



OPEN ACCESS

EDITED BY

Xuefeng Cui,
Shandong University, China

REVIEWED BY

Advait Balaji,
Occidental Petroleum Corporation,
United States
Zhi-Ping Liu,
Shandong University, China
Shihua Zhang,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE

Xingli Guo,
✉ xlguo@mail.xidian.edu.cn

RECEIVED 12 December 2023

ACCEPTED 30 January 2024

PUBLISHED 23 February 2024

CITATION

Wang Y, Guo X, Niu Z, Huang X, Wang B and
Gao L (2024), DeepCBS: shedding light on the
impact of mutations occurring at CTCF
binding sites.

Front. Genet. 15:1354208.

doi: 10.3389/fgene.2024.1354208

COPYRIGHT

© 2024 Wang, Guo, Niu, Huang, Wang and Gao.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

DeepCBS: shedding light on the impact of mutations occurring at CTCF binding sites

Yiheng Wang, Xingli Guo*, Zhixin Niu, Xiaotai Huang,
Bingbo Wang and Lin Gao

School of Computer Science and Technology, Xidian University, Xi'an, China

CTCF-mediated chromatin loops create insulated neighborhoods that constrain promoter-enhancer interactions, serving as a unit of gene regulation. Disruption of the CTCF binding sites (CBS) will lead to the destruction of insulated neighborhoods, which in turn can cause dysregulation of the contained genes. In a recent study, it is found that CTCF/cohesin binding sites are a major mutational hotspot in the cancer genome. Mutations can affect CTCF binding, causing the disruption of insulated neighborhoods. And our analysis reveals a significant enrichment of well-known proto-oncogenes in insulated neighborhoods with mutations specifically occurring in anchor regions. It can be assumed that some mutations disrupt CTCF binding, leading to the disruption of insulated neighborhoods and subsequent activation of proto-oncogenes within these insulated neighborhoods. To explore the consequences of such mutations, we develop DeepCBS, a computational tool capable of analyzing mutations at CTCF binding sites, predicting their influence on insulated neighborhoods, and investigating the potential activation of proto-oncogenes. Furthermore, DeepCBS is applied to somatic mutation data of liver cancer. As a result, 87 mutations that disrupt CTCF binding sites are identified, which leads to the identification of 237 disrupted insulated neighborhoods containing a total of 135 genes. Integrative analysis of gene expression differences in liver cancer further highlights three genes: ARHGEF39, UBE2C and DQX1. Among them, ARHGEF39 and UBE2C have been reported in the literature as potential oncogenes involved in the development of liver cancer. The results indicate that DQX1 may be a potential oncogene in liver cancer and may contribute to tumor immune escape. In conclusion, DeepCBS is a promising method to analyze impacts of mutations occurring at CTCF binding sites on the insulator function of CTCF, with potential extensions to shed light on the effects of mutations on other functions of CTCF.

KEYWORDS

CTCF, insulated neighborhoods, liver cancer, proto-oncogene, deep learning

1 Introduction

3D genomics is a rapidly growing field that investigates the complex folding and organization of chromosomes in eukaryotic cells. Various techniques have been developed to study the 3D structure of chromosomes, such as 3C (Dekker et al., 2002), Hi-C (Lieberman-Aiden et al., 2009), ChIA-PET (Fullwood et al., 2009), 4C (Simonis et al., 2009), and 5C (Dostie et al., 2006). The three-dimensional structure of chromosomes includes chromosomal domains, chromosome compartments,

topologically associated domains, insulator regions, and promoter-enhancer loops. In 3D genomics, insulated neighborhood is defined as a CTCF-CTCF homodimer that binds with cohesions and contains at least one gene's chromatin loop (Dowen et al., 2014). Miao have observed that this chromatin loop serves as a unit of gene regulation (Yu and Ren, 2017). When the CTCF binding site is disrupted, improper enhancer-promoter interactions can lead to the dysregulation of local genes (Hnisz et al., 2016b).

Recently, a study has revealed that mutations in CTCF binding sites occur frequently in cancer (Katainen et al., 2015). And CTCF/cohesin binding sites are a major mutational hotspot in the cancer genome. Some of these mutations can cause a decrease in CTCF binding, leading to the disappearance of insulated neighborhoods (Hnisz et al., 2016b; Umer et al., 2016). Non-coding mutations at CTCF binding sites have the potential to disrupt insulated neighborhoods, leading to altered gene expression within these regions. This, in turn, could potentially contribute to the development of diseases. Previous research has indicated that there are typically silent proto-oncogenes within insulated neighborhoods. Additionally, the anchoring regions of insulated neighborhoods containing proto-oncogenes undergo frequent somatic mutations in various types of cancer (Hnisz et al., 2016a). In summary, variations in CTCF binding sites in cancer may lead to the disappearance of insulated neighborhoods and the activation of oncogenes, ultimately promoting the development of cancer. So the identification of variants that have the potential to disrupt insulated neighborhood is a critical task. A few studies focus on this task. Zhang et al. has proposed a method, named CTCF-MP, to predict chromatin loops. This method utilizes a machine learning model based on word2vec and boosted trees (Zhang et al., 2018). CTCF-MP algorithm incorporates sequence variations caused by mutations and enables prediction of the influence of such mutations on the formation of chromatin loops. Sequence-based deep learning methods have shown great potential in predicting the impact of genetic variants on insulated neighborhoods. When provided with a pair of DNA sequences of anchors, this model generates a value ranging from 0 to 1, which can be used to determine the probability or strength of the chromatin loop (Zhang et al., 2018). DeepCTCFLoop takes a pair of DNA sequence containing CTCF motifs with flanking regions and encodes it into one-hot encoding as input, uses a neural network to predict whether this pair of sequences can form a DNA loop (Kuang and Wang, 2021). DeepMILO, a deep learning framework, utilizes one-hot encoding to represent DNA sequences, comprises of an anchor model and an anchor orientation model. It accurately predicts the effects of variants on CTCF/cohesion mediated insulator loops and reveals a novel mechanism for oncogene dysregulation in malignant lymphoma (Trieu et al., 2020).

