Tri-Clustering Analysis for Dissecting Epigenetic Patterns Across Multiple Cancer Types

Yanglan Gan[1], Zhiyuan Dong[1], Xia Zhang[2], and Guobing Zou[2(✉)]

[1] School of Computer Science and Technology, Donghua University, Shanghai, China
[2] School of Computer Engineering and Science, Shanghai University, Shanghai, China
gbzou@shu.edu.cn

Abstract. Tumor cells not only harbor genetic and epigenetic alterations, but also are regulated by various epigenetic modifications. Identification of tumor epigenetic similarities across different cancer types is useful for the discovery of treatments that can be extended to different cancers. Nowadays, abundant epigenetic modification profiles have provided good opportunity to achieve this goal. Here, we proposed a tri-clustering approach for integrative pan-cancer epigenomic analysis, named TriPCE. We applied TriPCE to uncover epigenetic mode among seven cancer types. This approach can identify significant cross-cancer epigenetic modification similarities. The associated gene analysis demonstrates strong relevance with cancer development and reveals consistent tendency among cancer types.

Keywords: Tri-clustering · Epigenetic pattern · Pan-cancer

1 Introduction

Aberrant epigenetic modification is a critical factor involving human diseases [1]. Tumor cells usually exhibit epigenetic abnormalities and further routinely use epigenetic processes to ensure their escape from various treatments [2]. Epigenetic modification patterns that lead to the corresponding dysregulation in cancers have become a critical research issue of cancer studies [3, 4].

BLUEPRINT, TCGA and the International Cancer Genome Consortium have integrated many epigenetic maps in normal and cancerous tissues [5–7]. It is urgent to decipher cancer common epigenetic patterns. Because DNA methylation in cancer is addressed elsewhere [8, 9], we focus on covalent histone modifications in cancers. Previous works mainly focus on identifying combinatorial epigenetic states. CoSBI captures epigenetic patterns based on correlations of histone signals [10]. ChromHMM and HiHMM apply a HMM model to annotate genomic sequences by co-occurrence of multiple epigenetic marks [11, 12]. RFECS is developed based on random forests [13]. IDEAS jointly characterizes epigenetic landscapes in many cell types and detects differential regulatory regions [14]. These methods successfully identify combinatorial

D.-S. Huang et al. (Eds.): ICIC 2018, LNCS 10955, pp. 330–336, 2018.
https://doi.org/10.1007/978-3-319-95933-7_40

epigenetic patterns among different cell types. However, the correlations among different regions are still need to be investigated.

Here, we proposed a tri-clustering approach TriPCE for integrative pan-cancer epigenomic analysis. We applied TriPCE to various epigenomic maps of seven cancer types and identified significant cross-cancer epigenetic modification similarities. Furthermore, the associated gene function analysis demonstrates strong relevance with cancer development and reveals consistent tendency among cancer types.

2 Materials and Methods

We analyzed the epigenomic maps of seven cancer types, including A549, K562, HepG2, HCT116, Hela-S3, multiple myeloma-Cell Line, sporadic Burkitt lymphoma-Cell Line. Totally, we obtained 42 datasets of six epigenetic modifications, including H3K4me1, H3K4me3, H3K9me3, H3K27ac, H3K27me3 and H3K36me3. RNA expression profiles of the seven cancer types were also collected. These dataset were downloaded from the website of NIH Roadmap Epigenome Project.

As shown in Fig. 1, the TriPCE model has three key components.

Step1. Preprocess the epigenetic modification data of different cancer types. Firstly, the genome was represented as consecutive genomic segments with size 200 bps. For each epigenetic mark, we computed the summary tag count of every segment. To remove noise, raw read counts were normalized by the total number of reads followed by arcsine transformation [15]. Further, the epigenetic profiles in the promoter regions were extracted. Then, for each epigenetic mark, the epigenetic profiles of different cancer types were represented as a matrix $E_{k,}$ where k is the index of the epigenetic mark ranging from 1 to K.

Step2. Identify BiClusters based on FP-growth algorithm for each epigenetic mark. We computed correlation coefficients of any two cancer types at every region and obtained a coefficient matrix. If the coefficient is higher than a given threshold, the epigenetic modifications of these two cancer types are regarded as coherent. Then we added the cancer type to the itemset. Based on the resulted itemset, we identified coherent epigenetic patterns using FP-growth algorithm. FP-growth is a data mining method that was originally developed for frequent itemset mining. Further, we inversely identified the corresponding gene set and determined the BiClusters.

