國立臺灣大學生命科學院分子與細胞生物學研究所

碩士論文

Institute of Molecular and Cellular Biology

College of Life Science

National Taiwan University

Master's Thesis

TERRA RNA 對基因組中 G4 與 CTCF 的影響

TERRA RNA impacts G-quadruplexes and CTCF occupancy across the genome

李姿瑢

TZU-RONG LEE

指導教授：朱雪萍 博士

Advisor: Hsueh-Ping Chu, Ph.D.

中華民國 114 年 7 月

July 2025

doi:10.6342/NTU202504256

國立臺灣大學碩士學位論文
口試委員會審定書
MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

TERRA RNA 對基因組中 G4 與 CTCF 的影響

TERRA RNA impacts G-quadruplexes and CTCF occupancy across the genome

本論文係 李姿瑢 R11B43007 在國立臺灣大學 分子與細胞生物學研究所 完成之碩士學位論文，於民國 114 年 7 月 24 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Department / Graduate Institute of Molecular and Cellular Biology on 24 July, 2025 have examined a Master's Thesis entitled above presented by TZU-RONG, LEE R11B43007 candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

（指導教授 Advisor）

系（所、學位學程）主管 Director:

致謝

沒想到在這裡蹲了四年，真是不可思議。這一路上遇到了許多人類：謝謝朱老師願意讓我加入 R807；謝謝妤宣帶我熟悉實驗室環境；謝謝辰嘉傳授細胞培養大法；謝謝歐文擔任把拔，能在把拔的保護傘下令人安心，離巢獨立後仍受到諸多照顧；謝謝仁龍傳授多種實驗心法，也很敢提供機會讓我在實戰中累積經驗。在大學生時期，很幸運能和強大的趙亮組隊，他真的會照亮未知的道路。

轉職成碩班生之後，認識了重要的同期和前輩：特別感謝安之和沛淇，努力實踐最理想的科學態度，在學術和生活上都提供了非常大的幫助，在實驗上也願意花時間理解並幫忙梳理研究方向，希望你們可以被世界溫柔以待；謝謝榮澤吃大餐都會分享；謝謝家樂一起成為 ES 同門，收細胞就是要配倉鼠影片。謝謝依恬擔任馬麻，手把手將 ES 小白養大，在實驗和研究方面都提供了非常多的協助；謝謝庭嘉作為師父，幫打了很多怪。馬麻和師父無微不至的關心，外加不定時投食，是碩班生活中非常重要的依靠。中途加點黑魔法技能時，獲得分析大老們的照顧：謝謝宥宏和依恬提供新手指南；謝謝博丞傾囊相授畢生絕學，並提供精良的自救大全；謝謝孟婷總是耐心為我解惑，是非常厲害的 NGS 多邊形戰士。在 R807 日常中，謝謝維茜讓枯燥生活變得有趣；謝謝家瑜施展大天使的祝福；謝謝建平不忘督促論文進度；謝謝可容的吐槽陪伴和日常關心，希望你的畢業道路更順利一點；謝謝涵雅分享高效率秘訣，你的笑聲讓人有前進的動力；謝謝進華會喊著要吃雞；謝謝保靜常常在週末做實驗時互相打氣；謝謝育昇常常比我更擔心我的未來；謝謝胤賢一起在 ES 苦海中沉浮；謝謝庭琪加入 ChIP 大隊，一起分享手瘦日常。

最後，我想特別謝謝家人願意花時間理解實驗室生活，習慣並支撐著手忙腳亂的每一天，是生活中最堅穩的後盾。

這一路上真的獲得太多幫助了，謝謝大家。

李姿瑢 2025.08.02

doi:10.6342/NTU202504256

中文摘要

G-四聯體是由富含鳥糞嘌玲的核酸序列摺疊形成的非典型二級結構，可透過去氧核醣核酸（DNA）或核醣核酸（RNA）構成。DNA G-四聯體廣泛分布於基因組中，並參與 DNA 複製、轉錄和基因組穩定性的調控。TERRA 是一類由 RNA 聚合酶II從次端粒和端粒區域轉錄而成的長片段非編碼 RNA，含有不等數量的 UUAGGG 重複序列，可摺疊形成 RNA G-四聯體結構。之前研究顯示 TERRA 可與 G4 解旋酶 BLM 結合。在小鼠胚胎幹細胞中進行 TERRA 下調處理後，BLM 在染色質的結合量上升，且 DNA G-四聯體的形成下降。

為了確認 TERRA 能否透過其富含鳥糞嘌呤的重複序列調控 DNA G-四聯體形成，我們將含有四個 UUAGGG 重複序列的 RNA（TERRA RNA）轉染至 HeLa 細胞中。免疫螢光染色結果顯示，BLM 在染色質的結合下降，而核內 DNA G-四聯體訊號上升，進一步支持 TERRA 具備促進 DNA G-四聯體形成的想法。

為了進一步探討 TERRA 的表觀遺傳功能，之前研究整合 ChIP-seq 與 RNA-seq 資料，發現 TERRA 下調會導致部分基因轉錄起始點（TSS）附近的 DNA G-四聯體訊號下降，而且這些基因與 TERRA 下調後表現量降低的基因有關。近期有其他研究指出 DNA G-四聯體可能影響 CTCF 的 DNA 結合調控。CTCF 是一種高度保守的染色質結構蛋白，參與染色質圈構（chromatin looping）的形成，並影響基因表現。為了確認 TERRA 能否藉由 DNA G-四聯體的調控進一步影響 CTCF 與染色質的結合，我們在 TERRA 下調的小鼠胚胎幹細胞中進行 CTCF ChIP-seq 實驗，觀察到大多數 CTCF 在結合位點都有訊號上升的趨勢，而且在 G4 訊號下降的位置，CTCF 訊號上升程度更為顯著。然而，這些位點附近的 G-四聯體訊號變化不一，顯示 CTCF 結合的改變可能並非單純由 DNA G-四聯體調控所致。此外，CTCF 與 cohesin 共結合的區域訊號上升更為顯著，暗示 TERRA 可能參與染色質三維結構的調控。

doi:10.6342/NTU202504256

綜上所述，本研究透過 TERRA RNA 轉染系統與 TERRA KD 系統，分別確認 TERRA 會影響染色質上的 BLM 結合與 DNA G-四聯體的形成，並進一步指出 TERRA 可能透過對 CTCF 的影響，揭示潛在的表觀遺傳調控機制。

關鍵字：TERRA、G-四聯體、DNA G-四聯體、BLM、CTCF

doi:10.6342/NTU202504256

Abstract

G-quadruplexes (G4) are noncanonical secondary structures formed by guanine-rich sequences in DNA and RNA. DNA G4s are widely distributed throughout the genome and play roles in regulating replication, transcription, and genome stability. TERRA, a long non-coding RNA transcribed by RNA polymerase Ⅱ, transcribed from subtelomeric regions to telomeres, with varying numbers of UUAGGG repeats capable of forming RNA G4 structures. TERRA has been shown to interact with BLM, an ATP-dependent helicase known to unwind G4 structure, and TERRA knockdown in mESCs leads to increased BLM occupancy and reduced DNA G4 levels.

To examine whether TERRA regulates DNA G4 through its G-rich repetitive sequences, we introduced synthetic TERRA RNA oligos containing four tandem UUAGGG repeats into HeLa cells. Immunofluorescence staining revealed that TERRA RNA reduced BLM occupancy and increased nuclear DNA G4 signals, suggesting that TERRA promotes G4 formation.

To further investigate the epigenetic role of TERRA, integrated analyses of G4 ChIP-seq and RNA-seq showed that TERRA knockdown led to G4 signal reduction near the transcription start sites (TSSs), particularly in genes downregulated upon TERRA depletion. Given that studies from other labs revealed that G4 formation may influence CTCF binding, we tested whether TERRA regulates CTCF binding to chromatin. CTCF

doi:10.6342/NTU202504256

ChIP-seq analysis revealed a genome-wide increase in CTCF occupancy in TERRA knockdown mESCs. Notably, G4 reduction correlates with the greatest increase in CTCF binding. However, changes in G4 signal near CTCF binding sites were not positively correlated with changes in CTCF occupancy in TERRA knockdown cells, suggesting that G4 modulation alone may not fully account for the increased in CTCF occupancy upon TERRA depletion.

Moreover, the increase in CTCF binding was more pronounced at sites co-occupied with cohesin complexes, implicating a potential role for TERRA in modulating chromatin looping via the CTCF-cohesin axis. Taken together, our findings suggest a previously unrecognized epigenetic mechanism by which TERRA influences chromatin architecture and regulates gene expression.

Key words: TERRA、G-quadruplex、DNA G-quadruplex、BLM、CTCF

doi:10.6342/NTU202504256

Contents

doi:10.6342/NTU202504256

doi:10.6342/NTU202504256

doi:10.6342/NTU202504256

doi:10.6342/NTU202504256

Contents of Figure

doi:10.6342/NTU202504256

doi:10.6342/NTU202504256

Contents of Table

doi:10.6342/NTU202504256

Contents of Supplementary Figure

doi:10.6342/NTU202504256

Contents of Supplementary Table

doi:10.6342/NTU202504256

Abbreviations

Abbreviation	Full name
5nc	Negative control on chromosome 5 (CTCF negative binding sites on chromosome 5)
ALS	Amyotrophic lateral sclerosis
APS	Ammonium persulfate
ASO	Antisense oligonucleotide
ATAC	Assay for Transposase-Accessible Chromatin
ATP	Adenosine Triphosphate
bp	base pair
C9orf72	Chromosome 9 Open Reading Frame 72
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
CHIRT	Chromatin Isolation by RNA Purification followed by Tagmentation
CLIP	Cross-Linking and Immunoprecipitation
Ctrl RNA	Control RNA
DAPI	4' ,6-Diamidino-2-Phenylindole
DeepLncCTCF	Deep-learning-based Predictor of CTCF Binding to Long Non-coding RNAs
DEG	Differentially Expressed Gene
DEPC	Diethyl Pyrocarbonate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide

doi:10.6342/NTU202504256

DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
dPBS	Dulbecco's Phosphate-Buffered Saline
DPR	Dipeptide Repeat
dsDNA	Double-Stranded DNA
DTT	Dithiothreitol
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic Acid
EGTA	Ethylene Glycol Tetraacetic Acid
EtOH	Ethanol
FBS	Fetal Bovine Serum
FISH	Fluorescence In Situ Hybridization
FTD	Frontotemporal Dementia
FUS	Fused in Sarcoma
G4	G-quadruplex
G4C2	GGGGCC Hexanucleotide Repeat
G4C2 RNA	RNA oligonucleotide containing four repeats of GGGGCC
HeLa	Human Cervical Cancer Cell Line
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
Hi-C	High-throughput Chromosome Conformation Capture
HRP	Horseradish Peroxidase
IAA	Indole-3-Acetic Acid
iDRIP	Identification of Direct RNA Interacting Proteins
IF	Immunofluorescence
IgG	Immunoglobulin G

doi:10.6342/NTU202504256

IP	Immunoprecipitation
IQR	Interquartile Range
ITR	Interstitial Telomeric Repeat
KD	Knockdown
KO	Knockout
Lab	Laboratory
LIF	Leukemia Inhibitory Factor
LLPS	Liquid-Liquid Phase Separation
LNA	Locked Nucleic Acid
lncRNA	Long Non-coding RNA
LowOCs	Low Occupancy CTCF Binding Sites
mChrY	Subtelomeric Regions on Mouse Chromosome Y
MEF	Mouse Embryonic Fibroblast
MEM NEAA	Minimum Essential Medium Non-Essential Amino Acids
mESC	Mouse Embryonic Stem Cell
MKN74	Human Gastric Cancer Cell Line MKN74
mRNA	Messenger RNA
MS	Mass Spectrometry
NGS	Next-Generation Sequencing
NP-40	Nonidet P-40 (non-ionic detergent)
Oligo	Oligonucleotide
PBS	Phosphate-Buffered Saline
PBST	Phosphate-Buffered Saline with Tween-20
PC-3	Human Prostate Cancer Cell Line PC-3
pen-strep	Penicillin-Streptomycin

doi:10.6342/NTU202504256

PFA	Paraformaldehyde
PIC	Protease Inhibitor Cocktail
PK	Proteinase K
PMSF	Phenylmethylsulfonyl Fluoride
PQS	Putative G-Quadruplex Sequence
PVDF	Polyvinylidene Difluoride
qPCR	Quantitative Polymerase Chain Reaction
RAN	Repeat-Associated Non-AUG Translation
RBP	RNA-Binding Protein
RIPA	Radioimmunoprecipitation Assay Buffer
RNA	Ribonucleic Acid
RQC	Ribosome Quality Control
RT	Reverse Transcription
SC	Scramble
SC KD	Scramble knockdown
SC unique CTCF peaks	CTCF peaks identified in scramble KD mESCs excluding the CTCF peaks identified in TERRA KD mESCs
SC/TE common CTCF peaks	CTCF peaks identified in both scramble KD mESCs and TERRA KD mESCs
SCE	Sister Chromatid Exchange
SDS	Sodium Dodecyl Sulfate
seq	Sequencing
SMARD	Single Molecule Analysis of Replicated DNA
SSIV	SuperScript IV Reverse Transcriptase

doi:10.6342/NTU202504256

STRING database	Search Tool for the Retrieval of Interacting Genes/Proteins database
TAD	Topologically Associating Domain
TAE	Tris-Acetate-EDTA
TBST	Tris-Buffered Saline with Tween-20
TE	TERRA
TE KD	TERRA knockdown
TE unique CTCF peaks	CTCF peaks identified in TERRA KD mESCs excluding the CTCF peaks identified in Scramble KD mESCs
TEN	Tris-EDTA-NaCl
TERRA	Telomeric Repeat-Containing RNA
TERRA RNA	RNA oligonucleotide containing four repeats of UUAGGG
TES	Transcription End Site
TF	Transcription Factor
TSS	Transcription Start Site
UV	Ultraviolet
WT	Wild Type
ZF	Zinc Finger

doi:10.6342/NTU202504256

Chapter 1 Introduction

1.1 G-quadruplex (G4)

G-quadruplexes (G4s) are non-canonical nucleic acid secondary structures formed in G-rich sequences of DNA or RNA, following the forming formula $G_{2-4}N_{3-7}G_{2-4}N_{3-7}G_{2-4}N_{3-7}G_{2-4}$. They are characterized by the stacking of G-quartets, which is a square planar assembly of four guanines held together by Hoogsteen hydrogen bond [1]. These structures are stabilized by monovalent cations, particularly K^+ and Na^+ [2]. DNA G4s were identified to fold in parallel, antiparallel, or hybrid form depending on the ionic conditions, while RNA G4s tend to fold parallel conformation and are generally more thermodynamic stable under similar environment conditions [3, 4].

