

## Disease bioinformatics/translational medicine

# Gaining insights into Alzheimer's disease by predicting chromatin spatial organization

Camilo Villaman<sup>1</sup> , Irene Cartas-Espinel<sup>2,3</sup>, Mauricio Saez<sup>2,\*</sup>, Alberto J.M. Martin<sup>1,4,\*</sup>

<sup>1</sup>Laboratorio de Redes Biológicas, Fundación Ciencia & Vida, Avenida del Valle Norte 725, Huechuraba, Santiago, 8580702, Chile

<sup>2</sup>Laboratorio de Investigación en Salud de Precisión, Departamento de Procesos Diagnósticos y Evaluación, Facultad de Ciencias de la Salud, Universidad Católica de Temuco, Manuel Montt 56, Temuco, 4813302, Chile

<sup>3</sup>Centro de Investigación, Innovación y Creación UCT, Universidad Católica de Temuco, Rudecindo Ortega 03694, Temuco, 4781509, Chile

<sup>4</sup>Escuela de Ingeniería, Facultad de Ingeniería, Universidad San Sebastián, Bellavista 7, Recoleta, Santiago, 8420524, Chile

\*Corresponding authors. Alberto J.M. Martin, Fundación Ciencia & Vida, Avenida del Valle Norte 725, Huechuraba, Santiago, 8580702, Chile.

E-mail: alberto.martin@uss.cl; Mauricio Saez, Universidad Católica de Temuco, Campus San Francisco CSF16\_2, Manuel Montt 56, Temuco, 4813302 Chile. E-mail: mauricio.saez@uct.cl.

Associate Editor: Vinícius Maracajá-Coutinho

Downloaded from https://academic.oup.com/bioinformaticsadvances/article/5/vba268/8301184 by guest on 24 November 2025

### Abstract

**Motivation:** CTCF is a conserved protein involved in the establishment and maintenance of topologically associating domains (TADs) and loops. Alzheimer's disease (AD) represents the most common form of dementia, affecting over 50 million elderly individuals. Epigenetic alterations are a hallmark of AD, and epigenetic disruptions are able to affect CTCF binding and looping. Understanding the dynamics of CTCF loops behind AD may lead to new, undiscovered contributions of CTCF to the etiology of AD. To understand the dynamics behind CTCF loops, we developed a CTCF loop predictor using different genomic and epigenomic features, such as CTCF motif information, CTCF protein binding information, and different histone marks.

**Results:** We obtained F-scores of over 0.9 in GM12878 and K562 cell lines. We reported the importance of each feature in classification, and compared the results with other loop predictors. After testing the predictor, we predicted loops in control and AD data, reported a score of loop disruption and selected the top disrupted loops on AD which were all previously linked with AD in bibliography. Our study contributes to a better understanding of the role of CTCF binding and CTCF loops in gene regulation, and highlights new clues about CTCF in the etiology and development of AD.

**Availability and implementation:** The method can be found in <https://github.com/networkbiolab/jalpy>.

### 1 Introduction

CTCF (CCCTC-binding factor) is a conserved zinc finger protein capable of DNA binding (Kim *et al.* 2015) involved on multiple different biological processes. CTCF was first described as a negative regulator of the myc gene in chicken (Klenova *et al.* 1993), and later it was described as a transcriptional regulator involved on many cellular events such as insulation (Jia *et al.* 2020), alternative splicing (Alharbi *et al.* 2021), and loop formation (Xi and Beer 2021). CTCF is reported as the most important insulator in mammals (Kim *et al.* 2007), and is capable of blocking interactions between enhancers and promoters. CTCF is involved on the establishment and maintenance of topologically associated domains (TADs) (Nanni *et al.* 2020), which are domains of increased self-interaction involved with gene regulation. CTCF is enriched at TAD boundaries and blocks interactions from elements inside the TAD with elements outside it (Jia *et al.* 2020).

CTCF is involved with tridimensional genome organization as it can interact with different proteins forming tridimensional functional domains called CTCF loops (Zhang *et al.*

2023). The mechanism behind CTCF loop formation is not completely described, but the loop extrusion model states that cohesin loads into DNA and extrudes the DNA through a cohesin ring, moving along the DNA until reaching bound CTCF anchors, which can form CTCF homodimers interacting with the cohesin ring at the base of the loop (Hansen 2020). Since CTCF and CTCF loops are involved in maintaining domain boundaries or blocking enhancer activities, CTCF binding is critical in regulation of cell gene expression (Liu *et al.* 2023), and alterations to CTCF binding patterns lead to transcriptional dysregulation (Fang *et al.* 2020).

The CTCF loop landscape is considered to be mostly invariant across similar cell types, and differences on the loop landscape can be linked with changes on the normal cell expression program (Grubert *et al.* 2020). There is evidence that epigenetic disruptions are able to affect CTCF binding and looping (Monteagudo-Sánchez *et al.* 2024), and thus, diseases which present epigenetic alterations as a feature may also feature abnormal CTCF binding and CTCF loop disruption, with Alzheimer's disease (AD) being an example of possible CTCF disruption (Burton *et al.* 2002).

Received: January 31, 2025; Revised: September 16, 2025; Accepted: September 23, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

AD is a neurodegenerative disease and represents the most common form of dementia, affecting over 50 million elderly individuals (“2024 Alzheimer’s disease facts and figures,” 2024). AD is characterized by chronic memory impairment and cognitive deficits related, but not limited to, language, orientation, spatial and temporal awareness, executive capacity, and behavior, a progressive loss of autonomy, dementia, and death (Lane et al. 2018). Alzheimer’s is classified into early-onset Alzheimer’s disease (EOAD) and late-onset Alzheimer’s disease (LOAD) (Atri 2019). EOAD accounts for 1%–2% of all AD cases and it usually presents in patients between 30 and 60 years old (Atri 2019). EOAD is inherited in an autosomal dominant fashion, with a rapid rate of progression, and the most common biomarkers related with EOAD are mutations on the APP, PSEN1, and PSEN2 genes (Atri 2019). In comparison, LOAD accounts for >97% of all AD cases and typically occurs after the age of 65, with many different genetic and environmental factors contributing to the apparition of the disease. In fact, most AD cases are caused by a combination of different risk factors, with age becoming one of the most relevant (Zhao and Huai 2023). AD is characterized histologically by two hallmarks, the intracellular deposition abnormally phosphorylated Tau protein (Maccioni et al. 2018), and the extracellular aggregates of Amyloid-beta peptide (AB) plaques (Viola and Klein 2015). AB oligomers may also activate microglial cell signaling cascades that lead to more hyperphosphorylation of Tau protein, which also leads to Tau protein aggregates (Viola and Klein 2015, Maccioni et al. 2018). Tau aggregates get released on cell death and also kickstart microglia activation, leading to a cyclic pathological cascade which culminates with cell death and neurodegeneration (Viola and Klein 2015, Maccioni et al. 2018). Age-related epigenetic disruptions may be contributing directly to the etiology of AD, since AD and especially LOAD, may be associated with the epigenetic alterations related with aging (López-Otín et al. 2023, Yang et al. 2023). The CTCF loop landscape in AD is not completely understood, and due to different disruptions in AD epigenetics, abnormal CTCF binding in AD contributing to the etiology of this disease (Patel et al. 2023). Currently, the full contribution of CTCF and CTCF loops to AD has not been entirely described, and since epigenetic alterations are a hallmark of AD, those alterations may be contributing to CTCF loop disruption, leading to expression changes on the transcriptional program of the cell (Patel et al. 2023).

