Tensor decomposition–based unsupervised feature extraction for integrated analysis of TCGA data on microRNA expression and promoter methylation of genes in ovarian cancer

Y-h. Taguchi
Department of Physics
Chuo University, 1-13-27 Kasuga
Bunky-ku, Tokyo 112-8551, Japan
Email: tag@granular.com

Ka-Lok Ng
Department of Bioinformatics and Medical Engineering
Asia University, Taichung, Taiwan, Email: ppiddi@gmail.com
Department of Medical Research, China Medical University Hospital
China Medical University, Taiwan, Email: ppiddi@gmail.com

***Abstract*—Integrated analysis of epigenetic profiles is important but difficult. Tensor decomposition–based unsupervised feature extraction was applied here to data on microRNA (miRNA) expression and promoter methylation of genes in ovarian cancer. It selected seven miRNAs and 241 genes by expression levels and promoter methylation degrees, respectively, such that they showed differences between eight normal ovarian tissue samples and 569 tumor samples. The expression levels of the seven miRNAs and the degrees of promoter methylation of the 241 genes also correlated significantly. Conventional Student's t test–based feature selection failed to identify miRNAs and genes that have the above properties. On the other hand, biological evaluation of the seven identified miRNAs and 241 identified genes suggests that they are strongly related to cancer as expected.**

***Keywords*-tensor decomposition; feature extraction; ovarian cancer; microRNA; promoter methylation**

I. Introduction

Multiomics analysis is key to understanding complicated regulation of gene expression by multiple factors. Examples of these factors are DNA methylation [1], histone modification [2], and chromatin structures [3]. Functional noncoding RNAs, including microRNA (miRNA) [4], are also often regarded as some of important regulators of gene expression. In spite of the importance of these regulators, it is rarely discussed how multiple factors regulate gene expression cooperatively.

Especially, the relation between methylation and miRNA as regulators is unclear, although how methylation affects miRNA expression is discussed [5]. The reason why this topic is not discussed much is possibly that methylation contributes to pretranscriptional regulation whereas miRNA contributes to post-transcriptional regulation. Because it is difficult to figure out how methylation and miRNA can regulate gene expression cooperatively from the biological point of view, data-driven approaches are the only possible strategy. To this end, we need to identify a set of genes to which the amount of methylation is attributed and miRNAs that fulfill the following conditions.

1) MiRNA should be expressed differentially between treated and control samples.
2) The degrees of promoter methylation of the genes should be different between treated and control samples.
3) Expression levels of these miRNAs and the degrees of promoter methylation of the above genes should significantly correlate.

If a set of miRNAs and genes fulfills these criteria, then they are candidates that regulate gene expression cooperatively, although further analysis will be necessary to see if they really work cooperatively. The purpose of this study was restricted to the identification of a set of miRNAs and genes that satisfy the above three conditions.

Nonetheless, even finding the sets of miRNAs and genes that fulfill these *weak* conditions (by expression levels and degrees of promoter methylation, respectively) is not easy. First of all, because we cannot restrict pairing of miRNA expression levels and degrees of promoter methylation of genes, all possible pairs must be tested. The number of possible pairs can easily exceed a few million. This means that P-values must be at least smaller than $1 \times 10^{-6} \times 0.05 \simeq 1 \times 10^{-8}$ if the required possible threshold P-value is 0.05. It is generally not easy to identify such highly significant pairs of differentially expressed miRNAs and differentially methylated genes. Second, miRNAs and genes showing differences in expression levels between treated and control samples do not always correlate. If the difference between controls and treated samples is not large enough, the correlation between expression of miRNAs and promoter methylation data on genes is not governed by the differences between control and treated samples but rather by correlation within each class: miRNA expression vs gene methylation in normal tissues or miRNA expression vs gene methylation in tumors. For example, The Cancer Genome Atlas [6] (TCGA) sample is not associated with equal numbers of treated and

[0]Accepted for IEEE BIBE2018 DOI: 10.1109/BIBE.2018.00045

control samples but is associated with a combination of very small numbers of control samples and large numbers of tumor samples; this means that the correlation between miRNA expression levels and promoter methylation degrees of genes is governed by that within tumor samples.

To overcome this difficulty, we employed tensor decomposition (TD)-based unsupervised feature extraction (FE) [7], [8], [9], [10], [11], [12], [13]. TD-based unsupervised FE was applied to ovarian methylation profiles and miRNA expression data retrieved from TCGA. We successfully obtained a set of methylation sites and miRNAs that significantly correlate and show a significant dissimilarity between controls and treated samples simultaneously. Enrichment analysis also identified biological significance of the differentially expressed miRNAs and differentially methylated genes.

