097C and YKL113C. Also, YLR288C is separated to other genes in both graphs. The YLR288C (MEC3) is a DNA damage and meiotic pachytene checkpoint protein (Kondo et al, 1999).

Although detailed interactions in both networks are differ

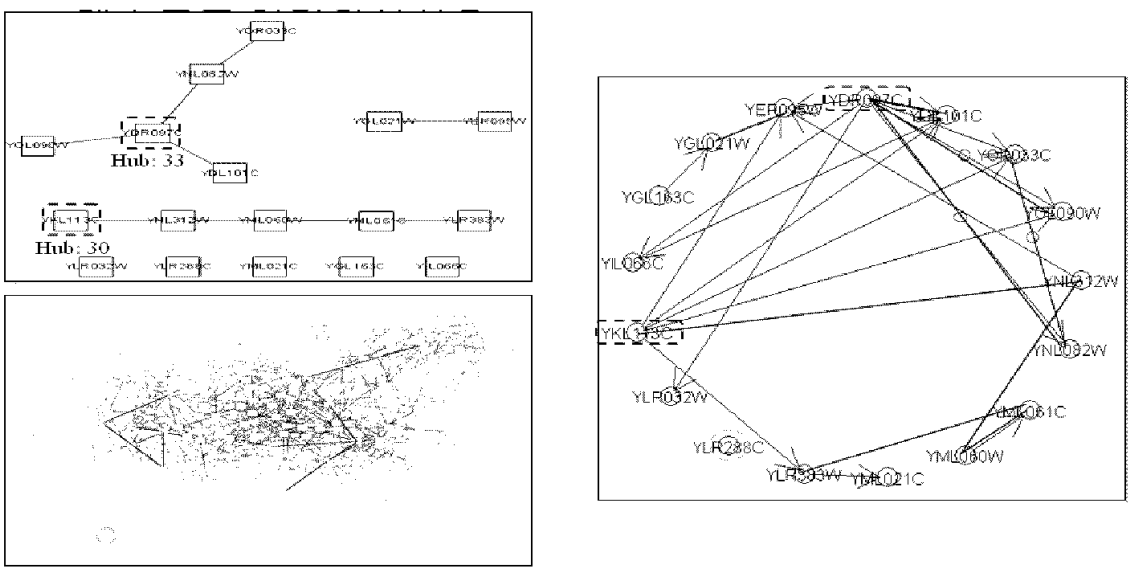

Figure 3. The networks of Bayesian and ARACNE approach The networks of ARACNE (left) and Bayesian (right) approach are shown. The blue rectangle means that the genes may play central roles in the networks because it has more relationships to others compare to other genes. The green circle means a separated gene in the networks, because it has no relationship to other genes. The red line in the right graph means the relationship which is estimated in ARACNE approach. The blue circle on the red line in right graph means the confirmed interaction which is mentioned in interaction information in S. cerevisiae database.

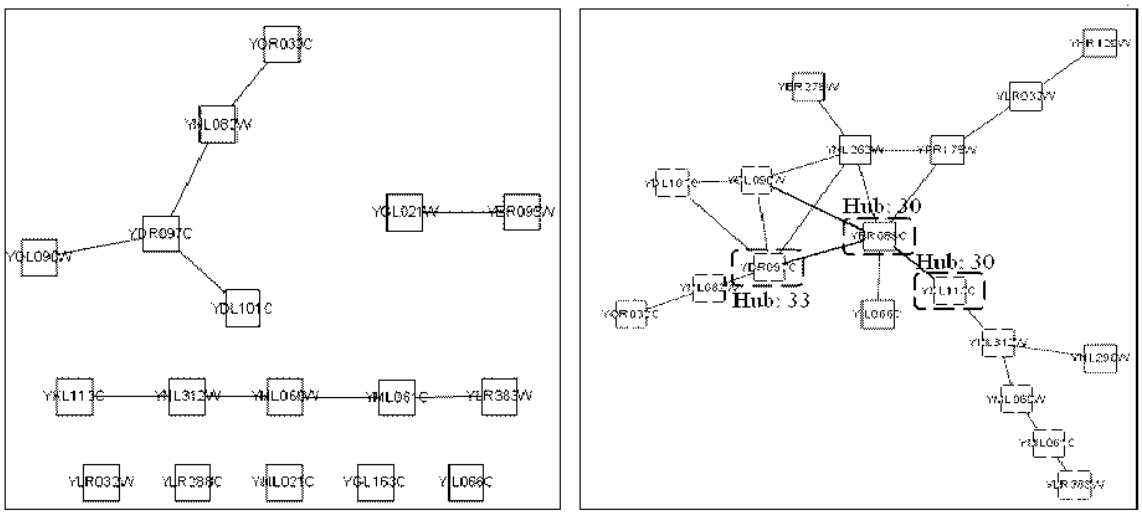

Figure 4. Comparison of previous 17 genes networks and newly made networks. The previous networks of 17 genes (left) and newly made networks of 18 genes (right) among 36 genes are shown. The yellow 10 nodes mean the nodes which are existed as interconnected sub networks in 17 genes networks. The yellow nodes are joined by YBR088C. Also, the interactions between them are confirmed in S. cerevisiae database. The confirmed interaction of YBR088C is colored as red lines.

ent, we can conclude that the key features are similar to each other based on the observation.

