

Extracting coordinated patterns of DNA methylation and gene expression in ovarian cancer

Je-Gun Joung,^{1,2,3} Dokyoon Kim,^{1,2} Kyung Hwa Kim,^{1,2} Ju Han Kim^{1,2}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/amiajnl-2012-001571>).

¹Division of Biomedical Informatics, Seoul National University Biomedical Informatics (SNUBI), Seoul, Republic of Korea

²Systems Biomedical Informatics National Core Research Center, Seoul National University College of Medicine, Seoul, Republic of Korea

³Institute of Endemic Diseases, Seoul National University College of Medicine, Seoul, Korea

Correspondence to

Professor Ju Han Kim, Seoul National University Biomedical Informatics (SNUBI), Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 110-799, Korea; juhan@snu.ac.kr

Received 13 December 2012

Revised 18 February 2013

Accepted 29 March 2013

Published Online First

18 April 2013

ABSTRACT

Objective DNA methylation, a regulator of gene expression, plays an important role in diverse biological processes including developmental process, carcinogenesis and aging. In particular, aberrant DNA methylation has been largely observed in several types of cancers. Currently, it is important to extract disease-specific gene sets associated with the regulation of DNA methylation.

Materials and methods Here we propose a novel approach to find the minimum regulatory units of genes, co-methylated and co-expressed gene pairs (MEGP) that are highly correlated gene pairs between DNA methylation and gene expression showing the co-regulatory relationship. To evaluate whether our method is applicable to extract disease-associated genes, we applied our method to a large-scale dataset from the Cancer Genome Atlas extracting significantly associated MEGP and analyzed their functional correlation.

Results We observed that many MEGP physically interacted with each other and showed high semantic similarity with gene ontology terms. Furthermore, we performed gene set enrichment tests to identify how they are correlated in a complex biological process. Our MEGP were highly enriched in the biological pathway associated with ovarian cancers.

Conclusions Our approach is useful for discovering coordinated epigenetic markers associated with specific diseases.

INTRODUCTION

Epigenetic regulation including DNA methylation and histone modification plays an important role in various regulatory cellular processes. DNA methylation is an important mechanism of gene silencing and control of gene expression.¹ There is a strong correlation between DNA methylation and transcriptional inhibition.² High levels of methylations in CpG islands of promoters lead to decreased gene expression, whereas low levels of methylation in promoters increase gene expression. In particular, pattern changes due to aberrant DNA methylation at some promoters have been shown to play a critical role in disease pathogenesis including carcinogenesis.³ In fact, significant differences in specific methylation and gene expression levels have been observed between normal cells and cancer cells.^{4,5}

Several epigenomic studies have recently reported that coordinated patterns of DNA methylation and gene expression are involved in a gene regulatory system. Most previous studies have focused on cancer and stem cell regulatory systems. For example, the coordinated alterations of promoter DNA methylation and gene expression were identified in a distinct subset of glioblastoma tumor

samples.⁶ A more interesting epigenomic pattern was observed in a regulatory circuit for major genes such as *Oct4*, *Sox2*, and *Nanog* with essential roles in the development of embryonic stem cells.⁷ In this circuit, dense methylation was markedly found in the *Oct4* and *Nanog* promoters, and they co-occupied a substantial portion of their target genes.⁸ These regulatory relationships raise several questions for the analysis of gene regulatory systems including: which genes show a negative correlation between gene expression and DNA methylation in their promoters; which genes are correlated simultaneously in gene expression as well as DNA methylation; and which biological processes are related to those genes.

Practical computational methods are needed to answer these questions from given profiles of methylation and gene expression. Previously, Louhimo and Hautaniemi⁹ introduced a method to find coordinated relationships by integrating multiple profiles such as DNA methylation and gene expression profiles. This method integrates copy number status and methylation profile with the expression level of a specific gene. However, it has limited function, that is, it does not provide the regulatory relationship between genes regulated by DNA methylation and expression. In detail, it does not extract gene pairs or sets that are co-expressed as well as co-methylated.