ChIA-PET is used to detect specific protein-mediated chromatin loops genome-wide at high resolution (Fullwood et al. 2009). ChIA-PET requires chromatin cross-linking, proximity ligation of the interacting fragments with linkers, and DNA sequencing of the fragments to estimate the frequency of chromatin interactions (Fullwood et al. 2009, Tang et al. 2015). Despite recent technical advances, experimental profiling of CTCF-mediated interactions remains difficult and costly (Fullwood et al. 2009, Tang et al. 2015), therefore, using computational predictions with integrated available datasets can be an interesting alternative to interrogate the CTCF-mediated interactome on a target of interest.

The dynamics of CTCF loops during the development of AD are not completely described, and understanding the dynamics of CTCF loops behind AD may lead to new, undiscovered contributions of CTCF to the etiology of this disease. To understand the dynamics behind CTCF loops, CTCF loop formation can be framed as a classification problem where

the presence or absence of certain features contribute to a loop score, such as the conservation of the CTCF binding site, the presence of proteins interacting with CTCF, or the presence of distinct epigenetic marks. Different machine learning approaches have been used in similar biological classification tasks (Greener et al. 2022), and to provide new insights about CTCF loop formation in AD we developed a new classifier based on XGBoost (Chen and Guestrin 2016), tested it on different cell lines, and compared it against state-of-the-art CTCF loop predictors (Kai et al. 2018, Xu et al. 2023). We next predicted loops in control and AD patients, using cell samples from dorsolateral prefrontal cortex. We report a score of loop disruption and report the top disrupted loops on AD. We expect to facilitate *in silico* prediction of CTCF loops, provide new insights about the role of different features on CTCF loop formation, and clarify the role of CTCF looping in AD.

## 2 Methods

### 2.1 Dataset download

We downloaded ChIA-PET data from ENCODE (<https://www.encodeproject.org/>) for the GM12878 and K562 cell lines and aligned it against the hg19 reference genome using the ChIA-PET2 (Li et al. 2017) pipeline which streamlines all the steps required for ChIA-PET data analysis, including trimming, mapping, removal of duplicates, calling of peaks, and calling of chromatin loops. ChIA-PET2 output was transformed to bedpe format, and loops with <2 PETs and FDR >0.05 were removed. We used FIMO (Grant et al. 2011) with default values and the MA0139.1 CTCF frequency matrix to identify CTCF binding sites (CTCFBS) on the hg19 reference genome. We also downloaded from the same ENCODE repository, CTCF and RAD21 hg19 ChIP-seq narrowPeak files to define CTCF loops, and h3k4me1, h3k9me3, h3k27me3 and h3k27ac hg19 narrowPeak bed files as extra features for the model for the same cell lines. For GM12878 and K562, the accession codes are: for CTCF ChIP-seq, ENCSR000DZN, ENCSR000EGM, for RAD21 ChIP-seq, ENCSR000EAC, ENCSR000FAD, for h3k4me1 ChIP-seq, ENCSR000AKF, ENCSR000AKS, for h3k9me3 ChIP-seq, ENCSR000AOX, ENCSR000APE, for h3k27me3 ChIP-seq, ENCSR000AKD, ENCSR000EWB and for h3k27ac ChIP-seq, ENCSR000AKC, ENCSR000AKP.

### 2.2 CTCF loop determination

From the bedpe loop file, for each loop, we generated a window of 1000 pb around the start and the end position reported by ChIA-PET2. We removed loops without CTCFBS inside the 1000 pb start and end windows. We only considered loops in the range between 2 kb and over 2 mb, as most of the loops in humans were reported to exist in that pair-bases range (Tang et al. 2015). We analyzed the area around three different intervals, the first interval was 1000 pb surrounding the start of the loop, the second interval was 1000 pb surrounding the end of the loop, and the last interval was the area inside the start and the end loop ranges mentioned (interloop).

### 2.3 Feature description and loop prediction

We assigned values to the start and end windows based on the FIMO scores of the CTCF frequency matrix described earlier. If a FIMO predicted CTCFBS was found on either of

the start or end intervals, we saved the FIMO predicted CTCFBS score as the start-interval motif score or as the end-interval motif score, respectively. For the rest of the analyzed features, we considered the start-loop interval, the end-loop interval, and the interloop interval windows earlier mentioned. Instead of using the FIMO score, we assigned scores based on the presence or absence of an analyzed feature in the evaluated window. We included CTCF binding, h3k4me1, h3k9me3, h3k27me3, and h3k27ac as analyzed features. After aggregating the loop data, we split the loops into 2/3 and 1/3, using 2/3 of the loops of GM12878 to train the predictor, and tested the predictor with the remaining 1/3 of the loops. We repeated the whole process using the K562 cell line. To confirm if our approach was able to generalize after being trained, we predicted CTCF loops on the K562 cell line using the GM12878 cell line, using the whole GM12878 loops to as input to train the predictor, and the whole K562 loops as a test set.

#### 2.4 High-confidence biological loop determination for loop prediction

In humans, the ring-like cohesin complex ring is composed by the SMC1-SMC3 dimer, the kleisin RAD21, and the two SA1 and SA2 cohesin subunits (Gruber *et al.* 2003). We classify a loop as positive when convergently oriented CTCF motifs were occupied by CTCF in any of the start-loop interval, or end-loop interval. We also required colocalization with RAD21 in any of the two CTCF loop anchors, as both interact for the maintenance of chromatin loop domains (Pugacheva *et al.* 2020). We used these loops as high-confidence CTCF loops for training and prediction. Importantly, RAD21 was only used to determine which loops were positive and which loops were negative, and it was excluded as a training feature in the pipeline. We included a diagram of the pipeline as Fig. 5, available as supplementary data at *Bioinformatics Advances* online.

#### 2.5 Loop prediction with simulated CTCF binding

Since CTCF ChIP-seq information is not always available, we developed a Random Forests CTCF binding predictor based on different genomic and epigenetic features (Villaman *et al.* 2023), with the objective of using known CTCF-related features on a known cell line to determine CTCF binding on a different cell line without CTCF ChIP-seq data. We evaluated if the predictions made with our predictor were also capable of contributing to the identification of CTCF loops. We generated CTCF binding scores for the GM12878 cell line and compared the loop values with the GM12878 RAD21 loop values, and repeated the same method using GM12878 with simulated values as a training set, and K562 as a test set. We reported precision-recall-area under the curve (PRAUC) and receiver operating characteristic-area under the curve (ROCAUC).

#### 2.6 Feature importance in loop prediction

To evaluate feature importance we generated bar plots for each feature used in classification in the GM12878 cell line, K562 cell line, and predicting across cell lines using GM12878 as a training set and K562 as a test set.

#### 2.7 Benchmarking with known approaches

We compared our approach with other published approaches (Kai *et al.* 2018, Xu *et al.* 2023), training with GM12878 cell

line and predicting on chr2, 20, 21, and 22 of K562 cell line. We reported performance using ROC and PR curves.

#### 2.8 AD dataset download

We downloaded information from the ENCODE RUSH AD Project with the objective of having a standard pipeline for data analysis and comparison. We downloaded three GRCh38-aligned control patients without AD, and one patient with AD and cognitive impairment. We downloaded GRCh38 CTCF ChIP-seq data, h3k4me3 ChIP-seq data, h3k27me3 ChIP-seq data, h3k27ac ChIP-seq data, and DNase-seq data from dorsolateral prefrontal cortex. We also downloaded Hi-C data from the same dataset. We used FIMO with default values and the MA0139.1 CTCF frequency matrix to identify CTCF binding sites (CTCFBS) on the GRCh38 reference genome downloaded from ENCODE, and saved both the FIMO score and the *P* value for posterior use.