Extended analysis with ARACNE

We used the 799 gene pool which are used in BNArray to compare the result of the two approaches. To understand the whole process of DNA repair more deeply and fill the gap between predicted regulatory networks and real cellular operation, some genes are added to the data set. The target genes are 23 genes which are included to mismatch repair in Gene Ontology. There are 12 genes among the 23 genes are not included in the original 799 gene pool. With the newly made 811 gene sets, we re-made regulatory networks by using ARACNE. From the networks, we selected 36 genes that is comprised of previous 17 original genes and newly added 23 genes. After analyzing the result we found newly made interconnected networks which contain 18 genes among the 36 genes. The newly made networks of 18 genes are shown in figure 4.

The two interconnected sub networks in previous networks of 17 DNA repair genes are joined by an added gene YBR088C. It is a Proliferating Cell Nuclear Antigen (PCNA). The gene is predicted to function as a docking site for other proteins required for DNA repair and mitotic and meiotic chromosomal DNA replication (Tsurimoto et al, 1999). The YBR088C has 30 relationships as first neighbors. It means one additional hub is found in extended analysis. In addition, the interaction between YBR088C and other genes are confirmed by annotated interaction information in S. cerevisiae database. Interestingly, the gene shows confirmed physical interactions to YDR097C and YKL113C which are predicted to show important roles during DNA repair processes.

Also, we found a transcription factor, or YLR183C, among first neighbors of the 18 genes. The transcription factor is related to 4 genes among the 18 genes. The YKL113C and YBR088C are also related to the transcription factor as shown in figure 5.

Discussion

After making some regulatory networks of genes which are related to DNA repair in S. cerevisiae and analyzing the meaning of relationship between each node in the networks, we can predict three genes play central roles in the repair processes. YDR097C and YKL113C are found by comparing the two networks which are made by using Bayesian network and ARACNE. The two genes have conserved interactions and existed as hubs in both networks. YBR088C is found by adding some genes to original data sets, looking for hubs, and validating relationship between YBR088C and other genes. In addition, we found a transcription factor which is related to YBR088C and YKL133C.

Our remaining thing is investigating roles of the three

genes (YOL090W, YBR088C and YKL133C) in DNA repair processes to validate and support meaning of our research while trying to make new gene regulatory networks by using other microarray data to find other important genes and interactions.

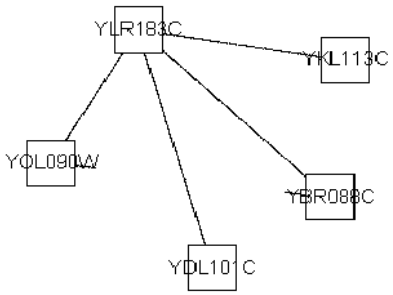

Figure 5. Transcription factor and related 4 genes among 18 selected genes. The transcription factor, or YLR183C, is related to 4 genes among the 18 genes.

References

[1] Liang,S. et al. (1998) Reveal, a general reverse engineering algorithm for inference of genetic network architectures. Pacific Symposium on Biocomputing 3:18-29.

[2] Arkin,A. et al. (1997) A Test Case of Correlation Metric Construction of a Reaction Pathway from Measurements. Science, 227: 1275-1279.

[3] Basso,K. et al. (2005) Reverse engineering of regulatory networks in human B cells. Nat Genet, 37(4):382-90.

[4] Chen,X. et al. (2006) BNArray: an R package for constructing gene regulatory networks from microarray data by using Bayesian network. Bioinformatics, 22(23):2952-4.

[5] Spellman,PT. et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell, 9(12):3273-97.

[6] Margolin,AA. et al. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics, 7 Suppl 1:S7.

[7] Michael,J. et al. (1998) SGD: Saccharomyces genome database. Nucleic Acid Research, 26(1):73 - 79.

[8] Marsischky,GT. et al. (1996) Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev 10(4):407-20.

[9] Bowers J, et al. (1999) A mutation in the MSH6 subunit of the Saccharomyces cerevisiae MSH2-MSH6 complex disrupts mismatch recognition. J Biol Chem 274(23):16115-25.

[10] Kao HI, et al. (2002) Cleavage specificity of Saccharomyces cerevisiae flap endonuclease 1 suggests a double-flap structure as the cellular substrate. J Biol Chem 277(17):14379-89.

[11] Ayyagari R, et al. (2003) Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 AND DNA2. J Biol Chem 278(3):1618-25.

[12] Friedman,N. (2004) Inferring cellular networks using probabilistic graphical models. Science, 303(5659):799-805.

[13] Kondo T, et al. (1999) Role of a complex containing Rad17, Mec3, and Ddc1 in the yeast DNA damage checkpoint pathway. Mol Cell Biol 19(2):1136-43

[14] Tsurimoto T (1999) PCNA binding proteins. Front Biosci 4:D849-58