However, CTCF is a multifunctional protein, associated with a number of vital cellular processes such as transcriptional activation, repression, insulation, imprinting and genome organization (Oh et al., 2017). CTCF not only regulates gene expression by forming loops but also can independently regulate gene expression.

While the discussed methods, such as DeepMILO, have certain limitations as they require paired data, making them effective in predicting the impact of mutations occurring at CTCF binding sites on their insulator function. These methods are not capable of predicting the effects of mutations at CTCF binding sites on other function.

Both DeepMILO and DeepCTCFLoop utilize the one-hot encoding method. One-hot encoding treats each position in the sequence as an independent feature, disregarding the sequential relationships between adjacent nucleotides. However, biological sequences often contain important sequence patterns or motifs that play a critical role in the functionality or structure of the sequence. In comparison, using only one-hot encoding may not fully capture the information conveyed by these patterns. In a recent study, a novel method named dna2vec, has been proposed. This method leverages the human genome sequences as the learning corpus and embeds k-mers into a 100-dimensional continuous vector space (Ng, 2017). By employing this encoding approach, the model can capture a more comprehensive set of information, enhancing its ability to capture relevant patterns and features in the sequences.

Considering the limitations of the discussed methods, we have developed a method named DeepCBS, which employs a DNA sequence as input instead of a paired sequence and utilizes the dna2vec encoding method for representation. Applying DeepCBS to somatic mutation data of liver cancer patients, we predicted the impact of these mutations on CTCF binding sites. Then, through analysis of differential gene expression, we identify three potential liver cancer oncogenes, providing potential therapeutic targets for the treatment of liver cancer. In our study, DeepCBS successfully predicts the impact of mutations occurring at CTCF binding sites on insulated neighborhoods. In the future, it can also be utilized to predict the effects of mutations on other functions of CTCF.