II. Materials and Methods

A. Methylation profiles and miRNA expression

Ovarian methylation profiles and miRNA expression data were downloaded from TCGA. They are composed of eight normal ovarian tissue samples and 569 tumor samples. Our dataset includes expression data on 723 miRNAs as well as promoter methylation profiles of 24906 genes.

B. TD-based unsupervised FE

Given that the method was described in detail in another paper [10], it is described here only briefly. Suppose that $x_{ij}^{\text{methyl}} \in \mathbb{R}^{N\times M}$ is the degree of promoter methylation of the ith gene of the jth sample whereas $x_{kj}^{\text{miRNA}} \in \mathbb{R}^{K\times M}$ is the expression level of the kth miRNA of the jth sample. $N(=\ 24906)$ is the number of genes whose promoter methylation status is known, and $K(=723)$ is the number of miRNAs whose expression has been measured, and $M(=577)$ is the number of samples. Both x_{ij} and x_{kj} were standardized such that they were associated with zero mean and unit variance, i.e., $\sum_i x_{ij}^{\text{methyl}} = \sum_k x_{kj}^{\text{miRNA}} = 0$, $\sum_i \left(x_{ij}^{\text{methyl}}\right)^2 = N$, and $\sum_k \left(x_{kj}^{\text{miRNA}}\right)^2 = K$.

Next, to generate a case II type I tensor [10], we define

$$x_{ijk} = x_{ij}^{\text{methyl}} x_{kj}^{\text{miRNA}} \tag{1}$$

x_{ijk} was subjected to Tucker decomposition as follows:

$$x_{ijk} = \sum_{\ell_1=1}^{N}\sum_{\ell_2=1}^{M}\sum_{\ell_3=1}^{K} G(\ell_1,\ell_2,\ell_3) x_{\ell_1 i} x_{\ell_2 j} x_{\ell_3 k} \tag{2}$$

where G is the core tensor and $x_{\ell_1 i} \in \mathbb{R}^{N\times N}, x_{\ell_2 j} \in \mathbb{R}^{M\times M}, x_{\ell_3 k} \in \mathbb{R}^{K\times K}$ are singular value matrices that are orthogonal. Because Tucker decomposition is not unique, we have to specify how we derive Tucker decomposition. In particular, we chose higher-order singular value decomposition (HOSVD) [14].

Given that x_{ijk} is too large to apply TD as is, we generate a case II type II tensor

$$x_{ik} = \sum_{j=1}^{M} x_{ijk} \tag{3}$$

Singular value decomposition (SVD) was applied to matrix $X \in \mathbb{R}^{N\times K}$ whose components are X_{ij}; thus, we get

$$X = U\Sigma V^T \tag{4}$$

where $U \in \mathbb{R}^{N\times K}$ and $V \in \mathbb{R}^{K\times K}$ are orthogonal matrices (here $N > K$), and $\Sigma \in \mathbb{R}^{K\times K}$ is a diagonal matrix. U^T should correspond to $x_{\ell_1 i}$. This means that $x_{\ell_1 i} = 0$ for $\ell_1 > K$. On the other hand, V^T should corresponds to $x_{\ell_3 k}$.

$x_{\ell_2 j}$ that corresponds to samples cannot be obtained by SVD. As shown in the previous study [10], we can obtain two $x_{\ell_2 j}$s that correspond to methylation and miRNA, respectively, in the following way:

$$x_{\ell_2 j}^{\text{miRNA}} = \sum_{k=1}^{K} x_{\ell_3 k} x_{kj} \tag{5}$$

$$x_{\ell_2 j}^{\text{methyl}} = \sum_{i=1}^{N} x_{\ell_1 i} x_{ij} \tag{6}$$

The selection of genes to which methylation profiles are attributed and miRNAs using the above results can be performed as follows. First, among singular value vectors attributed to samples, we select $x_{\ell_2 j}^{\text{methyl}}$ and $x_{\ell_2 j}^{\text{miRNA}}$ that show significant differences between normal ovarian tissues ($1 \le j \le 8$) and tumors ($j > 8$). This task can be accomplished, for example, with some statistical tests like Student's t test. Suppose that ℓ_2' turned out to show dissimilarity between control and treated samples. Then, P-values are attributed to k miRNAs and i genes, assuming that $x_{\ell_1 i}$ and $x_{\ell_3 k}$ obey a normal distribution,

