

Modeling and Oncogenic Characterization of HMGB1 Transcriptional Regulation by HTLV-1 Tax and Its M22/M47 Mutants in T Cells

Ruoting Lin¹, Jianxing Zeng², Shuang Zhou³, Hong-zhi Gao^{4*}, Ruoteng Xie^{2*}

Department of Neurology, Second Affiliated Hospital of Fujian Medical University, No.950, Donghai Street, Fengze District, Quanzhou, 362000 Fujian Province, China.

Department of Laboratory Medicine, Shanghai sixth People's Hospital Fujian Campus, No.16 Jin Guang Road, Luosheng Street, Jinjiang City, Quanzhou, 362200 Fujian Province, China.

Clinic center of Molecular Diagnosis and Therapy, Second Affiliated Hospital of Fujian Medical University, No.950, Donghai Street, Fengze District, Quanzhou, 362000 Fujian Province, China.

Department of Neurosurgery, Clinic center of Molecular Diagnosis and Therapy, Second Affiliated Hospital of Fujian Medical University, No.950, Donghai Street, Fengze District, Quanzhou, 362000 Fujian Province, China.

Department of Neurosurgery, Clinic center of Molecular Diagnosis and Therapy, Second Affiliated Hospital of Fujian Medical University, No.950, Donghai Street, Fengze District, Quanzhou, 362000 Fujian Province, China.

Department of Laboratory Medicine, Second Affiliated Hospital of Fujian Medical University, No.950, Donghai Street, Fengze District, Quanzhou, 362000 Fujian Province, China.

CITATION

Ruoting L, Jianxing L. et al. Modeling and Oncogenic Characterization of HMGB1 Transcriptional Regulation by HTLV-1 Tax and Its M22/M47 Mutants in T Cells *Psycho-Oncologie*.

ARTICLE INFO

Received:

Accepted:

Available online:

COPYRIGHT



Copyright © 2025 by author(s).

Psycho-Oncologie is published by PisecoMed Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.

<https://creativecommons.org/licenses/by/4.0/>

Abstract: Background: HMGB1 is a key nuclear protein involved in inflammation, cancer and viral infections. The HTLV-1 virus produces a protein called Tax, which can strongly influence the activity of host genes. This study explores how Tax and its two functional mutants (M22 and M47) affects the expression of HMGB1 in T cells. Methodology: Refined 3D structures of the normal (wild-type) Tax protein and its mutants, M22 (NF- κ B inactive) and M47 (CREB inactive) were modeled followed by apo molecular dynamics simulations to see how these proteins behave in a cellular environment. Protein- Protein docking, trajectory analysis and MD simulation was performed to check the interaction and stability of variant with NF- κ B and CREB proteins. The promoter region of the HMGB1 gene was scanned to identify potential binding sites for relevant transcription factors. Additionally, RNA-seq data from HTLV-1-infected cell lines (MT1 and KK1) were analyzed to evaluate the relationship between Tax expression and HMGB1 activity. Transcriptomic and pathway-level relevance of HMGB1 in cancer was further explored for enrichment, expression, survival and interaction analyses.

Results: The M22 and M47 mutants showed distinct structural changes around their mutation sites. Docking analysis revealed that M22 had reduced ability to bind NF- κ B while M47 showed stronger interactions with CREB. Further simulations confirmed that M22 formed more rigid but stable complexes with NF- κ B whereas M47 formed more flexible yet consistently interacting complexes with CREB. Seven strong NF- κ B and CREB1 binding motifs in the HMGB1 promoter region were identified. Gene expression analysis showed that HMGB1 levels were consistently higher in cells expressing Tax but the levels of CREB1 and RELA (a major NF- κ B subunit) remained stable. Enrichment analy-

sis supported CREB1 and RELA as the main transcription factors and HMGB1's involvement in NF- κ B signaling, autophagy and cancer-associated immune pathways. Pan-cancer datasets showed upregulation across multiple tumors and poor prognosis in LUAD, LIHC and SARC. Interaction analysis revealed TLR2/4, TP53 and BECN1 as consistently co-expressed HMGB1-binding partners across cancer types. Conclusion: These findings confirm that M22 mutation disrupts NF- κ B binding and the M47 mutation enhances CREB-mediated transcription. This highlights that these specific Tax mutations alter its ability to interact with key transcription factors leading to distinct patterns and potential immune-oncogenic role of HMGB1 gene regulation in HTLV-1-associated and epithelial cancers.

Keywords: HTLV-1, Transcriptional modulation, Protein-protein interaction, CREB signaling, Viral oncogenesis

Introduction

High-mobility group box 1 (HMGB1) is a multifunctional, non-histone chromatin-associated protein that plays a pivotal role in DNA architecture, transcriptional regulation and cellular stress responses. In the immune system, HMGB1 acts as both a nuclear regulator and an extracellular damage-associated molecular pattern (DAMP) molecule that can influence inflammatory cascades, cytokine release and immune cell activation ^[1, 2]. Within T lymphocytes, HMGB1 is increasingly recognized for its involvement in regulating gene expression related to cell survival, activation and proliferation. Its dysregulation has been implicated in several inflammatory and neoplastic diseases including autoimmune conditions, sepsis and hematological malignancies ^[3]. However, its regulation under the influence of viral oncoproteins, particularly in the context of retroviral infections, remains insufficiently characterized.

Human T-lymphotropic virus type 1 (HTLV-1) is an oncogenic retrovirus associated with adult T-cell leukemia/lymphoma (ATLL) and HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) ^[4]. HTLV-1 preferentially infects CD4⁺ T lymphocytes and promotes cellular transformation via its regulatory protein Tax. The Tax protein is a potent trans-activator that disrupts host transcriptional and signaling programs through aberrant modulation of key pathways including NF- κ B, CREB/ATF, AP-1 and p53 ^[5]. Through these interactions, Tax reprograms infected T cells toward uncontrolled proliferation, resistance to apoptosis and genomic instability. Tax-driven hyperactivation of transcriptional networks not only underpins the leukemogenic potential of HTLV-1 but also significantly perturbs immune cell homeostasis ^[6]. Tax exerts its transcriptional influence by functioning as a molecular bridge between host transcription factors and the cellular promoter/enhancer elements of specific genes. Among the most well-characterized targets of Tax are the NF- κ B and CREB/ATF signaling pathways ^[7]. Tax promotes constitutive NF- κ B activation through direct interaction with I κ B kinase (IKK) subunits, particularly IKK γ (NEMO), which leads to the

nuclear translocation of RELA/p65 and transcription of pro-survival genes. Concurrently, Tax associates with the CREB transcription factor and enhances the recruitment of transcriptional coactivators such as CBP/p300 to cAMP response elements (CREs) on viral and cellular promoters ^[8]. These dual functions enable Tax to simultaneously modulate multiple gene networks essential for T-cell transformation and viral persistence.

Given the central role of both NF- κ B and CREB in regulating HMGB1 transcription, it is hypothesized that Tax may influence HMGB1 expression by orchestrating the assembly of these transcriptional complexes on its promoter ^[9]. Although previous studies have demonstrated upregulation of HMGB1 in virally transformed and leukemic cells, the mechanistic underpinnings involving HTLV-1 Tax have not been fully elucidated. Importantly, mutations in the Tax protein that selectively impair its interaction with either NF- κ B or CREB pathways have served as powerful tools to dissect its functional domains. Among these, two well-characterized mutants, M22 and M47 have been instrumental in defining pathway-specific transactivation roles ^[10]. The M22 mutant selectively disrupts Tax's ability to activate the NF- κ B pathway while preserving CREB transactivation, whereas M47 abrogates CREB-mediated transcription without affecting NF- κ B activation ^[11]. These mutants provide a unique platform to delineate the individual contributions of NF- κ B and CREB in the Tax-mediated regulation of host genes such as HMGB1. A schematic model summarizing the proposed regulatory effects of these variants on HMGB1 transcription is shown in **Fig. 1**. Despite the availability of these functional mutants, little is known about their structural and transcriptional impact on HMGB1 regulation in the context of HTLV-1–infected T cells. No study to date has comprehensively examined the molecular docking interfaces, promoter binding potential, or transcription factor accessibility of these mutant complexes relative to wild-type Tax. Furthermore, experimental limitations in accessing mutant-specific transcriptomic data pose a challenge to mechanistic validation. In this context, *in silico* structural biology, molecular dynamics simulations, transcription factor motif analyses and transcriptome validation provide powerful complementary tools to investigate these regulatory mechanisms.

In this study, we present an integrative *in silico* approach to explore the transcriptional regulation of HMGB1 by HTLV-1 Tax protein and its functional M22 and M47 mutants in T lymphocytes. We first retrieved and modeled the three-dimensional structures of the wild-type and mutant Tax proteins, followed by protein-protein docking with key transcriptional mediators (NF- κ B and CREB). Complex stability was evaluated through molecular dynamics simulations and promoter-level interactions was assessed via motif scanning of the HMGB1 upstream regulatory region. Additionally, publicly available transcriptomic dataset from HTLV-1–infected cells was integrated to explore the transcriptional correlation between HMGB1 and its regulatory pathways. Our results provide mechanistic insight into how the bifurcation of NF- κ B and CREB

pathways by Tax mutants differentially influences HMGB1 regulation, thus offering a structural and functional basis for understanding HMGB1 dysregulation in HTLV-1–infected T cells.

Methodology

This study did not involve any human participants, animal models, or in vivo experimentation. All analyses were conducted using publicly available in silico transcriptomic dataset, computational tools and in vitro functional validation. Ethical approval was not required as the work exclusively utilized de-identified, open-access data (GSE108601) from the NCBI GEO.

Structural Modeling and Mutagenesis of HTLV-1 Tax Variants

The amino acid sequence of the HTLV-1 Tax protein was retrieved from the UniProt Knowledgebase (UniProt ID: P03409) in FASTA format. This full-length sequence comprises 353 amino acids, corresponding to the trans-activator protein Tax encoded by the HTLV-1. Tax is known to exert transcriptional regulation primarily via interaction with transcriptional coactivators including CREB and NF- κ B^[12]. To investigate the effect of well-characterized functional mutants, two mutant variants of Tax were generated. The three-dimensional structure of the wild-type HTLV-1 Tax protein was predicted using the AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk>) via its ColabFold implementation^[13]. The highest-confidence model was selected based on predicted Local Distance Difference Test (pLDDT) scores, with most residues scoring above 70, indicative of a high-confidence backbone conformation. The M22 mutant, known to impair NF- κ B activation, involves a double substitution at positions 130 and 131 of the wild-type sequence, where threonine (T) and leucine (L) are replaced by alanine (A) and serine (S), respectively (T130A and L131S)^[14]. The M47 mutant, which disrupts CREB-mediated transcription, involves a double mutation at positions 319 and 320, substituting leucine (L) with arginine (R) and leucine (L) with serine (S), respectively (L319R and L320S)^[15]. These mutations were manually introduced using UCSF ChimeraX (version 1.8), following literature-reported coordinates and functional validation. Subsequent energy minimization and atomic-level refinement were performed for all three models (WT, M22 and M47) using the GalaxyRefine module of GalaxyWEB^[16]. This method improves local backbone and side-chain conformations through iterative rebuilding and repacking. Among the five models generated per structure, model 1 was selected in each case based on overall structural quality and clash minimization. To assess the stereochemical quality of the predicted and mutant models, structural validation was performed using MolProbity (<http://molprobity.biochem.duke.edu/>)^[17]. Evaluation metrics included Ramachandran plot statistics, clash score and rotamer outlier frequencies were employed for comparison of the structural integrity across wild-type and mutant proteins. All models were visually inspected

in UCSF ChimeraX for residue-level anomalies or disruption in secondary structural elements.

Apo Molecular Dynamics Simulation

To evaluate the intrinsic structural stability and conformational dynamics of the wild-type (WT) HTLV-1 Tax protein and its two functional mutants (M22 and M47) in the absence of interacting partners, molecular dynamics simulations were conducted for each protein in their apo forms. All simulations were performed using the AMBER24 simulation package ^[18]. Each system was subjected to a 50ns production run within the isothermal-isobaric (NPT) ensemble at 310 K and 1 atm pressure. The proteins were first solvated in an explicit TIP3P water model using a truncated octahedral periodic box to ensure a minimum distance of 10 Å between the protein surface and the box edge. Counterions were added to neutralize the system and 150 mM NaCl was introduced to emulate physiological ionic strength. Long-range electrostatics were treated using the particle mesh Ewald (PME) method and all covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm, allowing for a 2fs time step. Prior to the production run, each system underwent a two-stage energy minimization protocol followed by gradual heating from 0 K to 310 K over 200 ps under constant volume (NVT), and subsequent equilibration for 500 ps in the NPT ensemble. The 50 ns production simulations were then carried out with trajectories saved every 10 ps for downstream analyses.