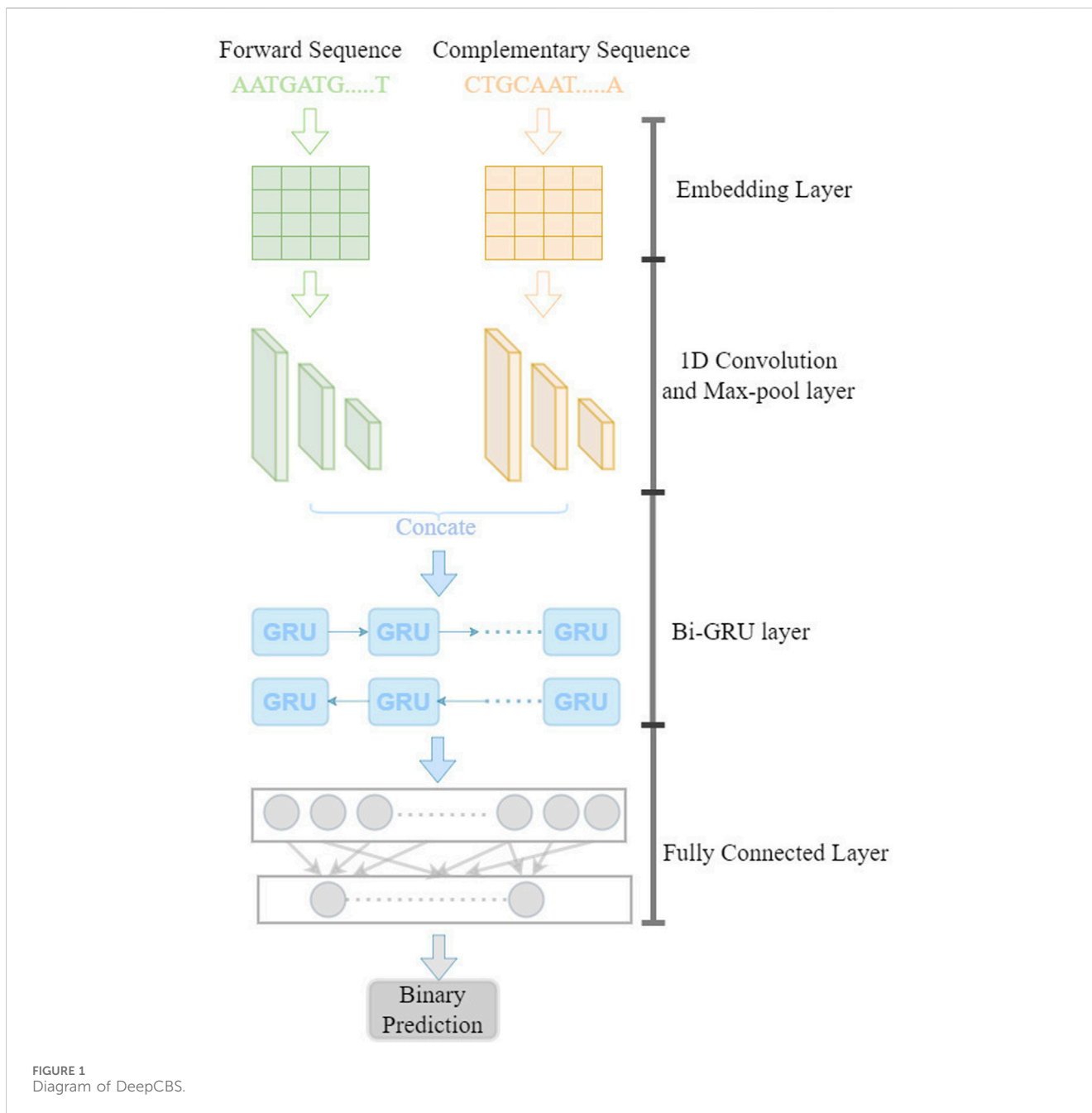
2 Materials and methods

2.1 Data collection and processing

CTCF ChIP-seq data for GM12878, HepG2, K562, MCF-7, and HMEC cell lines is downloaded from the Encode portal (accession: ENCFF710VEH, ENCFF237OKO, ENCFF738TKN, ENCFF738TKN, ENCFF288RFS). We also download RAD21 CHIP-seq raw data of GM12878 (accession: ENCFF002CPK) and CTCF CHIA-PET raw data of GM12878 (accession: ENCFF780PGS). We download comprehensive gene annotation data from GENCODE.

Positive samples are generated by selecting 100 base pairs from the summit of each ChIP-seq peak. Negative samples are generated using the R package gkmSVM by matching the repeat fraction, length, and GC content of the repetitive sequences in positive samples (Ghandi et al., 2016). Then we get 43,631 positive and 48,753 negative samples for GM12878 cell line, 60,229 positive and 56,099 negative samples for HepG2 cell line, 56,889 positive samples and 53,875 negative samples for K562 cell line.

We collect simple somatic mutations data of 1706 liver cancer patients from ICGC database, we also collected RNA-seq data of



liver cancer from this database, at the same time (see the [Supplementary Material](#)).

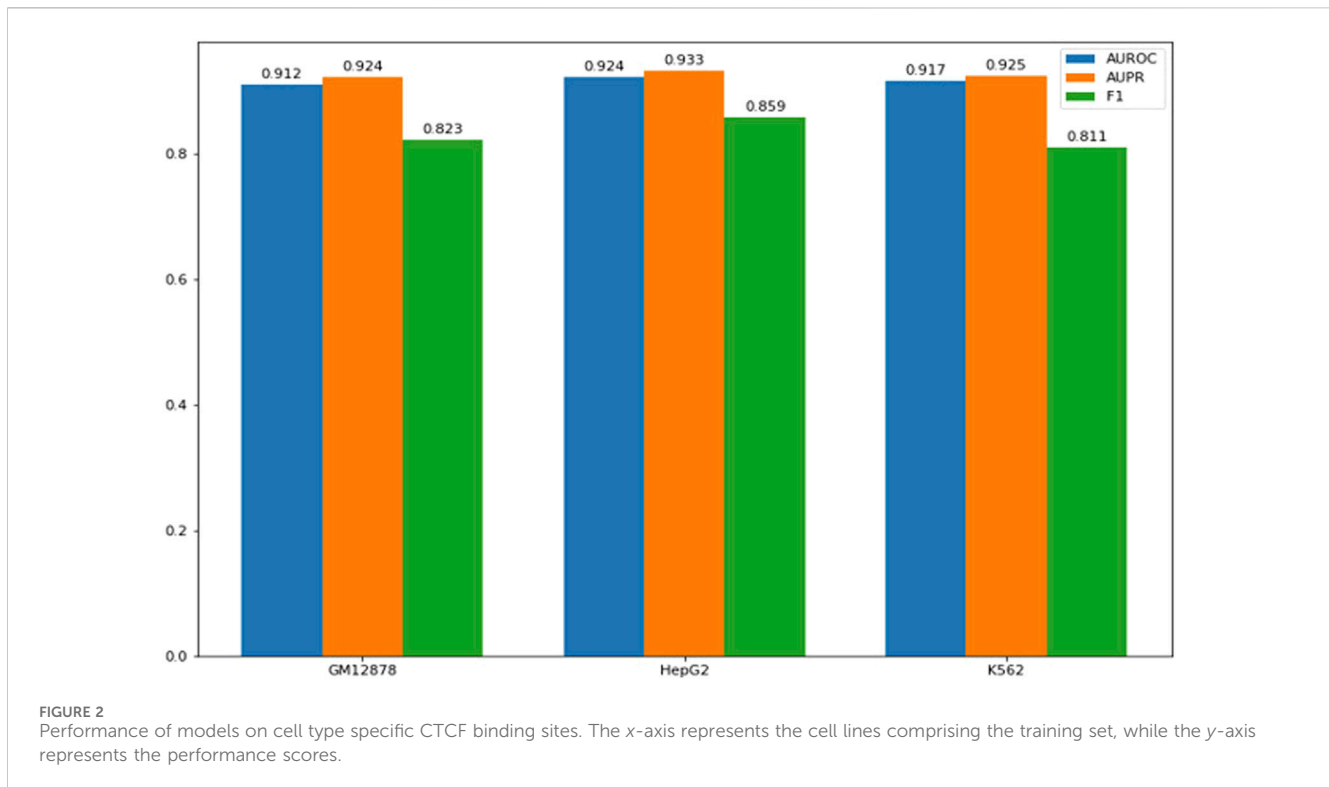
2.2 Construction of DeepCBS

The model is illustrated in [Figure 1](#). In this model, the forward and reverse DNA sequence with CTCF binding are taken as input by encoding into a matrix using the dna2vec (Ng, 2017) approach. Then, a three-layer convolutional neural network is used to learn the sequence motifs and high level features. The Bi-GRU (Bidirectional Gate Recurrent Unit) layer is used to learn the long-range dependencies between the high-level features. Next, two fully

connected layer is used to combine the output from the Bi-GRU layer and make the binary prediction.