Here we propose a novel method to find co-methylated and co-expressed gene pairs (MEGP) that are highly correlated with each other epigenetically. We identify gene pairs by comparing patterns between DNA methylation and gene expression and measuring the similarity of both profiles between two genes. Our method is based on the differential co-expression score that measures how much two genes are differentially methylated and expressed as well as how much their methylation and expression patterns are similar. To demonstrate the utility of our methodology, we analyzed the gene expression and methylation profiles from patients with ovarian cancer cells. Alternatively, our method finds MEGP that are likely to be involved in the epigenetic gene network and functionally correlated in specific biological processes. Using information such as physical interaction and semantic relationships between genes, we assessed whether these pairs are functionally correlated. Furthermore, we explored their functions in a complex biological process by performing the gene set enrichment test.

MATERIALS AND METHODS

DNA methylation and gene expression profiles

The DNA methylation and gene expression datasets in ovarian cancers were obtained from the

To cite: Joung J-G, Kim D, Kim KH, et al. *J Am Med Inform Assoc* 2013;**20**:637–642.

Cancer Genome Atlas (TCGA).¹⁰ Each dataset consisted of ovarian cancer (n=496) and healthy controls (n=9). We collected DNA methylation profiles using the Infinium HumanMethylation27 Beadchip that covers 27 578 CpG loci and collected gene expression profiles using 12 034 probes of the AffymetrixHT Human Genome U133 Array. Then, we constructed matrices for the DNA methylation and gene expression profiles. In each matrix, the row corresponded to 9129 genes that are overlapped between both sets and the column corresponded to a total of 505 samples. In the matrix of DNA methylation, each cell indicated a normalised methylation level (β value).

Co-methylated and co-expressed gene pairs

We measured the MEGP score for pairs of all the genes using the methylation and gene expression profiles. It was calculated by combining two scores, the differential score and correlation score, which measures the regulatory relationship between both profiles. Here the differential score measures how much two genes are differentially methylated as well as differentially expressed between the disease and control samples. On the other hand, the correlation score measures how much the DNA methylation and expression patterns are similar between two genes.

Let n denote the number of profile types and m the number of sample types. In our case, both numbers are fixed as follows: $n=2$ (ie, DNA methylation and gene expression) and $m=2$ (ie, cancer and control). When given a gene pair, X and Y , a vector of variables, X^{ij} (or Y^{ij}) belongs to one of four categories:

1. If $i=1$ and $j=1$, X^{ij} (or Y^{ij}) \in methylation and cancer samples,
2. else if $i=2$ and $j=1$, X^{ij} (or Y^{ij}) \in expression and cancer samples,
3. else if $i=1$ and $j=2$, X^{ij} (or Y^{ij}) \in methylation and control samples,
4. else if $i=2$ and $j=2$, X^{ij} (or Y^{ij}) \in expression and control samples.

Here the index i corresponds to the type of profile and the index j indicates the type of sample.

First, we describe the differential score that measures how much two genes, X and Y are differently methylated and expressed between the cancer and control group. When δ is defined as the mean fold change between the cancer and control group, it is transformed by a sigmoid function, making the value from 0 to 1.

$$P(X^i) = \frac{2}{1 + \exp(-\alpha|\delta(X^i)|)} - 1 \text{ and } P(Y^i) = \frac{2}{1 + \exp(-\alpha|\delta(Y^i)|)} - 1 \quad (1)$$

Here α is a parameter that controls the smoothness of the fold change value. In general, a high level of methylation in the CpG islands of promoters leads to decreased expression of genes, whereas a low level of methylation leads to increased expression. Therefore, we only perform further calculations when an inverse relationship is shown between methylation and expression, that is, the fold change of methylation and that of gene expression are opposite: $\text{sign}(\delta(X^1)) \neq \text{sign}(\delta(X^2))$. This step is able to decrease the total computing cost significantly for all pairwise calculations for scoring.

Using these sigmoid functions, the differential score is calculated by the following equation:

$$S_d(X, Y) = \frac{1}{2n} \sum_{i=1}^n (P(X^i) + P(Y^i)) \quad (2)$$

Here $P(X^i)$ means the difference between cancer and control in DNA methylation and gene expression level for a gene X . $P(X^1)$ thus indicates the modified fold change between cancer and control in DNA methylation, and $P(X^2)$ indicates the modified fold change in gene expression. $P(Y^i)$ is a value for gene Y similarly.

