

At CTCF Binding Sites Machine Learning Identifies Cancer-Driving Mutations

Researchers investigate whether machine learning can identify pan-cancer mutational hotspots at [persistent CCCTC-binding factor](#) (P-CTCF) binding sites (P-CTCFBSs).



Study

In the present study, researchers developed CTCF-In-Silico Investigation of PersistEnt Binding (INSITE), a computational tool capable of predicting the persistence of CTCF binding following knockdown in [cancer cells](#).

CTCF-INSITE is a machine learning tool that assesses both genetic and epigenetic characteristics accounting for the persistence of CTCF binding. The mutational load at PCTCF binding sites was determined using [International Cancer Genome Consortium](#) (ICGC) sequences from matched tumors by generating persistence metrics for the Encyclopedia of DNA Elements (ENCODE) CTCF ChIP-sequencing data from different tissue types. National Center for Biotechnology Information (NCBI) and GM12878 high-coverage whole-genome sequencing (WGS) data from the platinum genome initiative were also used for the analysis.

The researchers screened cohorts with fewer mutations per individual using CTCF ChIP-seq data from IMR-90, MCF7, and LNCaP cell lines isolated from [lung tissue](#), breast cancer, and prostate adenocarcinoma, respectively. After identifying and eliminating outliers using the Interquartile Range (IQR) method, 24 cohorts, including 3,218 patients, were available for the study.

Twelve distinct cancer types were then created by combining mutations from cohorts of the same cancer type. For IMR-90, LNCaP, and MCF7 cells, genomic features, [chromatin interactions](#), binding affinity, replication timing, constitutive binding, and conservation scores were investigated.

Random forest modelling was used because it has a superior success rate compared to [linear regression models](#) in predicting CTCF binding in silico. Data were divided into training and testing datasets using a 9:1 ratio.

Binding motif studies were also performed to determine the binding position inside a ChIP-seq peak from 200 to 2,000 base pairs (bp). A motif score was then calculated for each area of a [ChIP-seq peak](#).

Gene set enrichment analysis (GSEA) was used to determine the [trinucleotide mutational](#) context for every patient, and fluorescence polarization DNA binding (FPDB) assays were used to

compare the mutational burden between P—and L-CTCF-BSs. By aggregating these results, a background mutation rate of CTCFBSs was generated for every cancer.

Results

As compared to all CTCF binding sites, those for P-CTCF had significantly higher mutational rates in prostate and [breast cancers](#). In all 12 examined cancer types, projected P-CTCF binding sites exhibited a markedly increased mutational load. P-CTCF binding site mutations, predicted to have a functional effect on CTCF chromatin looping and binding, showed significantly more enrichment.

The in vitro experiments confirmed that the disruptively anticipated P-CTCF binding site cancer mutations reduced CTCF binding. [Mutations](#) in P-CTCF binding sites were more frequently observed than L-CTCF in 12 distinct cancer types. P-CTCF binding site mutations were related to loop disruption, thus indicating that these mutations contribute to three-dimensional genome dysregulation in cancer.

[Binding affinity](#) is crucial to P-CTCF-BS survival, especially at chromatin loop anchors, late replication timing regions, and TAD boundaries. Moreover, the co-location of chromosome loops indicates durability.

The researchers identified significant allelic imbalances in binding at 91 sites, wherein mutations reduced binding affinity. Breast cancer exhibited ultraviolet (UV) light-induced gene downregulation, whereas [prostate cancer](#) exhibited epithelial-to-mesenchymal transition gene enrichment. Compared to L-CTCF binding sites, P-CTCF-BSs were associated with a greater mutational rate and notable enrichment of disruptive mutations.

Conclusion

The study findings identify a novel subclass of cancer-specific CTCF-BS DNA mutations and provide important insights into the [crucial role](#) of these mutations in pan-cancer genomic structures. CTCF-INSITE showed significant enrichment for mutations across various cancer types. Due to the possible disruption of chromatin loops and decreased binding in in vitro binding tests, these mutations are considered functional.

Studying the mutational profiles of other types of cancer could be supported by the enhanced mutational signal at [P-CTCF binding sites](#). Thus, the predictive power of CTCF-INSITE for CTCF-BSs offers promising candidates for experimental modification that researchers must prioritize to better understand the etiology of cancer.

Source:

<https://www.news-medical.net/news/20240707/Machine-learning-identifies-cancer-driving-mutations-at-CTCF-binding-sites.aspx>