2.3 Identification of disrupted insulated neighborhoods

We obtain insulated neighborhoods by integrating CTCF ChIA-PET data, CTCF ChIP-seq data, RAD21 CHIP-seq data and comprehensive gene annotation data. Specifically, we define an insulated neighborhood as a CTCF loop whose loop anchors overlap with a CTCF CHIP-seq peak and a RAD21CHIP-seq peak, and which contains at least one gene.



We collect somatic mutation data from liver cancer patients, identify mutated insulated neighborhoods, and utilize the deep learning model proposed in the previous step to predict whether these insulated neighborhoods would be disrupted.

2.4 Differential gene expression analysis

We obtain gene expression data from liver cancer patients in ICGC databases. To analyze gene differential expression, we utilized 3 R packages, namely, limma, edgeR, and DESeq2, independently. In order to enhance the robustness of our findings, we obtain differentially expressed genes by taking the intersection of the results from the three packages.

3 Results

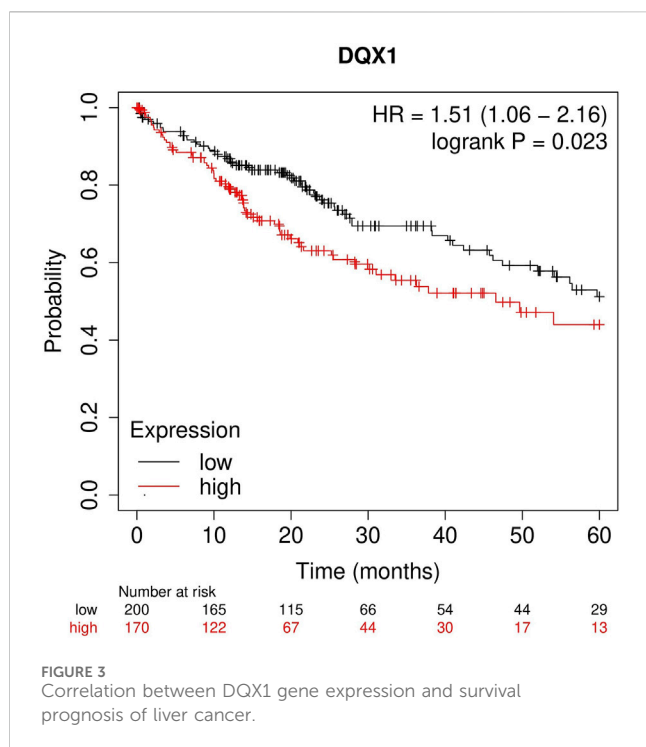
3.1 Workflow of DeepCBS

To elucidate the impact of non-coding mutations occurring at CTCF binding sites, we develop a method named DeepCBS, comprising the following main steps. Initially, we generate positive and negative samples from CTCF ChIP-seq data for 3 cell lines (GM12878, HepG2, K562). Using this data, we train a deep learning model to predict whether mutations on CTCF binding sites lead to the loss of CTCF binding at those sites. Subsequently, we obtain RAD21-mediated loops from RAD21 CHIA-PET data, defining a loop as an insulated neighborhood if both anchors of the loop overlap with CTCF CHIP-seq peaks. And if there are mutations within the CHIP-seq

TABLE 1 Cross cell performance evaluation.

Train set	Test set	Auroc
GM12878	HepG2	0.97945
GM12788	K562	0.97364
K562	HepG2	0.97965
K562	GM12878	0.98921
HepG2	K562	0.97507
HepG2	GM12878	0.99106

peak region that overlaps with loop anchors, then the insulated neighborhood is considered as a mutated insulated neighborhood. Leveraging the well-trained deep learning model, we predict whether mutations within the mutated insulated neighborhoods disrupt the binding of CTCF, resulting in the disruption of the insulated neighborhoods. In the next step, we observe a significant enrichment of proto-oncogenes in mutated insulated neighborhoods, suggesting that the disruption of these neighborhoods may play a crucial role in cancer development. Consequently, we identify the genes within the disrupted insulated neighborhoods and intersect these genes with the differentially expressed genes in liver cancer. This process yield three genes that may undergo upregulation due to the disruption of insulated neighborhoods. Notably, two out of the three genes have been previously reported as potential oncogenes in liver cancer. The remaining gene, DQX1, is identified as a potential liver cancer oncogene through bioinformatics analysis [Figure 2](#).



3.2 Performance of DeepCBS

To assess the model's performance across different cell types, we employ 1 cell type's samples as the training data and used samples from other cell types as the testing data, as shown in Table 1. As it can be seen, the model has achieved AUC values over 0.97, demonstrating that our method is a powerful tool for identifying CTCF binding sites.

Since some CTCF binding sites are cell-type specific, we collect CTCF CHIP-seq data from the MCF-7 cell line (breast cancer cell line) and HMEC cell line (normal breast epithelial cell line), and get cell-type specific CTCF binding sites in the MCF-7 cell line. As shown in Figure 2, our models have demonstrated excellent performance on cell-type-specific binding sites.