Genome-wide studies have revealed that DNA G4 structures are broadly distributed and particularly enriched in regulatory regions such as telomeres, promoters, and splice sites [5, 6], suggesting diverse regulatory roles in cells. At telomeres, G4 formation can interfere with the binding of telomere-associated proteins, impair replication, and contribute to telomere dysfunction [7]. G4s could also modulate telomerase activity, potentially through interactions with the POT1-TPP1 complex during repeat addition [8]. In gene regulation, G4 structures near promoters have been shown to affect transcription. For example, a G4 formed upstream of the c-myc promoter represses its transcription [9].

doi:10.6342/NTU202504256

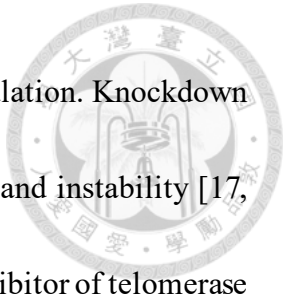
Putative G-quadruplex-forming sequences (PQS) analysis further suggest that the position and strand orientation of G4s relative to transcription start sites (TSSs) can influence gene expression, and that their effects may be modulated by G4 helicases, which were thought to unwind the G4 structures, such as BLM and WRN in the RecQ family [10]. Furthermore, G4 forming sequences near TSSs have also been identified as hubs for transcription factor (TF) binding [11], and are associated with enhanced recruitment of RNA polymerase II [12], supporting their potential roles in regulating gene expression.

Otherwise, RNA G4 have been implicated in various post-transcriptional regulatory processes. Several RNA G4-binding proteins, including FMRP, DHX36, FUS, and hnRNP-H, have been identified, suggesting roles in regulating mRNA translation and alternative splicing. Furthermore, several RNA G4 have also been shown to facilitate stress granule formation and participate in phase separation.

1.2 Telomeric Repeat-Containing RNA (TERRA)

TERRA are long-noncoding RNAs first identified in mammalian cells [13], with variable lengths ranging from a few hundred base pairs to several kilobases. They are transcribed by RNA polymerase II from subtelomeric regions toward telomeres [14], and typically contain multiple UUAGGG repeats at their 3’-ends. These G-rich sequences are capable of forming G4 structures [15], and hybridize with complementary telomeric DNA to form R-loop structures [16]. Due to their sequence specificity and transcriptional origin,

doi:10.6342/NTU202504256



several studies on TERRA have focused on its roles in telomere regulation. Knockdown of TERRA has been shown to cause increased telomere dysfunction and instability [17, 18]. Moreover, TERRA has been reported to act as an endogenous inhibitor of telomerase by directly binding and preventing its activity [18, 19], which indicates TERRA may contribute to telomere length regulation.

Genome-wide analyses of TERRA suggest its functions extend beyond telomere regulation [18]. RNA fluorescence in situ hybridization (FISH) and CHIRT-seq have shown that TERRA potentially binds across the genome, not just at telomere or subtelomeric regions. Furthermore, identification of direct RNA interacting proteins followed by mass spectrometry (iDRIP-MS) revealed that TERRA interacts not only with telomere-associated proteins but also with epigenetic regulators, and DNA replication factors. Notably, TERRA appears to antagonize ATRX, a chromatin remodeler belonging to the SWI/SNF family, exhibiting opposing effects on gene expression. In addition, immunofluorescence (IF) staining indicated that TERRA influences the occupancy of several G4 helicases, including BLM and RTEL1 [20]. G4 chromatin immunoprecipitation followed by sequencing (ChIP-seq) further demonstrated that TERRA may promote G4 formation near some TSS regions, which is associated with increased ATRX occupancy and differential gene expression [21]. These findings support the hypothesis that TERRA functions as an epigenetic regulator in cells.

doi:10.6342/NTU202504256

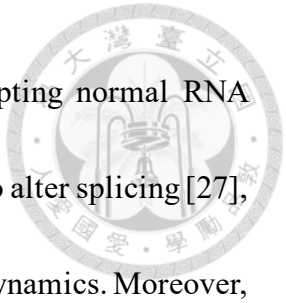
1.3 G4C2 repeats in C9orf72 gene

A hexanucleotide GGGGCC (G4C2) repeat expansion located in the first intron of C9orf72 gene on chromosome 9 has been reported in a large proportion of amyotrophic lateral sclerosis and frontotemporal dementia (ALS/FTD) patients [22, 23]. In healthy individuals, the repeat number is typically below 30, whereas patients may carry hundreds or thousands of repeats.

The toxicity associated with this expansion is considered to arise from both the production of dipeptide repeat proteins (DPRs) and the accumulation of repeat-containing RNA transcripts, with evidence suggesting that these two mechanisms can act independently. DPRs are translated from the expanded repeat region via a non-canonical process known as repeat-associated non-AUG (RAN) translation [24]. Five types of DPRs, including ploy-GA, GP, GR, PA, and PR, have been detected in patient tissues, among which poly-GA is the most aggregation-prone. The subcellular localization of DPRs, either nuclear or cytoplasmic, appears to influence their toxicity. Several studies have shown that these DPRs could interfere with nucleoli formation, nucleocytoplasmic transport, and motor neuron development [25, 26].

Additionally, expanded G4C2 RNA transcripts are prone to form RNA foci, both in patient-derived tissue and in cellular models transfected with G4C2 repeats [27]. These RNA foci have been reported to sequester specific RNA binding proteins (RBPs),

doi:10.6342/NTU202504256



including hnRNP-H, FUS, and G3BP1 [28, 29], potentially disrupting normal RNA metabolism. For example, hnRNP-H sequestration has been shown to alter splicing [27], while the interaction with G3BP1 implicates a role in stress granule dynamics. Moreover, the expanded G4C2 repeats in neuronal cells or zebrafish embryos has been associated with apoptosis and dysregulated proteostasis, further supporting the neurotoxic potential of these transcripts. The interactions between G4C2 RNA and RBPs is thought to be facilitated by the ability of the repeat-containing RNA to form secondary structures, such as hairpins [30] and G4 structures [31]. A study further demonstrated that RNA G4 enhance liquid-liquid phase separation (LLPS) of FUS through G4 binding, promoting a transition form liquid-like condensates to solid aggregates via liquid-to-solid transition (LST) [32], thereby potentially contributing to ALS pathology.

1.4 BLM (Bloom's syndrome helicase)

BLM is a member of the RecQ helicase family and belongs to the superfamily 2 (SF2) helicases. It unwinds DNA in an ATP-dependent manner with 3' to 5' directionality, playing a role in various aspects of DNA metabolism [33]. BLM contains several functional domains, including a pair of RecA-like domains that contribute the core helicase domain, and the RecQ C-terminal (RQC) domain. RQC domains is involved in the recognition of dsDNA and also specifically contributes to bind DNA G4 structures [34].

doi:10.6342/NTU202504256

Previous studies have demonstrated that BLM can directly interact with DNA G4 structures and unwind these non-canonical secondary structures in vitro. While some reports showed BLM-mediated G4 unwinding occurs in the absence of ATP [35, 36], others indicate ATP is required for efficient unwinding [37], suggesting that the ATP dependence of this activity may vary depending on experimental context or substrate. In addition, immunofluorescence using the BG4 antibody revealed that G4 signal intensity is reduced in BLM KO (BLM-/-) cells compared to heterozygous (BLM+/-) cells, suggesting that BLM may also unwind G4 DNA structure in vivo [38].

A previous study using single-cell DNA template strand sequencing (Strand-seq) to map the positions of sister chromatid exchanges (SCEs) shows that SCEs are more frequent in BLM-/- cells, particularly at G4 motif sites [39]. Another study using single molecule analysis of replicated DNA (SMARD) to trace telomere replication found that the replication rate is slower in BLM-/- cells compared to BLM+/- cells, and this defect could be rescued by treatment with G4 stabilizer [38]. These findings suggest that the ability of BLM to unwind DNA G4 structure may play a role in maintaining proper replication dynamics and genomic stability.

Additionally, BLM may also interact with RNA G4 structures. A recent study demonstrated that BLM could directly bind and unwind RNA G4s in an ATP-dependent manner [40]. Proximity labelling experiments using APEX revealed that the recruitment

doi:10.6342/NTU202504256

of BLM to stress granules is reduced upon treatment with an RNA G4 stabilizer, suggesting that RNA G4 formation may regulate BLM localization to stress granules. Furthermore, a previous ChIP-seq study in TERRA KD mESCs showed increased BLM occupancy at telomeres and interstitial telomeric repeat (ITR) regions, indicating a potential mechanism by which RNA G4s modulate BLM chromatin binding [20].

1.5 CCCTC-binding factor (CTCF)

CTCF is a well-known and evolutionarily conserved architectural protein characterized by 11 zinc-finger (ZF) domains. It plays a critical role in shaping chromatin architecture, particularly in mediating chromatin looping. Mutations in ZF2 to ZF7 are found to be lethal, while ZF4 to ZF7 are necessary for recognizing a core 12-15 bp DNA motif that serves as the canonical CTCF binding site [41]. In addition, other zinc fingers and non-ZF domains of CTCF have been reported to contribute to binding site specificity, potentially through interactions with upstream or downstream flanking sequences in mouse genomes [41, 42].

CTCF is known to act as an insulator, recognizing enhancer-blocking elements across vertebrates [43]. Using high-resolution in situ Hi-C and ChIP-seq, it was found that the vast majority of chromatin loop anchors are bound by CTCF and the cohesin complex, with CTCF motifs predominantly arranged in a convergent orientation, which is critical for loop formation [44-46]. While cohesin complex contribute to the loop

doi:10.6342/NTU202504256

extension, CTCF act as its anchor to stop extending looping size [47, 48]. Within a topologically associating domain (TAD), genomic loci tend to exhibit similar histone modification patterns, while those across domain boundaries do not. Many loops connect a promoter and an enhancer, and such interacting loops are associated with activated gene expression, highlighting a regulatory role of chromatin architecture [45]. Furthermore, acute CTCF depletion via an auxin-induceble degron system alters chromatin accessibility and slightly affects gene expression. Integrating ATAC-seq, RNA-seq, and mass spectrometry data, 40 transcription factors (TFs) were identified to be associated with CTCF occupancy [49], suggesting that CTCF not only shapes genome topology, but also indirectly regulates transcription by influencing TF binding, further impact on regulating gene expression.

Additionally, recent studies have explored the relationship between CTCF occupancy and G4 structures. While some evidence suggests that CTCF could not directly bind G4 structure [11], and that G4 formation at CTCF binding sites at the hTERT promoter may interfere with CTCF occupancy [50], a recent study presents another view. In vitro binding assays have demonstrated that G4 structure located near CTCF motifs may enhance CTCF binding [51], and genome-wide analysis have shown increased CTCF occupancy upon treatment with G4 stabilizers. While further investigation is required to

doi:10.6342/NTU202504256

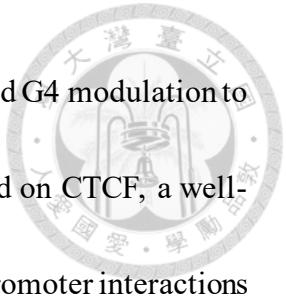
fully elucidate the relationship between CTCF and G4 structures, these findings suggest that DNA G4s may play a role in modulating CTCF occupancy.

Otherwise, recent studies have also explored the interaction between CTCF and RNAs. In vitro binding assays and crosslinking and immunoprecipitation followed by sequencing (CLIP-seq) analyses revealed that CTCF is capable of binding multiple non-coding RNAs [52-54]. For instance, Jpx was reported to regulate CTCF occupancy on its low-affinity binding site genome-wide through direct RNA-protein interaction and sequence similarity with the CTCF low-affinity binding motifs (LowOCs) [53]. However, whether such RNA-mediated regulation of CTCF occupancy is a widespread mechanism across the genome remains to be further investigated.

1.6 Hypothesis and aim in this study

Recent findings suggest that TERRA plays a regulatory role in modulating DNA G4 structures. Immunofluorescence staining revealed that nuclear G4 signals decreased upon TERRA KD and increased following the introduction of TERRA RNA [20, 21]. Consistently, G4 ChIP-seq analysis showed reduced G4 enrichment at TSSs upon TERRA depletion [21]. Integration with RNA-seq data further revealed that genes exhibiting G4 signal reduced at TSSs significantly overlapped with TERRA KD downregulated genes, suggesting that TERRA may influence gene expression through the regulation of G4 formation at promoter regions.

doi:10.6342/NTU202504256



However, the downstream mechanisms linking TERRA-mediated G4 modulation to transcriptional regulation remain unclear. To explore this, we focused on CTCF, a well-characterized chromatin architectural protein involved in enhancer-promoter interactions and transcriptional insulation. Given that CTCF plays a critical role in maintaining higher-order chromatin structure and gene regulation, previous studies have shown that CTCF can interact with multiple non-coding RNAs [53, 54] and can be regulated by a lncRNA Jpx at low-affinity binding sites [53]. In addition, recent studies have shown that G4 structures forming near CTCF motifs can promote its binding [51]. Based on these observations, we hypothesized that TERRA-mediated G4 formation near the CTCF binding sites may influence on CTCF occupancy, potentially at their lower binding sites, and further contributing to gene regulation.

doi:10.6342/NTU202504256

Chapter 2 Materials and Methods

2.1 HeLa cell culture

About 2×10^6 Human HeLa cells were thawed in a 37°C water bath, and the freezing medium containing DMSO was removed by centrifugation at 1000 rpm for 3 minutes. The cells were plated onto a 10 cm dish with DMEM containing 10% FBS, 2% L-glutamine, pen-strep, within the 37°C incubator containing 5% CO_2. For passage, cells were collected with trypsin-EDTA and maintained at a density of 1×10^6 per 10 cm dish, with passaging performed every two days. The cells were ready for small RNA transfection with at least one passage after thaw from liquid nitrogen.

2.2 Small RNA transfection

Small RNA transfections were performed in HeLa cells using Lipofectamine RNAiMAX (Thermo Fisher Scientific, #13778150) according to the manufacturer's protocol. Three distinct 24-mer small RNA oligos were used (see Supplementary Table S1).

For sample preparation, 18-mm circular coverslips were sterilized with boiled RO water and placed in each well of a 24-well plate. In each well, 4×10^4 HeLa cells were reverse-transfected with 0.25 ul of 100 μM small RNA oligos and 2 ul Lipofectamine RNAiMAX in medium containing 1/4 volumes of Opti-MEM.

doi:10.6342/NTU202504256

After 36 hours, cells were washed with dPBS, and transfection efficiency was assessed via live-cell microscopy in cells transfected with fluorescent labeled oligos. Cells on coverslips were then processed for downstream applications.

2.3 Slide preparation for staining

Cells attached to coverslips in 24-well plates were first washed with PBS on ice for 5 minutes, then permeabilized with 0.2% CSKT on ice for 12 minutes. Afterward, cells were fixed with 4% PFA at RT for 10 minutes. The cells were ready for immunofluorescence and could be stored at -20°C in 70% EtOH.