Identification of Regulatory Motifs in the HMGB1 Promoter Region

To investigate the potential regulatory influence of NF-κB and CREB transcription factors on the expression of HMGB1, a promoter and transcription factor binding site (TFBS) analysis was conducted. The upstream promoter region of the HMGB1 gene was retrieved using the UCSC Genome Browser (<https://genome.ucsc.edu/>), with reference to the GRCh38/hg38 human genome assembly. Based on the transcription start site (TSS) annotation, a 2 kb upstream region was extracted corresponding to the coordinates chr13:30,466,327–30,468,327 on the reverse strand. To scan the retrieved HMGB1 upstream region for known TFBS motifs of NF-κB and CREB, the Find Individual Motif Occurrences (FIMO) tool, part of the MEME Suite (<https://meme-suite.org/>), was employed ^[19]. A custom MEME-formatted motif file containing position-specific scoring matrices (PSSMs) for NF-κB (MA0105.4) and CREB1 (MA0602.1) was generated using publicly available matrix data from JASPAR. The matrix included 10-11 bp wide motifs with normalized position probability values and default nucleotide background frequencies. The motif and sequence files were uploaded to FIMO with statistical significance threshold parameters ($p\text{-value} \leq 1e-3$). The analysis was executed on MEME Suite version 5.5.8. The resulting output included all significant matches to NF-κB and CREB1 motifs within the 2 kb upstream region of

HMGB1.

Protein–Protein Docking and MD Simulation of Tax Transcription Factor Complexes

To investigate the structural basis of altered transcriptional regulation by HTLV-1 Tax mutants, protein–protein docking was performed between the modeled Tax variants and the transcription factors NF- κ B and CREB, both of which were identified as potential regulators of the HMGB1 gene in prior motif analysis. The M22 mutant was docked with NF- κ B (PDB ID: 1VKX), while the M47 mutant was docked with CREB (PDB ID: 1KDX) based on their respective roles in modulating the NF- κ B and CREB pathways.

Rigid-body docking was carried out using the ClusPro server, which employs fast Fourier transform (FFT)-based global docking followed by energy-based clustering of docked poses^[20]. For each complex, the top-ranked cluster based on balanced scoring (accounting for electrostatic, desolvation, and van der Waals interactions) was selected for further analysis. Docked complexes were visually inspected to ensure physiologically plausible interface orientations consistent with known transcriptional coactivator interactions. To evaluate the dynamic stability and interface integrity of the docked complexes under near-physiological conditions, 100-nanosecond molecular dynamics (MD) simulations were performed for the M22–NF- κ B and M47–CREB complexes using the AMBER24 simulation suite. Each complex was solvated in a TIP3P water box with a 10 Å buffer and neutralized with counterions. A salt concentration of 150 mM NaCl was added to mimic physiological ionic strength. After energy minimization and gradual heating to 310 K, systems were equilibrated under the NPT ensemble and subjected to 100 ns production simulations. Long-range electrostatics were handled via the Particle Mesh Ewald method, and covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm, allowing a 2fs time step. To elucidate the molecular interface between HTLV-1 Tax functional mutants and their respective transcription factor partners, protein–protein interaction diagrams were generated using LigPlot+.

Functional Transcriptomic Analysis

To validate the transcriptional regulation of HMGB1 by HTLV-1 Tax, we analyzed bulk RNA-seq data from FACS-sorted Tax(–) and Tax(+) ATLL cell populations (MT1 and KK1 cell lines; GSE108601). Both cell lines demonstrated differential expression of HMGB1, RELA, and CREB1 in Tax(+) cells, consistent with the modeled transcriptional activation via NF- κ B and CREB1 pathways. For each line, Tax-positive (GFP⁺) and Tax-negative (GFP[–]) cells were sorted and sequenced, resulting in four total conditions: MT1_Tax⁺, MT1_Tax[–], KK1_Tax⁺, and KK1_Tax[–]. CuffDiff output files for each cell line were downloaded and imported into R for processing. Gene expression values in fragments per kilobase of exon per million reads mapped (FPKM) and log₂ fold changes were extracted

for both MT1 and KK1. The datasets were then merged using RefSeq transcript accession identifiers. To annotate gene symbols, RefSeq IDs were mapped to HGNC symbols using the org.Hs.eg.db annotation package ^[21].

Expression values and log₂ fold changes were visualized to compare transcriptional activity between Tax⁺ and Tax⁻ conditions in both MT1 and KK1 cell lines. A scatterplot of global log₂ fold changes was generated to assess transcriptome-wide concordance and differential behavior across both lines. Subsequently, specific genes of interest were selected for targeted analysis. This included the direct target HMGB1, and transcription factors RELA (NF-κB subunit) and CREB1, all previously implicated in Tax-mediated transcriptional regulation. To further explore pathways influenced by Tax activity, genes upregulated specifically in KK1 Tax⁺ cells (log₂FC > 1 and FPKM > 1) were isolated for transcription factor enrichment analysis ^[22]. Top hits in the “TRRUST” and “ENCODE and ChEA Consensus TFs” libraries included RELA, CREB1, NFKB1 and ATF3 which aligned well with our functional hypotheses and docking analyses.

HMGB1 Functional Enrichment Analysis

To explore the functional landscape and oncogenic context of HMGB1, enrichment analysis was performed using the Enrichr web server (<https://maayan-lab.cloud/Enrichr/>) ^[23]. The human HMGB1 gene was used to identify functionally associated gene networks based on curated datasets of co-expression and protein–protein interactions. Enrichments were computed using four major libraries i.e., Reactome 2024, KEGG Human 2021, Gene Ontology (GO) Biological Processes 2025 and Tissues Curated 2025. Enrichr computes enrichment using Fisher’s exact test, and results were presented as adjusted p-values (Benjamini–Hochberg correction), odds ratios and combined scores (based on z-scores and p-values). Only significantly enriched terms (adjusted p < 0.05) were considered for interpretation.

HMGB1 Expression, Prognostic and Correlation Analysis

Gene expression, survival and gene–gene correlation analyses were performed using the GEPIA v1 web platform (<http://gepia.cancer-pku.cn>) ^[24]. For gene expression profiling, the “Expression DIY” module was used with default parameters, including log₂(TPM+1) normalization, comparison across all TCGA tumor types and matched GTEx normal tissue controls. For gene correlation analysis, Spearman’s correlation coefficient was calculated between HMGB1 and EGFR (as a positive control oncogene) across the combined TCGA tumor and GTEx normal tissue datasets. To assess the prognostic relevance of HMGB1 across multiple cancer types, we also utilized the Pan-cancer RNA-seq module of the KM Plotter online tool (<http://kmplot.com/analysis/>) ^[25]. The overall survival (OS) parameter was selected as the survival endpoint. No additional filters or subtype restrictions were applied to preserve pan-cohort comparability. Hazard ratios (HR), 95% confidence intervals (CI) and log-rank p-values were automat-

ically computed and plotted. The cancers analyzed include breast cancer (BC), head and neck squamous cell carcinoma (HNSC), liver hepatocellular carcinoma (LIHC), sarcoma (SARC) and lung adenocarcinoma (LUAD).

PPI Network Construction and Correlation Analysis

To analyze the protein interactome of HMGB1 and explore its context-specific regulatory partners in cancer, a protein–protein interaction (PPI) network was constructed using Pathway Commons (v12) (<https://www.pathwaycommons.org/>)^[26]. A total of 29 direct interactors of HMGB1 were extracted from the merged pathway data. Subsequently, we performed Spearman correlation analysis between HMGB1 and its interacting partners across different cancer types from the TCGA Pan-Cancer Atlas using the TIMER2.0 gene module (<https://timer.cistrome.org/>)^[27]. Normalized transcriptomic (TPM) data were used for each tumor type, and correlation significance was determined at $p \leq 0.05$. A correlation matrix was visualized as a heatmap, color-coded by Spearman's rho values, with statistically significant correlations highlighted by filled squares.

Cell Line Acquisition and Culture

The MT1 and KK1 T lymphocyte cell lines, known to be HTLV-1-infected, were purchased. These cell lines are commonly used to study the effects of HTLV-1 Tax protein in T cells. Upon receipt, cells were thawed rapidly in a 37°C water bath for 1–2 minutes and transferred to 10 mL of pre-warmed complete culture medium consisting of RPMI-1640 supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 2 mM L-glutamine. The cells were cultured at 37°C in a 5% CO₂ incubator, with subculture performed every 2–3 days to ensure optimal growth conditions.

Plasmid Construction and Transfection

To investigate the effects of HTLV-1 Tax on HMGB1 expression, the wild-type (WT) Tax construct and its M22 (NF-κB inactive) and M47 (CREB-selective) mutants were cloned into expression vectors (pcDNA3.1, Thermo Fisher Scientific). These constructs were then transfected into MT1 or KK1 cells using a transfection reagent, such as Lipofectamine® 3000 (Thermo Fisher Scientific), according to the manufacturer's protocol. To monitor transfection efficiency, a GFP reporter plasmid was co-transfected. After 24–48 hours post-transfection, the cells were harvested for subsequent analysis.

RNA Extraction and RT-qPCR Analysis

To measure HMGB1 mRNA expression, total RNA was extracted from Tax-(+) and Tax(-) cells using the TRIzol reagent or RNeasy Mini Kit (Qiagen), following the manufacturer's instructions. RNA quality and quantity were assessed using a NanoDrop spectrophotometer. Reverse transcription was performed using SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific) to synthesize cDNA. The cDNA was then used for qRT-PCR with HMGB1-specific primers

and a housekeeping gene (GAPDH). PCR amplification was performed using SYBR™ Green PCR Master Mix (Thermo Fisher Scientific) on an Applied Biosystems QuantStudio real-time PCR machine. The $\Delta\Delta C_t$ method was employed to quantify the relative expression of HMGB1 mRNA in Tax-(+) and Tax(-) cells. The results were expressed as fold changes in expression. Following primers were used for amplification purpose.

GAPDH-F 5'-ACCCACTCCTCCACCTTTGAC-3',

GAPDH-R 5'-CTGTTGCTGTAGCCAAATTCG-3'

HMGB1-F: 5'-GCGAAGAACTGGGAGAGATGTG-3'

HMGB1-R: 5'-GCATCAGGCTTTCCTTTAGCTCG-3'

Western Blot Analysis of HMGB1 Protein Expression

To assess HMGB1 protein levels, total protein was extracted from transfected cells using RIPA buffer (Thermo Fisher Scientific) supplemented with protease and phosphatase inhibitors (Thermo Fisher Scientific). The protein concentration was determined using the BCA Protein Assay Kit (Thermo Fisher Scientific). Equal amounts of protein (30 μ g per lane) were resolved by SDS-PAGE on a 10% gel and transferred onto a nitrocellulose membrane (Thermo Fisher Scientific). Membranes were blocked in 5% non-fat milk (Thermo Fisher Scientific) for 1 hour at room temperature, followed by overnight incubation at 4°C with a primary antibody against HMGB1 (Abcam, ab79823) and a GAPDH antibody (Thermo Fisher Scientific, MA5-15738) for normalization. After incubation with an appropriate HRP-conjugated secondary antibody (Thermo Fisher Scientific, 31460), protein bands were visualized using ECL chemiluminescent substrate (Thermo Fisher Scientific) and detected with a ChemiDoc imaging system (Bio-Rad). The HMGB1 protein expression levels were normalized to GAPDH levels and quantified using ImageJ software (National Institutes of Health).

Statistical Analysis

All transcriptomic data analyses were conducted using R (version 4.3.1) and Bioconductor packages. Gene expression values were reported as fragments per kilobase of exon per million mapped reads (FPKM), and differential expression between Tax⁺ and Tax⁻ groups was assessed using log₂ fold changes calculated from CuffDiff outputs. Transcription factor enrichment analysis was performed using the Enrichr web tool, with significance assessed using adjusted p-values based on the combined score metric. For motif enrichment, the FIMO tool (MEME Suite v5.5.8) was used with a p-value threshold of 0.001. Reported motif matches for CREB1 and NF- κ B were filtered by p-value and ranked by q-value for inclusion. All MD simulation results were analyzed using standard structural metrics (RMSD, interface contacts), and figures were visualized using UCSF ChimeraX (version 1.8). Statistical comparisons between multiple groups (HTLV-1 Tax(+), Tax(-), and HTLV-1-W) were conducted using one-way analysis of variance (ANOVA) followed by Tukey's post hoc test to identify significant differences between specific groups. For comparisons involving only two groups,

an unpaired two-tailed Student's t-test was used. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Results

Structural Modeling and Stereochemical Validation of HTLV-1 Tax Variants

The HTLV-1 Tax WT, M22, and M47 mutant structures were refined using GalaxyRefine to optimize side-chain packing and minimize steric clashes introduced during initial homology modeling. For each variant, five refined models were generated and Model 1 was selected based on superior backbone geometry, minimized steric hindrance and favorable global quality metrics. Stereochemical validation using MolProbity confirmed high model quality across all variants (Fig. 2A–C). The wild-type Tax protein exhibited excellent geometry, with 99.4% (349/351) of residues in the favored regions of the Ramachandran plot and only two outliers (Leu124 and Glu345). The MolProbity clashscore for the WT model was 3.28, indicating minimal steric interference. The M22 mutant showed 98.9% of residues in favored regions with a single outlier at Glu341 and a comparably low clashscore. Similarly, the M47 variant displayed 99.1% favored residues and one outlier at Phe13, with no evidence of major steric clashes or geometric distortions.