$$P_i = P_{\chi^2}\left[> \sum_{\ell_1=\ell_2'} \left(\frac{x_{\ell_1 i}}{\sigma_{\ell_1}}\right)^2 \right] \tag{7}$$

$$P_k = P_{\chi^2}\left[> \sum_{\ell_3=\ell_2'} \left(\frac{x_{\ell_3 k}}{\sigma_{\ell_3}}\right)^2 \right] \tag{8}$$

where $P_{\chi^2}[>x]$ is cumulative probability that the argument is greater than x in a χ^2 distribution. σ_{ℓ_1} and σ_{ℓ_3} are standard deviations for $x_{\ell_1 i}$ and $x_{\ell_3 k}$, respectively. After P-values are adjusted by means of the Benjamini–Hochberg (BH) criterion [15], miRNAs and genes that are associated with adjusted P-values less than 0.01 are selected as those showing differences in expression and promoter methylation, respectively, between controls (normal ovarian tissues) and treated samples (tumors).

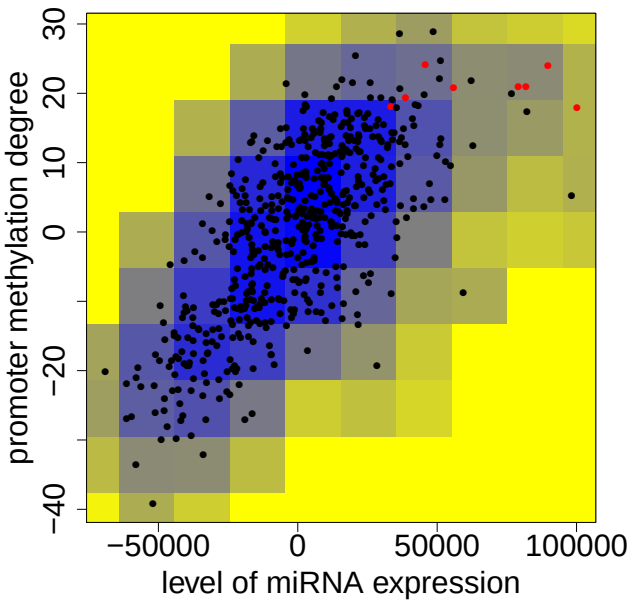

Figure 1. Scatter plot of $x^{\text{miRNA}}_{\ell_2 j}$ and $x^{\text{methyl}}_{\ell_2 j}$. The coefficient of correlation between them is 0.72 ($P = 2.0 \times 10^{-92}$). Red points and black points correspond to normal ovarian tissues and tumors, respectively. Color indicates local density of points (blue to yellow denote denser to sparser). All units are arbitrary.

C. Analysis of correlation between expression of the miRNAs and promoter methylation of the genes

Pearson's correlation coefficients were computed by the `corr` function in R software. P-vales were computed with the `cor.test` function of R.

III. Results

We applied TD-based unsupervised FE to an ovarian cancer dataset retrieved from TCGA. Then, we found that $x^{\text{miRNA}}_{\ell_2 j}$ and $x^{\text{methyl}}_{\ell_2 j}$ with $\ell_2 = 2$ are different between normal ovarian tissues and tumors. Student's t test was performed on the two sets, $j \leq 8$ and $j > 8$. The P-values obtained for miRNA and genes were 1.3×10^{-4} and 1.2×10^{-11}, respectively. To detect a correlation between $x^{\text{miRNA}}_{\ell_2 j}$ and $x^{\text{methyl}}_{\ell_2 j}$ for $\ell_2 = 2$, we constructed a scatter plot from these data (Fig. 1). It is obvious that they strongly correlate. P_i and P_k were computed by means of $x_{\ell_1 i}$ and $x_{\ell_3 k}$ of $\ell_1 = \ell_3 = 2$. Finally, we found seven miRNAs and 241 genes that showed differences in expression and promoter methylation, respectively, between control and treated samples. To confirm that the expression levels of seven miRNAs and the degrees of promoter methylation of 241 genes are correlated, we computed pairwise coefficients of correlation between them and computed adjusted P-values for each of these $7 \times 241 = 1687$ pairs. As presented in Table I, most of the pairs are associated with adjusted P-values less than 0.01, i.e., indicating statistical significance. Thus, TD-based unsupervised FE successfully identified pairs of miRNAs and genes having significant pairwise correlations.

