Identification and characteristic analysis of enhancers across 13 major cancer types

Mingming Qian*, Wenzhu Wang*, Yana Zhang, Yi Zhao, Huige Quan, Yuting Chen, Xinyue Dai, Zhiyun Guo✉

✉ Correspondence: zhiyunguo@swjtu.edu.cn

School of Life Sciences and Engineering, Southwest Jiaotong University, Chengdu 610031, China

* These authors have contributed equally to this work.

Abstract

Enhancers are often mutated and dysregulated in various diseases such as cancer. By integrating the FANTOM enhancers expression profiles and RNA-seq data from TCGA of 13 cancers and their corresponding para-cancerous tissues, we systematically identified a total of 4702 significantly differentially expressed enhancers (DE enhancers). Furthermore, a total of 1036 differentially expressed genes (DE genes) regulated by differentially expressed enhancers (DE enhancers) were identified. It was found that in these 13 cancers, most (61.13%) enhancers were ubiquitously expressed, whereas DE enhancers were more likely to be tissue-specific expressed, and the DE genes regulated by DE enhancers were significantly enriched in cancer-related pathways. Finally, it was manifested that 74 SNPs located in 37 DE enhancers, and these SNPs affected the gain and loss of functional transcription factor binding sites (TFBS) of 758 transcription factors, which had been shown to be highly correlated with tumorigenesis and development.

Keywords Cancer; TCGA; Enhancers; Transcription factors; SNPs
Introduction

Enhancers is a class of cis-acting elements in the genome of eukaryotes, which can positively regulate the gene expression and the dysregulation of enhancer can cause a variety of diseases such as cancer (Herz 2016, Kang et al. 2021). Transcriptomics and genomics studies have found that active enhancer can be used as a transcription unit to transcribe eRNA (enhancer RNA) (Liu 2017). In cell signaling and transcriptional regulation, the expression level of eRNA is highly correlated with the activity of functional enhancer (Li et al. 2016). In addition, mutations in the enhancers may affect the gain and loss of transcription factor binding sites (TFBS) on the enhancer (Pluta et al. 2018). These studies suggest that the enhancer mutations may be involved in biological processes and signaling pathways of cancer.

In order to analyze and identify the relationship between dysregulated enhancers and tumorigenesis, we identified a lot of differentially expressed enhancers (DE enhancers) in these 13 cancers and matched para-cancerous tissues of TCGA by using multi-omics approaches. Then, lots of differentially expressed genes (DE genes) which were co-expression regulated by DE enhancers in cancers were obtained. Finally, a series of SNPs which may affect the gain and loss of TFBS were identified, and TFs located in these TFBS are highly correlated with cancer, suggesting that they could be used as potential targets for future cancer drugs.

Materials and Methods

Identification of Differentially Expressed Enhancers and Genes

The enhancer region (hg19) and expression profile data (RPKM) of 13 cancers and matched para-cancerous tissues (5727 cancer samples and 619 normal samples) were obtained from a previous study (Chen et al. 2018). The 65423 FANTOM enhancers capable of transcribing eRNA were re-annotated. They then removed those enhancers that overlapped with known genes or intron regions. RNA-Seq data of 13 cancers and para-cancerous tissues (613 cancer
samples and 609 normal samples) were downloaded from TCGA (https://portal.gdc.cancer.gov/). The enhancers/genes with expression in at least 10% of cancers/normal tissue samples were identified as candidate enhancers/genes. Significantly differentially expressed enhancers/genes between cancers and para-cancerous tissues were determined according to the following criterion ($|\log_{2}\text{FC}| \geq 1$, q-value<0.05, t-test). In the same cancer, for each enhancer (gene as well), the average of its expression across samples was used to calculate log2FC and t-test p-values.

**Identification of DE Enhancers-DE Gene Interactions**

The TSSs of genes were retrieved by processing GENCODE annotations v19, and the 0.5 kb upstream and 1 kb downstream from TSS were used as promoter regions (Abugessaisa et al. 2017). If a gene promoter intersected with the region 100 kb upstream and downstream from the enhancer center, the gene was considered as a candidate target gene of the enhancer (Gaffney et al. 2019). Considering that the expression of both enhancers and genes differed in different tissues, we processed the RNA-seq data from different samples separately. The Spearman correlation between enhancer and gene expression was calculated and screened with $\text{corr} \geq 0.3$ and q-value<0.05 as the threshold.