Comparative Conformational Dynamics of HTLV-1 Tax WT and Mutants in Apo State

The Root Mean Square Deviation (RMSD) trajectories (Fig. 3A) revealed early stabilization of the wild-type (WT) structure, reaching equilibrium within the first 10 ns and maintaining low average fluctuations (~ 2.5 Å) throughout the simulation. In contrast, M22 exhibited higher structural drift (~ 3.5 Å) with late stabilization which indicate increased conformational rearrangement. M47 showed moderate RMSD values but with greater trajectory variance, suggesting localized instability in certain regions.

Root Mean Square Fluctuation (RMSF) analysis (Fig. 3B) highlighted differential flexibility across the three variants. WT showed limited fluctuations in loop regions, particularly near the C-terminus. M22 exhibited elevated RMSF peaks across residues 120–140, aligning with the mutation site and its putative disruption of the NF- κ B interaction interface. M47 displayed notable fluctuations around residues 310–330, corresponding to its CREB-interacting region, consistent with structural destabilization introduced by the L319R/L320S substitutions. The radius of gyration (R_g) plots (Fig. 3C) revealed that WT maintained a compact fold (~ 22.5 Å), while both mutants demonstrated increased R_g values, particularly M22 (~ 24.3 Å), reflecting a partial unfolding or loosening of the tertiary structure. Structural superposition of the final simulated conformations (Fig. 3D) visually confirmed these observations. The WT structure preserved its initial fold, while M22 showed marked deviation in helices and coil rearrangements near the

mutation site. M47 exhibited moderate structural displacement, particularly at the C-terminal region. Further analysis of intra-protein hydrogen bonding (Fig. S1A) indicated that WT retained the highest number of stable hydrogen bonds throughout the simulation. Both M22 and M47 showed reduced intra-H bonding, particularly M22. Principal component analysis (PCA) of the top two eigenvectors (Fig. S1B-D) illustrated broader conformational sampling in the mutants. The phase space distribution of M22 showed greater dispersion along PC1 and PC2, whereas WT occupied a more confined and energetically stable conformational space. M47 exhibited intermediate sampling between WT and M22.

Motif Discovery and Transcription Factor Binding Site Analysis

Motif scanning of the 2 kb upstream promoter region of the HMGB1 gene using FIMO v5.5.8 identified a total of seven high-confidence binding sites corresponding to the CREB1 (MA0602.1) and NF- κ B (RELA; MA0105.4) transcription factor motifs, each meeting the significance threshold of $p < 0.001$. Four CREB1-binding motifs were detected on both DNA strands, with the most statistically significant match located on the negative strand at chr13:30464923–30464932 ($p = 2.05 \times 10^{-4}$; $q = 0.404$). Additionally, three NF- κ B motifs were identified, all located on the positive strand, with strong matches at chr13:30464658–30464668 and chr13:30464969–30464979. Although q -values varied, all p -values were well below the defined cutoff, indicating robust motif enrichment (Table S1). These findings support the hypothesis that both CREB1 and NF- κ B may directly interact with the HMGB1 promoter. The presence of these binding sites within the examined sequence window provides a mechanistic basis for the subsequent structure-based modeling and molecular dynamics simulations exploring Tax-mediated transcriptional modulation via these transcription factors.

Docking Analysis

In the Tax WT–NF- κ B complex (Fig. 4A), the interface exhibited a compact and stable interaction surface, with well-defined contact zones involving the DNA-binding domain of NF- κ B and the activation domain of Tax. In contrast, the M22–NF- κ B complex (Fig. 4B) showed an altered binding orientation and more dispersed contact interface, suggesting possible steric or electrostatic incompatibility introduced by the M22 mutation. These structural differences are consistent with the known transcriptionally inactive phenotype of M22, which lacks NF- κ B pathway activation. For the Tax–CREB complexes, the WT–CREB structure (Fig. 4C) demonstrated moderate interface complementarity, whereas the M47–CREB complex (Fig. 4D) formed a more extensive and tightly packed interface, with improved surface burial and alignment along the transcriptional activation region. These features are suggestive of enhanced binding affinity and support the reported CREB-selective functionality of M47. As summarized in Supplementary Fig. S2, the M22–NF- κ B complex exhibited a less favorable Center Weighted Score (–1123.2) compared to the WT–NF- κ B complex (–1216.6)

which reflects weaker average binding. However, both displayed similar Lowest Energy Scores (−1360.5 for M22 and −1375.0 for WT), suggesting that while optimal binding poses are theoretically achievable for M22, they are less consistently sampled. For the CREB complexes, the M47 mutant demonstrated a markedly stronger interaction than WT, with both its Center Weighted Score (−1359.0) and Lowest Energy Score (−1449.3) significantly more negative than those of the WT–CREB complex (−1216.6 and −1375.0, respectively), indicating enhanced binding stability.

Interaction Analysis and MD Simulation Analysis

As shown in Fig. 5A, the M22–NF- κ B complex demonstrates a robust hydrogen bonding network involving key residues such as Gly268, Asp291 and Asn65. These contacts contribute to a compact interface and suggest a stable and potentially high-affinity interaction. In contrast, Fig. 5B illustrates the interaction landscape of the M47–CREB complex, where residues such as Ser360, Gly314 and Tyr305 participate in both hydrogen bonding and van der Waals interactions. The engagement of these residues at the docking interface supports the functional relevance of the M47 mutant in modulating CREB-dependent transcription. The presence of distinct yet consistent interaction profiles in both complexes supports the structural feasibility of altered transcriptional regulation mediated by Tax mutants.

Several structural parameters were evaluated over a 100 ns MD simulation to assess the stability and interaction dynamics of the mutant Tax–transcription factor complexes. In Fig. 6A, the M22–NF- κ B complex exhibited rapid stabilization within the first 20 ns, with the RMSD plateauing around ~0.25 nm and minimal fluctuation throughout the simulation, indicating a highly stable binding configuration. In contrast, the M47–CREB complex displayed slightly higher RMSD values, ranging from ~0.3 to 0.35 nm, and moderate fluctuations, suggesting a more flexible yet conformationally stable structure. The Rg profile of this complex also remained relatively stable, indicating no major structural collapse. Fig. 6B presents the RMSF profiles, showing residue-level fluctuations. The M47–CREB complex exhibited elevated RMSF values in several regions, particularly those corresponding to loops and unstructured segments. Conversely, the M22–NF- κ B complex demonstrated overall lower RMSF values, consistent with a more rigid and tightly packed interaction surface. As depicted in Fig. 6C, the radius of gyration (Rg) for both complexes remained relatively stable throughout the 100 ns simulation, indicating the absence of major conformational collapse. The CREB+M47 complex showed slightly higher Rg values and greater fluctuations showing increased flexibility and a higher rate of structural vibration compared to the more compact NF+M22 complex. In Fig. 6D, the intra-complex hydrogen bonding profile further highlights the differential stability between the two complexes. The M22–NF- κ B complex maintained a relatively constant number of hydrogen bonds (approximately 10–15) throughout the simulation,

indicating persistent polar interactions stabilizing the interface. The M47–CREB complex formed slightly fewer hydrogen bonds (mostly ranging from 7–12) but displayed a stable contact profile over time, suggesting a dynamic yet non-transient interaction. The M22–NF- κ B complex appears more rigid and structurally constrained, while the M47–CREB complex adopts a comparatively flexible yet stable conformation. These differences may underlie the differential transcriptional regulation of HMGB1 via NF- κ B and CREB signaling pathways.

Transcriptomic Analysis using Expression Datasets

Initial integration of the MT1 and KK1 datasets revealed widespread transcriptional modulation associated with HTLV-1 Tax expression. The scatterplot of $\log_2$ fold changes between MT1 and KK1 demonstrated a dense cluster near the origin, with several transcripts exhibiting asymmetric differential expression, suggesting both shared and cell line-specific Tax-responsive genes (Fig. 7A). Targeted expression analysis showed that HMGB1 exhibited consistently higher FPKM levels in Tax⁺ cells compared to Tax⁻ in both MT1 and KK1 lines (Fig. 7B). This transcriptional upregulation supports the hypothesis that Tax facilitates HMGB1 activation. In contrast, CREB1 and RELA mRNA levels were relatively stable or modestly downregulated in Tax⁺ cells, consistent with their regulation being primarily post-translational. These observations validate the selection of HMGB1 as a transcriptionally regulated output of Tax signaling, whereas CREB1 and RELA are likely downstream modulators involved at the protein-protein interaction or phosphorylation level. We further employed Enrichr-based transcription factor enrichment analysis to specifically investigate the regulatory influence of CREB and NF- κ B pathways on the Tax-driven gene expression profile. Fig. 7C displays the transcription factor enrichment results from the ENCODE and ChEA Consensus TFs database. This analysis highlights RELA, CREB1, and ATF3 as top enriched transcription factors based on ChIP-X evidence from genes upregulated in KK1_Tax⁺ cells, indicating strong transcriptional influence of NF- κ B and CREB pathways. Fig. 7D shows results from the TRRUST database, which independently confirms enrichment of NFKB1, RELA, and CREB1 as upstream regulators of the differentially expressed genes. This supports the hypothesis that HTLV-1 Tax may regulate HMGB1 expression via both NF- κ B and CREB transcription factors.

HMGB1 Links Inflammation, Cell Death and Cancer Signaling

Pathway enrichment analysis of HMGB1 via Enrichr revealed strong associations with inflammation, apoptosis regulation and tumor-related immune signaling across multiple gene-set databases. In the Reactome 2024 database (Fig. 8A), HMGB1 was associated with several signaling cascades involved in immune dysregulation and programmed cell death. Notably, “Advanced Glycosylation Endproduct Receptor Signaling” was the most significantly enriched pathway. Other enriched pathways included “Apoptosis Induced DNA Fragmentation,” “TRAF6-Mediated NF- κ B Activation” and “Diseases of Immune System,” fur-

ther confirms its involvement in transcriptional regulation linked to innate immunity and oncogenic inflammation. In the KEGG Human 2021 pathway database (Fig. 8B), HMGB1 showed significant enrichment in “Base Excision Repair” suggesting a potential role in maintaining genomic stability under stress conditions. In addition, it was linked to autophagy and necroptosis pathways which are often dysregulated in cancer cells to promote survival and treatment resistance. Tissue-specific enrichment based on the Tissues Curated 2025 dataset (Fig. 8C) revealed that HMGB1 is most highly expressed in several cancer-relevant tissues including breast cancer cells, uterine cervix, adenocarcinoma, and carcinoma cell types. Further insight into HMGB1’s biological role was gained through the GO Biological Processes 2025 library (Fig. 8D). Enriched terms included “CD4-positive, alpha-beta T cell activation”, “Positive regulation of leukocyte differentiation” and “Regulation of apoptotic cell clearance.” These processes align with HMGB1’s known activity as a DAMP molecule that orchestrates immune responses in cancer including T-cell activation, chemokine signaling, and apoptotic resolution.

Pan-Cancer Expression and Prognostic Value of HMGB1

Correlation analysis between HMGB1 and epidermal growth factor receptor (EGFR), a key oncogene, showed a weak but statistically significant positive correlation ($R = 0.098$, $p < 0.001$), suggesting potential co-activation or pathway convergence in tumor settings (Fig. 9A). HMGB1 was found to be overexpressed in multiple tumor types compared to normal tissues (Fig. 9B). Elevated expression was particularly notable in diffuse large B-cell lymphoma (DLBC), glioblastoma multiforme (GBM), esophageal carcinoma (ESCA), pancreatic adenocarcinoma (PAAD) and thymoma (THYM). This supports HMGB1’s transcriptional activation in diverse oncogenic contexts. Using KM plotter, LIHC patients with high HMGB1 expression had a significantly reduced survival ($HR = 1.47$, $p = 0.039$; Fig. 10C). Patients with SARC also showed significantly reduced survival with higher expression of HMGB1 ($HR = 1.59$; 95% CI: 1.07–2.36; $p = 0.022$; Fig. 10D). Similarly, LUAD cases demonstrated significantly worse prognosis associated with HMGB1 overexpression ($HR = 1.38$, $p = 0.041$; Fig. 10E). In BC, no significant survival difference was observed between high and low HMGB1 expression groups ($HR = 0.79$, $p = 0.15$; Fig. 10A), while HNSC showed a non-significant trend toward poorer survival with elevated HMGB1 expression ($HR = 1.34$, $p = 0.052$; Fig. 10B). Collectively, these findings highlight a cancer-type-specific prognostic role for HMGB1, with stronger associations observed in SARC, LUAD and LIHC.