3.3 Identification of potential oncogenes in disrupted insulated neighborhoods in liver cancer

We apply a hypergeometric distribution test to our data and find a significant enrichment of proto-oncogenes in mutated insulated neighborhoods ($p < 0.05$). We have also observed this phenomenon in the data provided by Ji (Ji et al., 2016). This suggests that the disruption of insulated neighborhoods may be a key driver of cancer development, as it can lead to the abnormal activation of proto-oncogene into oncogene. We identify 237 disrupted insulated neighborhoods, comprising a total of 135 genes. We perform differential gene expression analysis, then identify 1,218 differentially expressed genes using 3 R packages. To explore which genes among the 135 affected genes in the disrupted insulated neighborhoods are key genes related to cancer. Then, we take the intersection of the differentially expressed genes with the genes located within the

disrupted insulated neighborhoods, which resulted in the identification of three key genes: ARHGEF39, UBE2C, and DQX1. And all of them are upregulated genes, potentially activated due to the disruption of insulated neighborhoods.

ARHGEF39 is a novel member of the Dbl-family of guanine nucleotide exchange factors (Wang et al., 2012). Guanine nucleotide exchange factors are recognized as crucial activators of Rho GTPases, which play a significant role in cell migration (Cook et al., 2014; Goicoechea et al., 2014). Overexpress of ARHGEF39 promotes gastric cancer cell proliferation and migration through the Akt signaling pathway (Wang et al., 2018; Zhou et al., 2018). Previous literature has proposed that ARHGEF39 may act as an oncogene in the progression of liver cancer, and thus represents a potential prognostic indicator and therapeutic target for this disease (Gao and Jia, 2019). Ubiquitin-conjugating enzyme E2C(UBE2C), a member of the E2family, is encoded by the UbcH10gene situated on human chromosome20q13.12. Its function involves the degradation of various target proteins through catalysis. UBE2C has been found to be upregulated in various types of cancer, including breast cancer, and is considered a potent proto-oncogene associated with tumor malignancy (Chou et al., 2014; Han et al., 2015). In liver cancer, UBE2C has been identified as a potential oncogene that can promote cell proliferation, migration, invasion, and drug resistance (Xiong et al., 2019).

Based on the above, we speculate that in liver cancer, the overexpression of ARHGEF39 and UBE2C serves as activated oncogenes and is involved in liver cancer development due to the disruption of the insulated neighborhoods containing them. However, there is currently no literature exploring the role of the DQX1 in liver cancer.

3.4 Overexpression of DQX1 is oncogenic in liver cancer

The Kaplan-Meier plotter (<https://kmplot.com/analysis/>) is a powerful tool that enables the assessment of the impact of 54k genes (including mRNA, miRNA, and protein) on survival across 21 types of cancer (Györfy, 2023). In this study, we focus on the analysis of the relationship between DQX1 expression and survival in liver cancer. The result of survival analysis, as shown in Figure 3, show that highly expressed DQX1 is linked to poor prognosis of overall survival (OS) for cancers of liver cancer.

Gene set enrichment analysis (GSEA) is further performed to explore the signaling pathways and molecular mechanisms that were differentially affected by DQX1 in liver cancer. In this study, the tumor samples are grouped based on the mean expression level of DQX1. Samples with expression levels higher than the mean are assigned to the high-expression group (DQX1. Hi), while those with expression levels lower than the mean are assigned to the low-expression group (DQX1. Low). In our study, the Hallmark database is utilized for performing the gene set enrichment analysis. As depicted in Figure 4, the analysis reveal that high expression of DQX1 is significantly associated with the activation of cell proliferation-related pathways. This finding suggests that DQX1 may play a crucial role in promoting cell proliferation in liver cancer.

DQX1 is one of the RNA-binding protein genes and RNA-binding protein can regulate the infiltration degrees of immune cells (Sun et al., 2021). Therefore, we implement an immunological