Second, we describe the correlation score that measures how much a gene pair, X and Y , is similar in the pattern of DNA methylation and gene expression level simultaneously. The correlation score is calculated by the following function:

$$S_c(X, Y) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m r(X^{ij}, Y^{ij}) \quad (3)$$

This equation takes the average of four possible correlations between two genes. Here, $r(X^{ij}, Y^{ij})$ calculates the Pearson correlation coefficient between two genes.

Finally, the MEGP score between any two genes is calculated by integrating both scores, the differential score and the correlation score:

$$S(X, Y) = \sqrt{S_d(X, Y)^\beta + S_c(X, Y)^\beta} \quad (4)$$

Here β is a parameter that adjusts points scaled by the square root. Basically, the MEGP score catches gene pairs of which both the differential score and correlation score are high. The threshold for a significant MEGP score is estimated by the permutation test of both the DNA methylation and gene expression matrices. With a similar framework, we can also measure a single type of profile. When we measure differentially co-methylated gene pairs (MGP), the number of profile types is one and X^1 indicates a variable for methylation. The measurement for differentially co-expressed gene pairs (EGP) is also calculated in a similar way.

The controlling parameters have the following characteristics. If α is lower, the MEGP score will concentrate on the correlation score. On the other hand, if this value is higher, it will concentrate on the differential score. When β is lower, the MEGP score will reflect the effect of combining both scores. According to the change in parameters α and β , the trend, whether the MEGP scores are balanced, is also shown in supplementary figure S2 of additional file 2 (available online only). In our experiment, the two parameters α and β were set to 1.0 and 0.1, respectively. The proper balance between both scores was observed in these parameter values.

Measurement of MEGP by expected conditional F-statistic

We compared the performance of the proposed method with a previous method, expected conditional F-statistic (ECF). It recognises the correlated pattern by extending the traditional F-statistic so that it can measure the differential gene-gene co-expression across different groups.¹¹ The ECF statistic of a correlation between genes X and Y is defined as:

$$ECF(X, Y) = E_Y(F_{X|Y=y}) + E_X(F_{Y|X=x}) \quad (5)$$

Here the ECF statistic is not symmetric so that two ECF statistics $E_Y(F_{X|Y=y})$ and $E_X(F_{Y|X=x})$ may be different. The differential gene-gene co-expression patterns are ranked by combining

them. More details on the ECF statistic method are shown in the paper by Lai *et al.*¹¹

Currently, scoring of the ECF statistic method is only applied to the analysis of gene expression profiles. Therefore, we reformulated previous ECF statistics by combining the ECF statistic for DNA methylation and that for gene expression. The modified ECF score is calculated by the average ECF statistics of DNA methylation and gene expression as follows:

$$ECF^{me}(X, Y) = \frac{1}{2}[ECF(X^m, Y^m) + ECF(X^e, Y^e)] \quad (6)$$

Here we also performed this calculation when the relationship between the fold change of methylation and that of gene expression is negative.

Measurement of functional correlations

In order to assess the functional correlation of each gene pair, we measured their physical connectivity using human protein-protein interaction (PPI) data and semantic similarity based on gene ontology (GO) terms, respectively. Previous studies have reported that PPI revealed relationships between functionally linked genes.¹²⁻¹³ We expect that two genes lie close together in PPI networks if they are functionally correlated. It is more likely that the regulation of co-expression and co-methylation between two genes allows them to coordinate their functional coupling. Human PPI data were obtained from the protein interaction network analysis platform that contains 77 158 interactions from six public databases (MINT, IntAct, DIP, BioGRID, HPRD and MIPS/MPact) (updated 25 February 2011).¹⁴ The network distance between two proteins was measured by the NetworkX package (<https://networkx.lanl.gov/>). The semantic similarity

was measured by GOsemSim in which four information content methods and a graph-based method were implemented.¹⁵

RESULTS

Extracting MEGP

In order to extract MEGP, we applied our MEGP scoring method to DNA methylation and gene expression datasets of ovarian cancer samples in TCGA.¹⁰ Our scoring method measures how much two genes are differentially methylated and expressed between cancer and control samples as well as how much they are correlated over the methylation and gene expression profiles. In addition, it explores simultaneously specific regulatory patterns so that the methylation ratio and expression ratio have an inverse relationship (figure 1). The inputs of the proposed scoring method consisted of two matrices of DNA methylation and two matrices of gene expression of the same size. These two matrices consisted of 9129 genes×496 cancer samples and 9129 genes×9 healthy control samples. From these inputs, we extracted a total of 1133 significant MEGP containing 423 unique genes, with a cut-off corresponding to a p value less than 10⁻⁴ from the permutation test (see supplementary additional file 1, available online only).