Cancer-Specific Co-regulation Patterns of HMGB1 and Its Binding Partners

The HMGB1 PPI network retrieved from Pathway Commons revealed several well-characterized interactors involved in inflammation, chromatin re-

modeling and immune signaling including TLR2/4, TP53, RELA, NFKB1/2, MAPK1, JUN and TNF (Fig. 11A). To quantify their transcript-level co-regulation across cancers, we examined Spearman correlations in TCGA datasets (Fig. 11B). Notably, positive correlations were observed between HMGB1 and canonical inflammatory mediators such as RELA, TNF and MAPK1 in cancers like LUAD, LIHC and SARC, where HMGB1 was also prognostically significant. Correlation with nuclear factors like NFKB1/2 and TP53 was context-dependent, showing moderate-to-strong associations in immune-rich cancers (e.g., HNSC, SKCM and CESC) but not in immunologically cold tumors (e.g., GBM, UCS). In contrast, some partners such as SOD2 and CXCL12 showed weak or negative correlations. This supports the hypothesis that HMGB1 promotes tumor-associated inflammation via transcriptional modulation of its binding partners in specific cancer contexts.

In vitro expression validation of HMGB1 in Tax-(+) and Tax(-) HTLV-1 cells

In the final part of the study, the expression of HMGB1 in HTLV-1 Tax(+) and Tax(-) cells was evaluated at both the mRNA and protein levels. RT-qPCR was employed to assess the relative mRNA expression of HMGB1, revealing a significant upregulation of HMGB1 in HTLV-1 Tax(+) cells compared to HTLV-1 Tax(-) and HTLV-1-W cells (**Fig.12A**). The results indicated a marked difference, with the Tax(+) cells showing the highest expression, while the Tax(-) and HTLV-1-W cells exhibited lower and comparable levels (**Fig.12A**). To further validate these findings, Western blot analysis was conducted to measure the protein expression of HMGB1. This analysis confirmed the results observed in the RT-qPCR, showing that HMGB1 protein levels were substantially higher in the HTLV-1 Tax(+) cells (**Fig.12B-C** and **Fig. S3**). Both the Tax (-) and HTLV-1-W cells displayed significantly lower levels of HMGB1 protein expression (**Fig.12B-C** and **Fig. S3**).

Discussion

This study provides a systematic analysis into the mechanistic regulation of HMGB1 expression by the HTLV-1 Tax protein and its functional mutants M22 and M47 in T lymphocytes. It is documented that HTLV-1 Tax is a transactivating oncoprotein that drives persistent T-cell activation and proliferation by constitutively activating the NF- κ B pathway via I κ B kinase (IKK) complex recruitment and nuclear translocation of RelA/p65^[28]. Additionally, Tax forms a ternary complex with phosphorylated CREB (pCREB) and CBP/p300 at cyclic-AMP response elements (CREs), enhancing transcription of both viral and cellular genes, including cyclin D1 which promote cell-cycle progression in infected T lymphocytes^[29]. The structural modeling and validation of Tax wild-type and its mutants M22 and M47 provide a foundational framework for understanding functional

divergence. GalaxyRefine optimization followed by MolProbity assessment confirmed that all three models possess high stereochemical quality, with minimal Ramachandran outliers and low clashscores. The M22 mutation is known to abolish NF- κ B activation without impacting CREB-mediated transcription^[30] and the M47 mutation conversely enhances CREB binding while impairing NF- κ B activation^[11]. Apo-state molecular dynamics simulations further confirmed the conformational implications of these mutations. The WT Tax protein displayed early stabilization and minimal deviation from its initial conformation, consistent with a functionally optimized scaffold. In contrast, the M22 mutant exhibited higher RMSD and increased fluctuations around residues 120–140, a region proximal to the known NF- κ B interaction interface, suggesting structural destabilization at a functionally relevant site. Similarly, M47 showed elevated flexibility near residues 310–330, overlapping with the CREB interaction domain. These structural signatures reinforce the functional segregation of the mutants and validate their use in dissecting pathway-specific regulation. Our docking analysis demonstrated that the M22 mutation reduces the complementarity and interaction surface with NF- κ B, disrupting the stability of the transcriptional complex. In contrast, the M47 mutant formed an enhanced interaction with CREB, evidenced by improved interface packing and surface burial. The docking scores and center-weighted ClusPro metrics further supported these observations, with M22 showing reduced binding strength compared to WT in the NF- κ B context, and M47 exhibiting a substantially improved binding profile with CREB. These interaction dynamics are parallel with previous experimental data that the Tax mutants M22 (T130A/L131S) and M47 (L319R/L320S) have been shown to segregate functional pathways i.e., M22 fails to activate NF- κ B while preserving CREB activation, whereas M47 is CREB-defective and retains NF- κ B activity^[31].

Molecular dynamics simulations of the Tax–TF complexes extended these insights. The M22–NF- κ B complex demonstrated low RMSD values and stable hydrogen bonding, indicative of residual binding affinity but lacked the conformational flexibility typically associated with active transcriptional assemblies. Conversely, the M47–CREB complex exhibited higher RMSF values in loop regions and consistent intra-complex hydrogen bonding which suggests the dynamic yet stable interface suited for transcriptional activation. These patterns support the view that the M47 mutation enhances the structural compatibility of Tax with CREB, while M22 compromises NF- κ B engagement. Promoter motif analysis using FIMO identified multiple high-confidence binding sites for both NF- κ B and CREB1 within the 2 kb upstream region of the HMGB1 gene. The presence of binding motifs for both transcription factors provide a mechanistic anchor for the hypothesis that Tax-mediated transcriptional regulation of HMGB1 may occur via dual pathway convergence. Notably, several CREB1 motifs were located on both strands, enhancing the probability of functional binding, while NF- κ B motifs clustered on the positive strand, consistent with canonical promoter architecture.

Transcriptomic analysis of Tax-expressing versus non-expressing T-cell lines (MT1 and KK1) corroborated the structural findings. HMGB1 expression was significantly upregulated in Tax⁺ cells in both models, implicating it as a downstream effector of Tax activity. In contrast, the mRNA levels of CREB1 and RELA remained relatively stable, suggesting their regulatory roles occur through protein-protein interactions rather than transcriptional upregulation. Transcription factor enrichment analysis further supported this model, with NF- κ B and CREB1 emerging as top-ranked upstream regulators of Tax-induced gene expression, including HMGB1. Oncogenic analyses collectively reinforce HMGB1's dual role in immune modulation and oncogenic signaling. The enrichment of NF- κ B activation, autophagy and DNA repair pathways supports its involvement in tumor-promoting inflammation and cellular stress responses. Elevated HMGB1 expression across multiple tumor types particularly in LUAD, LIHC and SARC shows its relevance as a tissue-specific prognostic marker. Moreover, its consistent co-expression with key immune and apoptotic regulators such as TLR2/4, TP53 and BECN1 across diverse cancers highlights HMGB1 as a central hub mediating tumor-immune system interactions and stress response pathways.

Altogether, the integration of structural dynamics, docking specificity, transcription factor motif enrichment and gene expression profiles establishes a robust mechanistic link between HTLV-1 Tax and HMGB1 transcription. The M22 and M47 mutants serve as precise molecular probes to disentangle the contributions of NF- κ B and CREB pathways in this regulatory axis. Given HMGB1's known roles in tumor progression, chronic inflammation, and immune evasion^[32], our findings provide novel insights into how HTLV-1 may co-opt host transcriptional machinery for pathogenic advantage. These insights may inform future therapeutic strategies targeting Tax-TF interfaces to mitigate HTLV-1-associated diseases.

Conclusion

This study demonstrates that HTLV-1 Tax regulates HMGB1 transcription through discrete, mutation-dependent engagement of NF- κ B and CREB transcription factors. The M22 mutant, defective in NF- κ B activation, exhibits impaired binding to NF- κ B and reduced interface complementarity, while the CREB-selective M47 mutant shows enhanced structural and energetic compatibility with CREB. These pathway-specific interactions are supported by molecular dynamics simulations, which reveal differential interface stability and flexibility, and are functionally corroborated by transcriptomic evidence showing Tax-induced upregulation of HMGB1 alongside enrichment of NF- κ B and CREB1 transcriptional targets. Together, these findings define a mechanistic framework in which Tax exploits distinct host transcriptional circuits via conformationally tuned interactions, offering novel insights into HTLV-1-mediated transcriptional reprogramming. Together, these findings establish HMGB1 as a mechanistically and

clinically relevant immuno-oncogenic effector, linking HTLV-1 Tax activity to cancer-associated transcriptional and signaling networks.

Conflict of Interest

None

Funding

The project received funding from Fujian Province Natural Science Foundation general project.

Project Title: Application Research of Nucleic Acid Aptamer-Molecular Interaction Technology in Diagnosis and Treatment of HTLV-1 Virus Infection.

Project funding number: 2021J01265

Data Availability

The bulk RNA-seq data used in this study were obtained from the NCBI Gene Expression Omnibus (GEO) under accession number GSE108601. All other data supporting the findings of this study, including structural models, molecular dynamics trajectories and docking output files are available from the corresponding author upon reasonable request.

Clinical Trial Number

Not applicable

Human Ethics and Consent to Participate declarations

Not applicable

Consent to Publish Declaration

Not Applicable

Acknowledgment

None

References

- [1] S. Datta, M. A. Rahman, S. Koka, and K. M. Boini, “High mobility group box 1 (HMGB1): molecular signaling and potential therapeutic strategies,” *Cells*, vol. 13, no. 23, p. 1946, 2024.
- [2] L. Morbée, “The function of high mobility group box 1 (HMGB1) in the immune system.”
- [3] M. Magna and D. S. Pisetsky, “The role of HMGB1 in the pathogenesis of inflammatory and autoimmune diseases,” *Molecular medicine*, vol. 20, pp. 138-146, 2014.
- [4] C. R. Bangham, A. Araujo, Y. Yamano, and G. P. Taylor, “HTLV-1-associated myelopathy/tropical spastic para-

paresis,” *Nature reviews Disease primers*, vol. 1, no. 1, pp. 1-17, 2015.

[5] K. J. Ernzen and A. R. Panfil, “Regulation of HTLV-1 transformation,” *Bioscience reports*, vol. 42, no. 3, p. BSR20211921, 2022.

[6] S. Mohanty and E. W. Harhaj, “Mechanisms of oncogenesis by HTLV-1 Tax,” *Pathogens*, vol. 9, no. 7, p. 543, 2020.

[7] R. Currer et al., “HTLV tax: a fascinating multifunctional co-regulator of viral and cellular pathways,” *Frontiers in microbiology*, vol. 3, p. 406, 2012.

[8] A. Mukherjee, “CHARACTERIZATION OF THE CANCER-ASSOCIATED EFFECTS OF LOSS AND/OR MUTATION OF TUMOR SUPPRESSOR GENES ON NF- κ B SIGNALING IN BREAST AND PANCREATIC CANCERS,” 2024.

[9] I. Azran, Y. Schavinsky-Khrapunsky, and M. Aboud, “Role of Tax protein in human T-cell leukemia virus type-I leukemogenicity,” *Retrovirology*, vol. 1, pp. 1-24, 2004.

[10] M. R. Smith and W. C. Greene, “Identification of HTLV-I tax trans-activator mutants exhibiting novel transcriptional phenotypes,” *Genes & development*, vol. 4, no. 11, pp. 1875-1885, 1990.

[11] M. D. Robek and L. Ratner, “Immortalization of CD4+ and CD8+ T lymphocytes by human T-cell leukemia virus type 1 Tax mutants expressed in a functional molecular clone,” *Journal of virology*, vol. 73, no. 6, pp. 4856-4865, 1999.

[12] D. Saggioro, “Anti-apoptotic effect of Tax: an NF- κ B path or a CREB way?,” *Viruses*, vol. 3, no. 7, p. 1001, 2011.

[13] M. Varadi et al., “AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models,” *Nucleic acids research*, vol. 50, no. D1, pp. D439-D444, 2022.

[14] E. W. Harhaj and C. Z. Giam, “NF- κ B signaling mechanisms in HTLV-1-induced adult T-cell leukemia/lymphoma,” *The FEBS journal*, vol. 285, no. 18, pp. 3324-3336, 2018.

[15] H. Lu, *Studies on coactivator function during Tax-mediated HTLV-I transcription*. The George Washington University, 2002.

[16] 이규리, “Development of protein modeling methods for structure refinement in the context of unreliable environments,” *서울대학교 대학원*, 2017.

[17] P. P. Putra, A. Fauzana, K. A. Salva, M. Sofiana, I. P. Sari, and H. Lucida, “Homology modeling and mutation prediction of ACE2 from COVID-19,” *Editorial Team*, p. 201.

[18] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, “An overview of the Amber biomolecular simulation package,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 3, no. 2, pp. 198-210, 2013.

[19] T. Chau, P. Timilsena, and S. Li, “Gene Regulatory Network Modeling Using Single-Cell Multi-Omics in Plants,” in *Plant Gene Regulatory Networks: Methods and Protocols*: Springer, 2023, pp. 259-275.