#### 2.9 AD CTCF binding prediction

From each FIMO predicted AD CTCFBS, we generated matrixes of 500 pb around the AD CTCFBS with 25 pb windows following the same approach described in earlier steps (Villaman *et al.* 2023). We predicted CTCF binding using a Random Forest predictor and compared it with the AD CTCF ChIP-seq dataset. We predicted each patient sample by itself, splitting the sample dataset randomly and using 2/3 of the dataset to train and 1/3 to predict. Then, we trained using two patients as a training set and the remaining one as a test set. To finish, we used the three patients to predict the AD dataset, we used the three patients as a training set and predicted the AD patient as a test set. We reported ROCAUC and PRAUC curves for each.

#### 2.10 AD CTCF loop prediction

We used our earlier developed XGBoost CTCF loop predictor to generate prediction scores for every loop on the training set. From the downloaded Hi-C data, we divided loops into positive and negative. Positive loops had two convergent CTCF binding sites on their start and end anchors respectively, between 2 kb and 2 mb pb, with predicted CTCF binding confirmed by ChIP-seq on each. Negative loops were pairs of CTCF binding sites outside the earlier mentioned criteria. We considered an area of 500 pb around the start and end anchors, and the area inside the loop. We considered the presence of a CTCF binding motif, DNA accessibility, h3k4me3, h3k27me3, and h3k27ac as features. We also included the CTCF binding predictions as features for each control and AD sample. We used the three control samples as a training set and the AD sample as a test set. We reported precision, recall, F-score, ROCAUC, and PRAUC (Taha and Hanbury 2015).

#### 2.11 Discordant loop determination

We generated scores for every CTCF loop on the control samples under the assumption that CTCF loops in every control sample are relatively constant and related to a normal cell state, and disruptions to loops are related to abnormal cell states such as AD. We used the three control samples as a training set and predicted loop scores on the AD patient, using the same mentioned features for both control and AD dataset. After generating loop scores, we designated discordant loops following any of the two following conditions: (i)

Loops on the control dataset with a high predicted score (1) and a low predicted score (<0.0005) on the AD dataset were marked as discordant (looping is lost on the AD dataset), and we define those loops as “lost.” (ii) Loops on the control dataset with a low predicted score 0 and a high predicted score (>0.999) on the AD dataset were marked as discordant (looping is regained on the AD dataset), and we defined those loops as “gained.” We calculated a discordant loop score based on the square of the difference between the loop score obtained on the control dataset and the AD dataset, as shown in the following formula:

To finish, we annotated the loops using the ENCODE GRCh38 gtf, assigning each loop to the single gene with the longest overlap inside the loop, allowing only a single gene for each loop. We ranked genes by the sum of the discordant loop score of loops in which they are within and reported the top 5 genes with the highest discordant loop score.

## 3 Results

### 3.1 Prediction and across-cell prediction

We selected 25 000 and 19 070 positive loop examples for GM12878 and K562, defining a positive loop with 2 bound CTCF sites on its boundary and RAD21 on any of the three reported loop windows (start, end, and intraloop). Negative examples were selected by picking two unbound CTCFBS without RAD21 on any of the loop windows mentioned. We tested GM12878 against itself, and K562 against itself as a first approach, by randomly splitting the dataset using 2/3 to train and 1/3 to test. To test the performance of the predictor on different cell types, we used GM12878 as a training set and K562 as a test set. We reported precision, recall and f-score of over 0.83 in all of the predicted cell lines, however, prediction across cell lines has lower scores than training and prediction with the same cell line (Table 1 and Fig. 1A and B, available as supplementary data at *Bioinformatics Advances* online).

### 3.2 Prediction based on CTCFBS prediction

CTCF ChIP-seq experiments are not always available, and to deal with this problem we developed a CTCF binding predictor based on different genetic and epigenetic features (Villaman et al. 2023). Since CTCF binding is a very important feature on the determination of CTCF loops, we wanted to test if the results from our predictor were able to be used as features for CTCF loop prediction. We trained the predictor using a similar protocol to the one mentioned earlier, in this case we employed the GM12878 cell line, considering DNA accessibility, h3k4me1, h3k9me3, h3k27me3 and h3k27ac histone marks, DNA methylation, and FIMO motif score, and used the score

**Table 1.** Precision, recall, and F-score after predicting CTCF binding.<sup>a</sup>

Sample	Precision	Recall	F-score
Pred P1 Sample 1	0.92	0.62	0.74
Pred P2 Sample 2	0.93	0.58	0.72
Pred P3 Sample 3	0.91	0.54	0.67
Pred P1 Cross 1	0.93	0.58	0.71
Pred P2 Cross 2	0.91	0.53	0.67
Pred P3 Cross 3	0.93	0.62	0.74
AD Predict	0.75	0.59	0.66

<sup>a</sup> Sample 1, 2, and 3 uses information from said sample, splits randomly the CTCFBS on the sample using 2/3rds as a training set and 1/3rd as a test set. Cross 1, 2, and 3 represents using two samples as a training set and predicting on the remaining one as a test set. AD Predict uses the three control patients to predict a patient with AD and cognitive impairment.

of the predictor instead of the experimental CTCF ChIP-seq data. We used the same approach as the first prediction, we randomly split the dataset using 2/3 to train and 1/3 to test. We also used the scores for testing across cells, using the predicted GM12878 CTCF binding scores as the training set and K562 CTCF ChIP-seq data as test set. We reported precision, recall and f-score of over 0.86, with prediction across cells having a lower score than training and prediction with the same cell line (Table 2 and Fig. 2A and B, available as supplementary data at *Bioinformatics Advances* online).