Table I
THE NUMBER OF MIRNA–GENE PAIRS SHOWING A SIGNIFICANT CORRELATION (ADJUSTED P-VALES LESS THAN 0.01) WHEN THEY ARE IDENTIFIED WITH TD-BASED UNSUPERVISED FE.

		negative correlation	
		T	F
positive correlation	T	0	985
	F	607	95

Table II
THE NUMBER OF MIRNA–GENE PAIRS SHOWING A SIGNIFICANT CORRELATION (ADJUSTED P-VALES LESS THAN 0.01) WHEN THESE ARE IDENTIFIED BY STUDENT'S t TEST.

		negative correlation	
		T	F
positive correlation	T	0	329896
	F	225495	3595139

In order to see if other conventional methods can compete with TD-based unsupervised FE, we performed simple Student's t test on miRNA expression data and degrees of promoter methylation of the genes. This analysis resulted in 214 miRNAs and 19395 genes associated with adjusted P-values less than 0.01 (by the BH criterion). This finding suggests that simple Student's t test cannot select a reasonable number of miRNAs or genes associated with significant P-values. Next, to confirm the superiority of TD-based unsupervised FE toward simple Student's t test, we computed pairwise coefficients of correlation between these 214 miRNAs and 19395 genes (Table II). In contrast to Table I where the majority of pairs show significant intrapair correlations, only ∼13% pairs show a significant correlation. Thus, Student's t test cannot identify pairs of miRNAs and genes with significant intrapair correlation with a small number of false positives. One may still wonder if Student's t test can compete with TD-based unsupervised FE when correlation analysis is restricted to the miRNAs and genes with large differences between controls and treated samples. For this purpose, we repeated correlation analysis with top-ranked seven miRNAs and 241 genes according to P-values computed by Student's t test (Table III) . It is obvious that restricting the analysis to only top-ranked miRNAs and genes does not improve the identification of pairs with a significant correlation at all if Table III is compared with Table I.

Readers may still wonder whether selecting pairs with significant correlations prior to the identification of those showing a difference between controls and treated samples can select miRNAs and genes effectively. To test this idea, we computed P-values for all pairs of miRNA expression

Table III
THE NUMBER OF MIRNA–GENE PAIRS SHOWING A SIGNIFICANT INTRAPAIR CORRELATION (ADJUSTED P-VALES LESS THAN 0.01) WHEN ONLY SEVEN TOP-RANKED MIRNAS AND 241 TOP-RANKED GENES SELECTED BY STUDENT'S t TEST ARE CONSIDERED.

		negative correlation	
		T	F
positive	T	0	13
correlation	F	28	1646

Table IV
THE NUMBER OF MIRNA–GENE PAIRS (AN EXPRESSION LEVEL AND DEGREE OF PROMOTER METHYLATION, RESPECTIVELY) SHOWING A SIGNIFICANT INTRAPAIR CORRELATION (ADJUSTED P-VALES LESS THAN 0.01).

		negative correlation	
		T	F
positive	T	0	608989
correlation	F	588783	16809266

levels and the degrees of promoter methylation of genes and identified pairs with significant intrapair correlations (Table IV). Apparently, this approach is successful because only a limited number of pairs (less than 10%) were identified. Nevertheless, it cannot be used for identification of the miRNAs and genes with desired properties because it turned out that all the miRNAs and genes had at least one significant correlation.

The above result suggests that TD-based unsupervised FE can outperform the conventional methods when selecting miRNAs and genes satisfying the three conditions presented in the Introduction section.

Next, we wanted to evaluate biological significance of the selected miRNAs and genes. First, seven miRNAs (Table V) were uploaded to DIANA-miRPath [16] with TarBase specified as a target identification database. Although a total of 66 KEGG pathways are enriched among these miRNAs (Table VI), there are at least 15 pathways related to cancer directly (bold). Next, gene symbols (Table V) were uploaded to MSigDB (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp). C6: oncogenic signatures were tested and as many as 33 oncogenic expressed gene sets were found to significantly overlap (Only top twenty sets are included in Table VII because of lack of space).