[20] A. Alekseenko, M. Ignatov, G. Jones, M. Sabitova, and D. Kozakov, “Protein-protein and protein-peptide docking with ClusPro server,” *Protein Structure Prediction*, pp. 157-174, 2020.

[21] K. D. Stamm, *Gene set enrichment and projection: A computational tool for knowledge discovery in transcriptomes*. Marquette University, 2016.

[22] E. Y. Chen et al., “Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool,” *BMC bioinformatics*, vol. 14, pp. 1-14, 2013.

[23] J. E. Evangelista, Z. Xie, G. B. Marino, N. Nguyen, D. J. Clarke, and A. Ma’ayan, “Enrichr-KG: bridging enrichment analysis across multiple libraries,” *Nucleic acids research*, vol. 51, no. W1, pp. W168-W179, 2023.

[24] J. Zhao, C. Guo, Z. Ma, H. Liu, C. Yang, and S. Li, “Identification of a novel gene expression signature associated with overall survival in patients with lung adenocarcinoma: A comprehensive analysis based on TCGA and GEO databases,” *Lung Cancer*, vol. 149, pp. 90-96, 2020.

[25] B. Jia, X. Zhao, Y. Wang, J. Wang, Y. Wang, and Y. Yang, “Prognostic roles of MAGE family members in breast cancer based on KM-Plotter Data,” *Oncology letters*, vol. 18, no. 4, pp. 3501-3516, 2019.

[26]I. Rodchenkov et al., “Pathway Commons: 2019 Update,” bioRxiv, p. 788521, 2019.

[27]T. Li et al., “TIMER2. 0 for analysis of tumor-infiltrating immune cells,” Nucleic acids research, vol. 48, no. W1, pp. W509-W514, 2020.

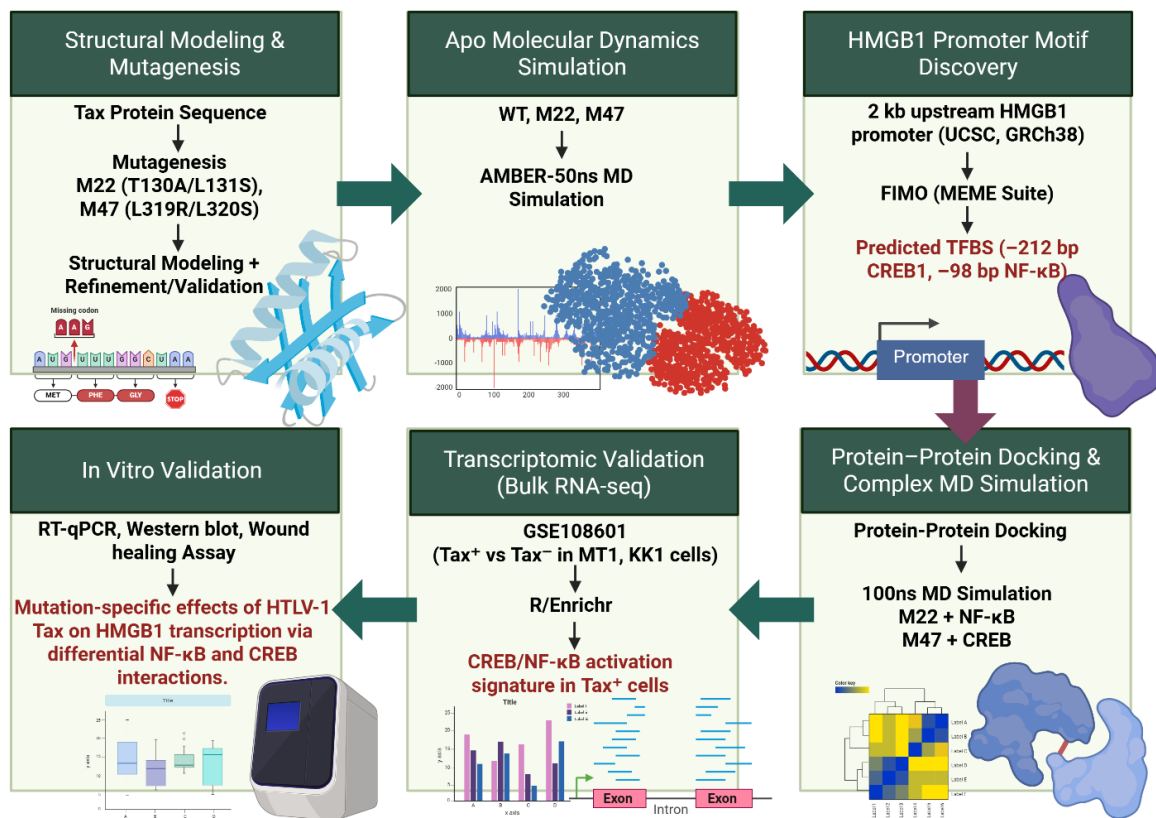
[28]R. Su, X. Kang, Y. Niu, T. Zhao, and H. Wang, “PCBP1 interacts with the HTLV-1 Tax oncoprotein to potentiate NF- κ B activation,” Frontiers in Immunology, vol. 15, p. 1375168, 2024.

[29]Y.-M. Kim, T. R. Geiger, D. Egan, N. Sharma, and J. K. Nyborg, “The HTLV-1 tax protein cooperates with phosphorylated CREB, TORC2 and p300 to activate CRE-dependent cyclin D1 transcription,” Oncogene, vol. 29, no. 14, pp. 2142-2152, 2010.

[30]M. Egbaria, Y. Tabakin-Fix, and M. Huleihel, “Dominant negative Tax double mutants as molecular inhibitors for wt Tax gene functions,” Leukemia research, vol. 33, no. 7, pp. 974-979, 2009.

[31]N. S. Harhaj, S.-C. Sun, and E. W. Harhaj, “Activation of NF- κ B by the human T cell leukemia virus type I Tax oncoprotein is associated with ubiquitin-dependent relocalization of I κ B kinase,” Journal of Biological Chemistry, vol. 282, no. 6, pp. 4185-4192, 2007.

[32]S. Martinotti, M. Patrone, and E. Ranzato, “Emerging roles for HMGB1 protein in immunity, inflammation, and cancer,” ImmunoTargets and therapy, pp. 101-109, 2015.



Graphical. Abstract. Schematic representation of the in silico and experimental workflow used to investigate the transcriptional regulation of HMGB1 by HTLV-1 Tax protein and its functional mutants. HTLV-1 Tax WT and mutant (M22, M47) structures were modeled using AlphaFold and ChimeraX, followed by apo-state and complex MD simulations with AMBER to assess conformational stability and TF-specific binding dynamics. CREB and NF- κ B motifs within the HMGB1 promoter were identified via FIMO, and ClusPro docking confirmed mutation-specific interactions. Transcriptomic validation using bulk RNA-seq (GSE108601) supported pathway activation, with in vitro assays planned for functional confirmation.

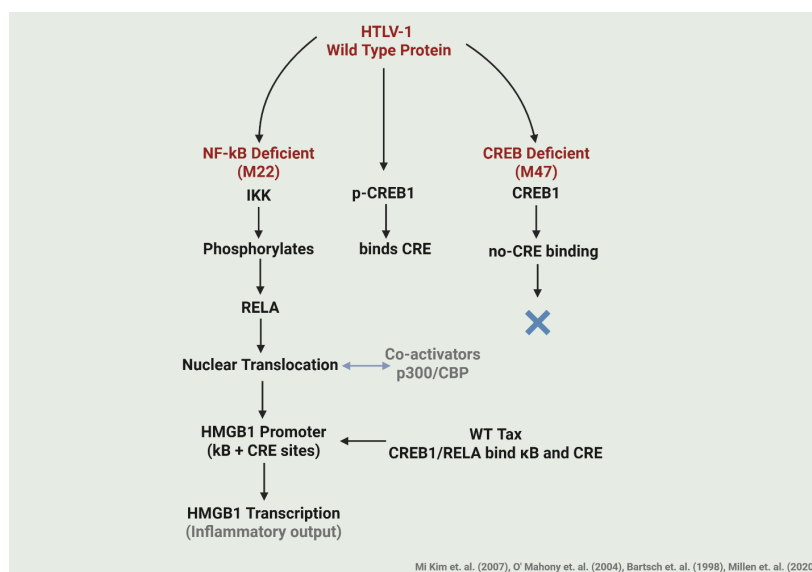


Figure 1. Schematic representation of pathway-specific regulation of HMGB1 transcription by HTLV-1 Tax and its M22 and M47 mutants. The HTLV-1 wild-type Tax protein activates both NF-κB and CREB pathways via interaction with RELA and phosphorylated CREB1 (p-CREB1), recruiting coactivators (p300/CBP) and promoting HMGB1 transcription. The M22 mutant is deficient in NF-κB signaling due to impaired IKK-mediated RELA activation, reducing nuclear translocation and transcriptional output. Conversely, the M47 mutant is CREB-defective, unable to bind CRE sites, and fails to engage the CREB axis. Both pathways converge on the HMGB1 promoter, which contains κB and CRE motifs, leading to inflammatory gene activation in the WT context.

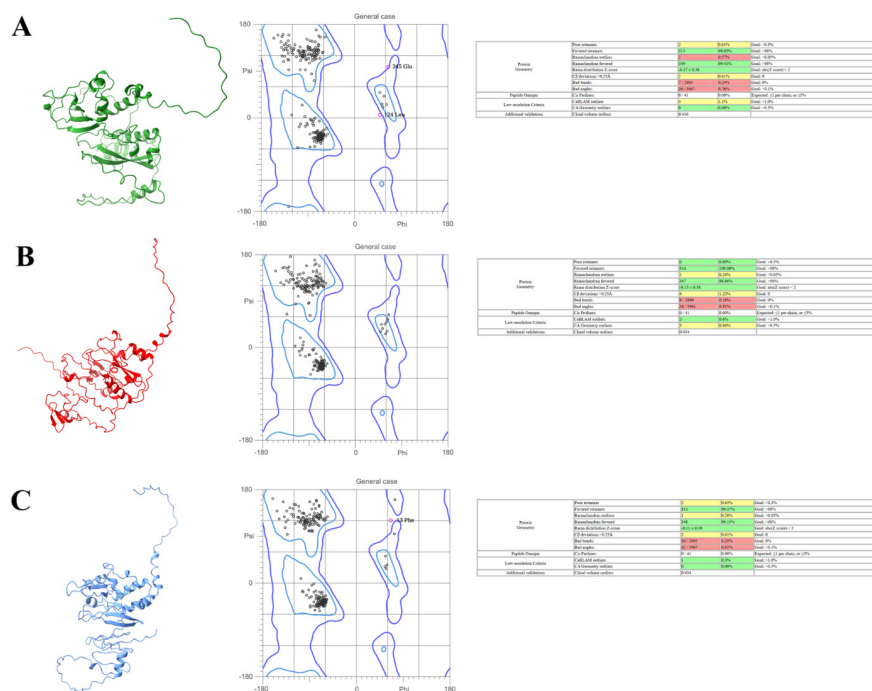


Figure 2. Stereochemical quality assessment of GalaxyRefine-optimized HTLV-1 Tax protein structures.

(A) The refined wild-type HTLV-1 Tax model demonstrates excellent stereochemical quality, with a Ramachandran plot

showing 99.4% of residues in favored regions and only two outliers (Leu124 and Glu345), along with a low MolProbity clashscore of 3.28. (B) The M22 mutant model retains high structural integrity, with 98.9% of residues located in favored Ramachandran regions and a single outlier at Glu341; the MolProbity analysis confirms the absence of significant steric clashes. (C) The M47 mutant also exhibits strong stereochemical parameters, with 99.1% of residues in favored regions and only one outlier (Phe13) accompanied by an acceptable clashscore.