2.4 Immunofluorescence (IF)

Immunofluorescence staining for DNA G4 and BLM was conducted on coverslips attached with small RNA oligo-transfected HeLa cells.

Slides were initially washed twice with PBS at RT for 5 minutes each. To ensure the BG4 antibody specificity for DNA G4 structures, cells were sequentially dehydrated using 70%, 80%, 90%, and 99% ethanol for 2 minutes each, air-dried at RT, and treated with 400 μg/ul RNase A at 37°C for 20 minutes to remove RNA. Cells were then washed three times with PBS for 5 minutes each and incubated with blocking buffer at RT for 30 minutes. Next, cells were incubated with primary antibodies diluted in blocking solution at RT for 2 hours (dilution details in Supplementary Table S4), followed by washed with

doi:10.6342/NTU202504256

0.2% PBST at RT three times for 5 minutes each time to remove excess primary antibodies. Subsequently, cells were incubated with secondary antibodies diluted in blocking buffer (1:500) at RT for 1 hour in the dark. Afterward, cells were washed with 0.2% PBST at RT for 5 minutes each time to remove unbound secondary antibodies.

Finally, cells were mounted with DAPI-containing mounting medium, sealed with nail polish, and stored at 4°C in the dark until imaging.

2.5 Microscopy data quantification

All immunofluorescence images were acquired using an Olympus IX80 microscope with a 60× oil-immersion objective. Image stacks consisted of 21 z-sections, each taken at 0.27 μm intervals. For quantification of DNA G4 and BLM signals, DAPI staining was used to defined nucleus boundaries, and total intensity within each nucleus was measured using Olympus CellSens (v2.3). For statistical analysis, difference in signal intensity between groups was analyzed using the non-parametric Mann-Whitney U test in GraphPad Prism (v8.0.2).

2.6 Mouse embryonic stem (ES) cell culture

About 3×10^7 mouse embryonic fibroblast (MEF) cells were thawed in a 37°C water bath, and the freezing medium containing 10% DMSO was removed by centrifugation at 1000 rpm for 3 minutes. The cells were seeded onto three 10 cm dishes coated with 0.2%

doi:10.6342/NTU202504256

gelatin. After attachment, the MEF medium was replaced with mES medium. 2×10^6 mESCs were thawed similarly, centrifuged to remove DMSO, and seeded onto one dish of MEF cells. The other two dishes were used to collect feeder-conditioned medium every 12 hours.

After 2 days of culture with medium changed every 12 hours, mESCs were isolated from the MEF cells. A total of 5×10^6 cells were seeded onto 15 cm dishes coated with 0.2% gelatin in mES medium mixed 1:1 with feeder-conditioned medium. Daily passages were performed, and cells were prepared for TERRA knockdown after at least one passage.

2.7 TERRA knockdown

TERRA knockdown was performed in mESCs using the Mouse Embryonic Stem Cell Nucleofector® Kit (Lonza, #VPH-1001), with either scramble and anti-TERRA LNA gapmers (see Supplementary Table S2).

Prior to nucleofection, the culture medium (a 1:1 mixture of mESC medium and feeder conditioned medium) was refreshed 3 hours in advance. For each nucleofection, 3×10^6 mESCs were harvested by centrifugation at 1000 rpm at RT for 3 minutes, followed by one wash with 1 mL cold dPBS. Cells were resuspended in 100 ul nucleofection mixture (4 ul of 100 uM scramble or anti-TERRA LNA gapmers, 18 ul Supplement 1,

doi:10.6342/NTU202504256

and 81 ul of mESC/MES Cell Nucleofector® Solution) thoroughly. Cells were transferred into a cuvatte, and nucleofection was carried out using nucleofector I device with program "A-23". Immediately following nucleofection, cells were transferred into a 6-well plate containing 2 mL of the medium (1:1 mESC medium and feeder-conditioned medium). Cells were collected for RNA extraction, ChIP, and western blotting after 12 hours.

2.8 RNA extraction

Cells were collected, washed twice with 1× PBS, and centrifuged at 400 g at 4°C for 5 minutes to remove supernatant after each wash. Cell pellets were lysed with 1 mL QIAzol Lysis Reagent and stored at -80°C. After thawing, 200 ul chloroform was added in to cell lysates, which were then inverted thoroughly for 30 seconds, and incubated at RT for 5 minutes. Followed by centrifugation at 21,100 × g at 4°C for 20 minutes, and the aqueous phase was collected and mixed with an equal volume of Acid-Phenol:Chloroform, pH 4.5 (with IAA, 125:24:1). After thoroughly inversion for 30 seconds and incubation at RT for 5 minutes, the mixture was centrifuged at 21,100 × g at 4°C for 15 minutes, and the aqueous phase was transferred into Lobind tubes containing 2 ul GlycoBlue. RNA was precipitated with 1.25× volume of isopropanol and stored overnight at -30°C.

RNA was pelleted by centrifugation at 21,100 × g at 4°C for 10 minutes, and the supernatant was removed. The pellet was washed twice with 1000 ul 80% EtOH, followed

doi:10.6342/NTU202504256

by the same centrifugation condition. The RNA pellet was then air-dried for 10 minutes, and resuspended in DEPC water for cDNA synthesis. For storage, 10 ul 3 M NaOAc and suitable volume of DEPC water were added to reach a total volume of 100 ul, followed by 1.25× volume of isopropanol (125 ul). RNA samples were mixed and stored at -30°C.

2.9 cDNA synthesis

Prepare 1 μg RNA sample in DEPC water to a total volume of 11 ul and mixed with 1 ul 100 mM dNTP and 1 ul 50 mM random hexamer, then incubated at 65°C for 5 minutes. Samples were kept samples on ice immediately after incubation. Add 4 ul SSIV buffer, 1ul 0.1 μM DTT, 1 ul RNaseOUT, and 1 ul reverse transcriptase (Thermo Fisher Scientific, #18090200) were added in the mixture. The samples were then incubated at 25°C for 10 minutes to allow primer annealing, 50°C for 50 minutes for reverse transcription, and 85°C for 5 minutes to inactivate the reverse transcriptase enzyme. The cDNA samples were ready for qPCR and could be store at -30°C.

2.10 qPCR (Real-time Quantitative Polymerase Chain Reaction)

qPCR was performed using iQ™ SYBR® Green Supermix (Bio-rad, #1708882) following the manufacturer's protocol. Each reaction contained 2 ul of DNA sample and 13 ul of master mix (0.1 ul of each forward and reverse primers, 5.3 ul of SYBR Green

doi:10.6342/NTU202504256

Supermix, and 7.5 ul of DEPC-treated water), and was loaded into Hard-Shell® 96-Well PCR Plates (Bio-Rad, #HSP9601). The qPCR program was set as follows: initial denaturation at 95°C for 3 minutes, followed by 40 cycles of 95°C for 10 seconds and 58°C for 1 minutes for annealing and extension. A final denaturation at 95°C for 10 seconds was performed, followed by a melting curve analysis from 65°C to 95°C to verify amplification specificity.

To assess TERRA knockdown efficiency, RNA samples isolated from mESCs were first reverse transcribed into cDNA, and subjected to qPCR using primers of Erdr1, mChrY, and GAPDH (see Supplementary Table S3).

To validate the CTCF ChIP, sonicated DNA samples pulled down by CTCF antibody (Cell Signaling Technology, #2899S) or IgG control (Cell Signaling Technology, #2729S) were analyzed by qPCR. Primers targeting H19 (positive control) and 5nc (negative control) genomic loci were used as a previously published paper mentioned [55] (see Supplementary Table S3). 10% input samples were serially diluted into 1%, 0.1%, 0.01%, and 0.001% input for generating standard curves.

2.11 Chromatin immunoprecipitation for CTCF

6.5×10^6 mESCs were collected into 15 mL tube and resuspended in 5 mL of mES medium. To crosslink DNA-protein complexes, 135 ul of 37% formaldehyde (final

doi:10.6342/NTU202504256

concentration: 1%) was added. The suspension was mixed thoroughly by pipetting and incubated at RT for 5 minutes. Crosslinking was quenched with 514 ul of 1.25 M glycine (final concentration: 0.1 M), followed by mixing and incubation at RT for 5 minutes. Cells were centrifuged at 1000 rpm for 3 minutes, and washed with 1 mL dPBS, centrifuged under the same conditions, and resuspended in 1 mL dPBS. The cell suspension was split into two Lobind tubes, centrifuged at 400 × g for 6 minutes at 4°C. After removing supernatant, cell pellets were snap-frozen in liquid nitrogen and stored at -30°C.

For chromatin immunoprecipitation, 6.5×10^6 fixed mESCs were resuspended in 500 ul of Buffer 1 (50 mM HEPES (pH 8.0), 150 mM NaCl, 1 mM EDTA (pH 8.0), 0.5% NP-40, 0.25% Triton X-100) supplemented with 2× protease inhibitor cocktail (PIC), mixed thoroughly, and rotated at 4°C for 10 minutes. Nuclei were pellet at by centrifugation at 13,800 × g at 4°C for 5 minutes. The pellets were resuspended in 500 ul Buffer 2 (10 mM Tris (pH 8.0), 200 mM NaCl, 5 mM EDTA (pH 8.0), 2.5 mM EGTA (pH 8.0)) supplemented with 2× PIC, mixed thoroughly, and rotated at 4°C for 10 minutes. After centrifugation at 13,800 g at 4°C for 5 minutes, the pellets were resuspended in 600 ul Buffer 3 (10 mM Tris (pH 8.0), 5 mM EDTA (pH 8.0), 2.5 mM EGTA (pH 8.0)) supplemented with 2× PIC. Then, 12.5 ul of 20 mg/mL RNase A (final concentration: 0.4 mg/mL) was added to each sample, followed by incubation at 37°C for 30 minutes to.

doi:10.6342/NTU202504256

Next, 32.5ul 10% N-lauroyl sarcosine (final concentration: 0.5%) was added, mixed thoroughly, and rotated at 4°C for at least 10 minutes.

Samples were sonicated using Covaris S-series S2 instrument under the following settings: 5% duty cycle, intensity 4, water level 12, 200 bursts/cycle, 20 cycles, and temperature between 4°C~7°C. For each round of sonication, 130 ul of sample was loaded into a microTUBE AFA Fiber Snap-Cap (Covaris, #520045). This step was repeated until all samples were processed. The sonicated samples were transferred into clean Lobind tubes and kept on ice. After sonication, samples were centrifuged at 21,100 g at 4°C for 10 minutes, and supernatants were collected.

To assess the size of sonicated DNA fragments, 20 ul of each sonicated sample was taken for reverse crosslinking. Each sample was treated with 180 ul TES buffer (50 mM Tris (pH 8.0), 10 mM EDTA (pH 8.0), 1% SDS) supplemented with 2 ul of 20 mg/mL Protease K (final concentration: 0.2 mg/mL), followed by incubation at 55°C for 1 hour. The reverse-crosslinked DNA was purified using FavorPrep™ GEL/PCR Purification Kit (Favorgen, #FAGCK 001-1). The purified DNA was loaded into 2.5% agarose gel for electrophoresis, post-stained with SYBR-green I, and visualized under UV light. DNA fragments were expected to be approximately 200 bp. DNA quality and concentration were also assessed using NanoDrop spectrophotometer.

doi:10.6342/NTU202504256

For immunoprecipitation, 5 μg of each sonicated chromatin sample (without reverse crosslinking) was added in Buffer 3 at a total volume of 150 ul, then mixed with 150 ul of 2× IP buffer (final concentration: 1×; 2% Triton X-100, 300 mM NaCl, 30 mM Tris-HCl (pH 8.0), 1× PIC) with 0.14 μg of CTCF or IgG antibody (see Supplementary Table S4). The antibody incubation was carried out with rotation overnight at 4°C. An additional 0.5 μg of each sonicated sample was set aside as a 10% input control.

To capture antibody-chromatin complexes, 20 ul Dynabeads™ Protein G (Thermo Fisher Scientific, #10004D) was added to each immunoprecipitated sample and rotated at 4°C for 2 hours. Beads were then isolated using a magnetic rack. The beads were then washed three times with 1 mL RIPA-1-500 mM NaCl (500 mM NaCl, 50 mM HEPES (pH 8.0), 10 mM EDTA (pH 8.0), 0.5% Sodium deoxycholate, 1% NP-40), followed by three washes with 1 mL RIPA-2-250 mM NaCl (250 mM NaCl, 50 mM HEPES (pH 8.0), 10 mM EDTA (pH 8.0), 0.5% Sodium deoxycholate, 1% NP-40). Each wash was performed with rotation at 4°C for 5 minutes.

Beads were subsequently washed with 500 ul TEN buffer (50 mM NaCl, 1 mM EDTA (pH 8.0), 10 mM Tris (pH 8.0)), and supernatants were removed promptly using a magnet rack. Elution was performed by adding 200 ul TES buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 1% SDS) to the beads, followed by incubation at 65°C for 15 minutes with inversion every 2 minutes. Supernatants were collected. For 10% input

doi:10.6342/NTU202504256

controls, TES buffer was added to reach a final volume of 200 ul, and samples were incubated under the same condition as immunoprecipitated samples.

For reverse crosslinking, both immunoprecipitated and input samples were treated with 1 ul of 20 mg/mL RNase A (final concentration: 0.1 mg/mL) and incubated at 37°C for 20 minutes, followed by addition of 2 ul of 20 mg/mL Protease K (final concentration: 0.2 mg/mL) treatment at 55°C overnight. DNA fragments were purified using FavorPrep™ GEL/PCR Purification Kit and stored at -30°C for subsequent qPCR analysis or library preparation for sequencing.

2.12 Library preparation for sequencing

Library preparation was performed using the NEBNext ULtra II DNA Library Prep Kit for Illumina (New England Biolabs, #E7103S), according to the manufacturer's protocol, with all reaction volumes reduced to 50% of the recommended amounts.

For end repair, 5’ phosphorylation, and dA-tailing, 25 ul of each DNA sample was mixed with 1.5 ul NEBNext ULtra II End prep Enzyme Mix (New England Biolabs, #E7646A) and 3.5 ul NEBNext ULtra II End prep Reaction Buffer (New England Biolabs, #E7647A). The reaction was incubated at 20°C for 30 minutes, followed by 65°C for 30 minutes.

doi:10.6342/NTU202504256

For adaptor ligation, the reaction was supplemented with 15 ul NEBNext ULtra II Ligation Master Mix (New England Biolabs, #E7648A), 0.5 ul NEBNext Ligation Enhancer (New England Biolabs, #E7374A), and 1.25 ul volume of 5× diluted NEBNext Adaptor for Illumina (New England Biolabs, #E7337A), then incubated at 20°C for 15 minutes.