It is known that the majority of ovarian cancers is derived from the ovarian surface epithelium [17]. It is evident from Table VII, among the first ten MSigDB records, five have "epithelium cell" descriptions, which accounted for 50% of the records. For instance, some of the oncogenic signatures were found in epithelial cells, the gene set names are WNT_UP.V1_DN, KRAS.600.LUNG.BREAST_UP.V1 UP, KRAS.LUNG.BREAST_UP.V1 UP, KRAS.300 UP.V1 UP, and KRAS.600 UP.V1 UP. Furthermore, clinical studies

Table V
DNA-METHYLATION-REGULATED MIRNAS (7) AND GENES (241).

hsa-miR-142-3p hsa-miR-142-5p hsa-miR-150 hsa-miR-21* hsa-miR-22 hsa-miR-224 hsa-miR-96
ABCG1 ACCN3 ACTN3 ADORA3 ADRA2B ANAPC13 APC2 APOL6 AZU1 BHMT BIK C10orf2 C11orf66 C13orf28 C14orf162 C15orf24 C1QTNF9 C20orf186 C21orf121 C2orf40 C2orf58 C3 C6orf204 C9orf41 CAMTA2 CAPS CARD10 CCKAR CCL21 CCRL2 CD1A CD1B CD274 CDO1 CEP63 CFTR CHI3L2 CLDN9 CLIC6 CNOT6 COL7A1 COQ3 COX6A2 CPNE8 CRYBB3 CRYGD CTAGE5 CTHRC1 CTNNBL1 CUL7 CYP2W1 CYP4F22 DAPP1 DENND2D DLG2 DOM3Z ECE1 ELF1 ELMO3 ELOVL2 ESM1 EVI2A EXOC3L2 FAM71F1 FANCG FBXO2 FBXO44 FERD3L FGD2 FKBP10 FLRT1 FLVCR2 GALP GBP4 GLIPR1L2 GLRX GNAS GNMT GPR12 GPR133 GPR32 GRIK2 GRIP1 GRM2 HBQ1 HCRTR1 HDAC11 HHATL HIST1H2BK HIST1H4I HLA-DMA HLA-DOB HNF1B HOXB5 HOXD4 HPS1 IGDCC3 IGFALS INHBE ITGBL1 JAKMIP3 KAZALD1 KCNAB1 KCTD12 KIAA0020 KIR3DX1 KLHL10 LCN12 LIPC LOC404266 LOC84931 LUC7L MACROD1 MAK16 MAP7D2 MIR10B MNDA MRPL2 MRPL43 MUC5B NA NAF1 NAGS NCL NEFM NF1 NFKBIL2 NLRP5 NLRP6 NRM NT5C3L NTM NTNG2 NUMBL NXN ODF3L2 OLFM1 OPRD1 PCDHA1 PCDHA10 PCDHA11 PCDHA12 PCDHA13 PCDHA2 PCDHA3 PCDHA4 PCDHA5 PCDHA6 PCDHA7 PCDHA8 PCDHA9 PCDHB12 PCDHB14 PCDHB15 PCDHB16 PCDHB4 PCDHB5 PCDHB7 PCDHB8 PCDHGA1 PCDHGA10 PCDHGA11 PCDHGA12 PCDHGA2 PCDHGA3 PCDHGA4 PCDHGA5 PCDHGA6 PCDHGA7 PCDHGA8 PCDHGA9 PCDHGB1 PCDHGB2 PCDHGB3 PCDHGB4 PCDHGB5 PCDHGB6 PCDHGB7 PDCD1LG2 PGBD4 PHACTR2 PICALM PKP3 POMC PPDPF PPIL6 PPP1CC PRG3 PRTN3 PSMB8 PUF60 PVRL4 PYY RENBP RGN RNASEH2A RNH1 S100A16 SCMH1 SEMA3B SERPINB5 SERPINB8 SLC35C1 SLC44A2 SMPD2 SP100 SPAG7 SPATA18 SRPX2 SRRM3 STARD8 STC2 STK19 STMN4 STXBP2 SULT1C4 SULT2A1 SYNE2 TAP1 TBX4 TBX5 TCL1A TEX264 TFF3 TMEM105 TMEM140 TMEM173 TMEM71 TPSAB1 TPSB2 TRIM22 TRIM63 TTLL7 UBB UCN UCN2 UCN3 VNN2 VPREB1 VPS28 VSTM1 VWA5B1 ZDHHC11 ZNF154 ZNF532 ZNF556 ZNF560 ZNF671 ZNF678 ZNHIT6

suggest that two hormones, estrogen and progesterone, are involved in ovarian cancer formation [18]. Table VIII lists the top six Gene Ontology (GO) molecular functional annotations of genes returned by MSigDB. Two of the molecular-function records are hormone related: "hormone activity" and "peptide hormone receptor binding"; the results were what we expected. In summary, the gene sets we identified were in line with the cell type and hormone records in the enrichment analysis.

Tables VI, VII, and VIII suggest that TD-based unsupervised FE successfully identified cancer-related miRNAs and genes as expected.