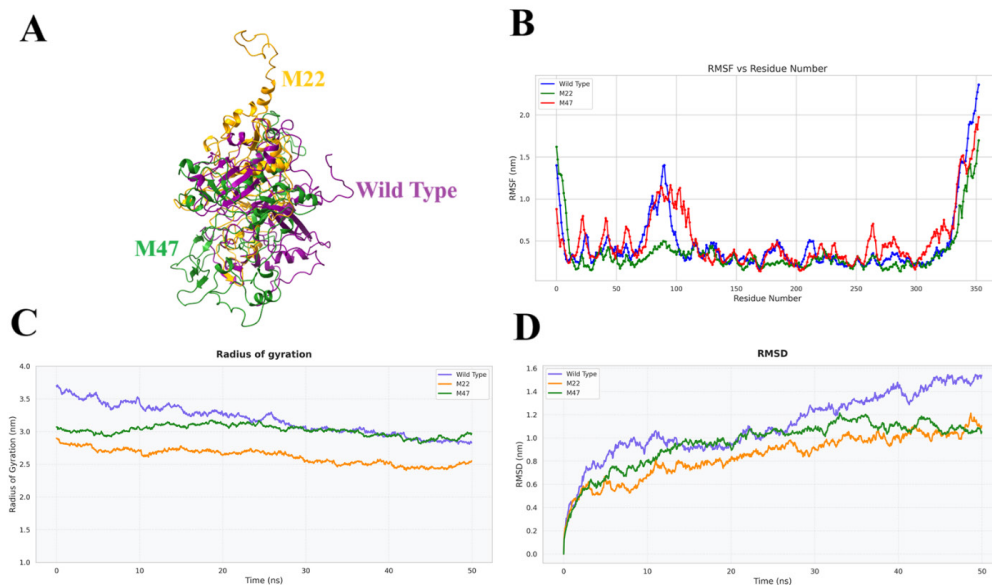


Figure 3. Apo Molecular Dynamics Simulations of HTLV-1 Tax WT and Functional Mutants (M22 and M47) (A) Structural superposition of WT (purple), M22 (orange), and M47 (green) after 50 ns simulation, revealing conformational divergence. (B) Root Mean Square Fluctuation (RMSF) per residue, indicating local flexibility. (C) Radius of Gyration (Rg) illustrating global compactness. (D) Root Mean Square Deviation (RMSD) of backbone atoms over 50 ns, highlighting structural stability.

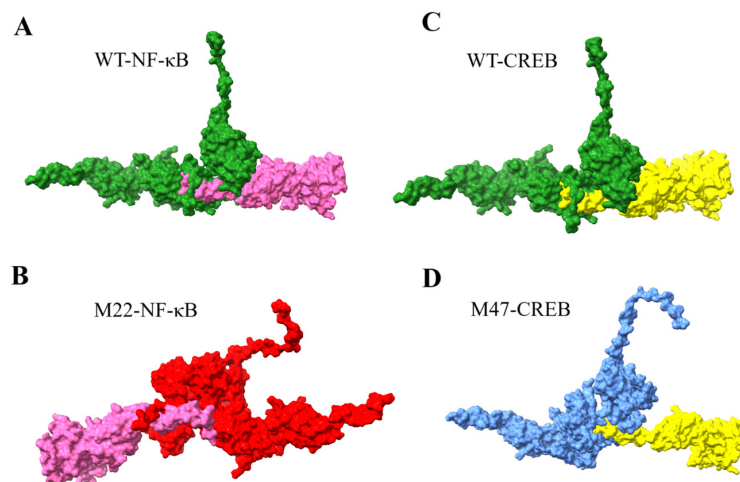


Figure 4. Structural representations of protein-protein docking between HTLV-1 Tax variants and transcription factors NF-κB and CREB. (A) Surface model of the WT Tax–NF-κB complex showing compact interfacial engagement. (B) M22 mutant complexed with NF-κB displays altered orientation and reduced surface complementarity. (C) WT Tax

docked with CREB, showing moderate binding interface. (D) M47 mutant docked with CREB forms a tighter, more extensive contact surface. All structures represent top-scoring Cluster 0 models from ClusPro under balanced docking mode. Green = Tax, Pink = NF- κ B, Yellow = CREB, Red = M22, Blue = M47.

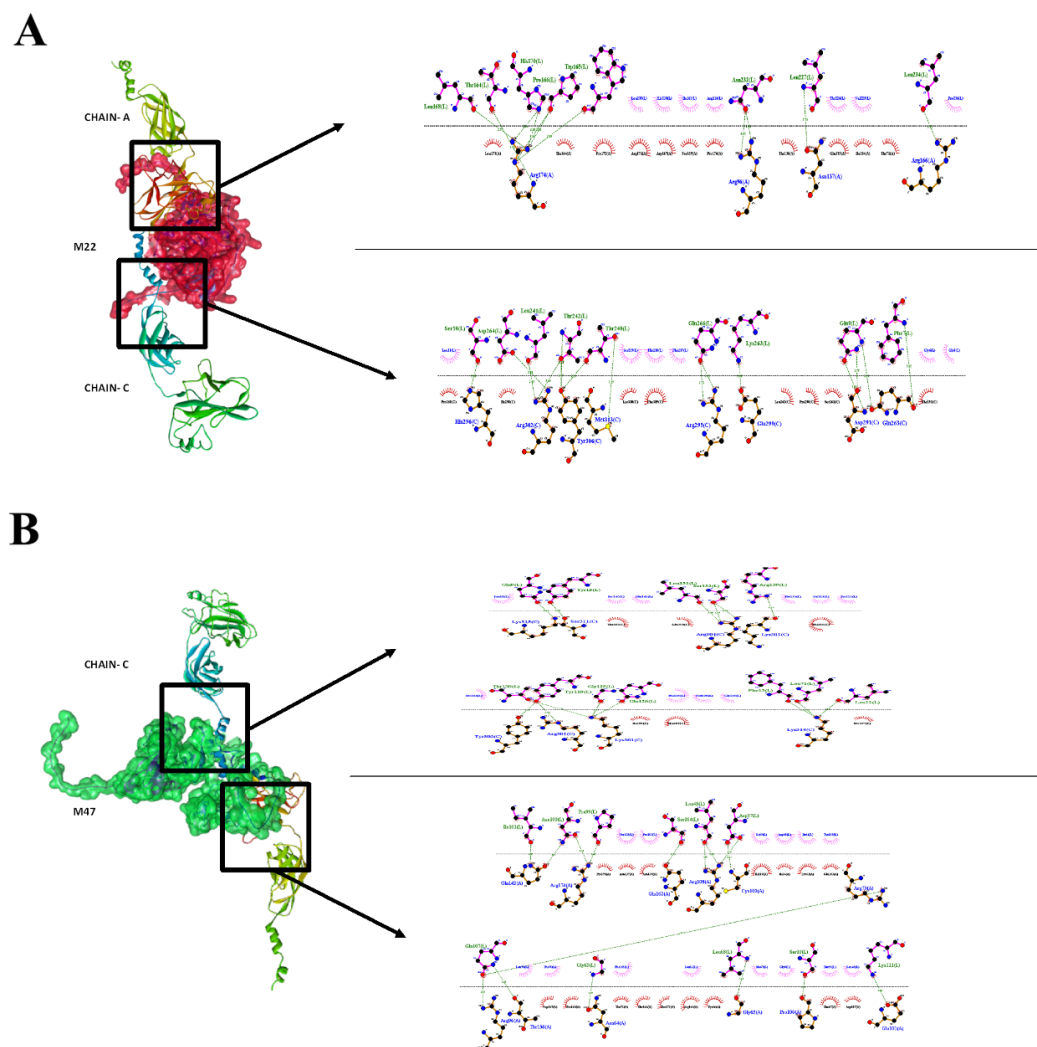


Figure 5. LigPlot interaction diagrams of docked Tax–transcription factor complexes. (A) Key residue-level interactions and hydrogen bonds between M22 and NF- κ B. (B) Ligand interaction map of M47 with CREB, highlighting polar and hydrophobic contacts that stabilize the protein–protein interface.

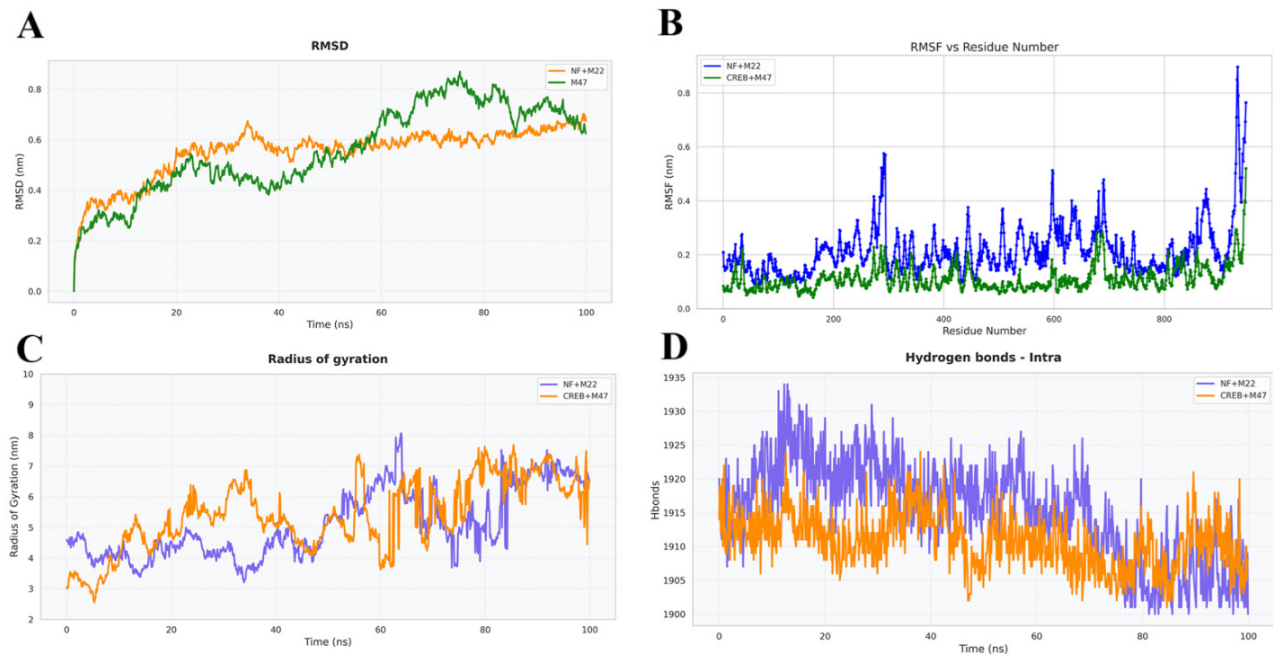


Figure 6. Structural dynamics and stability analysis of the M22–NF-κB and M47–CREB complexes during molecular dynamics simulations. (A) RMSD plot over a 100 ns simulation reveal early and sustained stabilization of the M22–NF-κB complex characterized by low RMSD values (~0.25 nm) and consistent compactness. In contrast, the M47–CREB complex exhibits moderate RMSD fluctuations (~0.3–0.35 nm) with a stable Rg profile. (B) RMSF profiles indicate enhanced residue-level mobility in the loop regions of the M47–CREB complex, whereas the M22–NF-κB complex displays uniformly lower fluctuations. (C) Radius of gyration analysis confirms that the M22–NF-κB complex maintains a consistently compact structure, while the M47–CREB complex displays slightly elevated Rg values and greater fluctuations. (D) Intra-complex hydrogen bond analysis shows that the M22–NF-κB complex sustains a higher number of hydrogen bonds (10–15) throughout the simulation, whereas the M47–CREB complex maintains a slightly lower but stable hydrogen bond count (7–12).

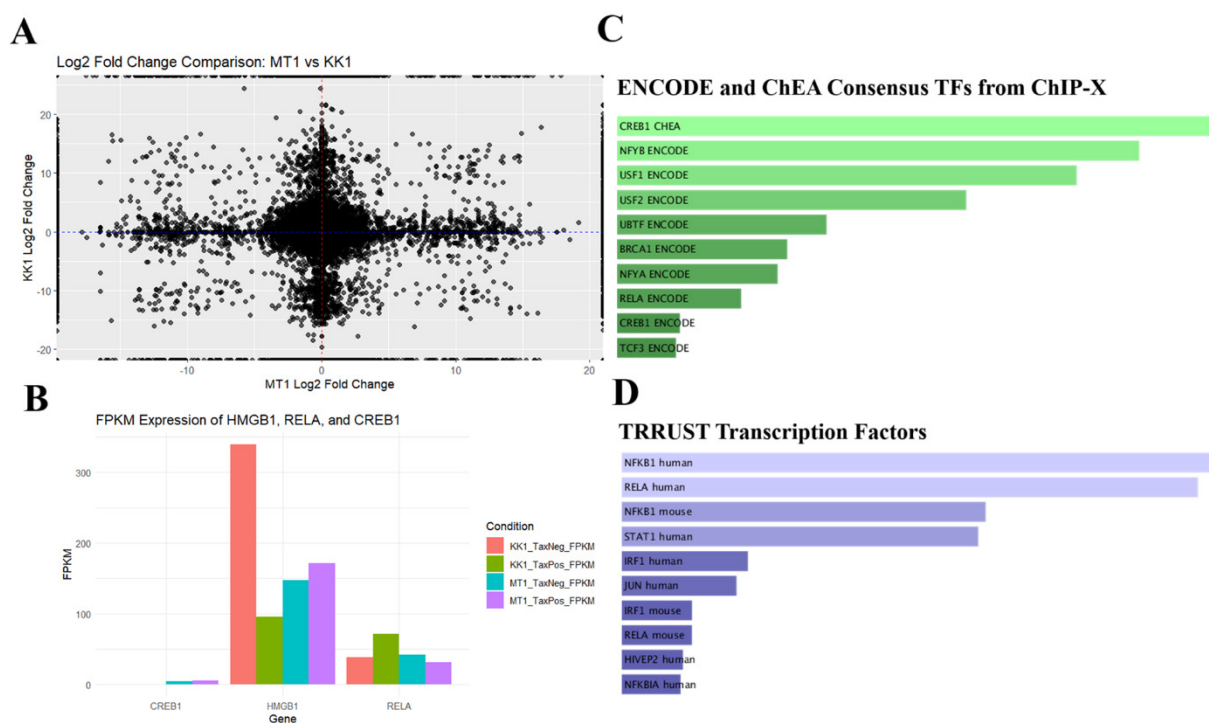


Figure 7. Transcriptomic analysis reveals Tax-dependent upregulation of HMGB1 and enrichment of NF- κ B and CREB regulatory pathways. (A) Log₂ fold change plot showing differential gene expression between KK1_Tax⁺ and KK1_Tax⁻ samples. (B) Correlation of log₂ fold changes between MT1 and KK1 cell lines, highlighting global transcriptomic similarity in Tax⁺ states. (C) Transcription factor enrichment analysis using the ENCODE and ChEA Consensus TFs dataset reveals significant enrichment of RELA, CREB1, and ATF3 in the upregulated gene set of KK1_Tax⁺ cells. (D) TRRUST-based enrichment analysis confirms NFKB1, RELA, and CREB1 as key upstream transcriptional regulators, reinforcing their functional involvement in Tax-mediated transcriptional reprogramming.