To remove the uracil-containing strand, 3 ul User enzyme (New England Biolabs, #E7428AA) was added to each ligated sample and incubated at 37°C for 30 minutes. Then, 53.25 ul of nuclease-free water (Thermo Fisher Scientific, #AM9938) was added to bring the total volume to 100 ul.

To select DNA fragments of the desired size, each sample was mixed with 0.3× volume of NEBNext Sample Purification Beads (New England Biolabs, #E7104S) and 20 ul nuclease-free water, and incubated at RT for 10 minutes. The tubes were then placed on a magnetic rack for 2 minutes to separate the beads, and the supernatants were added with 0.5× volume of beads (50 ul) followed by incubation at RT for 10 minutes. The tubes were placed on magnetic rack for 2 minutes to isolate the beads. The beads were washed twice with 200 ul freshly prepared 80% EtOH for 30 seconds each, and briefly centrifuged at 500 g at RT for 10 seconds to remove EtOH. After air dry the beads, DNA was eluted by resuspending the beads in 20 ul 0.1× TE buffer (Sigma-Aldrich, #93283) and incubating at RT for 10 minutes. The tubes were centrifuged at 100 g at RT for 1 minute,

doi:10.6342/NTU202504256

and the supernatants were collected with magnetic rack. The eluted samples were either used immediately for amplification or store at -20°C.

To assess the optimal cycle number for PCR amplification of each size-selected sample, a qPCR test was performed. For each sample, 5 ul NEBNext ULtra II Q5 Master Mix (New England Biolabs, #E7649A), 0.4 ul index primer (New England Biolabs, #E7335L), 0.4 ul universal primer (New England Biolabs, #S33102), 0.2 ul volume of 20× diluted SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific, #S33102), and 2 ul nuclease-free water were mixed with 2 ul of the samples per reaction. Each sample was tested in a single replicate. The qPCR program was set as follows: initial denaturation at 98°C for 30 seconds, followed by 25 cycles at 98°C for 10 seconds and 65°C for 75 seconds for annealing and extension. A final extension was performed at 65°C for 5 minutes. The cycle number for library amplification was determined based on the midpoint of the exponential phase of each sample's qPCR amplification curve. PCR amplification was then performed using 25 ul Q5 Master Mix, 2 ul index primer (sample-specific), 2 ul universal primer, and nuclease-free water to a total volume of 50 ul, following the same program as the qPCR test.

Following amplification, size selection was performed in two rounds. First, 0.9× volume of beads (45 ul) was added to each PCR product, incubated at RT for 10 minutes, and placed on magnetic rack for 2 minutes to isolate the beads. The beads were washed

doi:10.6342/NTU202504256

twice with 200ul freshly prepared 80% EtOH for 30 seconds, briefly centrifuged at 500 g at RT for 10 seconds, and EtOH was removed with magnetic rack. After air dry the beads, DNA was eluted in 52 ul 0.1× TE buffer and incubated at RT for 10 minutes. For the second round of size selection, 0.9× volume of beads (45 ul) was added again, followed by the same washing and elution steps. Final elution was performed in 15 ul 0.1× TE buffer, incubated at RT for 10 minutes. The final library samples were collected by centrifugation at 500 g at RT for 1 minute and placement on magnet rack for 5 minutes. The supernatants containing DNA libraries were store at -20°C until sequencing.

For quality control, the size distribution of library samples were accessed using an Agilent High Sensitivity DNA Kit (Agilent Technologies, #5067-4626), and DNA concentration was quantified by NEBNext Library Quant Kit for Illumina (New England Biolabs, #E7630S). The expected fragment size should be about 350 bp. For sequencing, 50 ng of each library sample, prepared with unique index primers, was pooled and submitted to Genomics company for Illumina sequencing.

2.13 NGS analysis for CTCF ChIP-seq

For CTCF ChIP-Seq analysis, the raw sequencing files (FASTQ format) were downloaded from a secure link provided by Genomics company. The integrity of the download files was verified using md5sum to ensure no corruption during data transfer.

doi:10.6342/NTU202504256

Raw sequencing reads were first trimmed to remove adaptor sequences using TrimGalore (v0.6.3), and aligned to the mm10 reference genome using Bowtie2 (v2.4.2). The resulting SAM files were converted to BAM format, sorted, deduplicated to remove PCR duplicates, and indexed using SAMtools (v1.13). To visualize the ChIP-seq signal, coverage tracks in BIGWIG format were generated using bamCoverage (deepTools v3.3.1) and viewed in IGV. Correlation analysis between replicates of scramble and TERRA KD groups was performed in R using Pearson's correlation.

To get the CTCF peaks, the ChIP signals were first normalized by subtracting their respective input signals using bamCompare (deepTools v3.3.1). Peaks were then called for each replicate individually using macs2 callpeak (MACS v2.2.9.1). To identify high-confidence peaks, the peaks shared between replicates were determined using intersectBed (BEDTools v2.31.1). Only the intersected peaks between replicates within either scramble or TERRA KD group, which called as SC KD and TE KD CTCF peaks, were used for subsequent analyses. To further assess the influence of TERRA KD, SC KD and TE KD CTCF peaks were then intersected using intersectBed (BEDTools v2.31.1), as the SC/TE common represents both in SC KD and TE KD CTCF peaks, while SC and TE unique CTCF peaks are non-intersected CTCF peaks for SC KD and TE KD CTCF peaks, respectively.

doi:10.6342/NTU202504256

For motif discovery, the DNA sequences corresponding to CTCF peaks were extracted from the mm10 genome fasta using Bedtools2. De novo motif analysis was then performed with meme-chip (MEME v5.5.3), setting the maximum motif length to 20 bp.

To analyze the genomic distribution of CTCF binding sites, CTCF peaks were annotated to mouse genome using HOMER. The resulting annotation tables were used to calculate distribution percentages and generate bar charts in R. To visualize CTCF signal coverage at common and unique peaks between scramble KD and TERRA KD groups, heatmaps and average profiles were generated using the computeMatrix reference-point, plotProfile, and plotHeatmap (deepTools v3.3.1), based on bigwig files produced by bamCompare and the peak regions defined in BED files.

To assess CTCF coverage, input-normalized bigwig files from CTCF ChIP-seq experiments upon scramble and TERRA KD in mESCs were analyzed. CTCF signal intensity within selected genomic regions was calculated using either computeMatrix scale-regions (deepTools_v3.3.1) or computeMatrix reference-point (deepTools_v3.3.1), and visualized as metaplots and heatmaps using plotProfile (deepTools_v3.3.1) and plotHeatmap (deepTools_v3.3.1), respectively.

To quantify the value of CTCF coverages upon scramble and TERRA KD, average CTCF coverage values within each selected genomic regions were calculated using the

doi:10.6342/NTU202504256

resulting CTCF coverage matrixes from computeMatrix scale-regions (deepTools_v3.3.1) and plotted as violin plots.

To better represent the central tendency of the data, outliers in each group were definded as values falling outside 1.5 times the interquartile range (IQR) below the first quartile or above the third quartile, and were removed.

2.14 NGS analysis for DNA G4 signal nearby CTCF binding sites

The G4 ChIP-seq data for scramble and TERRA KD mESCs were obtained from a previous study [21]. These data were processed using the same pipeline as described for the CTCF ChIP-seq data (see Materials and Methods 2.13), up to the generation of input-normalized bigWig files for visualization in IGV.

To examine the distribution of G4 signals around CTCF peaks, log2 fold changes of G4 ChIP-seq input-normalized signals between TERRA and scramble KD were calculated using bigwigCompare (deepTools v3.3.1). Heatmaps were generated centered on the summits of SC/TE common and TE unique CTCF peaks with ± 500 bp flanking regions. Based on G4 signal changes, these CTCF peaks were categorized into four clusters: cluster 1 is G4 signal remained unchanged or increased; cluster 2 and 3 are G4

doi:10.6342/NTU202504256

signal decreased on the downstream and upstream side of the center of CTCF peaks, respectively; and cluster 4 is G4 signal decrease at the center of CTCF peaks.

To evaluate whether G4-mediated changes affected CTCF occupancy, the average log2 fold changes of CTCF ChIP-seq input-normalized signals between TERRA and scramble KD calculated for each CTCF peak in the four clusters using computeMatrix scale-regions (deepTools v3.3.1). Violin plots were used to visualize the distribution of CTCF signal changes.

To better represent the central tendency of the data, outliers in each group were definded as values falling outside 1.5 times the interquartile range (IQR) below the first quartile or above the third quartile, and were removed. The percentage of removed outliers for each group was as below: cluster 1 CTCF peaks: 0.92% (G4 replicate 1 with CTCF replicate 1), 0.94% (G4 replicate 1 with CTCF replicate 2), 1.02% (G4 replicate 2 with CTCF replicate 1), 1.06% (G4 replicate 2 with CTCF replicate 2); cluster 2 CTCF peaks: 0.94% (G4 replicate 1 with CTCF replicate 1), 0.76% (G4 replicate 1 with CTCF replicate 2) , 0.82% (G4 replicate 2 with CTCF replicate 1), 0.77% (G4 replicate 2 with CTCF replicate 2); cluster 3 CTCF peaks: 1.01% (G4 replicate 1 with CTCF replicate 1), 0.72% (G4 replicate 1 with CTCF replicate 2) , 0.92% (G4 replicate 2 with CTCF replicate 1), 0.86% (G4 replicate 2 with CTCF replicate 2); cluster 4 CTCF peaks: 1.92% (G4 replicate 1 with CTCF replicate 1), 2.49% (G4 replicate 1 with CTCF replicate 2) ,

doi:10.6342/NTU202504256

1.81% (G4 replicate 2 with CTCF replicate 1), 2.03% (G4 replicate 2 with CTCF replicate 2).

2.15 NGS analysis of CTCF signal at CTCF binding sites with or without cohesin co-binding

The Smc1 and Smc3 ChIP-seq data in wild-type mESCs were obtained from a previous study [56]. The ChIP-seq analysis method was the same as the method described in CTCF ChIP-seq until the peak calling (see Materials and Methods 2.13). The Smc1 and Smc3 peaks were also determined by intersecting the peaks between replicates.

To visualize the increase in CTCF signal coverage after TERRA KD at cohesin-CTCF co-binding sites, CTCF peaks were first categorized into two groups, those with or without cohesin binding, using intersectBed (BEDTools v2.31.1). The fold change in CTCF signal was determined by comparing the log2 ratio of TERRA KD to scramble KD groups, calculated using bigwigCompare (deepTools_v3.3.1) with input-subtracted bigwig files.

To compute the average log2 CTCF signal within each CTCF peak, a coverage matrix was generated using computeMatrix (deepTools_v3.3.1), and the mean value across each row was taken as the representative CTCF signal. Violin plots were generated

doi:10.6342/NTU202504256

to visualize the distribution of CTCF signal increase at CTCF binding sites with or without cohesin binding.

To better represent the central tendency of the data, outliers in each group were definded as values falling outside 1.5 times the interquartile range (IQR) below the first quartile or above the third quartile, and were removed. The percentage of removed outliers for each group was as below: SC KD CTCF with Smc1: 0.74% (replicate 1), 0.81% (replicate 2); SC KD CTCF without Smc1: 0.91% (replicate 1), 1.36% (replicate 2); SC KD CTCF with Smc3: 0.54% (replicate 1), 0.88% (replicate 2); SC KD CTCF without Smc3: 0.95% (replicate 1), 1.37% (replicate 2).

2.16 Cell fractionation

2×10^6 mESCs were collected and washed with 1 mL Buffer A (10 mM HEPES (pH=7.9), 5 mM $MgCl_2$, 0.25 M sucrose, 1 mM EDTA (pH=8.0)) containing 0.5 mM DTT, 1 mM PMSF, and 2× PIC. Cells were centrifuged at 25,000 g at 4°C for 5 minutes, and the supernatant was discarded. The pellet was lysed with Buffer A containing 0.2% NP-40, and supplemented with 2× PIC and 1 mM PMSF on ice for 5 minutes. After centrifugation at 25,000 g at 4°C for 10 minutes, the supernatant was collected as cytosolic fraction.

doi:10.6342/NTU202504256

The pellet was washed in 200 ul Buffer A supplemented with 2× PIC and 1 mM PMSF, followed by centrifugation at 25,000 g at 4°C for 5 minutes. The pellet was resuspend in Buffer B (10 mM HEPES (pH=7.9), 1 mM EDTA (pH=8.0), 1.5 mM $MgCl_2$, 25% glycerol, 300 mM NaCl) supplemented with 2× PIC and 1 mM PMSF, and incubated on ice for 30 minutes. After centrifugation at 18,000 g at 4°C for 20 minutes, the supernatant was collected as nucleus-soluble fraction.

The pellet was resuspended in 25 ul Benzonase buffer (25 mM Tris-HCl (pH=7.5), 20 mM NaCl, 10% Glycerol, 0.1 mM EDTA (pH=8.0)) supplemented with 2× PIC, 1 mM PMSF,and 0.5 ul Benzonase. The samples were incubated on ice for 30 minutes to digest the DNA. Subsequently, 25 ul 2 M KCl and 0.5 ul 100× PIC were added to each sample, followed by incubation on ice for 10 minutes to promote protein dissociation from chromatin. The final supernatant was prepared as chromatin-bound fraction.

2.17 Western blotting

All the different concentrations of sample buffer were diluted from 6× sample buffer (0.3M Tris-HCl (pH=6.8), 6% SDS, 50% Glycerol, 0.6M DTT, and 0.05% Bromophenol Blue).

For preparation of total protein sample, 8×10^5 mESCs were collected and centrifuged at 400 g at 4°C for 5 minutes to remove supernatant. The pellet was washed

doi:10.6342/NTU202504256

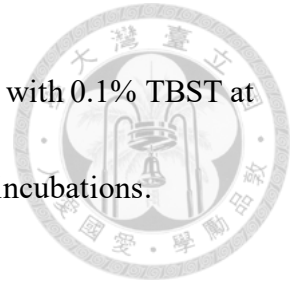
with 1 mL of PBS, followed by centrifugation under the same conditions. The pellet was resuspended in 62 μL of 2× sample buffer containing 1× PIC.

For preparation of the fractionated cell samples, 125 μL and 80 μL of 4× sample buffer were added to the cytosolic and nucleus-soluble fractions, respectively. For the chromatin-bound fraction, 200 μL of 2.5× sample buffer was added to reduce the salt concentration.

All protein samples were denatured at 95°C for 5 minutes before performing western blotting. 15 ul of each sample were loaded, and proteins were separated based on their molecular weights using 12% SDS-polyacrylamide gel electrophoresis. After electrophoresis, proteins were transferred from the gel onto an Immobilon-P PVDF membrane (Merck, #IPVH85R), which was pre-activated with methanol, at 300 mA on ice for 2.5 hours or at 33 V at 4°C overnight.