First of all, we focused on the annotation of the highest ranked gene pairs among the MEGP. The top-ranked MEGP among significant pairs (the q values are close to zero) are shown in table 1. Many genes were annotated as functions related to oncogenes. As shown in table 1, oncogene proteins such as epithelial cell transforming sequence 2 oncogene and differentially expressed genes in cancers such as microtubule-associated homolog were observed. In particular, lymphocyte-specific protein 1 associated with breast cancer¹⁶ was significantly correlated with SPI-1 proto oncogene that

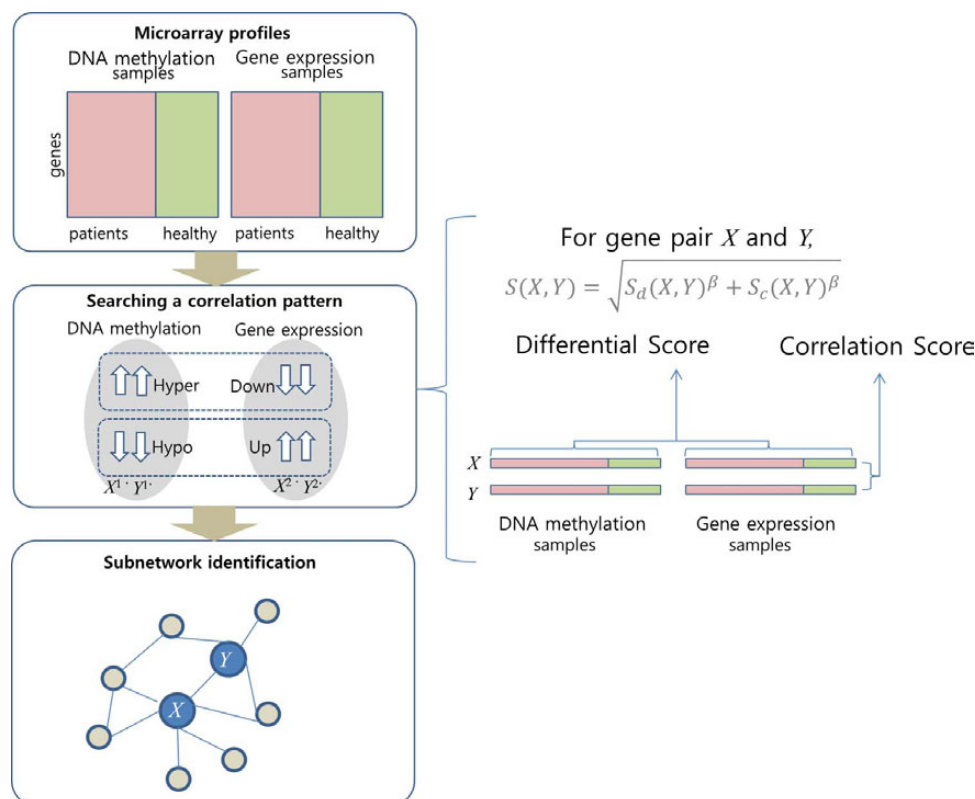


Figure 1 Schematic view of our approach to find co-methylated and co-expressed gene pairs (MEGP). It is the workflow of the MEGP analysis, which finds gene pairs correlated between DNA methylation and gene expression by the scoring function. This figure is only reproduced in colour in the online version.

Table 1 The top 10 ranked MEGP that are extracted from the proposed scoring method

Gene pair	Description (first gene)	Description (second gene)	MEGP score	
BUB1	CENPE	Mitotic checkpoint serine/threonine-protein kinase BUB1	Centromere-associated protein E	1.374
MSH2	RFC3	Truncated mutS homolog 2 protein	Unknown	1.374
BUB1	CCNA2	Mitotic checkpoint serine/threonine-protein kinase BUB1	Cyclin-A2	1.374
HSPA14	MCM10	Heat shock 70 kDa protein 14	Minichromosome maintenance protein 10	1.373
BUB1	CHEK2	Mitotic checkpoint serine/threonine-protein kinase BUB1	CHK2 checkpoint homolog	1.373
ANP32E	CHEK1	Unknown	Checkpoint kinase 1	1.372
CCNA2	FBXO5	Cyclin-A2	F-box protein 5	1.372
BUB1	FBXO5	Mitotic checkpoint serine/threonine-protein kinase BUB1	F-box protein 5	1.372
CDCA8	OIP5	Borealin	Unknown	1.372
C12orf48	OIP5	PARP1-binding protein	Unknown	1.372