A

Reactome Pathway Analysis

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Advanced Glycosylation Endproduct Receptor Signaling	0.0006500	0.008293	19987.00	146675.88
2	Apoptosis Induced DNA Fragmentation	0.0006500	0.008293	19987.00	146675.88
3	MyD88 Deficiency (TLR2 4)	0.0009500	0.008293	19981.00	139049.21
4	IRAK4 Deficiency (TLR2 4)	0.001000	0.008293	19980.00	138017.41
5	Regulation of TLR by Endogenous Ligand	0.001100	0.008293	19978.00	136099.47
6	TRAF6 Mediated NF- κ B Activation	0.001250	0.008293	19975.00	133255.54
7	Pyroptosis	0.001350	0.008293	19973.00	131975.02
8	Diseases of Immune System	0.001750	0.008361	19965.00	126740.99
9	Diseases Associated With the TLR Signaling Cascade	0.001750	0.008361	19965.00	126740.99
10	TAK1-dependent IKK and NF- κ B Activation	0.002250	0.008457	19955.00	121662.50

B

KEGG Human

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Base excision repair	0.001650	0.006600	19967.00	127928.56
2	Autophagy	0.006850	0.009450	19863.00	98987.63
3	Necroptosis	0.007950	0.009450	19841.00	95923.19
4	Neutrophil extracellular trap formation	0.009450	0.009450	19811.00	92353.95

C

Tissues Curated

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	BREAST CANCER CELL	0.01125	0.1806	19775.00	88738.27
2	UTERINE CERVIX	0.02050	0.1806	19590.00	76152.94
3	SMALL INTESTINE	0.02525	0.1806	19495.00	71720.85
4	ADENOCARCINOMA CELL	0.03690	0.1806	19262.00	63555.91
5	CARCINOMA CELL	0.03935	0.1806	19213.00	62159.13
6	EPITHELIOMA CELL	0.03935	0.1806	19213.00	62159.13
7	CEREBELLUM	0.04105	0.1806	19179.00	61237.96
8	METENCEPHALON	0.04110	0.1806	19178.00	61211.42
9	HINDBRAIN	0.04150	0.1806	19170.00	61000.22
10	BRAIN STEM	0.04515	0.1806	19097.00	59158.11

D

GO Biological Processes

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	CD4-positive, Alpha-Beta T Cell Activation (GO:0035710)	0.0003000	0.003219	19994.00	162186.52
2	Regulation of Endodeoxyribonuclease Activity (GO:0032071)	0.0003000	0.003219	19994.00	162186.52
3	T-helper 1 Cell Differentiation (GO:0045063)	0.0003000	0.003219	19994.00	162186.52
4	DNA Geometric Change (GO:0032392)	0.0003500	0.003219	19993.00	159096.45
5	Positive Regulation of Leukocyte Differentiation (GO:1902107)	0.0003500	0.003219	19993.00	159096.45
6	Regulation of Apoptotic Cell Clearance (GO:2000425)	0.0003500	0.003219	19993.00	159096.45
7	Positive Regulation of Chemokine (C-X-C Motif) Ligand 2 Production (GO:2000343)	0.0004000	0.003219	19992.00	156418.91
8	Negative Regulation of Alpha-Beta T Cell Differentiation (GO:0046639)	0.0004000	0.003219	19992.00	156418.91
9	DNA Topological Change (GO:0006265)	0.0004500	0.003219	19991.00	154056.47
10	Negative Regulation of CD4-positive, Alpha-Beta T Cell Activation (GO:2000515)	0.0005000	0.003219	19990.00	151942.59

Figure 8. HMGB1-centered functional enrichment analysis using Enrichr. (A) Reactome pathway enrichment showed significant associations with immune signaling, apoptosis, and NF- κ B activation. (B) KEGG Human 2021 analysis revealed enrichment in cancer-relevant processes such as base excision repair, autophagy, and necroptosis. (C) Tissue-specific enrichment from the Tissues Curated 2025 library identified high HMGB1 expression in breast cancer, uterine cervix, and adenocarcinoma cells, as well as multiple neural tissues. (D) GO Biological Processes highlighted HMGB1's role in T-cell activation, chemokine signaling, and apoptotic regulation, supporting its functional involvement in immune modulation and tumor-associated inflammation.

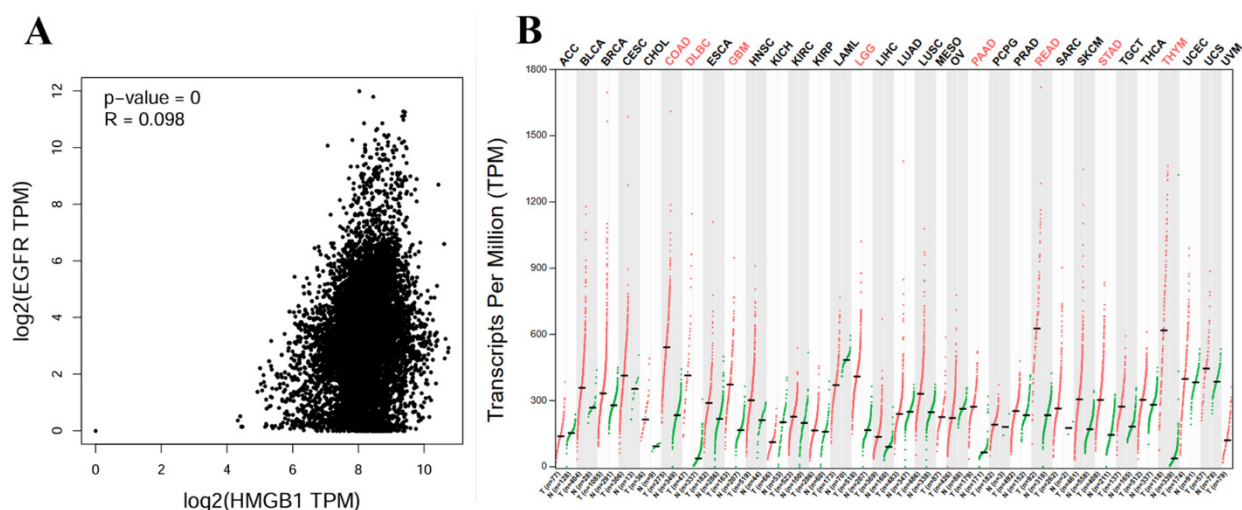


Figure 9. HMGB1 pan-cancer expression, survival, and gene correlation analysis using GEPIA2. (A) Spearman correlation between HMGB1 and EGFR expression across TCGA and GTEx datasets ($R = 0.098$, $p < 0.001$). (B) HMGB1 expression profile across 33 tumor types showing upregulation in several cancers including DLBC, GBM, ESCA and PAAD.

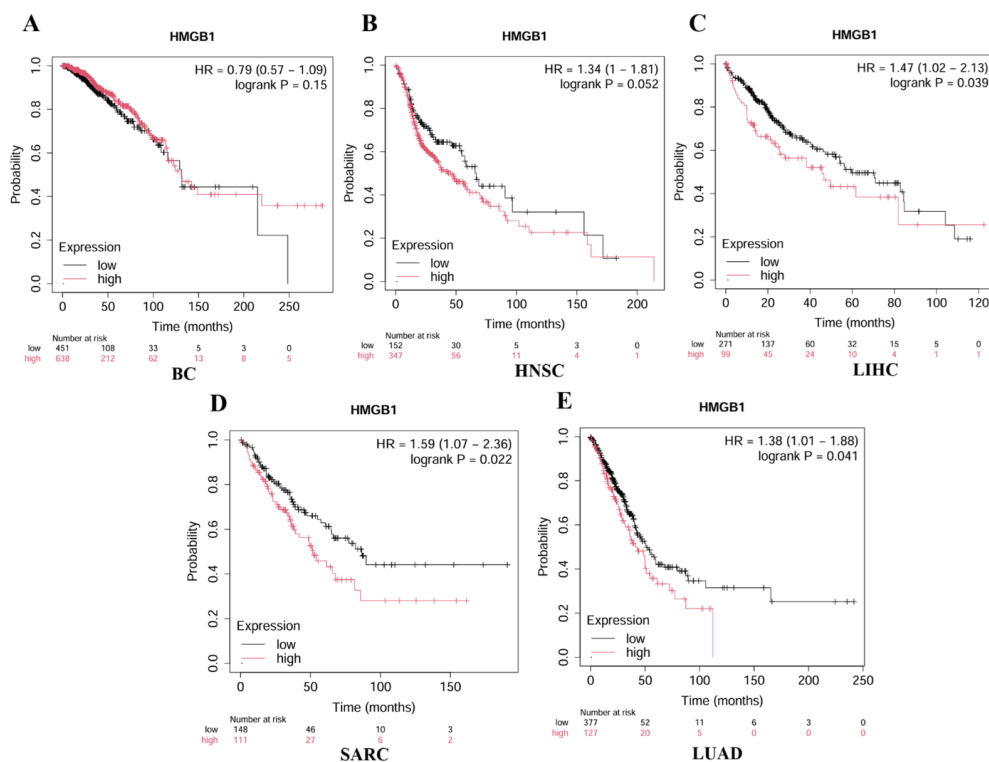


Figure 10. Prognostic impact of HMGB1 expression across five cancer types using KM Plotter pan-cancer module. Kaplan–Meier survival plots depict overall survival stratified by HMGB1 expression levels (high vs. low) in (A) Breast Cancer (BC), (B) Head and Neck Squamous Cell Carcinoma (HNSC), (C) Liver Hepatocellular Carcinoma (LIHC), (D) Sarcoma (SARC), and (E) Lung Adenocarcinoma (LUAD). Hazard ratios (HR), 95% confidence intervals, and log-rank p-values are shown. Survival was significantly poorer in patients with high HMGB1 expression in SARC, LUAD and LIHC cohorts.