Step3. Mine TriClusters with coherent epigenetic patterns across different cancers. Based on the BiClusters of each epigenetic mark, we enumerated the subsets of these epigenetic marks to obtain TriClusters. Each TriCluster represents as a gene set with similar epigenetic changes in different cancer types, which indicate conserved epigenetic signatures that shared by multiple cancer types.

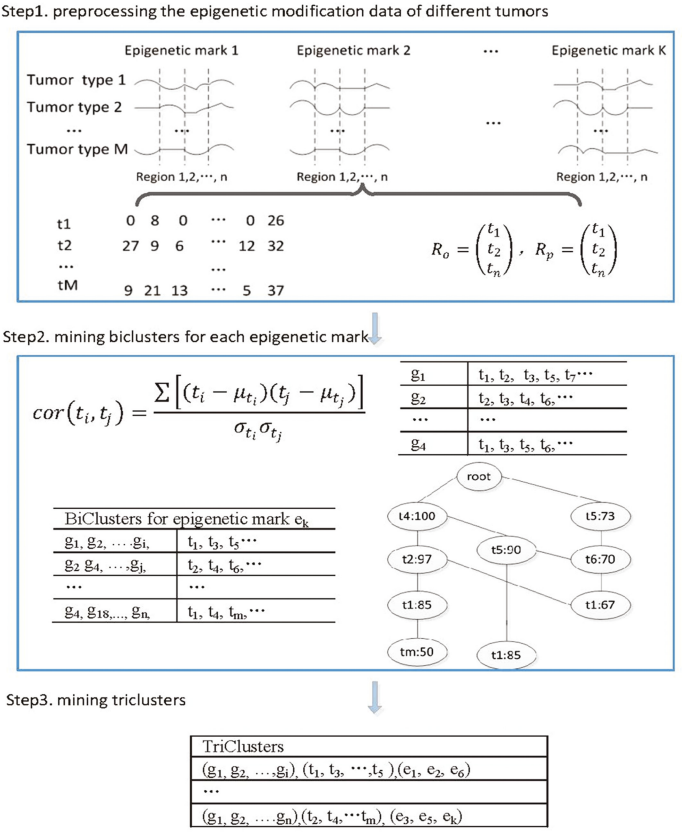

Fig. 1. The flowchart of the proposed TriPCE approach.

3 Results

3.1 Identifying Similar Epigenetic Patterns Across Different Cancer Types

We developed a tri-clustering approach, TriPCE, to capture similar epigenetic patterns among different cancer types. For each epigenetic mark, TriPCE first groups the regions based on the epigenetic modification profiles among different cancer types. Figure 2 shows a typical BiCluster of epigenetic mark H3K4me1, a gene set with similar modification pattern in cancer type Hela-S3, HepG2, K562 and A549. From this figure, we notice that the epigenetic profiles of these genes are similar in these cell types. Meanwhile, different cancers share similar epigenetic patterns. For examples, cancers (HepG2 and HCT116) are clustered together and share larger number of epigenetic marks, implying that they share more similar epigenetic regulation mechanisms. To get significant modification patterns, we set the minima support as 10% of the investigated genes. With diverse correlation coefficient thresholds, we respectively

gained different numbers of BiClusters for the epigenetic marks. Among these epigenetic marks, H3K4me3 and H3K9me3 vary most. On the contrary, there are more similar epigenetic patterns of H3K4me1 and H3k27me3. This result is consistent with previous finding that H3K9me3/me2 and H3K36me3/me2 frequently observed in breast cancer [16], esophageal cancer [17] and MALT lymphoma [18]. As the threshold slightly affects the trend among different epigenetic marks, we chose the BiClusters with threshold 0.7 for further analysis.