### 3.3 Feature importance determination

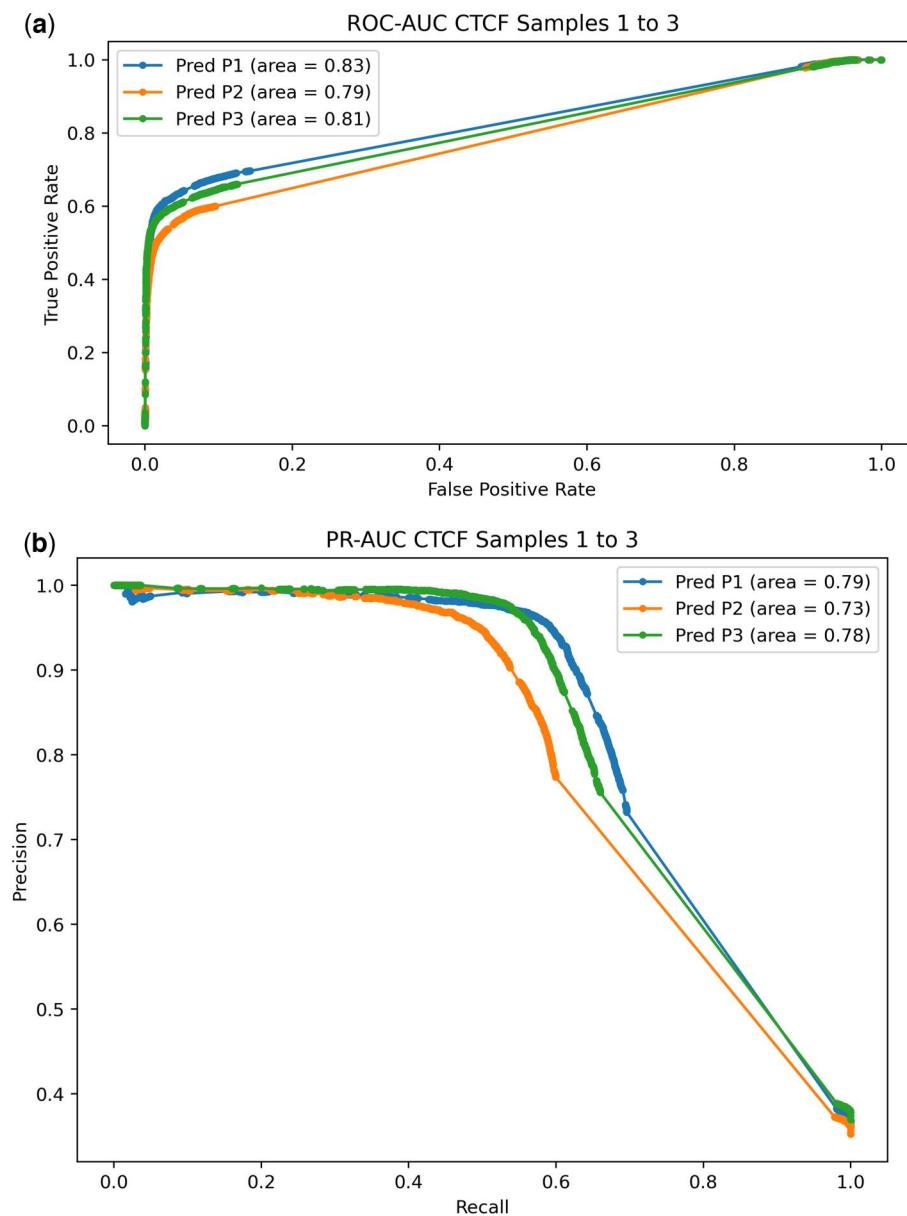
To gain insights about the contribution of the different evaluated features into the prediction, we extracted and reported the feature importance of every feature used in classification. The most important feature was the FIMO score at the start of the loop, followed by the presence of bound CTCF at the end of the loop. This is expected and it accounts for how the CTCF loop is being formed; if the start of the loop is conserved or bound, we can expect to find a loop if the end of the loop is also conserved or bound. Following these observations, after binding on the end of the loop the next important feature is FIMO score on the end, and then CTCF binding inside the loop, which may consider unaccounted CTCF binding nearby the start or the end of the loop. After those features, the following features were the loop distance and the histone marks, on any of the reported starts and ends of loops (Fig. 3A–C, available as supplementary data at *Bioinformatics Advances* online).

### 3.4 Comparison with other methods

We compared the performance of this method against two other state-of-the-art methods (Kai et al. 2018, Xu et al. 2023). We picked Lollipop (Kai et al. 2018) due to being one of the legacy CTCF loops predictor, and we used LoopAnchor (Xu et al. 2023) due to being one of the latest published CTCF loops predictor by the time of writing. We trained each predictor with the default features for each, tested training with GM12878 using default values and features for each method, and used chromosomes 2, 20, 21, and 22 of K562 cell line as a test set. We report ROCAUC and PRAUC of each predictor, finding that our predictor has a slightly lower ROCAUC than the other predictors, and a better performance in terms of PRAUC than Lollipop, but slightly worse than LoopAnchor (Fig. 4A and B, available as supplementary data at *Bioinformatics Advances* online).

### 3.5 AD CTCF binding prediction

We used FIMO to predict CTCF binding sites on the GRCh38 reference genome, and used the FIMO scores with four other different features (CTCF binding motif, DNA accessibility, h3k4me3, h3k27me3, and h3k27ac) to train a Random Forest predictor for CTCF binding on three different control AD patients, using CTCF ChIP-seq information as gold standard; 52 435 CTCF binding sites were predicted by FIMO. We first analyzed each patient by itself, splitting the whole dataset into 2/3 used for training and 1/3 used for testing and reported precision, recall, and F-score, and plotted ROCAUC and PRAUC (Table 1, Fig. 1A and B). After predicting each patient by itself, we used two patients to predict the third one, to increase the number of available examples and to confirm if the predictor was able to generalize. We reported precision, recall, and F-score, and plotted

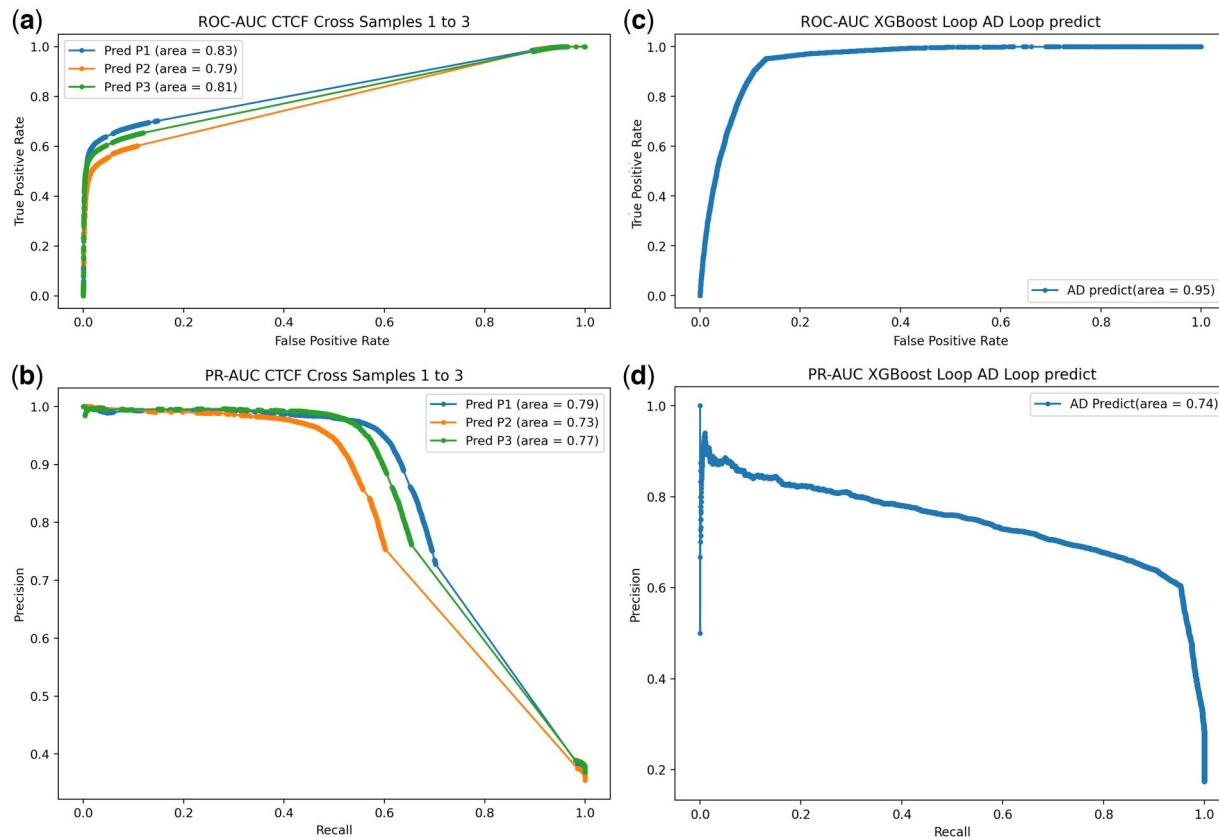


**Figure 1.** CTCF binding prediction on patients without cognitive impairment, using 2/3rds as a training set and 1/3rd as a testing set, using each patient to predict itself. (A) ROC-AUC. (B) PR-AUC.