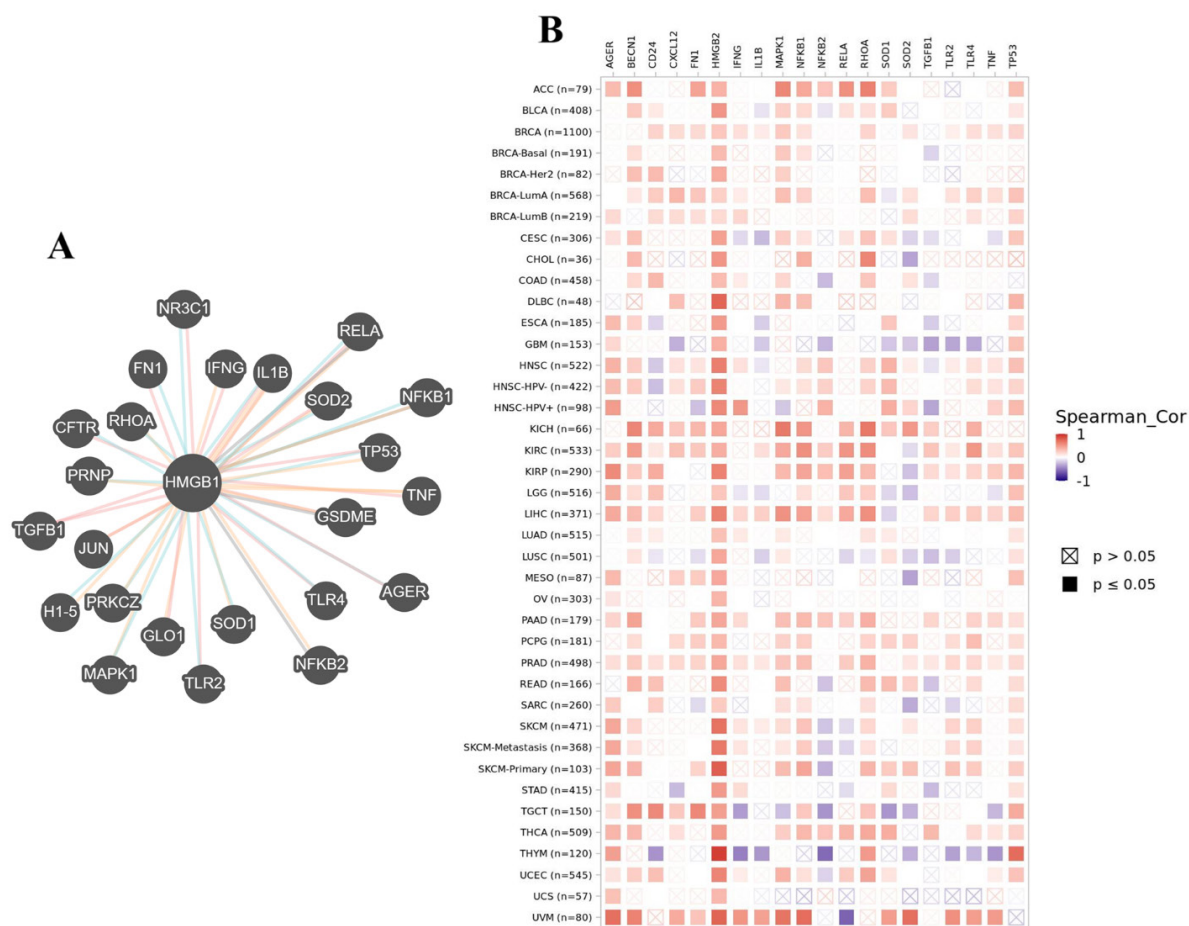


Figure 11. HMGB1 binding partners and cancer-type-specific correlation landscape. (A) Protein–protein interaction (PPI) network of HMGB1 constructed using curated datasets from Pathway Commons. (B) Heatmap showing the Spearman correlation coefficients between binding partners of HMGB1 across TCGA cancer types using TIMER2.0 RNA-seq data. Red indicates positive correlation and blue indicates negative correlation.

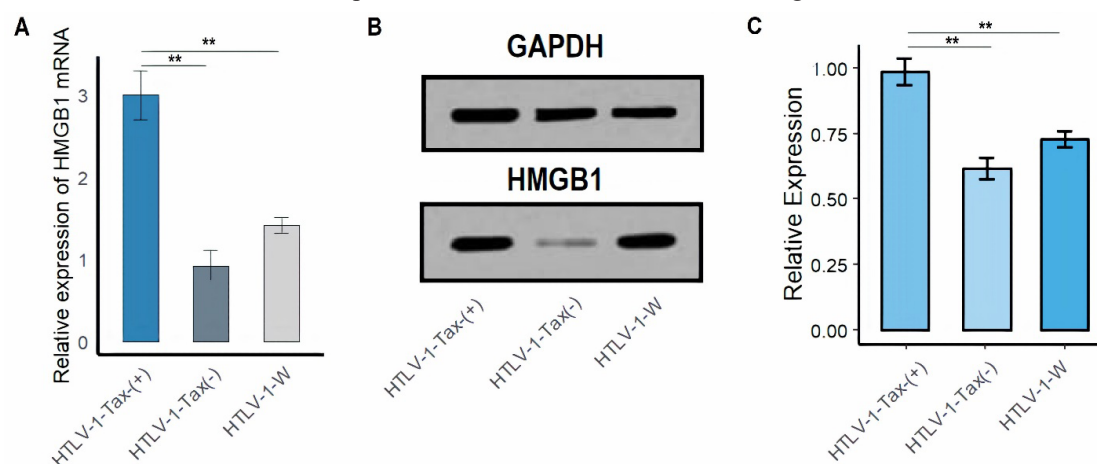


Figure 12. HMGB1 expression is up-regulated in HTLV-1 Tax(+) cells at mRNA and protein levels. (A) Relative HMGB1 mRNA expression levels were measured using RT-qPCR in HTLV-1 Tax(+), HTLV-1 Tax(-), and HTLV-1-W cells. (B) Representative immunoblot showing increased HMGB1 protein levels in HTLV-1 Tax(+) cells, while Tax(-)

and HTLV-1-W cells showed substantially lower expression. (C) Quantification of HMGB1 protein levels normalized to a loading control (GAPDH), confirming the elevated expression in Tax(+) cells. **P < 0.01.

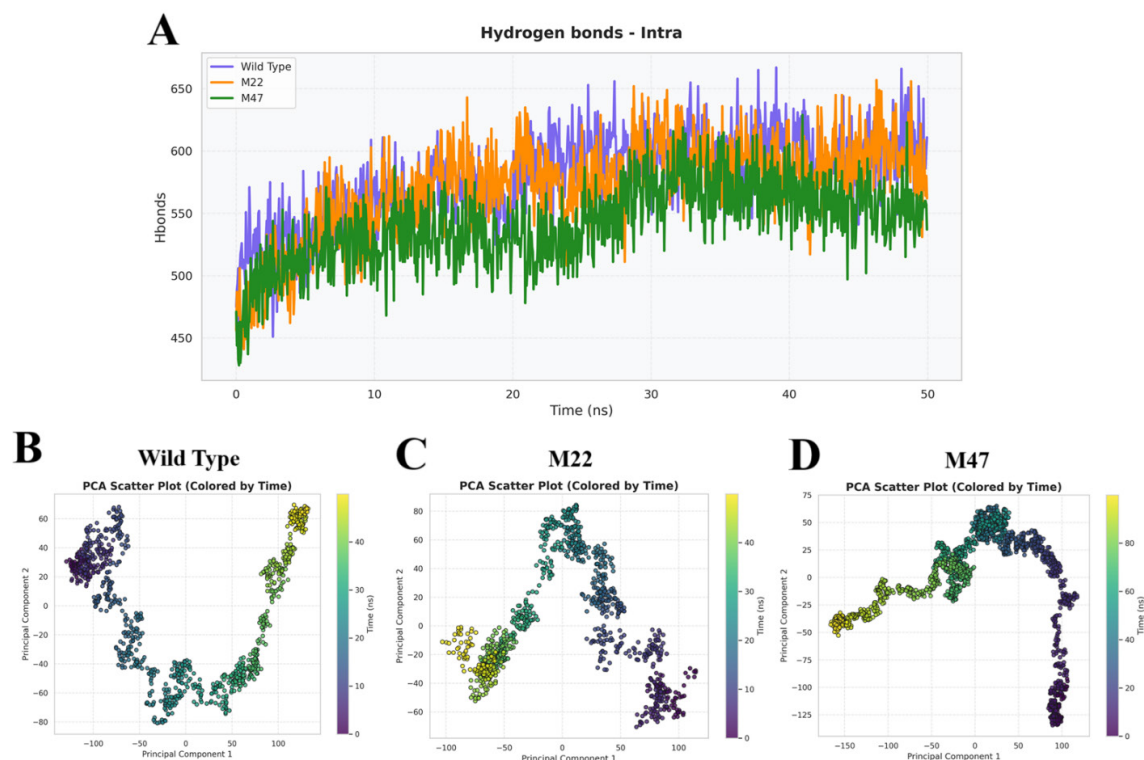


Figure S1. Intra-Protein Hydrogen Bonding and Principal Component Analysis (PCA) of HTLV-1 Tax Variants. (A) Intra-molecular hydrogen bond count over simulation time for WT, M22, and M47. (B) PCA plot showing the conformational landscape sampled during MD simulations along the first two principal components (PC1 vs. PC2) of WT. (C) PCA plot of M22. (D) PCA plot of M47.

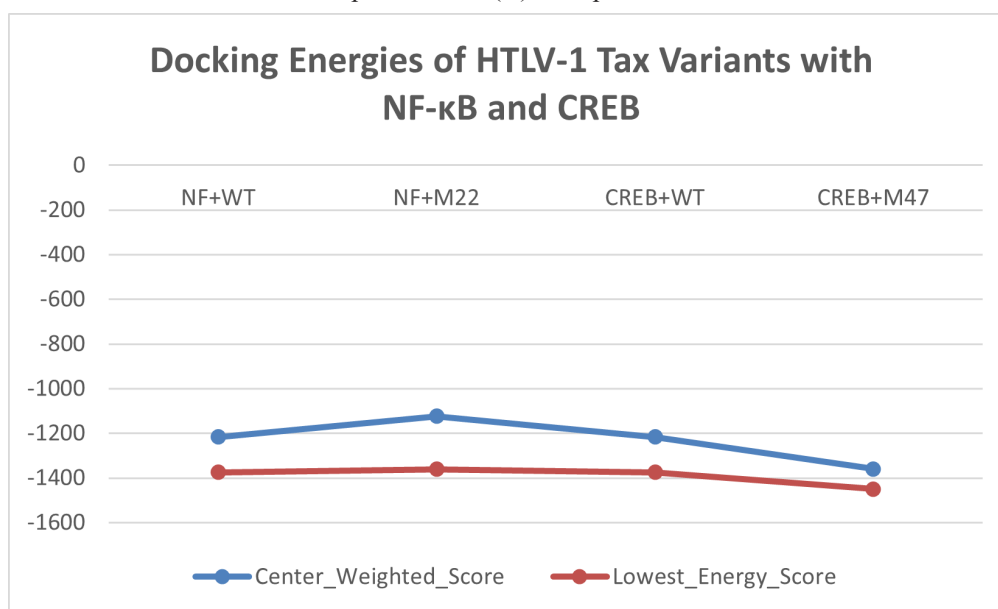


Figure S2. ClusPro docking energy comparison of HTLV-1 Tax WT and mutant complexes with NF-κB and CREB.

The graph shows Center Weighted Score and Lowest Energy Score for each complex derived from Cluster 0. M22 shows reduced center-weighted affinity for NF- κ B compared to WT, while M47 exhibits a significantly stronger binding profile with CREB than WT.

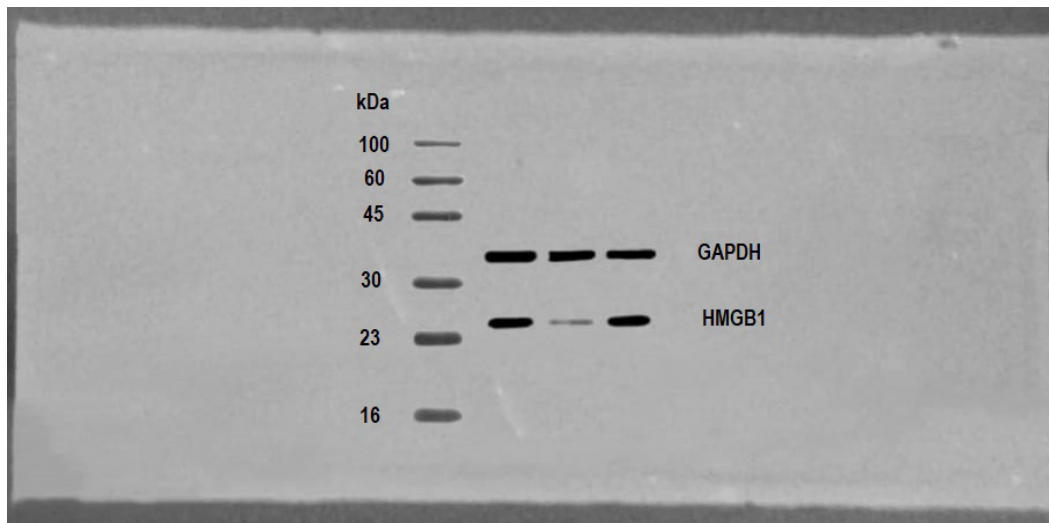


Figure S3. Uncut Western blot bands of HMGB1 and GAPDH.

Motif ID	Chromosome	Strand	Start	End	p-value	q-value	Matched Sequence
MA0602.1_CREB1	chr13	-	30464923	30464932	0.000205	0.404	GGGATCCGGC
MA0105.4_NFKB1_RELA	chr13	+	30464658	30464668	0.000442	0.842	GGCCAATGAGC
MA0105.4_NFKB1_RELA	chr13	+	30464969	30464979	0.000445	0.842	GGCAAATGAGG
MA0602.1_CREB1	chr13	+	30465440	30465449	0.000529	0.404	CGGTCGCGGC
MA0602.1_CREB1	chr13	+	30465661	30465670	0.000597	0.404	GGGAGACGGG
MA0602.1_CREB1	chr13	-	30464498	30464507	0.000753	0.404	CTGCTACGGG
MA0105.4_NFKB1_RELA	chr13	+	30465774	30465784	0.00097	1	CCGTAATGGCC

Table S1. FIMO-predicted transcription factor binding sites (TFBSs) for CREB1 and NF- κ B (RELA) within the 2 kb upstream promoter region of the HMGB1 gene.