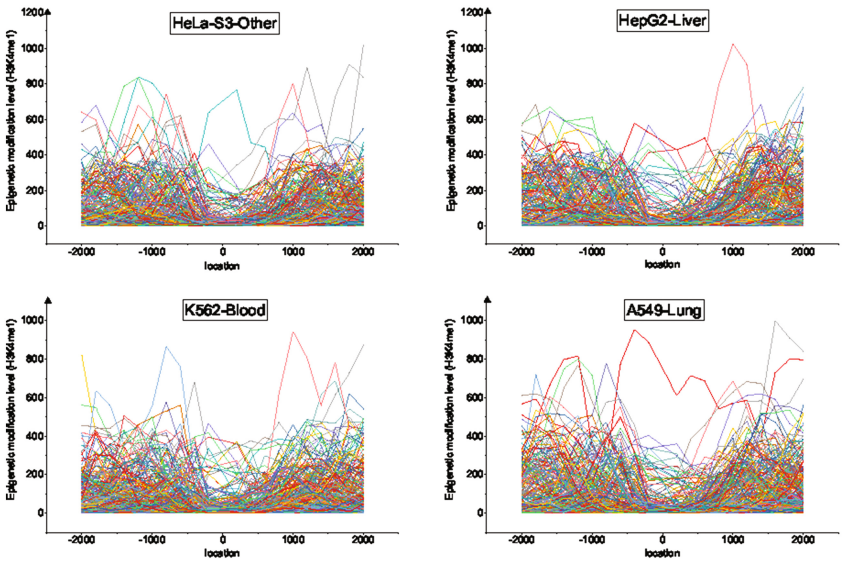

Fig. 2. Profiles of epigenetic modification H3K4me3 in a typical BiCluster display a similar pattern in four cancer types, including Hela-S3, HepG2, K562 and A549.

3.2 Identifying Coherent Patterns Among Different Epigenetic Marks

To identify conserved epigenetic states, we further clustered epigenetic marks based on the identified BiClusters. The TriClusters are represented as triples ('genomic regions', 'tumor types', 'epigenetic marks'). Initially, we obtained 175 TriClusters. Figure 3 shows the epigenetic marks, cancer types and supports of 15 typical clusters. There exist coherent epigenetic states across different cancers types. For example, the variation pattern of H3K4me1, H3K9me3, H3kK27me3 and H3K36me3 is shared in A549, HepG2 and K562. On the contrary, there are some epigenetic patterns are only coherent in certain cell types. We observed similar patterns of H3K36me3, H3K27ac and H3K27me3 among HepG2 and sporadic Burkitt lymphoma-Cell Line.

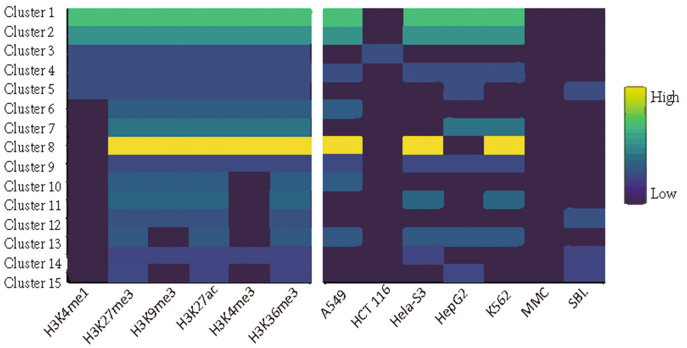

Fig. 3. Typical epigenetic TriClusters. (A) The epigenetic marks (column) in each cluster (row). (B) The cancer types (column) in each cluster (row).

3.3 Analyzing the Potential Roles of Associated Genes

To examine the potential functions of these genes, we performed systematic gene ontology enrichment analysis using DAVID tools. Overall, we found that the TriClusters enriched genes exhibited enrichment for cancer-related functions. Table 1 lists the result of a typical TriCluster (P value < 0.05). In this TriCluster, these genes exhibit coherent epigenetic pattern of H3K4me1, H3K4me3, H3K9me3, H3K27ac, H3K27me3 for HeLa-S3, HepG2, Multiple myeloma-Cell Line and Sporadic Burkitt lymphoma-Cell Line. In the table, term 'positive regulation of cell proliferation' and 'negative regulation of apoptotic process' are enriched in these gene sets. This result implies that the identified gene sets in the TriCluster are essential for cell proliferation and apoptotic process. Meanwhile, term 'negative regulation of gene transcription' is also enriched in the gene set, indicating these genes perform important regulation role in these cancers.

Table 1. Functional enrichment of genes in the identified TriClusters.