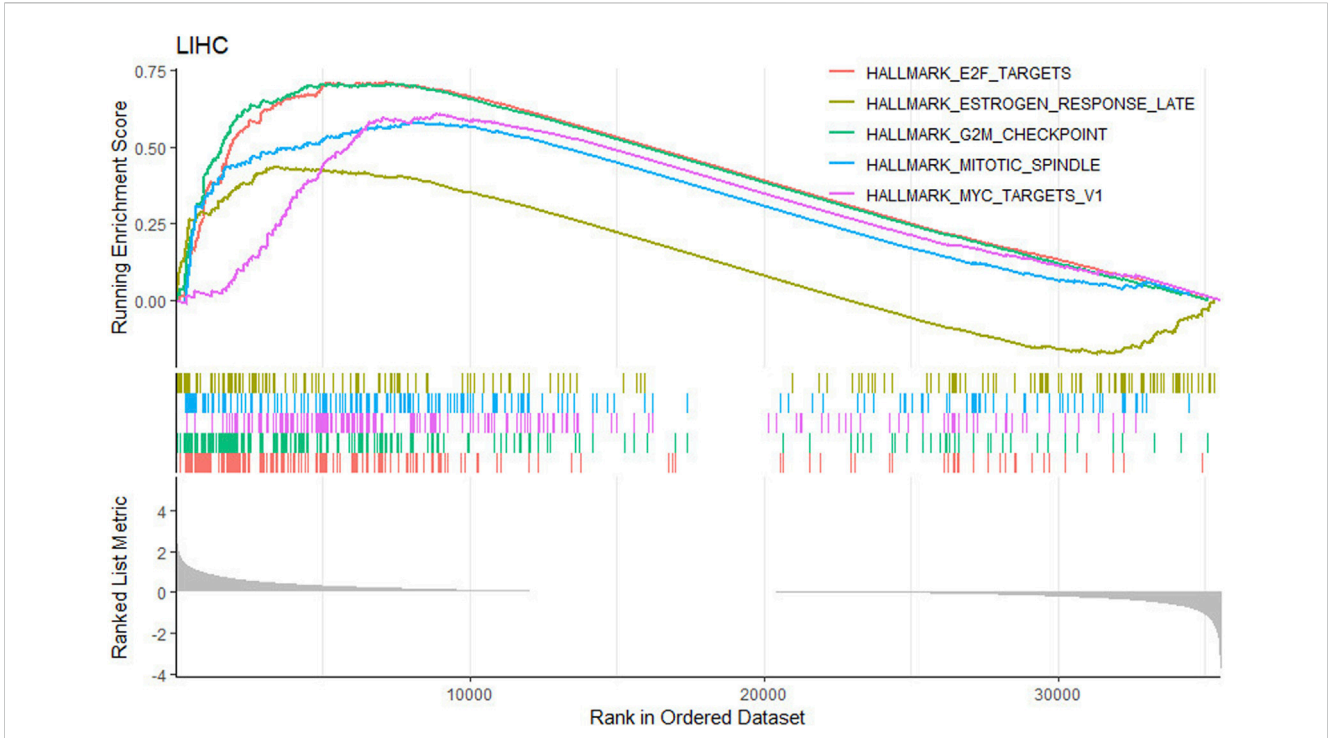


FIGURE 4 Signaling pathways associate with DQX1 in liver cancer. The depicted signaling pathways in the figure are all cell proliferation-related pathways.

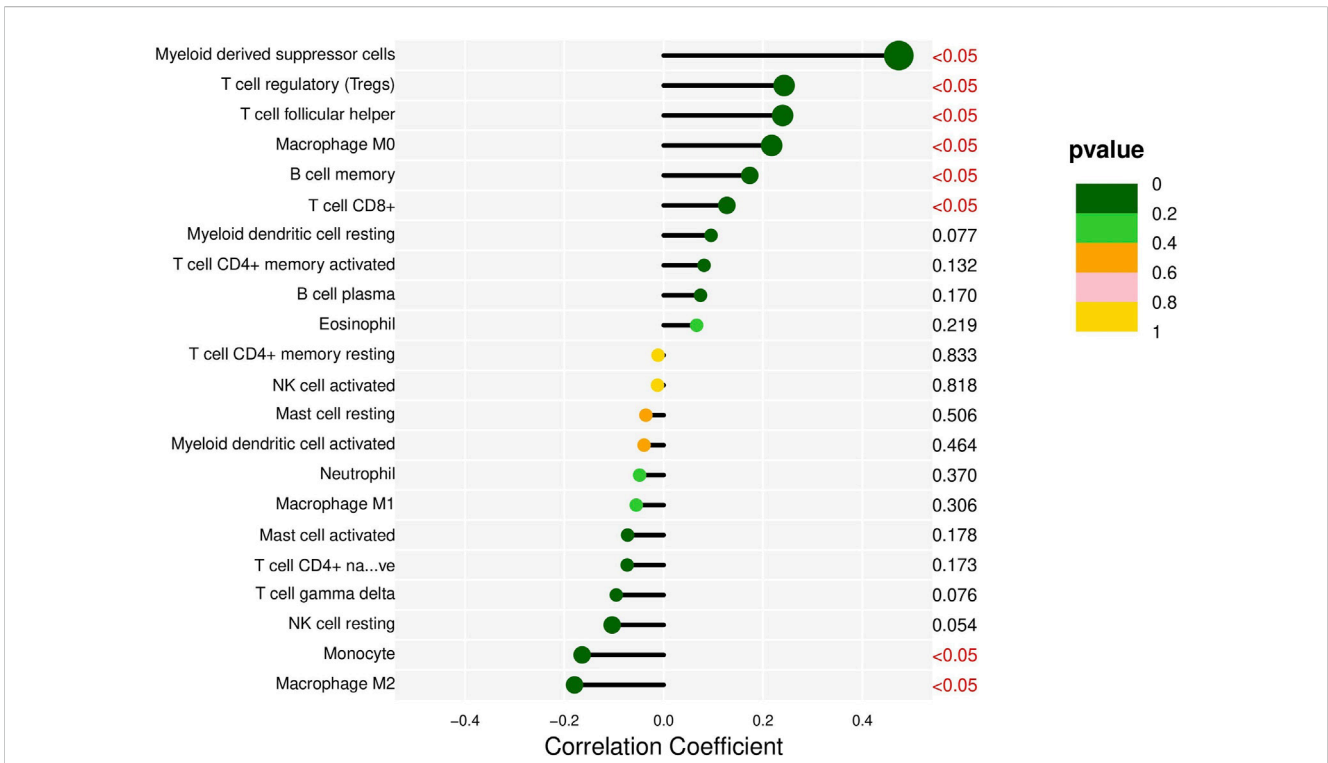
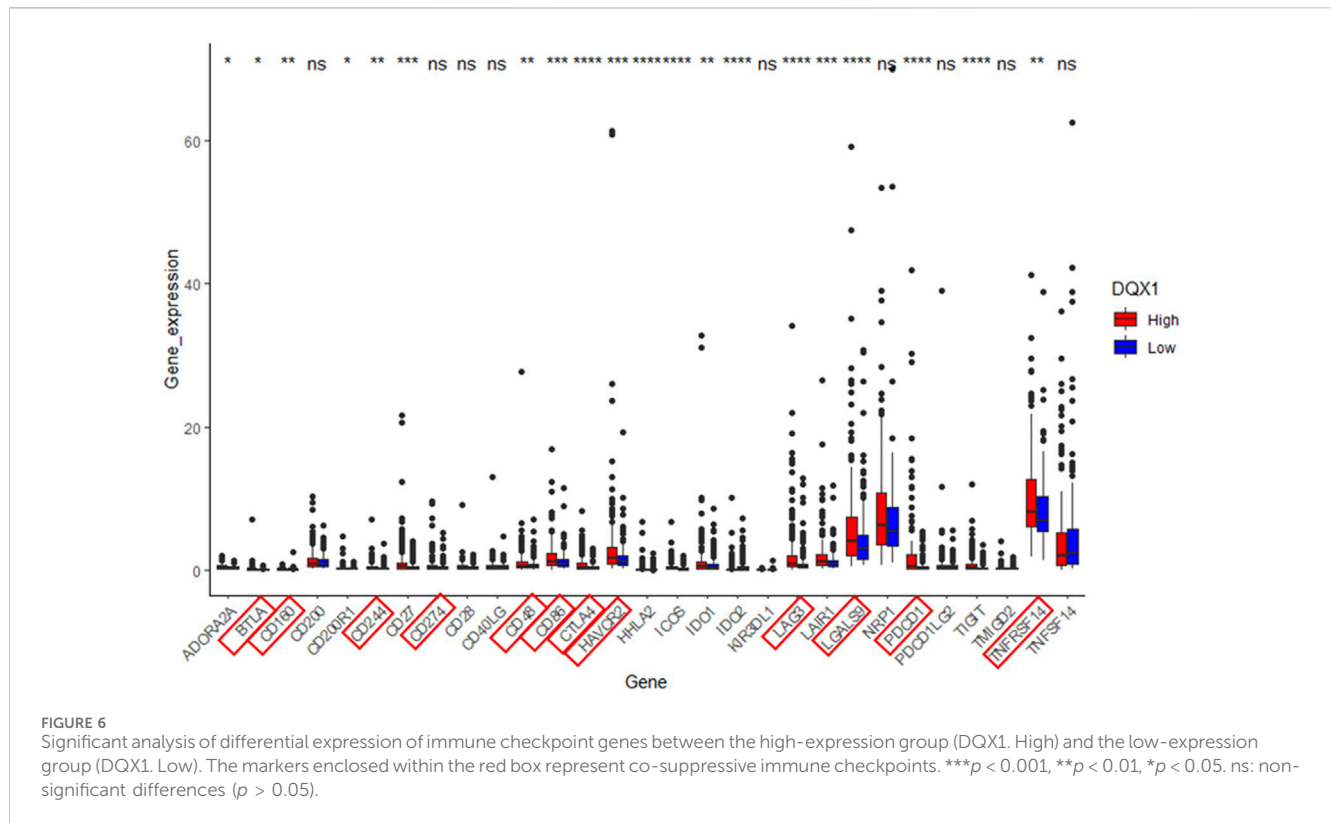