MEGP, co-methylated and co-expressed gene pairs.

promotes cancer progression via DNA replication.¹⁷ These findings suggest that a number of genes belonging to MEGP have substantial epigenetic changes and may cooperate with each other in cancers.

We also obtained MEGP with the previous method, ECF statistic, which measures the differential gene–gene co-expression across different groups.¹¹ Because it can measure only the co-expression of a gene pair, we modified it by adding gene–gene co-methylation to the previous measurement. After sorting by average significance of ECF, we extracted 1133 gene pairs. We could not find oncogene-related genes in the top-ranked gene pairs. Only 7.3% among the MEGP by scoring (83 gene pairs) overlapped with the MEGP from the ECF statistic. The list of MEGP from the ECF statistic and the shared pairs are shown in supplementary additional files 1 and 2, supplementary table S1 (available online only). We further compared functional similarity between the different methods using network connectivity with the PPI network and semantic similarity by GO.

Comparison of network connectivity and semantic similarity

We measured how closely the MEGP are connected using human PPI data obtained from the protein interaction network analysis platform.¹⁴ As most proteins do not directly interact with each other, we measured the network distance for all MEGP including direct and indirect protein interactions. We compared the shortest distance distributions of gene pairs

extracted by seven different methods (figure 2). MEGP extracted by scoring methods were more closely connected while those extracted by the ECF statistic were not. The distances of gene pairs extracted by single scoring methods such as MGP or EGP were higher than those extracted by the scoring method combining both. This result suggests that MEGP extracted by scoring methods have a higher probability of being functionally linked.

Furthermore, we extracted MEGP by direct physical interactions. Among the total MEGP, 2.7% (MEGP consisting of 39 unique genes) were found in the PPI network (see supplementary additional file 2, supplementary figure S1, available online only). Although this percentage of interactions is still low, it is substantially greater than expected from the random distribution. The rate of PPI gene pairs extracted by single scoring methods is not far from the random distribution. As more interactions are observed in the proteins of EGP than in those of MGP, it suggests that the patterns of gene expression profiles contribute more to finding functionally related gene pairs. Gene pairs extracted by the ECF statistic tend to have less interaction.

In addition, we assessed the functional similarity of gene pairs, using five methods measuring the semantic similarity, including four information content methods and a graph-based method, proposed by Wang *et al.*¹⁸ These methods measure the similarity between given two GO terms or two sets of terms. We performed the semantic comparisons of GO annotations in

Figure 2 Comparison of the shortest distance distributions of gene pairs extracted by seven different methods using protein–protein interaction data. ECF, expected conditional F-statistic; EGP, co-expressed gene pairs; MEGP, co-methylated and co-expressed gene pairs; MGP, co-methylated gene pairs. This figure is only reproduced in colour in the online version.