IV. CONCLUSION

In this paper, we applied TD-based unsupervised FE to miRNA expression and gene promoter methylation data (on ovarian tumors) retrieved from TCGA. TD-based unsupervised FE successfully identified genes with differential promoter methylation and differentially expressed miRNAs between normal ovarian tissues and tumors as well as significant correlations between the expression levels and methylation data. Student's t test failed to identify the sets of miRNAs and genes satisfying these criteria. Biological evaluation of the identified miRNAs by DIANA-miRPath

Table VI
ENRICHED KEGG PATHWAYS DETECTED BY DIANA-MIRPATH AMONG SEVEN MIRNAS (TABLE V). BOLD ONES ARE CANCER RELATED. G #: THE NUMBER OF GENES, M #: THE NUMBER OF RELATED MIRNAS; *p*-VALUES ARE ADJUSTED.

KEGG pathway	p-value	g #	m #
Viral carcinogenesis	4.17E-11	91	7
Proteoglycans in cancer	4.17E-11	86	7
Prion diseases	4.87E-09	10	6
Adherens junction	4.87E-09	41	7
Renal cell carcinoma	4.87E-09	38	7
Bacterial invasion of epithelial cells	2.79E-08	41	7
Central carbon metabolism in cancer	4.84E-08	37	7
Hippo signaling pathway	5.90E-08	57	7
Cell cycle	7.20E-08	62	7
TGF-beta signaling pathway	8.55E-08	37	7
Fatty acid biosynthesis	1.61E-07	4	4
Glycosaminoglycan biosynthesis - keratan sulfate	2.71E-07	8	6
Hepatitis B	1.69E-06	60	7
Prostate cancer	3.75E-06	46	7
Shigellosis	5.12E-06	33	6
Pathogenic Escherichia coli infection	8.67E-06	33	7
Pancreatic cancer	1.51E-05	34	7
Fatty acid metabolism	2.08E-05	14	5
FoxO signaling pathway	3.02E-05	59	7
Protein processing in endoplasmic reticulum	5.58E-05	74	7
Regulation of actin cytoskeleton	5.58E-05	83	7
p53 signaling pathway	5.72E-05	36	7
HIF-1 signaling pathway	6.18E-05	50	7
2-Oxocarboxylic acid metabolism	2.20E-04	9	5
Lysine degradation	2.20E-04	19	7
Oocyte meiosis	2.31E-04	45	7
Ubiquitin mediated proteolysis	3.30E-04	55	7
Endocytosis	4.41E-04	81	7
SNARE interactions in vesicular transport	4.44E-04	17	7
Colorectal cancer	6.24E-04	31	7
Endometrial cancer	9.95E-04	25	6

and of genes by MSigDB suggests that TD-based unsupervised FE identified genes and miRNAs related to cancers as expected.

ACKNOWLEDGMENT

This study was supported by KAKENHI 17K00417. Dr. Ka-Lok Ngs work is supported by the Ministry of Science and Technology (MOST) under the grants of MOST 106-2221-E-468-017, MOST 106-2632-E-468-002, and also supported by the grant from Asia University, 106-asia-06.

REFERENCES

[1] D. H. K. Lim and E. R. Maher, "DNA methylation: a form of epigenetic control of gene expression," *The Obstetrician & Gynaecologist*, vol. 12, no. 1, pp. 37–42, jan 2010.

[2] X. Dong and Z. Weng, "The correlation between histone modifications and gene expression," *Epigenomics*, vol. 5, no. 2, pp. 113–116, apr 2013.

[3] L. M. Almassalha, A. Tiwari, P. T. Ruhoff, Y. Stypula-Cyrus, L. Cherkezyan, H. Matsuda, M. A. D. Cruz, J. E. Chandler, C. White, C. Maneval, H. Subramanian, I. Szleifer, H. K. Roy, and V. Backman, "The global relationship between chromatin physical topology, fractal structure, and gene expression," *Scientific Reports*, vol. 7, p. 41061, jan 2017.

[4] C. Catalanotto, C. Cogoni, and G. Zardo, "MicroRNA in control of gene expression: An overview of nuclear functions," *International Journal of Molecular Sciences*, vol. 17, no. 10, p. 1712, oct 2016.

[5] F. Parodi, R. Carosio, M. Ragusa, C. D. Pietro, M. Maugeri, D. Barbagallo, F. Sallustio, G. Allemanni, M. P. Pistillo, I. Casciano, A. Forlani, F. P. Schena, M. Purrello, M. Romani, and B. Banelli, "Epigenetic dysregulation in neuroblastoma: A tale of miRNAs and DNA methylation," *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1859, no. 12, pp. 1502–1514, dec 2016.