FIGURE 5 The correlation between DQX1 and immune cell infiltration in liver cancer.



Analysis of DQX1 in liver cancer. TIMER, a comprehensive online resource for the systematic analysis of immune infiltrates in various cancer types, is employed in this study to explore the correlation between DQX1 expression in liver cancer and different immune infiltrates (Li et al., 2020). The results are presented in Figure 5.

Also, we perform differential expression analysis of immune checkpoint genes in relation to DQX1 using the same grouping approach as in GSEA. The result is depicted in Figure 6.

Overall, the expression of DQX1 shows significant positive correlations with immune infiltration levels of regulatory T cells (Tregs), myeloid-derived suppressor cells (MDSCs), and expressions of co-suppressive immune checkpoints, contributing to immune escape. This suggests that we can develop immunotherapies targeting DQX1 for the treatment of liver cancer, in the future.

Based on the bioinformatics analysis, we have been inferred that DQX1 may potentially act as an oncogene and be involved in the development of liver cancer.

4 Conclusion

In summary, the CTCF play an crucialrole in maintaining these insulated neighborhoods. The disruption of CTCF binding sites can lead to dysregulation of contained genes, potentially resulting in the activation of oncogenes and promoting cancer development. It is important to shed light on the impact of mutations occurring at CTCF binding sites. So we develop a novel method, DeepCBS, to analyze the impact of mutations occurring at CTCF binding sites. Our analysis has identified three potential oncogenes, ARHGEF39, UBE2C, and DQX1 of liver cancer. All three

genes play an oncogenic role in the development of liver cancer. And overexpression of DQX1 is associated with poor prognosis and tumor immune escape. Our findings demonstrate the potential of DeepCBS to analyze the impact of mutations occurring at CTCF binding sites, as well as providing valuable insights for the diagnosis and treatment of liver cancer. Over all, this study emphasizes the importance of understanding the 3D organization of the human genome and its impact on gene regulation, as well as highlights the potential of computational methods to identify new targets for cancer therapy.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