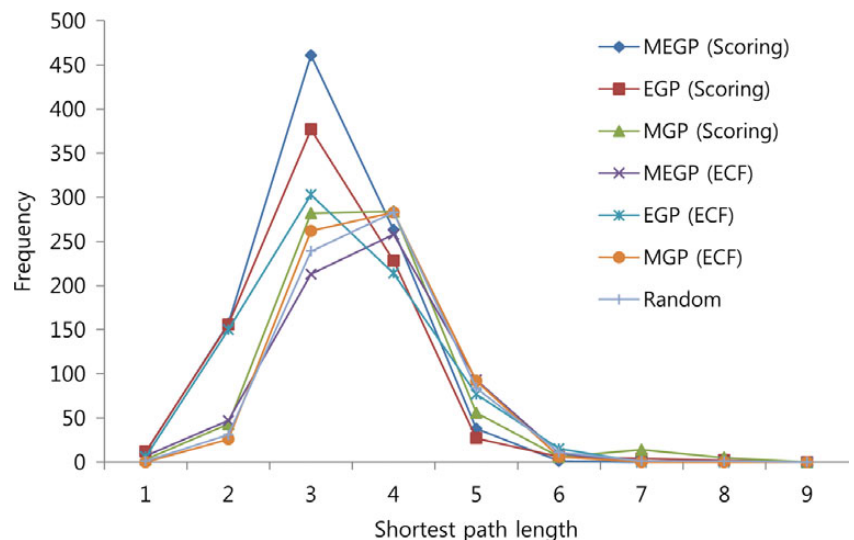


Table 2 Comparison of three different methods in semantic similarity for gene pairs

Categories	Methods for semantic similarity	Gene pairs		
		MEGP: scoring	MEGP: ECF statistics	Random pairs
MF (molecular function)	Resnik ¹⁹	0.114 (±0.035)	0.087 (±0.037)	0.077 (±0.035)
	Lin	0.652 (±0.171)	0.452 (±0.199)	0.366 (±0.186)
	Relevance ²⁰	0.487 (±0.131)	0.336 (±0.151)	0.271 (±0.141)
	Jiang and Conrath's	0.742 (±0.137)	0.570 (±0.171)	0.486 (±0.172)
	Wang ¹⁸	0.762 (±0.119)	0.610 (±0.146)	0.556 (±0.144)
BP (biological process)	Resnik	0.204 (±0.076)	0.169 (±0.037)	0.120 (±0.035)
	Lin	0.351 (±0.128)	0.309 (±0.199)	0.232 (±0.186)
	Relevance	0.319 (±0.133)	0.272 (±0.151)	0.198 (±0.141)
	Jiang and Conrath's	0.330 (±0.121)	0.314 (±0.171)	0.255 (±0.172)
	Wang	0.341 (±0.101)	0.331 (±0.146)	0.265 (±0.144)
CC (cellular component)	Resnik	0.198 (±0.052)	0.147 (±0.037)	0.098 (±0.035)
	Lin	0.553 (±0.127)	0.419 (±0.199)	0.307 (±0.186)
	Relevance	0.470 (±0.123)	0.351 (±0.151)	0.239 (±0.141)
	Jiang and Conrath's	0.657 (±0.108)	0.586 (±0.171)	0.555 (±0.172)
	Wang	0.655 (±0.100)	0.562 (±0.146)	0.501 (±0.144)

The values for the mean and SD for the semantic similarity are shown. Random pairs mean gene pairs randomly selected, having the same number as that of the MEGP. ECF, expected conditional F-statistic; MEGP, co-methylated and co-expressed gene pairs.

three different categories, including 'BP' (biological process), 'MF' (molecular function) and 'CC' (cellular component). Gene pairs extracted by the MEGP scoring method had overall higher semantic similarities than random pairs in the measurements including 'Resnik', 'Lin', 'Jiang', 'Rel' and 'Wang' (table 2).^{18–20} Furthermore, semantic similarities of those pairs were higher than the similarities measured from the EGP and MGP, shown in supplementary table S2 of additional file 2 (available online only). It suggests that our method could find close functional correlation between gene pairs. It is more efficient than when DNA methylation and gene expression profiles are independently used. Gene pairs extracted by MEGP ECF statistics had

overall higher semantic similarities than random pairs, whereas those similarities were not always higher than similarities measured by the EGP ECF statistics.

Enriched ovarian cancer-associated gene sets

To identify the functional role of our extracted MEGP in ovarian cancer, we performed enrichment analysis using gene sets of chemical and genetic perturbations. We extracted significantly enriched gene sets among a total of 2392 chemical and genetic perturbation gene sets in the MSigDB collections (<http://www.broadinstitute.org/gsea/msigdb/>). As shown in figure 3, we observed that two gene sets were associated with ovarian