**Table 2.** Top five most disrupted loops for gained and lost loop instances.<sup>a</sup>

chr	startloop	endloop	Predicted score	Discordant score	Gene name
chr10	124025000	124095000	8.93E-05	0.999821	CHST15
chr17	59440000	59570000	0.000123	0.999752	DHX40
chr11	70467000	70756000	0.000143	0.999713	SHANK2
chr12	40970000	41070000	0.000174	0.999651	CNTN1
chr2	74915000	75405000	0.000182	0.999635	POLE4
chr7	71750000	72640000	0.999581	0.999163	CALN1
chr7	71750000	72640000	0.999581	0.999163	TYW1B
chr6	33410000	33580000	0.999782	0.999564	SYNGAP1
chr6	33410000	33580000	0.999782	0.999564	GGNBP1
chr8	130290000	130680000	0.999925	0.999851	ASAP1

<sup>a</sup> Top five are lost loops, Bottom five are gained loops.



**Figure 2.** CTCF binding prediction on patients without cognitive impairment and patients with AD and cognitive impairment. (A) ROC-AUC predicting CTCF binding on three controls using two controls as a training set and the remaining one as a test set. (B) PR-AUC of the same controls. (C) ROC-AUC, using three control samples to predict the AD sample. (D) PR-AUC using three control samples to predict the AD sample.

ROCAUC and PRAUC (Table 1, Fig. 2A and B). For the AD dataset we trained the model, used the three patients datasets and used it to predict the patient with cognitive impairment and AD, and reported performance afterwards (Table 1, Fig. 2C and D).

### 3.6 CTCF loop prediction

For CTCF loop prediction we downloaded Hi-C loop data from ENCODE and assigned positive and negative loops considering positive loops as loops with convergent, CTCF-bound CTCF binding sites with a distance between 2 kb and 2mb, considering earlier mentioned features. We trained a XGBoost loop predictor and generated predictions for each loop using controls as a training set and AD patient as a test set. To finish, we generated scores for loops on the AD sample and saved them for posterior use.

### 3.7 Discordant loop determination

We compared the obtained loop scores between the control sample and the AD sample, and divided the scores between gained loops (loops with a low score on the control sample and high score on the AD sample) and lost loops (loops with a high score on the control sample and low score on the AD sample), and ranked them based on their discordant loop score (Table 2). The top predicted lost loops involved the genes CHST15, DHX40, SHANK2, CNTN1, and POLE4, and the top predicted gained loops contained SYNGAP1, GGNBP1, TYW1B, CALN1, and ASAP1. According to bibliography, four of the five top lost loops have been linked to

AD with supported evidence, suggesting a possible link between loop disruption and AD (Table 3).

## 4 Discussion

CTCF is related with multiple functions inside the cell and determining the full extension of its phenotypic and mechanistic roles remains to be done. Over 60% of the CTCF binding sites are conserved across different cell lines (Chen et al. 2012), and Rao et al. (2014) reported from 55% to a 75% of shared conserved peaks in HiC experiments across them. Under this context, FIMO score can be considered as a proxy of CTCF binding site conservation, and FIMO score is consistently one of the most important features in classification reported by our predictor. Both FIMO score and CTCF binding were reported as the most important features in CTCF loop classification, and due to the conservation of CTCF binding sites between different tissues and cell lines, the detection of conserved CTCF loops is possible using only the CTCF motif scores with our approach. Other methods to predict CTCF loops exist, however, they differ in both methods and performance (Kai et al. 2018, Zhang et al. 2018). One advantage of this method is prediction across cell lines, however, a performance reduction in our method is to be expected due to differences between each particular cell CTCF landscape. Nevertheless, other approaches have focused on the importance of the motif sequence and have predicted loops only training with the CTCF binding sequences (Zhang et al. 2018) and our results confirm the importance of these observations.

**Table 3.** Genes in disrupted loops and their link with AD.

Gene	AD link	Citation	DOI
CHST15	Neuropathological AD-related gene	Silva <i>et al.</i> (2012)	10.1371/journal.pone.0048751
DHX40	Age-dependent differentially methylated	Pellegrini <i>et al.</i> (2021)	10.3389/fnagi.2021.639428
SHANK2	Decreased in AD	Patel <i>et al.</i> (2023)	10.1016/j.nbd.2023.106192
CNTN1	Aggravates neuroinflammation in AD	Li <i>et al.</i> (2023)	10.14336/AD.2023.0228
POLE4	Reduced CTCF binding in AD	Patel <i>et al.</i> (2023)	10.1016/j.nbd.2023.106192
CALN1	Differentially expressed in AD	Li and De Munck (2021)	10.1016/j.bbih.2021.100227
TYW1B	Detected in AD neural networks	Ghose <i>et al.</i> (2024)	10.1093/bib/bbae704
SYNGAP1	Risk factor in GWAS AD	Mez <i>et al.</i> (2017)	10.1016/j.jalz.2016.09.002
GGNBP1	Gametogenitin-binding protein 1, unknown function, no known AD link.	—	—
ASAP1	Genetic modulator for Tau pathologies	Xu <i>et al.</i> (2024)	10.1007/s00401-024-02703-3

Cell-specific CTCF loops with lower conservation scores may require different considerations than conserved CTCF loops, for example, less conserved loops may determine specific cell processes related directly with lineage specification (Liu *et al.* 2023), and training on different cell lines may not be able to provide proper examples for training, interfering with the identification of loops across cell lines. Since CTCF cooperates with lineage-specific pioneer transcription factors (TFs), such as MyoD, FOXA, and PU.110, including expression data of these factors may improve cell-specific loop resolution. We considered only the presence of features inside the loop, or around the start or end of the loop, since we deemed the presence of a feature to be more relevant than its exact location on the sequence. However, we can increase the resolution by creating bins inside the analyzed regions of the CTCF loop to improve performance, especially if features related to opposite biological phenomena occupy the same region, such as DNA methylation and DNA expression, or h3k27ac and h3k27me3.

We reported relatively good performance using *in silico* predicted CTCF binding scores using two different cell lines. The two cell lines were selected according to the assays available at the time of the study. K562 is a widely used human leukemia cell line with an extensive genomic and epigenomic characterization (Zhou *et al.* 2019), while GM12878 is an EBV-transformed human B-lymphoblastoid cell line (Lin *et al.* 2022). It is well known that bypassing cellular senescence, either due to the tumoral cellular origin or via the immortalization process of primary cells, required additional mutation and/or epigenetic alterations. These mechanisms contribute to both inter- and intra-cell line heterogeneity and may have an influence on results outcomes (Zhu *et al.* 2023). Nonetheless, our work proposes a useful model to predict CTCF loops, which will be further strengthened as more human cell lines are characterized by multiomics profiling assays.

We considered overfitting as a possible issue, as we have evaluated the use of GM12878 multiple times during the development of this work. However, in earlier work by the same authors (Villaman *et al.* 2023), and other methods of TF binding prediction such as msCentipede (Raj *et al.* 2015) and Catchitt (Keilwagen *et al.* 2019) also report improved results on GM12878 when compared to other cell lines from the ENCODE dataset. We believe that the results are not related to overfitting, and they are related with better quality of the hg19 GM12878 dataset when compared to the rest of the ENCODE datasets.