Author contributions

YW: Methodology, Visualization, Writing–original draft. XG: Conceptualization, Writing–review and editing. ZN: Writing–review and editing. XH: Writing–review and editing. BW: Writing–review and editing. LG: Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (61672407, 62132015, U22A2037, Nos. 62172318) and the Fundamental Research Funds for the Central Universities (ZYTS23208).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Chou, C.-P., Huang, N.-C., Jhuang, S.-J., Pan, H.-B., Peng, N.-J., Cheng, J.-T., et al. (2014). Ubiquitin-conjugating enzyme UBE2C is highly expressed in breast microcalcification lesions. *PLoS One* 9, e93934. doi:10.1371/journal.pone.0093934
- Cook, D. R., Rossman, K. L., and Der, C. J. (2014). Rho guanine nucleotide exchange factors: regulators of Rho GTPase activity in development and disease. *Oncogene* 33, 4021–4035. doi:10.1038/ncr.2013.362
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *science* 295, 1306–1311. doi:10.1126/science.1067799
- Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16, 1299–1309. doi:10.1101/gr.5571506
- Downen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387. doi:10.1016/j.cell.2014.09.030
- Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., et al. (2009). An estrogen-receptor- α -bound human chromatin interactome. *Nature* 462, 58–64. doi:10.1038/nature08497
- Gao, J., and Jia, W.-D. (2019). Expression of rho guanine nucleotide exchange factor 39 (ARHGGEF39) and its prognostic significance in hepatocellular carcinoma. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* 25, 7826–7835. doi:10.12659/MSM.918270
- Ghandi, M., Mohammad-Noori, M., Ghareghani, N., Lee, D., Garraway, L., and Beer, M. A. (2016). gkmSVM: an R package for gapped-kmer SVM. *Bioinformatics* 32, 2205–2207. doi:10.1093/bioinformatics/btw203
- Goicoechea, S. M., Awadia, S., and Garcia-Mata, R. (2014). I'm coming to GEF you: regulation of RhoGEFs during cell migration. *Cell Adhesion Migr.* 8, 535–549. doi:10.4161/cam.28721
- Györfy, B. (2023). Discovery and ranking of the most robust prognostic biomarkers in serous ovarian cancer. *Geroscience* 45, 1889–1898. doi:10.1007/s11357-023-00742-4
- Han, Q., Zhou, C., Liu, F., Xu, G., Zheng, R., and Zhang, X. (2015). MicroRNA-196a post-transcriptionally upregulates the UBE2C proto-oncogene and promotes cell proliferation in breast cancer. *Oncol. Rep.* 34, 877–883. doi:10.3892/or.2015.4049
- Hnisz, D., Day, D. S., and Young, R. A. (2016a). Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* 167, 1188–1200. doi:10.1016/j.cell.2016.10.024
- Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A.-L., Bak, R. O., Li, C. H., et al. (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 351, 1454–1458. doi:10.1126/science.aad9024
- Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., et al. (2016). 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* 18, 262–275. doi:10.1016/j.stem.2015.11.007
- Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., et al. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* 47, 818–821. doi:10.1038/ng.3335
- Kuang, S., and Wang, L. (2021). Deep learning of sequence patterns for CCCTC-binding factor-mediated chromatin loop formation. *J. Comput. Biol.* 28, 133–145. doi:10.1089/cmb.2020.0225
- Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science* 326, 289–293. doi:10.1126/science.1181369
- Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 48, W509–W514. doi:10.1093/nar/gkaa407
- Ng, P. (2017). dna2vec: consistent vector representations of variable-length k-mers. *arXiv preprint arXiv:1701.06279*.
- Oh, S., Oh, C., and Yoo, K. H. (2017). Functional roles of CTCF in breast cancer. *BMB Rep.* 50, 445–453. doi:10.5483/bmbrep.2017.50.9.108
- Simonis, M., Klous, P., Homminga, I., Galjaard, R.-J., Rijkers, E.-J., Grosveld, F., et al. (2009). High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat. Methods* 6, 837–842. doi:10.1038/nmeth.1391
- Sun, D., Yang, K.-S., Chen, J.-L., and Wang, Z.-B. (2021). Identification and validation of an immune-associated RNA-binding proteins signature to predict clinical outcomes and therapeutic responses in colon cancer patients. *World J. Oncol.* 19, 314–413. doi:10.1186/s12957-021-02411-2
- Trieu, T., Martinez-Fundichely, A., and Khurana, E. (2020). DeepMILO: a deep learning approach to predict the impact of non-coding sequence variants on 3D chromatin structure. *Genome Biol.* 21, 79–11. doi:10.1186/s13059-020-01987-4
- Umer, H. M., Cavalli, M., Dabrowski, M. J., Diamanti, K., Kruczyk, M., Pan, G., et al. (2016). A significant regulatory mutation burden at a high-affinity position of the CTCF motif in gastrointestinal cancers. *Hum. Mutat.* 37, 904–913. doi:10.1002/humu.23014
- Wang, H., Li, M., Tao, X., Qian, Y., Chen, L., and Tao, G. (2018). ARHGGEF39 promotes gastric cancer cell proliferation and migration via Akt signaling pathway. *Mol. Cell. Biochem.* 440, 33–42. doi:10.1007/s11010-017-3153-3
- Wang, H., Li, Y., Wang, Y., Han, Z.-G., and Cai, B. (2012). C9orf100, a new member of the Dbl-family guanine nucleotide exchange factors, promotes cell proliferation and migration in hepatocellular carcinoma. *Mol. Med. Rep.* 5, 1169–1174. doi:10.3892/mmr.2012.783
- Xiong, Y., Lu, J., Fang, Q., Lu, Y., Xie, C., Wu, H., et al. (2019). UBE2C functions as a potential oncogene by enhancing cell proliferation, migration, invasion, and drug resistance in hepatocellular carcinoma cells. *Biosci. Rep.* 39, 82384. doi:10.1042/BSR20182384
- Yu, M., and Ren, B. (2017). The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.* 33, 265–289. doi:10.1146/annurev-cellbio-100616-060531
- Zhang, R., Wang, Y., Yang, Y., Zhang, Y., and Ma, J. (2018). Predicting CTCF-mediated chromatin loops using CTCF-MP. *Bioinformatics* 34, i133–i141. doi:10.1093/bioinformatics/bty248
- Zhou, H., Cai, L., Zhang, X., Li, A., Miao, Y., Li, Q., et al. (2018). ARHGGEF39 promotes tumor progression via activation of Rac1/P38 MAPK/ATF2 signaling and predicts poor prognosis in non-small cell lung cancer patients. *Lab. Invest.* 98, 670–681. doi:10.1038/s41374-018-0022-y

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2024.1354208/full#supplementary-material>