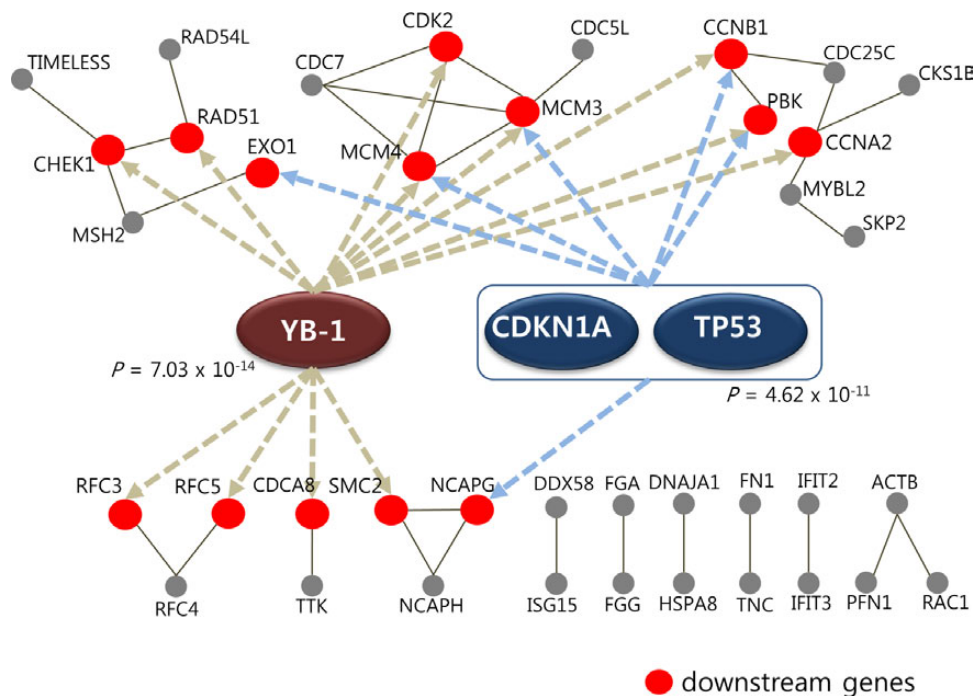


Figure 3 Ovarian cancer-related subnetwork identified from the gene set enrichment analysis. The interacting protein pairs are linked with lines and the downstream genes of Y-box-binding protein 1, CDKN1A and TP53 in ovarian cancer are linked to them with arrows. This figure is only reproduced in colour in the online version.

cancers among 56 significantly enriched gene sets ($p < 10^{-10}$) from the total gene sets. Among 39 physically interacting MEGP genes, 12 genes are involved in two gene sets that are upregulated under Y-box-binding protein 1 knockout conditions.²¹

Moreover, six genes are downstream of the *CDKN1A* gene, encoding p21WAF1/CIP1 and the *TP53* gene, encoding tumor protein p53 (figure 3).²² In particular, minichromosome maintenance deficient 4 (*MCM4*) and *MCM3* involved in the initiation of eukaryotic genome replication and cyclin A2 (*CCNA2*), a *RAD51* homolog (*RAD51*), and *CHK1* checkpoint homolog (*CHEK1*) are functionally associated with the cell cycle process. Consequently, this observation significantly suggests that these genes associated with the cell cycle process could lead to a defect in the DNA replication process and cause mutations and abnormal tissue growth in ovarian cancer.

DISCUSSION

We proposed a novel approach to find MEGP that are tightly correlated in DNA methylation as well as gene expression, regarded as the units of genes in a regulatory relationship. Our method found many MEGP that are more likely to be functionally correlated in specific biological processes and PPI networks, when we applied it to the datasets of gene expression and methylation profiles from patients with ovarian cancer cells. Furthermore, we found a number of downstream genes, perturbed by mutations in ovarian cancer, which were significantly overlapped with the set of MEGP.

To the best of our knowledge, there is no method to extract easily differentially expressed as well as differentially methylated gene pairs. Our method provides a unified scoring function to extract simultaneously the pairwise patterns of altered DNA methylation and gene expression between normal and disease samples. Although the proposed function is simply formulated by integrating the fold change and the correlation coefficient, it clearly extracts a number of significantly correlated gene pairs associated with epigenetic events. Employing our approach, we can rank the gene pairs and are able to determine which gene pairs are more epigenetically correlated. Despite this advantage, it has a weakness in the parameter setting. Our method needs parameters to control the balance between the differential score and the correlation score. Therefore, the fine tuning of the parameters for other datasets may be required.