Term type	Term name	P-value	Term type	Term name	P-value
BP	positive regulation of cell proliferation	2.84E − 06	MF	glutathione binding	7.85E − 04
BP	protein targeting to Golgi	8.87E − 05	MF	glutathione transferase activity	8.00E − 03
BP	nitrobenzene metabolic process	1.14E − 04	MF	histone binding	1.16E − 02
BP	xenobiotic catabolic process	1.00E − 03	MF	peptidyl-prolyl cis-trans isomerase activity	1.35E − 02
BP	negative regulation of gene expression, epigenetic	1.39E − 03	MF	protein heterodimerization activity	3.32E − 02
BP	negative regulation of apoptotic process	1.88E − 03	CC	extracellular exosome	1.13E − 02

4 Discussion

Identifying epigenetic pattern is important to understand epigenetic mechanisms in various cancers. Our knowledge about the patterns of epigenetic modification and the cause and consequence of them are still limited. Computational approach that exploits the complex epigenomic landscapes and discovers significant signatures out of them are required. In this paper, we developed a tri-clustering approach for integrative pan-cancer epigenomic analysis, named TriPCE. We applied TriPCE to uncover epigenetic patterns of six epigenetic marks among seven cancer types. This approach identifies significant cross-cancer epigenetic modification similarities. The associated gene analysis demonstrates strong relevance with cancer development and reveals consistent tendency among cancer types. Different from existing methods, our approach enable researchers to explore the epigenetic patterns among different cancer types as well as the combinational mode of multiple epigenetic marks.

Acknowledgment. This work was supported in part by the Fundamental Research Funds for the Central Universities (2232016A3-05), the National Natural Science Foundation of China (61772128), and Shanghai Natural Science Foundation (17ZR1400200).

References

1. Jones, P.A., Issa, J.P.J., Baylin, S.: Targeting the cancer epigenome for therapy. Nat. Rev. Genet. **17**(10), 630–641 (2016)
2. You, J.S., Jones, P.A.: Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell **22**(1), 9 (2012)
3. Dawson, M.A.: The cancer epigenome: concepts, challenges, and therapeutic opportunities. Science **355**(6330), 1147–1152 (2017)
4. Kelly, A.D., Issa, J.P.J.: The promise of epigenetic therapy: reprogramming the cancer epigenome. Curr. Opin. Genet. Dev. **42**, 68–77 (2017)
5. Kundaje, A., et al.: Integrative analysis of 111 reference human epigenomes. Nature **518** (7539), 317–330 (2015)
6. Weinstein, J.N., et al.: The cancer genome atlas pan-cancer analysis project. Nat. Genet. **45**(10), 1113–1120 (2015)
7. Beck, S., et al.: A blueprint for an international cancer epigenome consortium. a report from the AACR cancer epigenome task force. Can. Res. **72**(24), 6319–6324 (2012)
8. Kretzmer, H., et al.: Dna-methylome analysis in burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. Nat. Genet. **47**(11), 1316–1325 (2015)
9. Yang, X., et al.: Comparative pan-cancer dna methylation analysis reveals cancer common and specific patterns. Brief. Bioinform. **18**(5), 761 (2016)
10. Ucar, D., Hu, Q., Tan, K.: Combinatorial chromatin modification patterns in the human genome revealed by subspace clustering. Nucleic Acids Res. **39**(10), 4063–4075 (2011)
11. Ernst, J., et al., Coyne, M., et al.: Mapping and analysis of chromatin state dynamics in nine human cell types. Nature **473**(7345), 43 (2011)
12. Sohn, K.A., et al.: hiHMM: Bayesian non-parametric joint inference of chromatin state maps. Bioinformatics **31**(13), 2066–2074 (2015)

13. Rajagopal, N., Xie, W., Li, Y., Wagner, U., Wang, W., Stamatoyannopoulos, J., Ernst, J., Kellis, M., Ren, B.: RFECS: a random-forest based algorithm for enhancer identification from chromatin state. PLoS Comput. Biol. **9**(3), e1002968 (2013)
14. Zhang, Y., et al.: Jointly characterizing epigenetic dynamics across multiple human cell types. Nucleic Acids Res. **44**(14), 6721–6731 (2016)
15. Pinello, L., et al.: Analysis of chromatin-state plasticity identifies cell-type-specific regulators of H3K27me3 patterns. PNAS **111**(3), E344 (2014)
16. Liu, G., et al.: Genomic amplification and oncogenic properties of the GASC1 histone demethylase gene in breast cancer. Oncogene **28**(50), 4491 (2009)
17. Yang, Z.Q., et al.: Identification of a novel gene, GASC1, within an amplicon at frequently detected in esophageal cancer cell lines. Can. Res. **60**(17), 4735–4739 (2000)
18. Vinatzer, U., et al.: Mucosa-associated lymphoid tissue lymphoma: novel translocations including rearrangements of ODZ2, JMJD2C, and CNN3. Clin. Cancer Res. **14**(20), 6426–6431 (2008)