Testing across multiple cell lines would improve the number of available examples, improving performance on shared

loops, however, an impaired performance could be expected on cell-specific loops. Including additional features such as methylation, gene expression, or DNA accessibility would contribute to identify cell-specific loops if the conservation of their CTCF binding sites is low, nonetheless loops with weaker CTCF binding would most likely be exclusive to the cell and related with its own transcriptional landscape, making possible that the inclusion of extra features does not improve the performance of the predictor. Nevertheless, the relationship between performance improvement versus performance impairment remains to be studied. We were able to predict with over 0.9 F-score on the analyzed cell lines, with the worst performance obtaining when predicting training across cell lines, highlighting the idea that conserved loops share similar features while cell-specific loops present singular features. We plan to expand the array of both features and cell lines included on CTCF loop prediction to improve prediction performance, and at the same time to provide new insights about the role of CTCF loops across the myriad of existing human cell types.

Importantly, our results suggest the existence of two different kind of CTCF binding sites, conserved ones shared across cells, and cell-exclusive CTCF binding sites. In fact, there is evidence sustaining that conserved CTCF sites have a structural function while non-conserved ones have a regulatory function (Plasschaert *et al.* 2014, Marina-Zárate *et al.* 2023). Since there is a direct relation between CTCF binding site conservation and loop presence (Zhang *et al.* 2018), loops can be divided between shared loops with high CTCF binding site conservation between cell lines, and less conserved cell-specific loops. The samples analyzed from dorsolateral prefrontal cortex were sequenced from a combination of many different cell types on different states, in both the controls and the AD and cognitive impairment samples, and it is plausible that our predictor was capable of resolving shared CTCF binding sites but not cell-exclusive CTCF binding sites. This also supports the idea that there are also two kinds of CTCF loops, ones formed between conserved CTCF binding sites and cell-exclusive loops formed between less conserved CTCF binding sites. This idea is confirmed by many different authors (Martin *et al.* 2011, Wang *et al.* 2021, Xu *et al.* 2023). If the predictor is not able to properly identify CTCF binding, it would also not be able to properly resolve the cell-exclusive CTCF loops since they are dependent on the CTCF binding score feature. Another factor is the idea that changes in control patients and AD samples are related with the cell composition and cell death inherent with AD (Johnson *et al.* 2020), however, since the CTCF loop landscape is mostly invariant, it is possible to suggest that the changes in CTCF

loops, especially those with a high CTCF binding score, account directly for AD-related abnormalities and are shared across the analyzed tissue cell types.

Another issue to consider is the fact that changes in CTCF binding and looping may not affect gene expression directly, however, loss of CTCF binding may prevent the formation of CTCF loops in response to cellular events, affecting the ability of the cell to maintain homeostasis (Kaushal *et al.* 2021, Xu *et al.* 2021), this suggesting that changes to CTCF in AD may be a cumulative trait during the development of this disease. A proper progression of CTCF loop disruption in AD cannot be established due to the lack of available data, but we expect that as more AD datasets become available, more information would be available about AD and CTCF loop during the development of the disease.

There are reports of lower CTCF binding in AD samples, suggesting altered methylation as a possible mechanism of CTCF binding disruption (Patel *et al.* 2023). Conserved CTCF binding sites are shared across cell lines, and impaired CTCF binding on conserved CTCF sites may be a hallmark of this disease. However, if the loss of CTCF binding is random or gene-specific remains to be determined. Abnormal DNA methylation is a hallmark linking CTCF and different kinds of cancer (Damaschke *et al.* 2020), this connection can be explained by the abrogation of CTCF binding by methylation and the development of cancer-specific expression landscapes leading to cancer-specific hypermethylation (Damaschke *et al.* 2020), but if the loss of CTCF in AD leads to AD-specific methylation and gene expression is not yet verified.

Importantly, we found discordant loops and divided them in two groups, loops gained on AD and loops lost in AD. We suggest that loops gained in AD are representative of the different cell composition during the development of AD while comparing AD and control samples, and loops lost on AD are directly related with impaired CTCF binding shared across cells in the dorsolateral prefrontal cortex. CTCF binds to unmethylated DNA sequences (Holwerda and De Laat 2013, Prickett *et al.* 2013, Maurano *et al.* 2015) and it can bind to DNA damage sites and activate a cascade reaction resulting in DNMT1 inactivation and DNA demethylation (Zampieri *et al.* 2012). Moreover, CTCF is also related to patterns of histone modifications and is especially required to implement both H3K27ac and H3K27me3 (Wang *et al.* 2021). While their defects may rewire genome-wide chromatin accessibility and have serious implications on the normal cell expression program (Xu *et al.* 2021). Thus suggesting that the decline in CTCF binding with age may be an initial mechanism in AD pathogenesis (Wang *et al.* 2020, Hou *et al.* 2021). One of our main issues was data availability, as the Hi-C experiments required to confirm CTCF loops are not readily available for the conditions we wanted to evaluate, and we have not evaluated our approach in other datasets again due to the earlier lack of data mentioned. While it may not be possible to extend the results to the whole spectrum of AD etiology due to an inherent lack of robustness, we can also acknowledge the fact that the genes involved in the loops predicted as lost have also been experimentally described as relevant in AD, such as SHANK2 (Patel *et al.* 2023) and POLE4 (Ferrer *et al.* 2021, Patel *et al.* 2023). Patel and collaborators also analyzed the RUSH ENCODE dataset, however, while there is an overlap in controls, there is no overlap in AD cases, and while our results could be describing an isolated case, we believe that the similarity in results highlights

and underlying mechanism and not a singular phenomenon. Nevertheless, caution should be advised while interpreting these results, and we expect to expand on them on further analyses.

We report CTCF loop loss on genes directly related with AD on AD sample dorsolateral prefrontal cortex, however, we lack enough information to suggest the mechanism behind it. There is also evidence of interactions between APOE and CTCF and it is possible for CTCF to make protein complexes directly with AB peptides or Tau hindering its ability to bind DNA (Del Moral-Morales *et al.* 2023), however, more studies on the topic are required to confirm the existence of either a pattern or a mechanism confirming the role of CTCF and CTCF looping in AD. Finally, our evidence shows concrete links between epigenetics, CTCF, and the etiology of LOAD, and the exploration of those links may provide new clues about development of LOAD and possible therapeutic targets to this disease in the future.

## 5 Conclusion

We developed a CTCF loop predictor that can use CTCF binding predictions with similar performance to other state-of-the-art predictors and used it to gain new insights on AD. We identified that CTCF binding site conservation and CTCF binding are the most important features in loop classification, consistent with observations on different cell lines. We identified that conserved CTCF binding sites and conserved CTCF binding are related with CTCF loops, and CTCF loops with weaker conservation scores and CTCF binding may be related with cell-specific CTCF binding, requiring the inclusion of more features to be properly resolved. We predicted CTCF binding and CTCF loops in three control samples and an AD sample using genetic and epigenetic features. We reported loop disruption when comparing the control and the AD sample and found that the top disrupted loops were genes that were reported as AD-related in bibliography. We expect to expand on both the amount of analyzed cell lines and features included in the nearby future to provide new knowledge about CTCF function across different cell lines. Importantly, with the currently available information it is not possible to elucidate a clear biological mechanism besides correlation, and more studies are required to confirm the extent of disruption of CTCF binding and CTCF loops in AD. Nevertheless, our observations support the fact that there is a link between CTCF and the etiology of AD.

## Acknowledgements

We thank all members in our research groups for all useful discussions, feedback and funny jokes that helped during the time spent working on this article.

## Supplementary data

Supplementary data are available at *Bioinformatics Advances* online.

## Conflict of interest

None declared.