The membrane was then blocked with 5% milk/0.1% TBST (nonfat dried milk powder dissolved in 1× TBS containing 0.1% Tween-20) at RT for 1 hour to reduce non-specific binding of the primary antibody. It was then incubated with the primary antibody, diluted in 5% milk/0.1% TBST (antibodies and their dilutions see Supplementary Table S4) at 4°C overnight. Afterward, the membrane was incubated with the secondary antibody (see Supplementary Table S4), diluted in 5% milk/0.1% TBST, at RT for 1 hour.

doi:10.6342/NTU202504256



To remove excess antibodies, the membrane was washed three times with 0.1% TBST at RT for 10 minutes each after both primary and secondary antibody incubations.

To detect the protein signals, the membrane was incubated with ECL mixture using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific, #34580) at RT. The signals were detected using CCD camera, UVP ChemStudio Plus.

2.18 Quantification for western blotting

For quantification of chromatin-bound and whole-cell CTCF protein levels, CTCF signals from western blotting were measured using ImageJ and normalized to the corresponding H3 signal levels. The quantification results were plotted using GraphPad Prism (v8.0.2).

2.19 NGS analysis for CTCF CLIP-seq

The CTCF CLIP-Seq data in wild-type mESCs were obtained from a previous paper [57]. Raw data were downloaded from the NCBI SRA database and converted to FASTQ format using SRAtools (v3.0.3). Adaptor sequences were trimmed using TrimGalore (v0.6.3), and reads containing four consecutive TTAGGG repeats were filtered using bbduk (BBTools v37.62). rRNA and tRNA sequences were subsequently removed using Bowtie2 (v2.4.2), and the remaining reads were aligned to mm10 genome using STAR (v2.7.10a). PCR duplicates were removed using SAMtools (v1.13).

doi:10.6342/NTU202504256

The number and percentage of reads containing four TTAGGG repeats reads were recorded and used to generate quantitative comparisons between the CLIP-seq groups.

To provide another control for repeat-containing transcript levels, RNA-seq data from scramble KD mESCs, obtained from a previously published paper [18], were analyzed using the same processing as described above for CLIP-seq data.

2.20 NGS analysis for TERRA KD RNA-seq

The RNA-seq data from TERRA KD mESCs were obtained from a previously published paper [18]. Raw sequencing files were downloaded from the NCBI SRA database and converted to FASTQ format using SRAtools (v3.0.3). Adaptor sequences were trimmed using TrimGalore (v0.6.3), and aligned to mm10 genome using STAR (v2.7.10a). Aligned reads were sorted and converted into BAM format using STAR (v2.7.10a). Read counts were then calculated using Htseq (v0.13.5). The resulting count table was used to identify differentially expressed genes (DEGs) using DESeq2 in R, and the DEG table for TERRA KD was generated.

doi:10.6342/NTU202504256

Chapter 3 Results

3.1 RNA G4 promotes DNA G4 formation and reduce BLM occupancy.

Since our previous investigation into the epigenetic roles of TERRA was conducted using a TERRA KD system in mESCs [21], we sought to establish an effective RNA transfection system to complement and validate those findings. To this end, we referred to a published system [58] and subsequently developed transient RNA transfection systems in both mESCs [20] and HeLa cells. Details of these three experimental systems are summarized in Table 1. HeLa cells were chosen due to their relatively low endogenous TERRA expression [14].

In this study, we focus on the system established in HeLa cells. The RNA oligonucleotide (oligo) used was a 24-nucleotide sequence containing four tandem UUAGGG repeats, hereafter referred to as TERRA RNA. As a negative control, we designed a sequence based on a commonly used siRNA targeting the Luciferase gene with minor modifications, referred to as Ctrl RNA. To further distinguish between TERRA-specific effects and general RNA G4-mediated effects, an additional RNA oligo was designed containing four tandem GGGGCC repeats, referred to as G4C2 RNA. The G4C2 repeat is located in the C9orf72 gene and has been associated with ALS/FTD pathology. All synthetic RNA oligos were labeled with a Cy3 fluorescent dye at the 5' end. RNA

doi:10.6342/NTU202504256

oligos were introduced into HeLa cells, and transfection efficiency was assessed 36 hours post-transfection using live-cell fluorescence microscopy (Figure 1A). Most cells show Cy3 signals in the cytoplasm indicating a successful transfection efficiency. However, the subcellular localization patterns varied among the three RNA oligos. For Ctrl RNA, Cy3 signals were predominantly cytoplasmic. In contrast, approximately 30-40% of TERRA RNA transfected cells exhibited Cy3 signal accumulation in the nucleus (visual estimation; no quantification performed). Otherwise, G4C2 RNA formed multiple foci within the cells. The observation of G4C2 oligos is consistent with previous studies indicating that G4C2 repeated-containing RNA form RNA foci or stress granules [27, 59].

Although the RNA transfection system may require further optimization, we proceeded with downstream experiments to investigate the epigenetic roles of TERRA. In our previous study using a TERRA KD system in mESCs, depletion of TERRA reduced nuclear DNA G4 levels and increased the chromatin occupancy of several G4 helicases. Conversely, introducing TERRA RNA into mESCs was found to increase DNA G4 levels and decrease BLM occupancy [20]. To validate these findings in RNA transfection system in HeLa cells, I performed immunofluorescence staining targeting DNA G4 and BLM, which is a well-characterized G4 helicase. To specifically detect the DNA G4 structures (Figure 1B, C; Supplementary Figure 1A-D), cells were treated with RNase A prior to immunostaining to eliminates RNA G4, thereby avoid false-positive signals.

doi:10.6342/NTU202504256

Quantification of nuclear DNA G4 intensity revealed a consistent increase in the TERRA RNA-transfected cells compared to ctrl RNA transfected cells across three independent replicates. In contrast, transfection with G4C2 RNA yielded variable results, showing increased DNA G4 in two replicates and a decreased in another one.

For BLM staining (Figure 2; Supplementary Figure 2A-D), quantification of nuclear BLM intensity showed a consistent decrease in TERRA RNA transfected cells across three independent replicates. In the G4C2 RNA transfected cells, two replicates exhibited a decreased in BLM occupancy, while one showed no significant change. These results are consistent with the findings from both the TERRA KD system and introducing small RNA oligos in mESCs, supporting TERRA promotes DNA G4 structure formations and expels BLM from chromatin.

3.2 Investigation of CTCF occupancy upon TERRA depletion in mESCs.

To elucidate whether CTCF occupancy on chromatin is altered upon TERRA depletion, CTCF ChIP-seq was performed in TERRA KD mESCs. TERRA KD was conducted using antisense oligonucleotides (ASOs) fused with locked nucleic acid (LNA) sequence [18]. The ASOs was designed to bind TERRA RNA and induce endonuclease RNase H-mediated degradation specific to TERRA RNA (Supplementary Table S2). The KD efficiency in the two replicates used for CTCF ChIP-seq was evaluated by RT-qPCR

doi:10.6342/NTU202504256

(Figure 3A). After TERRA KD, expression levels of Erdr1 and mChrY were reduced to 36.9% and 63.9% of control levels, respectively. To validate the CTCF ChIP results, ChIP-qPCR was performed using primers targeting known CTCF-positive and -negative binding sites (Figure 3B). The result showed higher enrichment at the positive binding regions in the CTCF ChIP groups, indicating successful pulldown of CTCF-bound genomic regions.

To access the correlation between the two replicates of CTCF ChIP-seq, scatter plots of CTCF coverages were generated. The Pearson's correlation coefficients were calculated to be 0.912 and 0.955 for the scramble and TERRA KD groups, respectively (Figure 3C). CTCF coverage at known positive and negative binding regions was visualized using IGV (Figure 4). CTCF peaks were identified by intersecting the two replicates within each group, which were defined as SC KD and TE KD CTCF peaks for downstream analysis (Figure 3D). These peaks were further validated by comparing them with wild-type (WT) CTCF peaks (Figure 3E), which were determined by overlapping CTCF peaks identified in three independent WT CTCF ChIP-seq datasets [57, 60, 61]. While the number of CTCF peaks in SC KD and TE KD groups (in this study) exceeded those in the WT group (published by other labs), a large proportion of the WT CTCF peaks overlapped with both SC KD and TE KD peaks. These results indicate that CTCF ChIP-seq data in this study is consistent with the data from previous studies.

doi:10.6342/NTU202504256

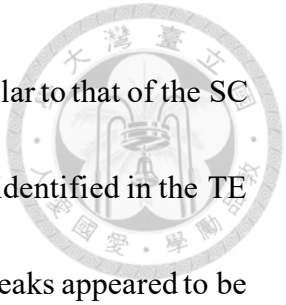
To investigate whether the CTCF binding motif changes after TERRA KD, meme-chip analysis was performed. The CTCF binding motifs identified in both SC and TERRA KD mESCs were highly similar (Figure 3F), and also resembled the well-characterized canonical CTCF motif [41]. To examine the genomic distribution of CTCF binding, histograms and summarized tables were generated to show the annotation result of the SC KD and TE KD CTCF peaks (Figure 3G). The genomic annotations of CTCF peaks were comparable between SC KD and TE KD mESCs. These results suggest that both binding motif and genomic distribution of CTCF are not significantly altered upon TERRA KD.

3.3 CTCF occupancy increases upon TERRA depletion.

To further investigate the impact of TERRA KD on CTCF, the CTCF peaks were classified into three groups: SC/TE common, SC unique, and TE unique CTCF peaks (Figure 4A). SC/TE common CTCF peaks refer to those present in both SC KD and TE KD groups, whereas SC unique and TE unique peaks are those identified exclusively in the SC KD or TE KD group, respectively. The number of TE unique CTCF peaks largely exceeds those in SC unique CTCF peaks, indicating that TERRA depletion gains more peaks compared to the control cells.

To examine potential changes in the CTCF binding motif after TERRA KD, meme-chip analysis was performed on common or unique CTCF peaks (Figure 4B). The CTCF

doi:10.6342/NTU202504256



binding motif identified in the SC/TE common CTCF peaks was similar to that of the SC unique peaks, and appeared as the reverse complement of the motif identified in the TE unique peaks. Additionally, the genomic distribution of these CTCF peaks appeared to be similar among SC/TE common, SC unique, and TE unique groups (Figure 4C). Taken together, these findings suggest that TERRA KD does not substantially alter CTCF binding motifs or its genomic distribution.

Since the number of SC unique peaks was relatively small (Figure 4A), and the CTCF ChIP-seq signals appeared to increase upon TERRA KD based on bigwig files visualized on IGV (Figure 6), we speculated that TERRA KD might enhance CTCF binding at its major binding sites. To test this hypothesis, CTCF coverage in SC KD and TE KD samples was visualized as metaplots and heatmaps, plotted separately for SC/TE common, SC unique, TE unique CTCF peaks (Figure 5A-C). Because most CTCF peaks are shorter than 500 bp, regions selected for plotting metaplots and heatmaps were defined as ± 500 bp flanking the summit of each peak. The heatmaps revealed that CTCF binding signals were generally increased at most CTCF binding sites after TERRA KD.

To better illustrate the differences in CTCF binding upon TERRA KD, The CTCF coverage of SC KD and TE KD groups were calculated for each peak within either SC/TE common or TE unique CTCF peaks, and the distribution of these values was visualized using violin plots (Figure 5D). In addition, the log2 fold change of CTCF signals between

doi:10.6342/NTU202504256

TE KD and SC KD groups were analyzed and plotted as metaplots and heatmaps (Figure 7). In replicate 1, the increase in CTCF signal after TERRA KD appeared comparable between SC/TE common and TE unique sites. In replicates 2, however, the increase was higher at SC/TE common sites compared to TE unique sites. Despite the different increase level of CTCF signals between replicates, the log2 fold change values indicate that CTCF signals increase after TERRA KD, suggesting that TERRA depletion increases CTCF occupancy at most CTCF binding sites.

3.4 CTCF binding increases significantly in regions with G4 reduction.

To investigate whether G4 structures influence CTCF occupancy upon TERRA depletion, I analyzed changes in G4 ChIP-seq signals near CTCF peaks obtained from a previously published paper [21]. Log2 fold changes of G4 signal between TE KD and SC KD groups were calculated and plotted as metaplots and heatmaps at the selected regions around summits of SC/TE common and TE unique CTCF peaks ± 500 bp (Figure 8A). Based on G4 signal dynamics, CTCF peaks were classified into four clusters using k-means clustering. Cluster 1 showed stable G4 signal, while clusters 2 and 3 exhibited moderate G4 reduction near the peak center. Notably, cluster 4 displayed a pronounced G4 decreased precisely at the summit of CTCF peaks.

doi:10.6342/NTU202504256

To determine whether changes in G4 signal correlate with CTCF occupancy, the log2 fold change of CTCF ChIP-seq signals between TE KD and SC KD groups were calculated and plotted as violin plots at these clustered regions (Figure 8B). All four clusters showed increased CTCF occupancy upon TERRA KD, but the increase was more significant in cluster 2 to 4. Among them, three of four quantitative results shows that cluster 4 exhibited the highest increase in CTCF binding, suggesting that loss of G4 at the center of CTCF peaks may enhance CTCF occupancy. This finding supports the previous reports showing that G4 formation at specific loci, such as the hTERT promoter, can interfere with CTCF binding [50].

Given that a recent in vitro study suggested that G4 structures adjacent to CTCF motifs might instead facilitate CTCF binding [51], direct comparison is limited by the absence of precise positional definitions in our dataset regarding the spatial relationship between G4 structures and CTCF motifs. Without this resolution, it remains difficult to compare the in vitro binding results with the CTCF data in this study.