Furthermore, our method can be applied to other types of datasets integrated with gene expression profiles, such as copy number variants, single nucleotide variants and micro RNA. As diverse high-throughput technologies for genome analysis have developed rapidly, the integration of multiple types of genome data has become an important problem. Among many genome projects, the TCGA project provides a large amount of multiple-dimensional datasets. Our future study will be more focused on identifying significant gene patterns from these multiple genome datasets with a similar strategy.

As epigenetic biomarkers can be utilised to support clinical decision making, it is important to discover useful epigenetic patterns associated with specific diseases. In conclusion, our approach could contribute to understanding functional modules or pathways associated with diverse diseases such as certain

malignant tumors by identifying key regulatory components of DNA methylation and gene expression.

Contributors JGJ conducted the study and wrote the manuscript. DK built the experimental datasets and revised the manuscript. KHK contributed to the biological interpretation and wrote the manuscript. JHK supervised the study.

Funding This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0000994). DK's educational grant was supported by the Ministry of Health and Welfare, Republic of Korea (A112020).

Competing interests None.

Patient consent Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 Wu H, Tao J, Sun YE. Regulation and function of mammalian DNA methylation patterns: a genomic perspective. *Brief Funct Genomics* 2012;11:240–50.
- 2 Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012;13:484–92.
- 3 Gonzalo V, Lozano JJ, Munoz J, et al. Aberrant gene promoter methylation associated with sporadic multiple colorectal cancer. *PLoS One* 2010;5:e8777.
- 4 Van der Auwera I, Yu W, Suo L, et al. Array-based DNA methylation profiling for breast cancer subtype discrimination. *PLoS One* 2010;5:e12616.
- 5 Houshdaran S, Hawley S, Palmer C, et al. DNA methylation profiles of ovarian epithelial carcinoma tumors and cell lines. *PLoS One* 2010;5:e9359.
- 6 Noushmehr H, Weisenberger DJ, Diefes K, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* 2010;17:510–22.
- 7 Boyer LA, Lee TI, Cole MF, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005;122:947–56.
- 8 Blöchl R, Wang Z, Meissner A, et al. Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. *Stem Cells* 2006;24:2007–13.
- 9 Louhimo R, Hautaniemi S. CNAMet: an R package for integrating copy number, methylation and expression data. *Bioinformatics* 2011;27:887–8.
- 10 The Cancer Genome Atlas Network. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011;474:609–15.
- 11 Lai Y, Wu B, Chen L, et al. A statistical method for identifying differential gene–gene co-expression patterns. *Bioinformatics* 2004;20:3146–55.
- 12 Ge H, Liu Z, Church GM, et al. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat Genet* 2001;29:482–6.
- 13 Jansen R, Greenbaum D, Gerstein M. Relating whole-genome expression data with protein–protein interactions. *Genome Res* 2002;12:37–46.
- 14 Wu J, Vallenius T, Ovaska K, et al. Integrated network analysis platform for protein–protein interactions. *Nat Methods* 2009;6:75–7.
- 15 Yu G, Li F, Qin Y, et al. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 2010;26:976–8.
- 16 Chen MB, Li C, Shen WX, et al. Association of a LSP1 gene rs3817198T>C polymorphism with breast cancer risk: evidence from 33,920 cases and 35,671 controls. *Mol Biol Rep* 2011;38:4687–95.
- 17 Rimmel P, Komatsu J, Hupe P, et al. Spi-1/PU.1 oncogene accelerates DNA replication fork elongation and promotes genetic instability in the absence of DNA breakage. *Cancer Res* 2010;70:6757–66.
- 18 Wang JZ, Du Z, Payattakool R, et al. A new method to measure the semantic similarity of GO terms. *Bioinformatics* 2007;23:1274–81.
- 19 Resnik P. Using information content to evaluate semantic similarity in a taxonomy. IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence. 1995;1:448–53.
- 20 Schlicker A, Domingues FS, Rahnenfuhrer J, et al. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics* 2006;7:302.
- 21 Basaki Y, Hosoi F, Oda Y, et al. Akt-dependent nuclear localization of Y-box-binding protein 1 in acquisition of malignant characteristics by human ovarian cancer cells. *Oncogene* 2007;26:2736–46.
- 22 Wu Q, Kirschmeier P, Hockenberry T, et al. Transcriptional regulation during p21WAF1/CIP1-induced apoptosis in human ovarian cancer cells. *J Biol Chem* 2002;277:36329–37.