**KEGG Enrichment Analysis of Target Genes**

In order to further understand the biological mechanism of the DE enhancer-DE gene interactions, KEGG functional enrichment analysis of the target genes of DE enhancers was performed by using “clusterProfiler” package (Yu et al. 2012) (R 3.6.0). The threshold of a statistically significant difference was q-value<0.05.

**Identification of Enhancers SNPs and Analysis of TFBS**

The SNPs data (GRCh38) of 13 TCGA cancers was obtained from the ncRNA-eQTLs database (Li et al. 2020), including SNP loci, rs ID, etc. First, the enhancer region was changed from hg19 to GRCh38 by using LiftOver tool of UCSC (Zweig et al. 2008). Then, the enhancer region was extended to 0.5kb upstream and 0.5kb downstream from boundary of enhancer, and the cancer SNPs loci were mapped to DE enhancers region to obtain enhancer SNPs. By using atSNP tool (Shin et al. 2019), the TFs affected by SNPs were identified, and the effects of SNPs on TFBS were analyzed. GEPIA (Tang et al. 2017) was used to identify
the survival curve of genes in cancer, with q-value<0.05 as the threshold.

**Results and Discussion**

**Differential Expression Analysis of Enhancers in 13 Cancers**

The expression profiles of the enhancers were analyzed in 13 cancers and matched para-cancerous tissues, yielding in 10,639 and 10,039 enhancers respectively. Based on the expression in 13 cancers and para-cancerous tissues, the enhancers were divided into three categories: tissue-specific enhancers (enhancers expressed in only one tissue, see blue block labelled 1 on Fig. 1a), ubiquitously expressed enhancers (enhancers expressed in all tissues, see green block labelled 13 on Fig. 1a), and other enhancers (see colour blocks labelled 2-12 on Fig. 1a). Statistics analysis on the types of all enhancers revealed that the distribution of enhancers was obviously tissue-specific and ubiquitously expressed, whether in the cancers or para-cancerous tissues (Fig. 1a). In a certain type of cancer or normal tissue, the ubiquitously expressed and tissue-specific enhancers occupied the largest and smallest proportion, respectively (Supplementary Figure S1). These results might indicate that most enhancers participated in the tissue regulatory network as widely regulatory factors, apart from a few tissue-specific enhancers which were responsible for tissue identification and development.
**Fig. 1** Differential expression analysis of enhancers in 13 cancers. (a) The distribution of enhancers across 13 cancers or para-cancerous tissues. The numbers and color blocks represent the proportion of all enhancers expressed in different numbers of cancers or para-cancerous tissues. (b) KEGG analysis of differentially expressed genes which were participating in the co-expression regulation with differentially expressed enhancers. The grey bars indicate the \( -\log_{10}(P \text{ value}) \) of these KEGG pathways. (c) Signature difference between regulated DE enhancers and non-regulated DE enhancers. Comparison of the expression level of regulated DE enhancers and non-regulated DE enhancers in 13 types of cancers, \( p\text{-value}<0.05 \).

**Target Genes of DE Enhancers are Significantly associated with Cancer**

Based on differential expression, a total of 4,702 significantly differentially expressed enhancers (DE enhancers) were identified in these 13 cancers (\( |\log_2 FC| \geq 1 \), \( q\text{-value}<0.05 \)), among which 1,646 and 1,589 enhancers were differentially up-regulated or down-regulated respectively (Supplementary Table S1). The function of enhancers is mainly achieved by
regulating the downstream target genes. Finally, a total of 14,461 DE genes were obtained, among which 5,357 and 2,933 enhancers were differentially up-regulated or down-regulated respectively (Supplementary Table S1). Although Hi-C data had been used on identification of enhancer-target genes, however, due to the limited available Hi-C data and low resolution, we identified the enhancer targets refer to previous study (Chen et al. 2018). By integrating distance analysis (within the enhancer center ± 100kb region) and co-expression analysis (Spearman correlation corr≥0.3 and q-value<0.05), a total of 1,036 DE genes were identified in these 13 cancers, these genes are potentially subject to regulation by DE enhancers. (Supplementary Table S2). We divide DE enhancers into two categories: DE enhancers that regulates DE genes (regulated DE enhancers) and DE enhancers that does not regulate DE genes (non-regulated DE enhancers). It was found that most (3662 out of 4540/80.66%) of the DE enhancers that regulated DE genes were ubiquitously expressed (Supplementary Figure S2). This result suggested that dysregulated enhancers were mainly involved in the fundamental biological process of tumors. In order to explore the association between cancer and target genes of DE enhancers, KEGG enrichment analysis was performed on the target genes. The DE genes that regulated by DE enhancers were significantly enriched in cancer-related pathways such as hippo signaling pathway, proteoglycans cancer and other cancer-related KEGG pathways (Fig. 1b), indicating that these DE enhancers may also play an important role in cancer. We also compared the enrichment pathways of DE genes that are not regulated by DE enhancers (Supplementary Figure S4), and we could see that they are enriched in pathways that are almost unrelated to cancer. We next investigated the expression levels of target genes regulated by the DE enhancer and those not regulated by the DE enhancer, found that the expression level of regulated DE enhancers was significantly higher than that of non-regulated DE enhancers in these 13 cancers (Fig. 1c, p-value<0.05).