Although the specific role of G4 structures in regulating CTCF occupancy remains unresolved, the global increase in CTCF signal upon TERRA depletion suggests that TERRA may modulate CTCF binding through additional G4-independent mechanisms. Further studies are required to elucidate these mechanisms.

doi:10.6342/NTU202504256

3.5 Increased CTCF occupancy after TERRA KD is more pronounced at cohesin co-binding sites

Given that CTCF is recognized as cohesin stalling factor essential for chromatin looping and TAD boundary formation, it is of interest whether the increase in CTCF occupancy following TERRA KD is more prominent at CTCF binding sites that are co-occupied by cohesin. To test this hypothesis, Smc1 and Smc3 ChIP-seq data from WT mESCs were obtained from a previously published paper [62], as Smc1 and Smc3 represent core components of cohesin complex. CTCF peaks identified in the SC KD group were categorized into groups based on whether they overlapped with Smc1 or Smc3 peaks (Figure 9A). The log2 fold change of CTCF coverage between TE KD and SC KD groups was calculated for each peak, and the distribution of these values was visualized using violin plots (Figure 9B). The results revealed that CTCF peaks co-bound by Smc1 or Smc3 exhibited significantly higher log2 fold changes in signal compared to peaks lacking cohesin binding. These findings suggest that the increase in CTCF occupancy observed upon TERRA KD preferentially occurs at cohesin-associated CTCF binding sites.

doi:10.6342/NTU202504256

3.6 TERRA KD leads to a modest increase in CTCF chromatin occupancy

To further validate the ChIP-seq observation that CTCF occupancy increases at most of its binding sites upon TERRA KD (Figure 5-7), cell fractionation followed by western blotting was performed using TERRA KD mESCs (Figure 10B, C; Supplementary Figure S3). Cells were fractionated into chromatin-bound, nucleus-soluble, and cytosolic compartments. The efficiency of TERRA KD was confirmed by qPCR, showing that Erdr1 and mChrY transcript level were reduced to 0.22% and 0.62% of control levels, respectively (Figure 10A). To assess whether TERRA KD affects not only the chromatin association of CTCF but also its overall protein expression, whole-cell lysate fractions were included in the analysis. The quantification revealed a consistent increase in chromatin-bound CTCF following TERRA KD, whereas the total CTCF protein level showed variability across replicates. These findings support the ChIP-seq results and suggest that TERRA knockdown may lead to a slightly increase in CTCF chromatin occupancy.

doi:10.6342/NTU202504256

Chapter 4 Discussion

In this study, we established an RNA transfection system in HeLa cells to investigate the regulatory role of TERRA on DNA G4 structures and BLM. Immunofluorescence staining revealed that introducing TERRA RNA led to increased nuclear DNA G4 signals and decreased BLM occupancy (Figure 1, 2; Supplementary Figure S1, S2), consistent with previous results observed in a similar system in mESCs, and in contrast to findings from TERRA KD system [20]. These results suggest that TERRA may modulate G4 helicase occupancy and regulate DNA G4 formation.

Given prior observations that G4 signals reduced at TSSs upon TERRA KD are associated with downregulated genes [21], we aimed to investigate the downstream mechanisms by which TERRA-mediated G4 modulation might influence gene expression. Considering recent studies implicating G4 structures in the regulation of CTCF binding [51], we hypothesized that TERRA might modulate gene expression by altering CTCF occupancy through changes in G4 formation.

To test this idea, I performed CTCF ChIP-seq in TERRA KD mESCs. The results revealed a global increase in CTCF occupancy across most binding sites (Figure 5-7), without major changes in either genomic annotation or binding motifs (Figure 3, 4). This finding was further supported by cell fractionation followed by western blotting, which showed slightly increased levels of chromatin-bound CTCF in TERRA KD mESCs

doi:10.6342/NTU202504256

(Figure 10; Supplementary Figure S3). Integration with previously published G4 ChIP-seq data [21] CTCF occupancy increased in both G4 remaining and reduced sites (Figure 8), suggesting that G4 modulation alone may not fully account for the observed effects. Notably, when comparing with previously published cohesin ChIP-seq data in WT mESCs [56], I noticed that TERRA KD-induced increases in CTCF signal were more pronounced at CTCF-cohesin co-binding sites (Figure 9). These findings suggest that TERRA may influence CTCF chromatin binding through a cohesin-associated mechanism, potentially contribute to the regulation of chromatin looping and gene expression. However, further studies are required to confirm the idea.

Before interpreting the biological relevance of our findings, it is essential to evaluate the robustness and limitations of the RNA transfection system used in this study.

In this study, we employed a synthetic RNA oligo containing four tandem repeats of the UUAGGG telomeric sequences to mimic TERRA. For the control, we altered the UUACCC RNA oligo design to a well-characterized non-targeting RNA oligo previously used in siRNA studies. This substitution was made after observing that the original C-rich control RNA exhibited significantly lower nuclear transfection efficiency in HeLa cells compared to the telomeric RNA. Despite the change, the new control oligo also showed poor nuclear uptake. In addition, we included a G4-forming RNA derived from G4C2 repeat sequence, which is known to form RNA G4 structures and is implicated in

doi:10.6342/NTU202504256

neurodegenerative disease. However, G4C2 RNA exhibited a tendency to form nuclear foci, which may reflect distinct intracellular bevaviors and limit its utility as a direct comparator to telomeric RNA in this context.

While the RNA transfection system successfully introduced synthetic RNA oligos into HeLa cells, accurately quantifying transfection efficiency remains challenging. In this study, Cy3-labeled RNA oligos were used to assess intracellular delivery via fluorescence microscopy, which enabled semi-quantitative evaluation of nuclear localization. However, no certain intensity threshold could distinguish transfected from untransfected cells at single-cell level. Although northern blotting can detect the overall abundance of transfected RNA, it lacks single-cell resolution and does not permit absolute quantification of intracellular RNA levels. Moreover, alternative validation methods such as RNA FISH are incompatible with subsequent G4 immunostaining, as RNase treatment is required prior to antibody incubation, and the CSKT buffer used in our protocol selectively retain chromatin-bound structures. These limitations should be considered when interpreting the observed changes in nuclear G4 levels and BLM occupancy following RNA transfection.

The observed increase in nuclear DNA G4 signal and reduction in BLM occupancy following introducing TERRA RNA uncover a potential role for TERRA in stabilizing DNA G4 structures by recruiting G4 helicases through its G4 forming capacity. BLM is

doi:10.6342/NTU202504256

a well-characterized DNA helicase that actively unwinds G4 structures in an ATP-dependent manner to maintain genome stability. The reduction of BLM signal in the presence of exogenous TERRA RNA suggests that TERRA may either compete with BLM for binding sites, or alter the chromatin environment in a way that limits BLM recruitment. Given that DNA G4s are enriched at regulatory genomic elements such as promoters and replication origins, the accumulation of G4 structures could have functional consequences for transcriptional regulation or DNA replication dynamics. Furthermore, the agreement between HeLa and mESC systems implies that this effect may be consistent within multiple cell lines. These findings support the emerging view that TERRA may act as a regulator of modulating DNA G4 homeostasis and G4 helicase occupancy across the genome.

As for the other part of this study to investigate the regulatory mechanism of TERRA-mediated G4 alternation influence on gene expression, the results suggest that TERRA may not regulate CTCF occupancy primarily through G4 modulation. Instead, ChIP-seq data point toward a more direct regulatory role for TERRA on CTCF binding, potentially independent of G4 structure dynamics.

To investigate the potential mechanisms through which TERRA regulates CTCF occupancy, we considered two possible directions: (1) whether TERRA can directly bind to CTCF, and (2) whether TERRA regulates CTCF indirectly through other proteins.

doi:10.6342/NTU202504256

To explore the first possibility, I examined the RNA-binding ability of CTCF. Previous studies have shown that CTCF could directly bind RNA [52-54]. A recent study developed a predictive model, DeepLncCTCF, to identify lncRNAs that are likely to interact directly with CTCF [54]. However, because TERRA transcripts originate from various subtelomeric regions and extend to telomeric repeats, their sequences are not consistent and are not included in the list of annotated lncRNAs. As a result, TERRA was not detected by this predictive approach. To evaluate potential direct interaction between TERRA and CTCF, I analyzed experimental data from TERRA iDRIP-MS [18] and CTCF CLIP-seq [57]. CTCF was not identified in the TERRA interactome generated by iDRIP-MS. Furthermore, analysis of CTCF CLIP-seq data in mESCs revealed no specific enrichment of the UUAGGG repeat sequence among RNA fragments bound by CTCF after UV crosslinking. However, as several studies have shown that CTCF does not directly bind to G4-forming regions [11], this may suggest that if CTCF interacts with TERRA, it is more likely to bind to the subtelomeric regions rather than the telomeric repeat sequences. Further direct binding assays are necessary to reach a definitive conclusion.

To investigate the possibility that TERRA might regulate CTCF through indirect interactions, I considered two potential mechanisms: (1) TERRA directly binds to CTCF cofactors, or (2) TERRA regulates the expression levels of CTCF factors.

doi:10.6342/NTU202504256

To evaluate these hypothesis, I utilized experimental data from TERRA iDRIP-MS and RNA-seq within TERRA KD mESCs [18]. For the selection of candidate CTCF interactomes, I first referred to a supplementary table from a review paper [63], which compiled CTCF interactomes from three sources: literature curation, the STRING database [64], and the Integrated Interactions Database [65]. In addition, I included data from a systematic ChIP-seq analysis study [66], in which they compiled a set of highly conserved CTCF binding sites across multiple human cell types and ranked co-binding protein factors based on their ChIP-seq overlap with these conserved CTCF peaks. From their ranked list, I selected the top 30 proteins as additional CTCF interactome candidates. I then integrated these four datasets and retained only those proteins that appeared in at least two sources, resulting in a final list of 34 condidate CTCF interactomes.

Comparison with the TERRA iDRIP-MS data revealed that only CHD8 was detected among the candidates. DEG analysis from RNA-seq further identified three candidates, including *Brd2*, *Chd8*, and *Smarca4*, as significantly altered upon TERRA KD (Table 3A). BRD2 has previously been shown to colocalize extensively with CTCF binding sites, although its KO does not affect CTCF occupancy [67]. CTCF binding at the mouse H19 locus is required for CHD8 recruitment, whereas CHD8 KD does not reduce occupancy [68]. For SMARCA4, MNase-seq analysis showed no significant changes in nucleosome positioning following auxin-induced degradation of SMARCA4 in human CD4+ T cells

doi:10.6342/NTU202504256

[69], indicating SMARCA4 may not alter the nucleosome pattern near CTCF binding sites and further influence CTCF occupancy. While these observations suggest that TERRA may influence the expression of certain CTCF cofactors, direct functional evidence linking CHD8 or SMARCA4 to CTCF occupancy remains lacking. Further mechanistics studies are needed to clarify their roles in TERRA mediated chromatin regulation.

Since CTCF is closely associated with the cohesin complex, and ChIP-seq data shows a more pronounced increase in CTCF coverage at CTCF-cohesin co-binding sites than that at CTCF-only sites (figure 9), I hypothesized that TERRA might influence CTCF chromatin occupancy by affecting cohesin-associated proteins. To investigate this, I focused on 11 well-characterized proteins involved in cohesin dynamics [70]: the core cohesin subunits SMC1A, SMC1B, SMC3, RAD21, STAG1, and STAG2; the cohesion loader NIPBL and its co-factor MAU2; the cohesin pause factor PDS5A and PDS5B; and WAPL, which facilitates cohesin release from chromatin. When comparing these proteins with the TERRA iDRIP-MS data, NIPBL and WAPL were identified as potential TERRA interactomes. In addition, RNA-seq analysis revealed Nipbl and Smc3 as upregulated genes, while Wapl was downregulated upon TERRA KD. Given that NIPBL promotes cohesin loading and WAPL mediates its release, these observations suggest that TERRA may modulate the chromatin association of cohesin. A previous study showed that CTCF

doi:10.6342/NTU202504256

KD reduced RAD21 occupancy, whereas RAD21 KD did not significantly affect CTCF occupancy [71]. However, it remains unclear whether enhanced cohesin loading or retention increase CTCF chromatin binding. Although additional experiments are required to confirm this hypothesis, this observation offers a possible explanation and potential research direction for the elevated CTCF occupancy observed in my ChIP-seq data following TERRA KD.

Based on the findings from both TERRA RNA transfection and TERRA KD system, we propose a potential model for how TERRA may function as an epigenetic regulator (Figure 11). When TERRA expression is elevated, particularly its G-rich repetitive sequences, it may recruit BLM helicase to the RNA, thereby limiting BLM availability on chromatin. This reduction in BLM chromatin occupancy would in turn hinder the unwinding of DNA G4 structures, leading to an increased level of DNA G4s in the genome. Conversely, upon TERRA depletion, we observed a global increase in CTCF occupancy, especially at sites co-occupied by cohesin complexes. While the mechanistic link between TERRA depletion and enhanced CTCF binding remains to be elucidated, our current findings suggest several possible directions for future investigation. First, G4 structures forming at the center of the CTCF binding sites may inhibit CTCF binding. Second, TERRA might directly interact with CTCF and thereby modulate its chromatin binding. Third, TERRA could regulate a CTCF cofactor that influence its occupancy.

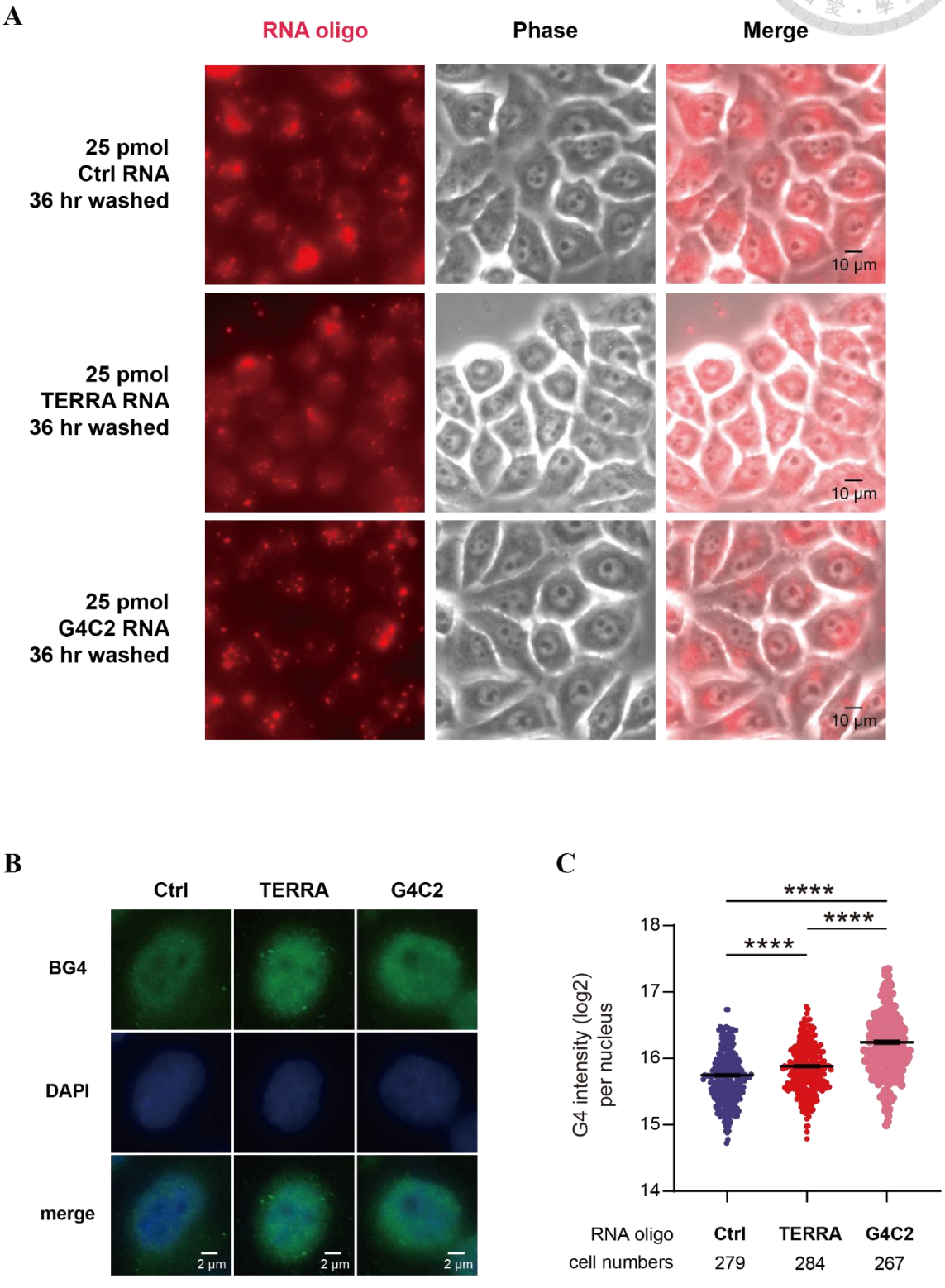
doi:10.6342/NTU202504256

Fourth, TERRA may affect cohesin-associated cofactors, thereby promoting cohesin loading and strengthening CTCF binding at co-bound sites. Otherwise, these observation raises the potential role of TERRA contributes to the modulation of chromatin architecture, potentially through regulating cohesin-associated CTCF binding and chromatin looping. Further investigation is necessary to elucidate whether this regulation is direct or mediated through additional cofactors.

doi:10.6342/NTU202504256

Chapter 5 Figures and Tables

Figure 1.



doi:10.6342/NTU202504256

Figure 1. Introducing G4 RNA results in increased DNA G4 structures.