## Funding

This work was funded by: ANID BASAL FB210008, ANID FONDECYT Exploración 13240042, 2023FIAS-MS-02, 2023FEQUIP-RB-01 (Proyectos internos VRI UCT). Proyecto Fondecyt de Postdoctorado 2025-ANID 3250632. Vicerrectoría de Investigación y Doctorados de la Universidad San Sebastián—Fondo USS-FIN-25-APCS-43. Powered@NLHPC: this research was partially supported by the supercomputing infrastructure of the NLHPC (CCSS210001).

## Data availability

The data employed in this article is available in ENCODE data repository (<https://www.encodeproject.org/>), accession IDs ENC SR000DZN, ENC SR000EGM, ENC SR000EAC, ENC SR000 FAD, ENC SR000AKF, ENC SR000AKS, ENC SR000AOX, ENC SR000APE, ENC SR000AKD, ENC SR000EWB, ENC SR 000AKC, and ENC SR000AKP. All scripts employed can be found at <https://github.com/networkbiolab/jalpy>.

## References

2024 Alzheimer's disease facts and figures. *Alzheimers Dementia* 2024; 20:3708–821. <https://doi.org/10.1002/alz.13809>

Alharbi AB, Schmitz U, Bailey CG *et al.* CTCF as a regulator of alternative splicing: new tricks for an old player. *Nucleic Acids Res* 2021; 49:7825–38. <https://doi.org/10.1093/nar/gkab520>

Atri A. The Alzheimer's disease clinical spectrum. *Med Clin North Am* 2019;103:263–93. <https://doi.org/10.1016/j.mcna.2018.10.009>

Burton T, Liang B, Dibrov A *et al.* Transforming growth factor- $\beta$ -induced transcription of the Alzheimer  $\beta$ -amyloid precursor protein gene involves interaction between the CTCF-complex and smads. *Biochem Biophys Res Commun* 2002;295:713–23. [https://doi.org/10.1016/S0006-291X\(02\)00725-8](https://doi.org/10.1016/S0006-291X(02)00725-8)

Chen H, Tian Y, Shu W *et al.* Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS One* 2012;7:e41374. <https://doi.org/10.1371/journal.pone.0041374>

Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, 785–94. <https://doi.org/10.1145/2939672.2939785>

Damaschke NA, Gawdzik J, Avilla M *et al.* CTCF loss mediates unique DNA hypermethylation landscapes in human cancers. *Clin Epigenet* 2020;12:80. <https://doi.org/10.1186/s13148-020-00869-7>

Del Moral-Morales A, Salgado-Albarrán M, Sánchez-Pérez Y *et al.* CTCF and its Multi-Partner network for chromatin regulation. *Cells* 2023;12:1357. <https://doi.org/10.3390/cells12101357>

Fang C, Wang Z, Han C *et al.* Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation. *Genome Biol* 2020;21: 247. <https://doi.org/10.1186/s13059-020-02152-7>

Ferrer I, Andrés-Benito P, Ausín K *et al.* Dysregulated protein phosphorylation: a determining condition in the continuum of brain aging and Alzheimer's disease. *Brain Pathol* 2021;31:e12996. <https://doi.org/10.1111/bpa.12996>

Fullwood MJ, Liu MH, Pan YF *et al.* An oestrogen-receptor- $\alpha$ -bound human chromatin interactome. *Nature* 2009;462:58–64. <https://doi.org/10.1038/nature08497>

Ghose U, Sproviero W, Winchester L *et al.* A. Genome-wide association neural networks identify genes linked to family history of Alzheimer's disease. *Brief Bioinform* 2024;26:bbae704. <https://doi.org/10.1093/bib/bbae704>

Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 2011;27:1017–8. <https://doi.org/10.1093/bioinformatics/btr064>

Greener JG, Kandathil SM, Moffat L *et al.* A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 2022;23:40–55. <https://doi.org/10.1038/s41580-021-00407-0>

Gruber S, Haering CH, Nasmyth K. Chromosomal cohesin forms a ring. *Cell* 2003;112:765–77. [https://doi.org/10.1016/S0092-8674\(03\)00162-4](https://doi.org/10.1016/S0092-8674(03)00162-4)

Gruber F, Srivastava R, Spacek DV *et al.* Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* 2020;583:737–43. <https://doi.org/10.1038/s41586-020-2151-x>

Hansen AS. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* 2020;11: 132–48. <https://doi.org/10.1080/19491034.2020.1782024>

Holwerda SJB, De Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci* 2013;368:20120369. <https://doi.org/10.1098/rstb.2012.0369>

Hou Y, Song Q, Gao S *et al.* CTCF mediates replicative senescence through POLD1. *Front Cell Dev Biol* 2021;9:618586. <https://doi.org/10.3389/fcell.2021.618586>

Jia Z, Li J, Ge X *et al.* Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biol* 2020;21:75. <https://doi.org/10.1186/s13059-020-01984-7>

Johnson TS, Xiang S, Helm BR *et al.* Spatial cell type composition in normal and Alzheimer's human brains is revealed using integrated mouse and human single cell RNA sequencing. *Sci Rep* 2020;10: 18014. <https://doi.org/10.1038/s41598-020-74917-w>

Kai Y, Andricovich J, Zeng Z *et al.* Predicting CTCF-mediated chromatin interactions by integrating genomic and epigenomic features. *Nat Commun* 2018;9:4221. <https://doi.org/10.1038/s41467-018-06644-6>

Kaushal A, Mohana G, Dorier J *et al.* CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. *Nat Commun* 2021;12:1011. <https://doi.org/10.1038/s41467-021-21366-2>

Keilwagen J, Posch S, Grau J. Accurate prediction of cell type-specific transcription factor binding. *Genome Biol* 2019;20:9. <https://doi.org/10.1186/s13059-018-1614-y>

Kim S, Yu N-K, Kaang B-K. CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med* 2015;47:e166. <https://doi.org/10.1038/emm.2015.33>

Kim TH, Abdullaev ZK, Smith AD *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 2007; 128:1231–45. <https://doi.org/10.1016/j.cell.2006.12.048>

Klenova EM, Nicolas RH, Paterson HF *et al.* CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-Finger protein differentially expressed in multiple forms. *Mol Cell Biol* 1993;13:7612–24. <https://doi.org/10.1128/mcb.13.12.7612-7624.1993>

Lane CA, Hardy J, Schott JM. Alzheimer's disease. *Eur J Neurol* 2018; 25:59–70. <https://doi.org/10.1111/ene.13439>

Li G, Chen Y, Snyder MP *et al.* ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. *Nucleic Acids Res* 2017;45:e4. <https://doi.org/10.1093/nar/gkw809>

Li S-J, Ma M-H, Li J-M *et al.* CNTN1 aggravates neuroinflammation and triggers cognitive deficits in male mice by boosting crosstalk between microglia and astrocytes. *Aging Dis* 2023;14:1853–69. <https://doi.org/10.14336/AD.2023.0228>

Li QS, De Muynck L. Differentially expressed genes in Alzheimer's disease highlighting the roles of microglia genes including OLR1 and astrocyte gene CDK2AP1. *Brain Behav Immun Health* 2021;13: 100227. <https://doi.org/10.1016/j.bbih.2021.100227>

Lin G, Zhang K, Han Y *et al.* Preparation of multiplexed control materials for cancer mutation analysis by genome editing in GM12878 cells. *J Clin Lab Anal* 2022;36:e24139. <https://doi.org/10.1002/jcl.24139>

Liu Y, Wan X, Li H *et al.* CTCF coordinates cell fate specification via orchestrating regulatory hubs with pioneer transcription factors. *Cell Rep* 2023;42:113259. <https://doi.org/10.1016/j.celrep.2023.113259>