[6] J. N. Weinstein, , E. A. Collisson, G. B. Mills, K. R. M. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, and J. M. Stuart, "The cancer genome atlas pan-cancer analysis project," *Nature Genetics*, vol. 45, no. 10, pp. 1113–1120, oct 2013.

[7] Y.-H. Taguchi, "Tensor decomposition-based unsupervised feature extraction can identify the universal nature of sequence-nonspecific off-target regulation of mRNA mediated by microRNA transfection," *Cells*, vol. 7, no. 6, p. 54, 2018. [Online]. Available: http://www.mdpi.com/2073-4409/7/6/54

[8] ——, "Tensor decomposition/principal component analysis based unsupervised feature extraction applied to brain gene expression and methylation profiles of social insects with multiple castes," *BMC Bioinformatics*, vol. 19, no. Suppl 4, p. 99, 2018.

[9] ——, "One-class differential expression analysis using tensor decomposition-based unsupervised feature extraction applied to integrated analysis of multiple omics data from 26 lung adenocarcinoma cell lines," in *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, Oct 2017, pp. 131–138.

[10] ——, "Tensor decomposition-based unsupervised feature extraction applied to matrix products for multi-view data processing," *PLoS ONE*, vol. 12, no. 8, p. e0183933, 2017.

[11] ——, "Identification of candidate drugs using tensor-decomposition-based unsupervised feature extraction in integrated analysis of gene expression between diseases and drugmatrix datasets," *Sci Rep*, vol. 7, no. 1, p. 13733, 2017.

[12] ——, "Tensor decomposition-based unsupervised feature extraction identifies candidate genes that induce post-traumatic stress disorder-mediated heart diseases," *BMC Med. Genomics*, vol. 10, no. Suppl 4, p. 67, 2017.

[13] ——, "Identification of candidate drugs for heart failure using tensor decomposition-based unsupervised feature extraction applied to integrated analysis of gene expression between heart failure and DrugMatrix datasets," in *Intelligent Computing Theories and Application*. Springer International Publishing, 2017, pp. 517–528.

[14] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, Jan 2000.

Table VII

OVERLAPS BETWEEN C6: ONCOGENIC SIGNATURES IN MSIGDB AND GENE SYMBOLS (TABLE V) ASSOCIATED WITH GENES IDENTIFIED BY TD-BASED UNSUPERVISED FE AS SHOWING DIFFERENTIAL PROMOTER METHYLATION BETWEEN NORMAL OVARIAN TISSUES AND TUMOR TISSUES. #G1 (K): THE NUMBER OF GENES IN EACH OVEREXPRESSED GENES SET. #G2 (K): OVERLAPS WITH GENES SELECTED BY TD-BASED UNSUPERVISED FE.