(A) Live-cell microscopy images captured 36 hours after reverse transfection of 50 μM small RNA oligos (Ctrl RNA, TERRA RNA, and G4C2 RNA) into HeLa cells using Lipofectamine RNAiMAX (see Materials and Methods 2.2). All RNA oligos were labeled with Cy3 at the 5' end (see Supplementary Table S1); the Cy3 fluorescence signal indicates the localization of RNA oligos. Phase-contrast images show live-cell morphology and positioning. (B) Representative immunofluorescence microscopy images showing nuclear DNA G4 signals in HeLa cells, 36 hours post-transfection with the small RNA oligos (Ctrl RNA, TERRA RNA, and G4C2 RNA) (see Materials and Methods 2.4, 2.5). Prior to staining, cells were treated with RNase A to remove RNA. G4 structures were detected using the BG4 antibody. The FITC channel displays the G4 signal, and DAPI staining marks the nuclear regions. (C) Quantification of total nuclear G4 signal intensity from the immunofluorescence staining shown in (B); at least 200 cells were analyzed per group. **** $p < 0.0001$ by Mann-Whitney U test. Representative immunofluorescence images and quantitative analyses of the other two replicates are shown in Supplementary Figure S1.

doi:10.6342/NTU202504256

Figure 2.

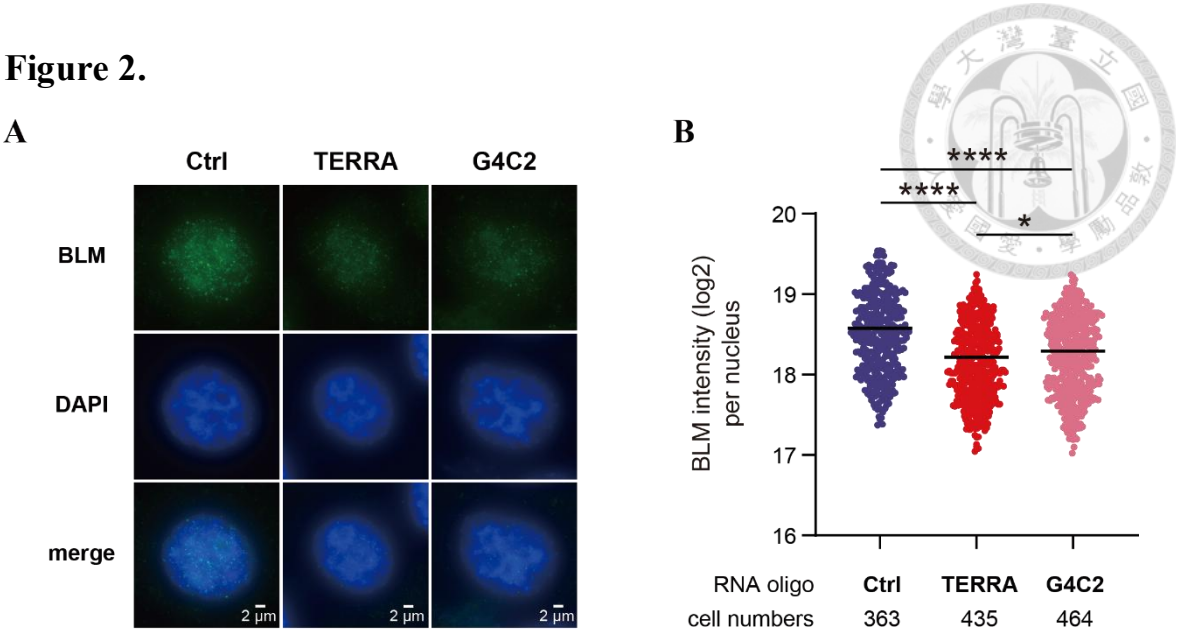
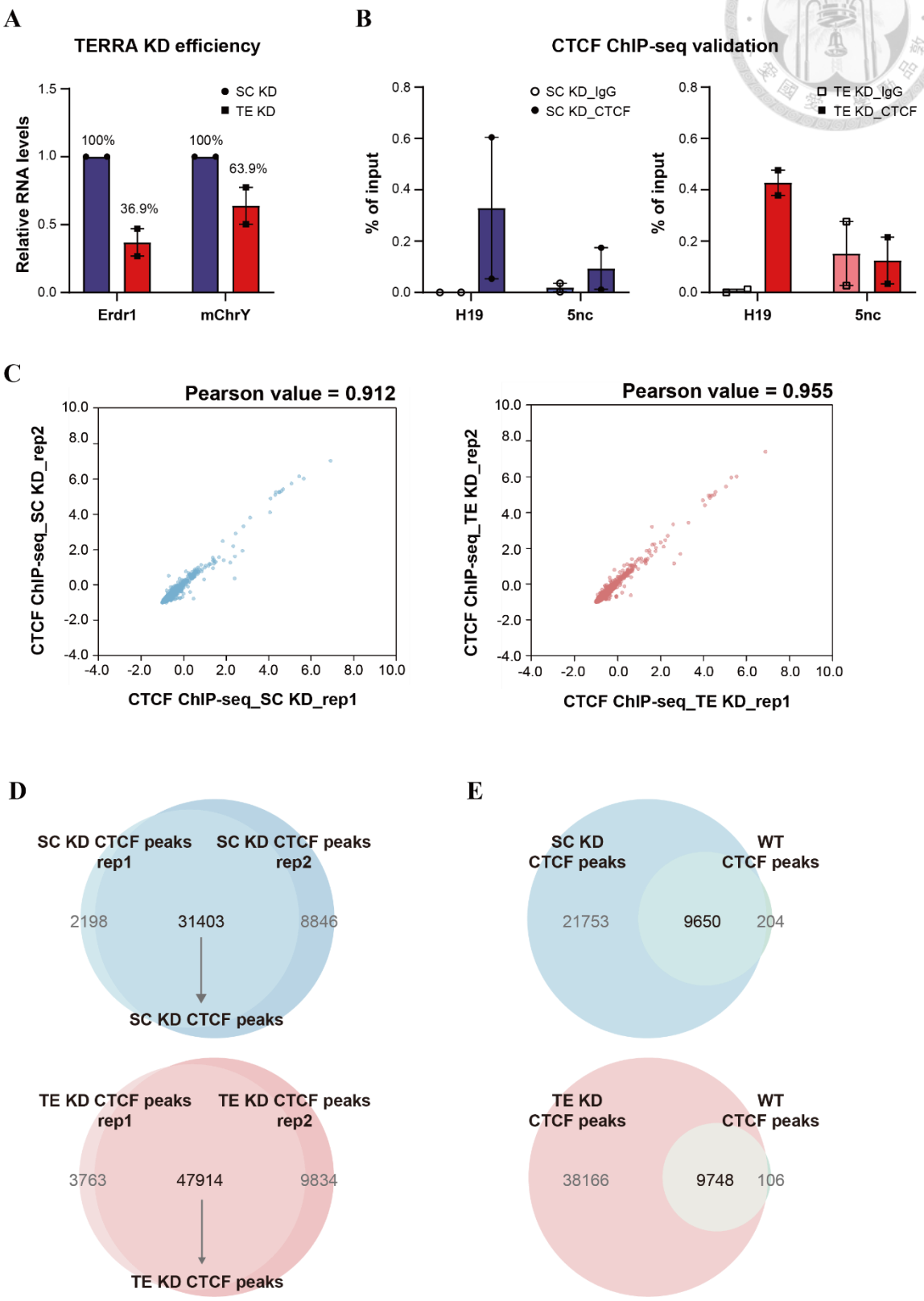


Figure 2. Introducing G4 RNA leads to increased BLM occupancy in the nucleus.

(A) Representative immunofluorescence microscopy images showing nuclear BLM signals in HeLa cells, 36 hours post-transfection with small RNA oligos (Ctrl RNA, TERRA RNA, and G4C2 RNA) (see Materials and Methods 2.4). The FITC channel displays the BLM signal, and DAPI staining marks the nuclear regions. (B) Quantification of total nuclear BLM signal intensity from the immunofluorescence staining shown in (A); at least 200 cells were analyzed per group. **** $p < 0.0001$ by Mann-Whitney U test. Representative immunofluorescence images and quantitative analyses of the other two replicates are shown in Supplementary Figure S2.

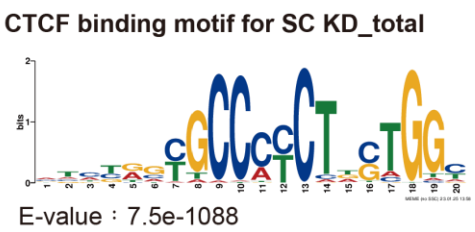
doi:10.6342/NTU202504256

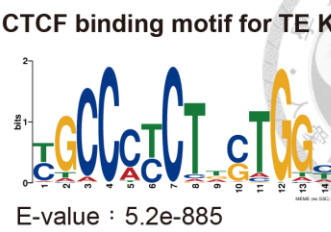
Figure 3.



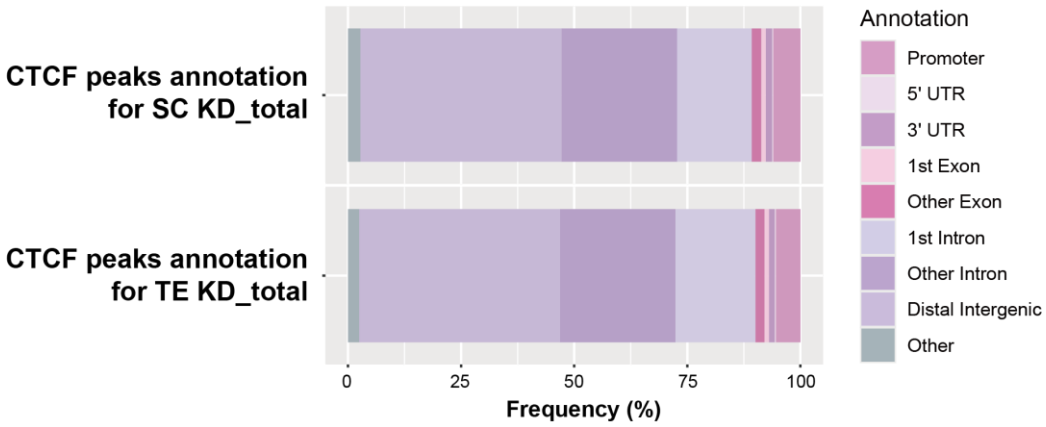
doi:10.6342/NTU202504256

F





G



	SC KD_total		TE KD_total	
Annotation	Reads_count	Frequency	Reads_count	Frequency
Promoter	1902	6.05%	2611	5.44%
5' UTR	62	0.20%	118	0.25%
3' UTR	447	1.42%	605	1.26%
1st Exon	313	1.00%	484	1.01%
Other Exon	658	2.09%	956	2.00%
1st Intron	5149	16.52%	8488	17.68%
Other Intron	8041	25.57%	12245	25.51%
Distal Intergenic	13972	44.43%	21305	44.39%
Other	855	2.72%	1188	2.48%

Figure 3. CTCF ChIP-seq analysis upon TERRA depletion.