The Identification of DE Enhancers Sequence Variants and Transcription-Factor Dysregulation affected by these Variants

Previous studies have proved that mutations in enhancers are highly related to the development of cancer, and SNPs could participate in the regulation of multiple biological
processes and trigger multiple diseases by affecting the gain and loss of TFBS (Fan et al. 2013). To further investigate this, we identified a total of 74 SNPs that were located in 37 regulated DE enhancers in 10 cancers through extracting ncRNA-eQTL database (Supplementary Table S3). By using atSNP tool, 758 TFs which had gain or loss of functional TFBS influenced by these SNPs were identified, and we found that most of these TFs were well-known cancer-related genes such as MYC, EP300 and REST (Fig. 2a, Supplementary Table S4). In particular, as a histone acetyltransferase, EP300 regulates transcription by chromatin remodeling and plays an important role in cell proliferation, transformation and differentiation. It has become a key transcription factor for identifying enhancers (Asaduzzaman et al. 2017). These results indicate that mutations in enhancers could interfere with the role of enhancers in cancer by affecting the binding of cancer-related TFs to enhancers.

We noticed that the highest proportion of SNPs (28.4%) were found in kidney renal clear cell carcinoma (KIRC) compared with other cancers. A total of 21 SNPs were identified in 15 enhancers, which affected 511 TFs (Supplementary Table S3). The enhancer located in chr20:59164329-59165752 contained five SNPs (rs73306874, rs6026739, rs6026740, rs6026742, rs73306876) that affected 241 TFs gaining or losing TF binding sites on this enhancer (Fig. 2b). The top three of TFs most affected by SNPs were SOX9 (effected by SNP rs73306874, rs6026739, rs6026740, rs73306876), POU2F1 (effected by rs73306874, rs6026739, rs73306876), and FOXO3 (effected by rs73306874, rs6026740, rs73306876). Survival analysis showed that the expression of POU2F1 and FOXO3 was significantly related to the survival time of patients with KIRC (Fig. 2c). Although the correlation between SOX9 and the survival time of KIRC patients was not significant, a previous study has shown that SOX9 could inhibit cell proliferation and invasion of renal cell carcinoma, in the way of being targeted by microRNA-138 (Hu et al. 2017). In summary, these SNPs influence the survival time of cancer patients by affecting TF binding to enhancers and can be used as potential enhancer-targeted drugs target sites.
Fig. 2 The gain and loss of functional TFBS and survival analysis of TFs. (a) The top 50 TFs that most affected by the SNPs which may relate to enhancer activity. The larger the logo of TF name, the more SNPs influencing this TF. (b) The gain and loss of functional TFBS of 241 TFs influenced by SNPs located in enhancer chr20: 59164329-59165752. (c) The survival analysis of the transcription factors POU2F1 and FOXO3, q-value<0.05

Conclusion

In summary, we integrated RNA-seq data from FANTOM enhancer expression profiles and TCGA to systematically identify a total of 4,702 significantly differentially expressed enhancers in 13 cancers and matched para-cancerous tissues. In addition, a total of 1,036 differentially expressed genes (DE genes) regulated by differentially expressed enhancers (DE enhancers) were identified by integration distance and co-expression analysis. We found that
DE enhancers were more likely to be tissue-specific in their expression and that DE genes regulated by DE enhancers were significantly enriched in cancer-related pathways. Finally, the results showed that 74 SNPs located in 37 DE enhancers and affected the gain and loss of functional TFBS of 758 transcription factors, which had been proved to be highly related to the occurrence and development of tumors. Taken together, these results provide informative data and methods of dysregulated enhancers mutations for future research in cancer treatment.