López-Otín C, Blasco MA, Partridge L et al. Hallmarks of aging: an expanding universe. *Cell* 2023;186:243–78. <https://doi.org/10.1016/j.cell.2022.11.001>

Maccioni RB, Tapia JP, Guzman-Martinez L. Pathway to tau modifications and the origins of Alzheimer's disease. *Arch Med Res* 2018;49:130–1. <https://doi.org/10.1016/j.arcmed.2018.05.002>

Marina-Zárate E, Rodríguez-Ronchel A, Gómez MJ et al. Low-affinity CTCF binding drives transcriptional regulation whereas high-affinity binding encompasses architectural functions. *iScience* 2023;26:106106. <https://doi.org/10.1016/j.isci.2023.106106>

Martin D, Pantoja C, Fernández Miñán A et al. Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* 2011;18:708–14. <https://doi.org/10.1038/nsmb.2059>

Maurano MT, Wang H, John S et al. Role of DNA methylation in modulating transcription factor occupancy. *Cell Rep* 2015;12:1184–95. <https://doi.org/10.1016/j.celrep.2015.07.024>

Mez J, Chung J, Jun G et al.; Alzheimer's Disease Genetics Consortium. Two novel loci, COBL and SLC10A2, for Alzheimer's disease in African Americans. *Alzheimers Dement* 2017;13:119–29. <https://doi.org/10.1016/j.jalz.2016.09.002>

Monteagudo-Sánchez A, Noordermeer D, Greenberg MVC. The impact of DNA methylation on CTCF-mediated 3D genome organization. *Nat Struct Mol Biol* 2024;31:404–12. <https://doi.org/10.1038/s41594-024-01241-6>

Nanni L, Cieri S, Logie C. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biol* 2020;21:197. <https://doi.org/10.1186/s13059-020-02108-x>

Patel PJ, Ren Y, Yan Z. Epigenomic analysis of Alzheimer's disease brains reveals diminished CTCF binding on genes involved in synaptic organization. *Neurobiol Dis* 2023;184:106192. <https://doi.org/10.1016/j.nbd.2023.106192>

Pellegrini C, Pirazzini C, Sala C et al. A Meta-Analysis of brain DNA methylation across sex, age, and Alzheimer's disease points for accelerated epigenetic aging in neurodegeneration. *Front Aging Neurosci* 2021;13:639428. <https://doi.org/10.3389/fnagi.2021.639428>

Plasschaert RN, Vigneau S, Tempora I et al. CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. *Nucleic Acids Res* 2014;42:774–89. <https://doi.org/10.1093/nar/gkt910>

Prickett AR, Barkas N, McCole RB et al. Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. *Genome Res* 2013;23:1624–35. <https://doi.org/10.1101/gr.150136.112>

Pugacheva EM, Kubo N, Loukinov D et al. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc Natl Acad Sci USA* 2020;117:2020–31. <https://doi.org/10.1073/pnas.1911708117>

Raj A, Shim H, Gilad Y et al. msCentipede: modeling heterogeneity across genomic sites and replicates improves accuracy in the inference of transcription factor binding. *PLoS One* 2015;10:e0138030. <https://doi.org/10.1371/journal.pone.0138030>

Rao SSP, Huntley MH, Durand NC et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;159:1665–80. <https://doi.org/10.1016/j.cell.2014.11.021>

Silva ART, Grinberg LT, Farfel JM et al. Transcriptional alterations related to neuropathology and clinical manifestation of Alzheimer's disease. *PLoS One* 2012;7:e48751. <https://doi.org/10.1371/journal.pone.0048751>

Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* 2015;15:29. <https://doi.org/10.1186/s12880-015-0068-x>

Tang Z, Luo OJ, Li X et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* 2015;163:1611–27. <https://doi.org/10.1016/j.cell.2015.11.024>

Villaman C, Pollastri G, Saez M et al. Benefiting from the intrinsic role of epigenetics to predict patterns of CTCF binding. *Comput Struct Biotechnol J* 2023;21:3024–31. <https://doi.org/10.1016/j.csbj.2023.05.012>

Viola KL, Klein WL. Amyloid  $\beta$  oligomers in Alzheimer's disease pathogenesis, treatment, and diagnosis. *Acta Neuropathol* 2015;129:183–206. <https://doi.org/10.1007/s00401-015-1386-3>

Wang G, Xia B, Zhou M et al. MACMIC reveals a dual role of CTCF in epigenetic regulation of cell identity genes. *Genomics Proteomics Bioinformatics* 2021;19:140–53. <https://doi.org/10.1016/j.gpb.2020.10.008>

Wang W, Gao L, Ye Y et al. CCIP: predicting CTCF-mediated chromatin loops with transitivity. *Bioinformatics* 2021;37:4635–42. <https://doi.org/10.1093/bioinformatics/btab534>

Wang Y, Zhang X, Song Q et al. Characterization of the chromatin accessibility in an Alzheimer's disease (AD) mouse model. *Alzheimers Res Ther* 2020;12:29. <https://doi.org/10.1186/s13195-020-00598-2>

Xi W, Beer MA. Loop competition and extrusion model predicts CTCF interaction specificity. *Nat Commun* 2021;12:1046. <https://doi.org/10.1038/s41467-021-21368-0>

Xu B, Wang H, Wright S et al. Acute depletion of CTCF rewires genome-wide chromatin accessibility. *Genome Biol* 2021;22:244. <https://doi.org/10.1186/s13059-021-02466-0>

Xu H, Yi X, Fan X et al. Inferring CTCF-binding patterns and anchored loops across human tissues and cell types. *Patterns (NY)* 2023;4:100798. <https://doi.org/10.1016/j.patter.2023.100798>

Xu H, Qiu Q, Hu P et al. MSUT2 regulates tau spreading via adenosine-nergic signaling mediated ASAP1 pathway in neurons. *Acta Neuropathol* 2024;147:55. <https://doi.org/10.1007/s00401-024-02703-3>

Yang J-H, Hayano M, Griffin PT et al. Loss of epigenetic information as a cause of mammalian aging. *Cell* 2023;186:305–26.e27. <https://doi.org/10.1016/j.cell.2022.12.027>

Zampieri M, Guastafierro T, Calabrese R et al. ADP-ribose polymers localized on ctcf–Parp1–Dnmt1 complex prevent methylation of ctcf target sites. *Biochem J* 2012;441:645–52. <https://doi.org/10.1042/BJ20111417>

Zhang H, Shi Z, Banigan EJ et al. CTCF and R-loops are boundaries of cohesin-mediated DNA looping. *Mol Cell* 2023;83:2856–71.e8. <https://doi.org/10.1016/j.molcel.2023.07.006>

Zhang R, Wang Y, Yang Y et al. Predicting CTCF-mediated chromatin loops using CTCF-MP. *Bioinformatics* 2018;34:i133–41. <https://doi.org/10.1093/bioinformatics/bty248>

Zhao J, Huai J. Role of primary aging hallmarks in Alzheimer's disease. *Theranostics* 2023;13:197–230. <https://doi.org/10.7150/thno.79535>

Zhou B, Ho SS, Greer SU et al. Comprehensive, integrated, and phased whole-genome analysis of the primary ENCODE cell line K562. *Genome Res* 2019;29:472–84. <https://doi.org/10.1101/gr.234948.118>

Zhu Q, Zhao X, Zhang Y et al. Single cell multi-omics reveal intra-cell-line heterogeneity across human cancer cell lines. *Nat Commun* 2023;14:8170. <https://doi.org/10.1038/s41467-023-43991-9>