Gene Set Name	#G1 (K)	Description	# G2 (k)	k/K	p-value	FDR q-value
WNT_UP.V1_DN	170	Genes downregulated in C57MG cells (mammary epithelium) by overexpression of WNT1 [Gene ID=7471] gene.	9	0.0529	3.63E-07	4.58E-05
KRAS.600.LUNG.BREAST _UP.V1_UP	288	Genes upregulated in epithelial lung and breast cancer cell lines overexpressing an oncogenic form of KRAS [Gene ID=3845] gene.	11	0.0382	4.85E-07	4.58E-05
RPS14_DN.V1_UP	192	Genes upregulated in CD34+ hematopoietic progenitor cells after a knockdown of RPS14 [Gene ID=6208] by RNA interference (RNAi).	9	0.0469	1.01E-06	5.64E-05
VEGF_A_UP.V1_UP	196	Genes upregulated in HUVEC cells (endothelium) by treatment with VEGFA [Gene ID=7422].	9	0.0459	1.19E-06	5.64E-05
MEL18_DN.V1_UP	141	Genes upregulated in DAOY cells (medulloblastoma) upon a knockdown of PCGF2 [Gene ID=7703] by RNAi.	7	0.0496	1.13E-05	4.26E-04
KRAS.LUNG.BREAST _UP.V1_UP	145	Genes upregulated in epithelial lung and breast cancer cell lines overexpressing an oncogenic form of KRAS [Gene ID=3845] gene.	6	0.0414	1.32E-04	3.37E-03
KRAS.300_UP.V1_UP	146	Genes upregulated in four lineages of epithelial cell lines overexpressing an oncogenic form of KRAS [Gene ID=3845] gene.	6	0.0411	1.37E-04	3.37E-03
BMI1_DN.V1_UP	147	Genes upregulated in DAOY cells (medulloblastoma) upon a knockdown of BMI1 [Gene ID=648] by RNAi.	6	0.0408	1.43E-04	3.37E-03
KRAS.600_UP.V1_UP	287	Genes upregulated in four lineages of epithelial cell lines overexpressing an oncogenic form of KRAS [Gene ID=3845] gene.	8	0.0279	1.64E-04	3.44E-03
CAHOY_ASTROGLIAL	100	Genes upregulated in astroglia cells.	5	0.05	2.02E-04	3.82E-03
ATF2_UP.V1_DN	187	Genes downregulated in myometrial cells overexpressing ATF2 [Gene ID=1386] gene.	6	0.0321	5.19E-04	7.71E-03
CYCLIN_D1_UP.V1_UP	188	Genes upregulated in MCF-7 cells (breast cancer) overexpressing CCND1 [Gene ID=595] gene.	6	0.0319	5.33E-04	7.71E-03
SRC_UP.V1_UP	188	Genes upregulated in primary epithelial breast cancer cell culture overexpressing SRC [Gene ID=6714] gene.	6	0.0319	5.33E-04	7.71E-03
PTEN_DN.V1_UP	191	Genes upregulated upon a knockdown of PTEN [Gene ID=5728] by RNAi.	6	0.0314	5.80E-04	7.71E-03
MTOR_UP.N4.V1_DN	193	Genes downregulated in CEM-C1 cells (T-CLL) by rapamycin (sirolimus) [PubChem = 6610346], an mTOR pathway inhibitor.	6	0.0311	6.12E-04	7.71E-03
KRAS.LUNG_UP.V1_UP	141	Genes upregulated in epithelial lung cancer cell lines overexpressing an oncogenic form of KRAS [Gene ID=3845] gene.	5	0.0355	9.74E-04	1.15E-02
ALK_DN.V1_UP	145	Genes upregulated in DAOY cells (medulloblastoma) after a knockdown of ALK [Gene ID=238] by RNAi.	5	0.0345	1.10E-03	1.16E-02
BMI1_DN_MEL18 _DN.V1_UP	145	Genes upregulated in DAOY cells (medulloblastoma) upon a knockdown of BMI1 and PCGF2 [Gene ID=648, 7703] by RNAi.	5	0.0345	1.10E-03	1.16E-02
P53_DN.V2_UP	148	Genes upregulated in HEK293 cells (kidney fibroblasts) upon a knockdown of TP53 [Gene ID=7157] by RNAi.	5	0.0338	1.21E-03	1.20E-02
KRAS.50_UP.V1_UP	48	Genes upregulated in four lineages of epithelial cell lines overexpressing an oncogenic form of KRAS [Gene ID=3845] gene.	3	0.0625	2.14E-03	2.03E-02

Table VIII

THE TOP SIX GO MOLECULAR FUNCTIONAL ANNOTATIONS OF THE GENES.

Gene Set Name [# of Genes (K)]	# of Genes in Overlap (k)	p-value	FDR q-value
GO_CALCIUM_ION_BINDING [697]	46	3.12 E-36	2.81 E-33
GO_HORMONE_ACTIVITY [119]	9	1.63 E-8	7.32 E-6
GO_RECEPTOR_BINDING [1476]	27	2.60 E-8	7.80 E-6
GO_G_PROTEIN_COUPLED_RECEPTOR_BINDING [259]	11	1.62 E-7	3.65 E-5
GO_NEUROPEPTIDE_RECEPTOR_BINDING [29]	5	4.25 E-7	7.65 E-5
GO_PEPTIDE_HORMONE_RECEPTOR_BINDING [17]	4	1.72 E-6	2.58 E-4

[15] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995. [Online]. Available: http://www.jstor.org/stable/2346101

[16] I. S. Vlachos, K. Zagganas, M. D. Paraskevopoulou, G. Georgakilas, D. Karagkouni, T. Vergoulis, T. Dalamagas, and A. G. Hatzigeorgiou, "DIANA-miRPath v3.0: deciphering microRNA function with experimental support," *Nucleic Acids Research*, vol. 43, no. W1, pp. W460–W466, may 2015.

[17] S.-M. Ho, "Estrogen, progesterone and epithelial ovarian cancer," *Reproductive Biology and Endocrinology*, vol. 1, no. 1, p. 73, 2003.

[18] H. A. Risch, "Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone," *JNCI Journal of the National Cancer Institute*, vol. 90, no. 23, pp. 1774–1786, dec 1998.