Authors’ contributions

Zhiyun Guo designed the research. Mingming Qian, Wenzhu Wang, Yi Zhao, Huige Quan, Yuting Chen, Xinyue Dai analyzed the data. Zhiyun Guo and Wenzhu Wang helped to draft the manuscript. All authors read and approved the final manuscript.

Author Disclosure Statement

We declare we have no competing interests.

Acknowledgement

This work was supported by Basic Research Cultivation Support Program of Fundamental Research Funds for the Central Universities (2682021ZTPY016) and National Science and Technology Major Project of Infectious Diseases (R111320H01004).

References

Asaduzzaman M et al. (2017) Tumour suppressor EP300, a modulator of paclitaxel resistance and stemness, is downregulated in metaplastic breast cancer Breast Cancer Res Treat 163:461-474 doi:10.1007/s10549-017-4202-z

Fan J, Dy JG, Chang CC, Zhou X (2013) Identification of SNP-containing regulatory motifs in the myelodysplastic syndromes model using SNP arrays and gene expression arrays Chin J Cancer 32:170-185 doi:10.5732/cjc.012.10113

Chen H, Li C, Peng X, Zhou Z, Weinstein JN, Liang H (2018) A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples Cell 173:386-399.e312 doi:10.1016/j.cell.2018.03.027
Herz HM (2016) Enhancer deregulation in cancer and other diseases Bioessays 38:1003-1015 doi:10.1002/bies.201600106

Hu B, Wang J, Jin X (2017) MicroRNA-138 suppresses cell proliferation and invasion of renal cell carcinoma by directly targeting SOX9 Oncology letters 14:7583-7588 doi:10.3892/ol.2017.7160

Li J et al. (2020) ncRNA-eQTL: a database to systematically evaluate the effects of SNPs on non-coding RNA expression across cancer types Nucleic acids research 48:D956-d963 doi:10.1093/nar/gkz711

Li W, Notani D, Rosenfeld MG (2016) Enhancers as non-coding RNA transcription units: recent insights and future perspectives Nat Rev Genet 17:207-223 doi:10.1038/nrg.2016.4

Liu F (2017) Enhancer-derived RNA: A Primer Genomics Proteomics Bioinformatics 15:196-200 doi:10.1016/j.gpb.2016.12.006

Pluta A, Rola-Luszczak M, Douville RN, Kuzmak J (2018) Bovine leukemia virus long terminal repeat variability: identification of single nucleotide polymorphisms in regulatory sequences Virol J 15:165 doi:10.1186/s12985-018-1062-z

Shin S, Hudson R, Harrison C, Craven M, Keles S (2019) atSNP Search: a web resource for statistically evaluating influence of human genetic variation on transcription factor binding Bioinformatics 35:2657-2659 doi:10.1093/bioinformatics/bty1010

Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses Nucleic acids research 45:W98-w102 doi:10.1093/nar/gkx247

Xu TP et al. (2017) E2F1 induces TINCR transcriptional activity and accelerates gastric cancer progression via activation of TINCR/STAU1/CDKN2B signaling axis Cell Death Dis 8:e2837 doi:10.1038/cddis.2017.205

Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters Omics : a journal of integrative biology 16:284-287 doi:10.1089/omi.2011.0118

Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ (2008) UCSC genome browser tutorial Genomics 92:75-84 doi:10.1016/j.ygeno.2008.02.003

Abugessaisa, Imad, Shuhei Noguchi, Piero Carninci, and Takeya Kasukawa. 2017. ‘The FANTOM5 Computation Ecosystem: Genomic Information Hub for Promoters and Active Enhancers’. Methods in Molecular Biology (Clifton, N.J.) 1611: 199–217. https://doi.org/10.1007/978-1-4939-7015-5_15.

Gaffney, Daniel J. 2019. ‘Mapping and Predicting Gene-Enhancer Interactions’. Nature Genetics 51 (12): 1662–63. https://doi.org/10.1038/s41588-019-0540-6.

Kang, Ran, Zhentang Tan, Mei Lang, Linqi Jin, Yin Zhang, Yiming Zhang, Tailin Guo, and Zhiyun Guo. n.d. ‘EnhFFL: A Database of Enhancer Mediated Feed-Forward Loops for Human and Mouse’, 7.

Appendices

Supplementary Figures and Tables are available in the supplementary materials for review.