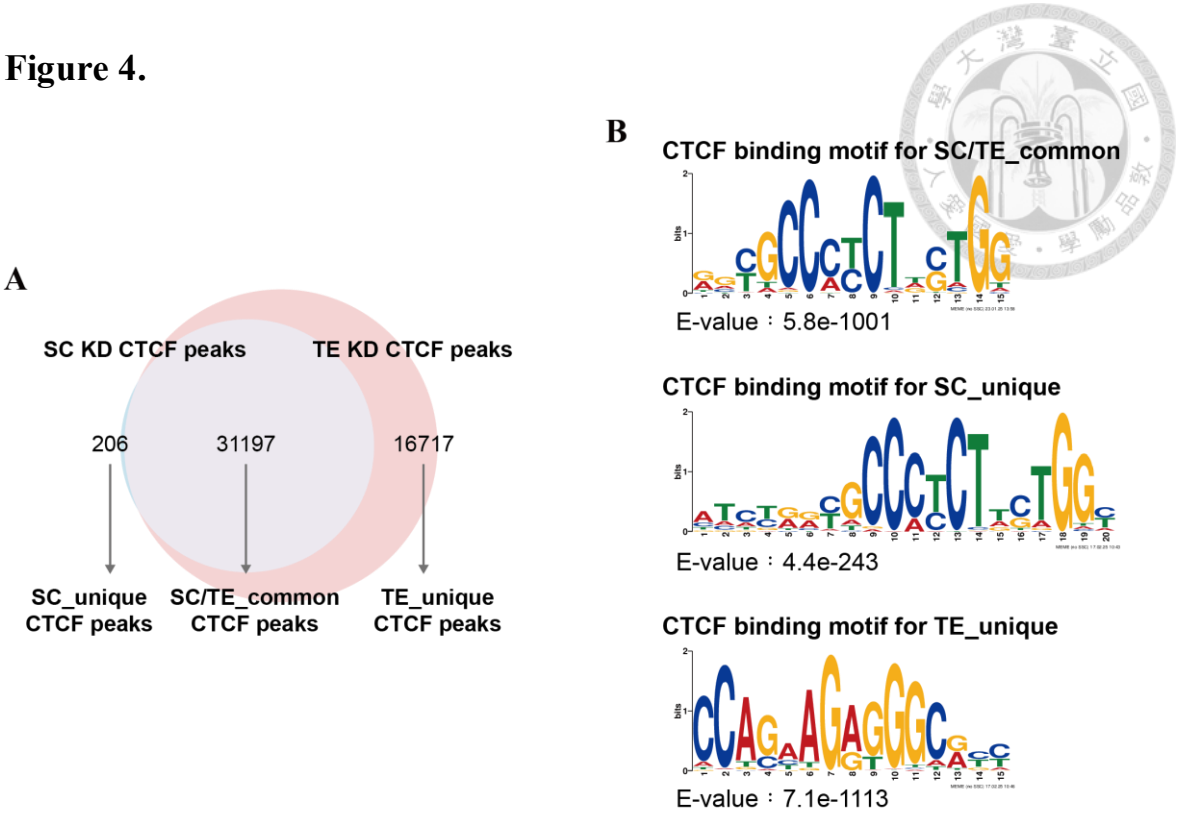
(A) RT-qPCR analysis result of TERRA expression in scramble (SC KD) and TERRA KD (TE KD) mESCs, using primers targeting *Erdr1* and the subtelomeric region on mouse chromosome Y. TERRA KD experiments were performed as described in Materials and Methods 2.7. Expression levels were normalized to *Gapdh*, with the SC KD group set to 1 for normalization. Data from two replicates corresponding to CTCF ChIP-seq samples are shown. (B) ChIP-qPCR validation of CTCF binding in SC KD and TE KD mESCs using primers targeting a known CTCF-positive binding site (*H19*) and a CTCF-negative control site (*5nc*). CTCF enrichment is presented as a percentage of input, calculated from a standard input curve. Two replicates were shown. (C) Scatter plots

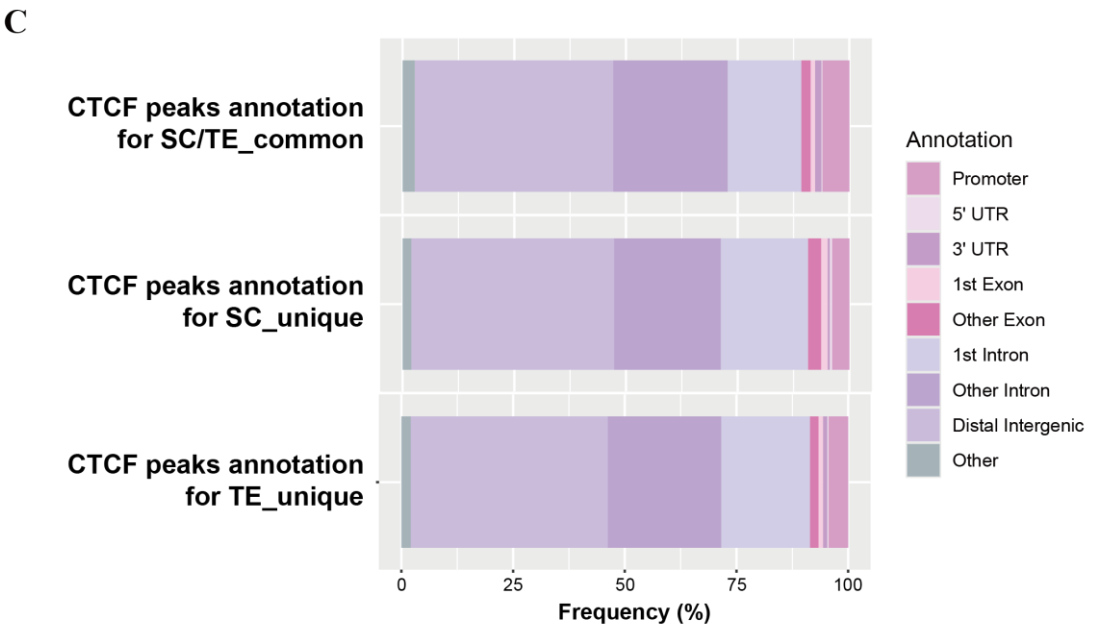
doi:10.6342/NTU202504256

showing the correlation of CTCF ChIP-seq signal between two replicates in SC KD and TE KD groups (see Material and Methods 2.13). Pearson's correlation coefficient (r) were 0.912 for SC KD and 0.955 for TE KD. (D) Venn diagrams showing the overlap of CTCF peaks between two replicates in SC KD and TE KD groups. Peaks shared by both replicates were defined as SC KD or TE KD CTCF peaks. A total of 31,403 peaks were identified in SC KD, and 47,914 peaks in TE KD. (E) Venn diagrams showing the overlap between SC KD or TE KD CTCF peaks and previously published wild-type (WT) CTCF peak datasets [57, 60, 61]. (F) DNA motifs enrichment analysis performed on SC KD and TE KD CTCF peaks (see Materials and Methods 2.13) defined in (D). (G) Genomic distribution of SC KD and TE KD CTCF peaks, which are defined in (D), annotated based on the mm10 reference genome (see Materials and Methods 2.13). The number and percentage of CTCF peaks in each genomic category are shown as histograms and summarized in the accompanying quantitative table.

doi:10.6342/NTU202504256

Figure 4.





	SCTE_common		SC_unique		TE_unique	
Annotation	Reads_count	Frequency	Reads_count	Frequency	Reads_count	Frequency
Promoter	1890	6.06%	8	3.90%	738	4.39%
5' UTR	61	0.20%	1	0.49%	50	0.30%
3' UTR	446	1.43%	1	0.49%	162	0.96%
1st Exon	308	0.99%	3	1.46%	171	1.02%
Other Exon	651	2.09%	6	2.93%	318	1.89%
1st Intron	5147	16.50%	40	19.51%	3343	19.88%
Other Intron	7983	25.59%	49	23.90%	4265	25.36%
Distal Intergenic	13859	44.43%	93	45.37%	7420	44.12%
Other	850	2.72%	4	1.95%	351	2.09%

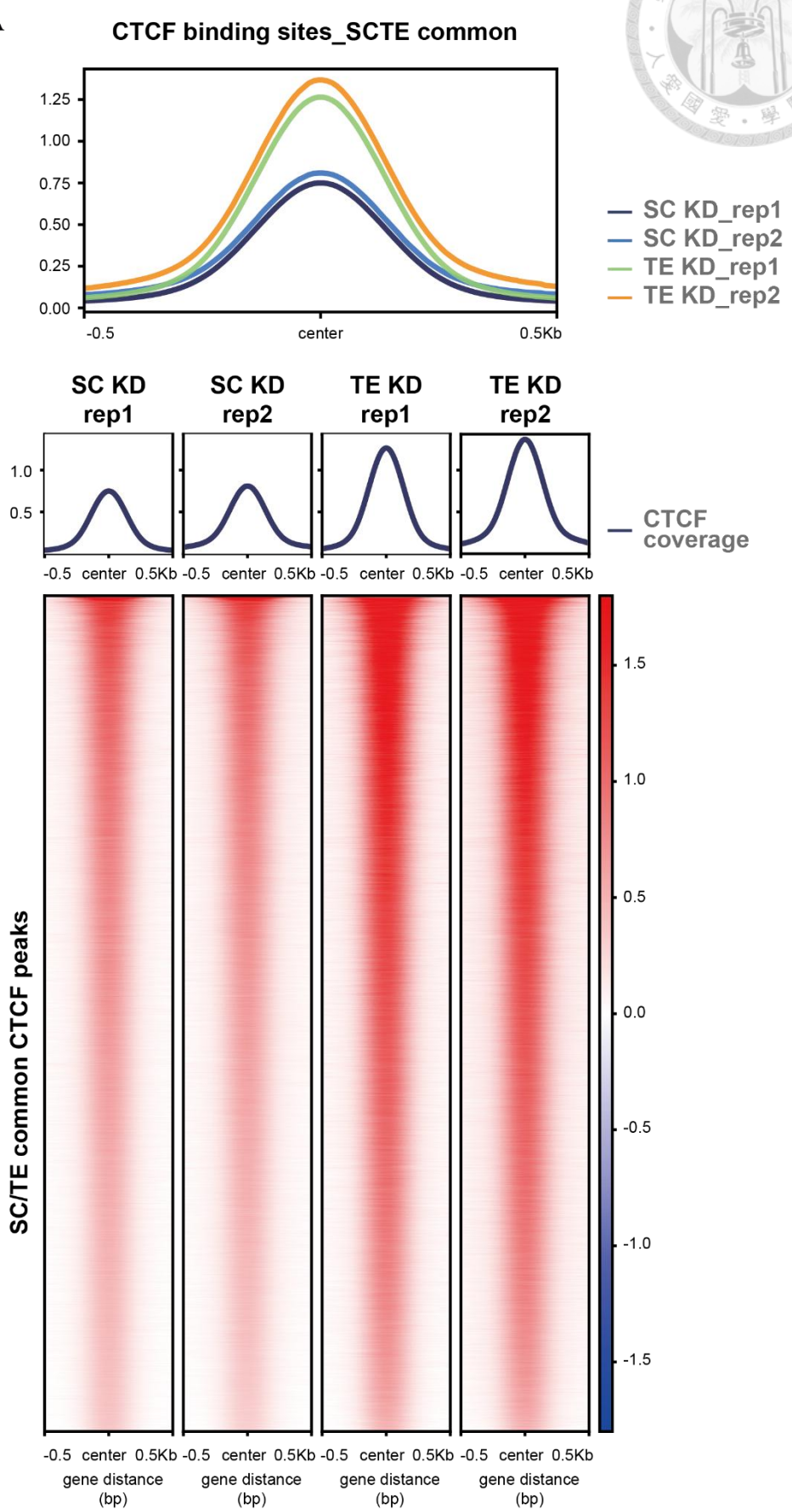
doi:10.6342/NTU202504256

Figure 4. Classification and analysis of common and unique CTCF peaks upon TERRA depletion.

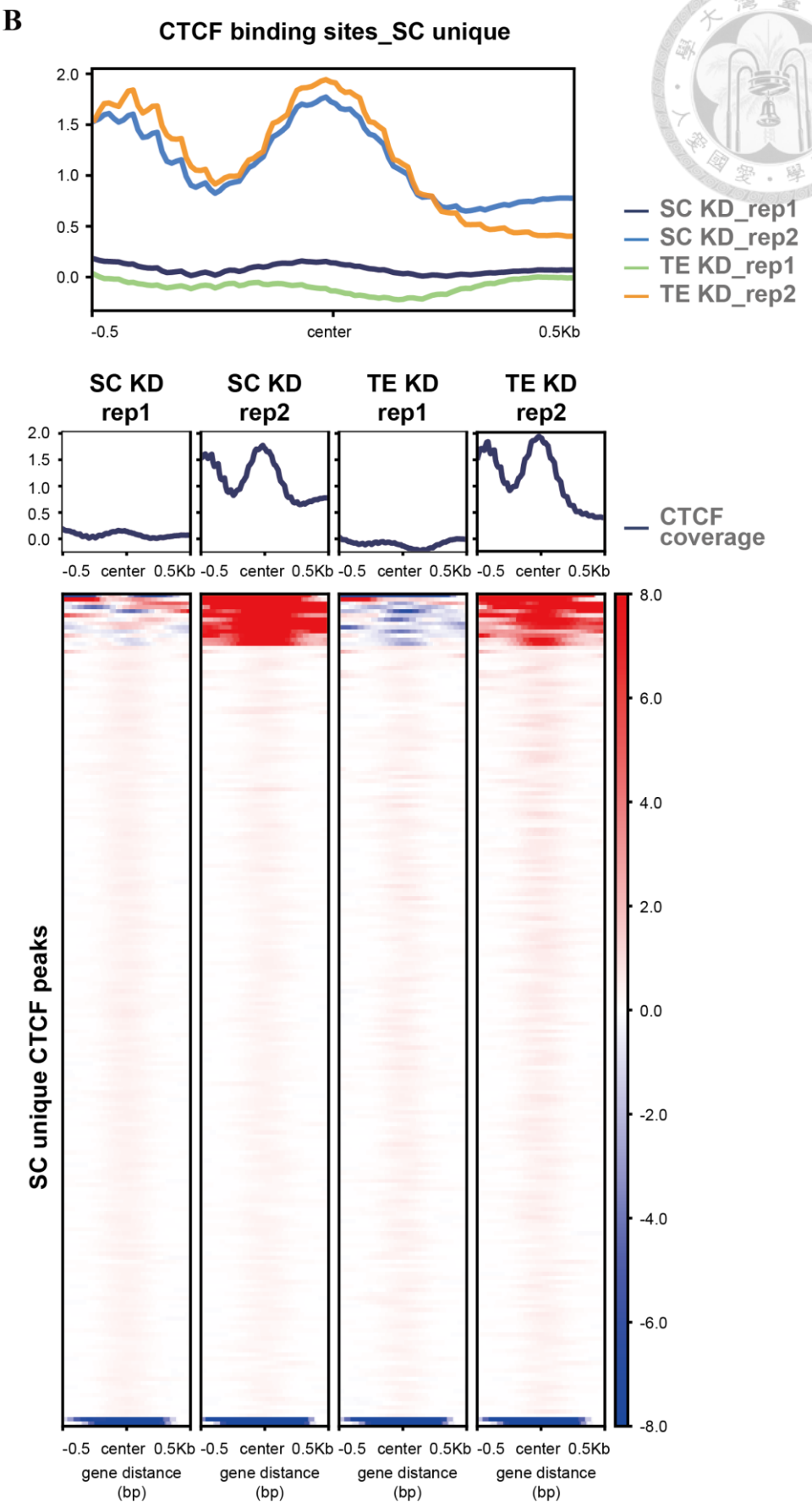
(A) Venn diagrams showing the overlap of CTCF peaks between scramble (SC KD) and TERRA KD (TE KD) mESCs. SC KD and TE KD CTCF peaks were identified as the intersecting peaks between two replicates shown in Figure 3D. Peaks present in both groups were defined as SC/TE common CTCF peaks, while non-overlapping peaks were classified as SC unique or TE unique CTCF peaks, respectively. A total of 31,197 peaks were shared between SC KD and TE KD groups (SC/TE common CTCF peaks), with 206 SC unique and 16,717 TE unique CTCF peaks identified. (B) DNA motifs enrichment analysis performed on SC/TE common, SC unique, and TE unique CTCF peaks (see Materials and Methods 2.13). Peaks groups were defined in Figure 4A. (C) Genomic distribution of SC/TE common, SC unique, and TE unique CTCF peaks (see Materials and Methods 2.13). The common and unique CTCF peaks are determined in Figure 4A. CTCF peaks were annotated using the mm10 reference genome. The number and percentage of CTCF peaks within each genomic category are shown in the histogram and summarized in the accompanying quantitative table.

doi:10.6342/NTU202504256

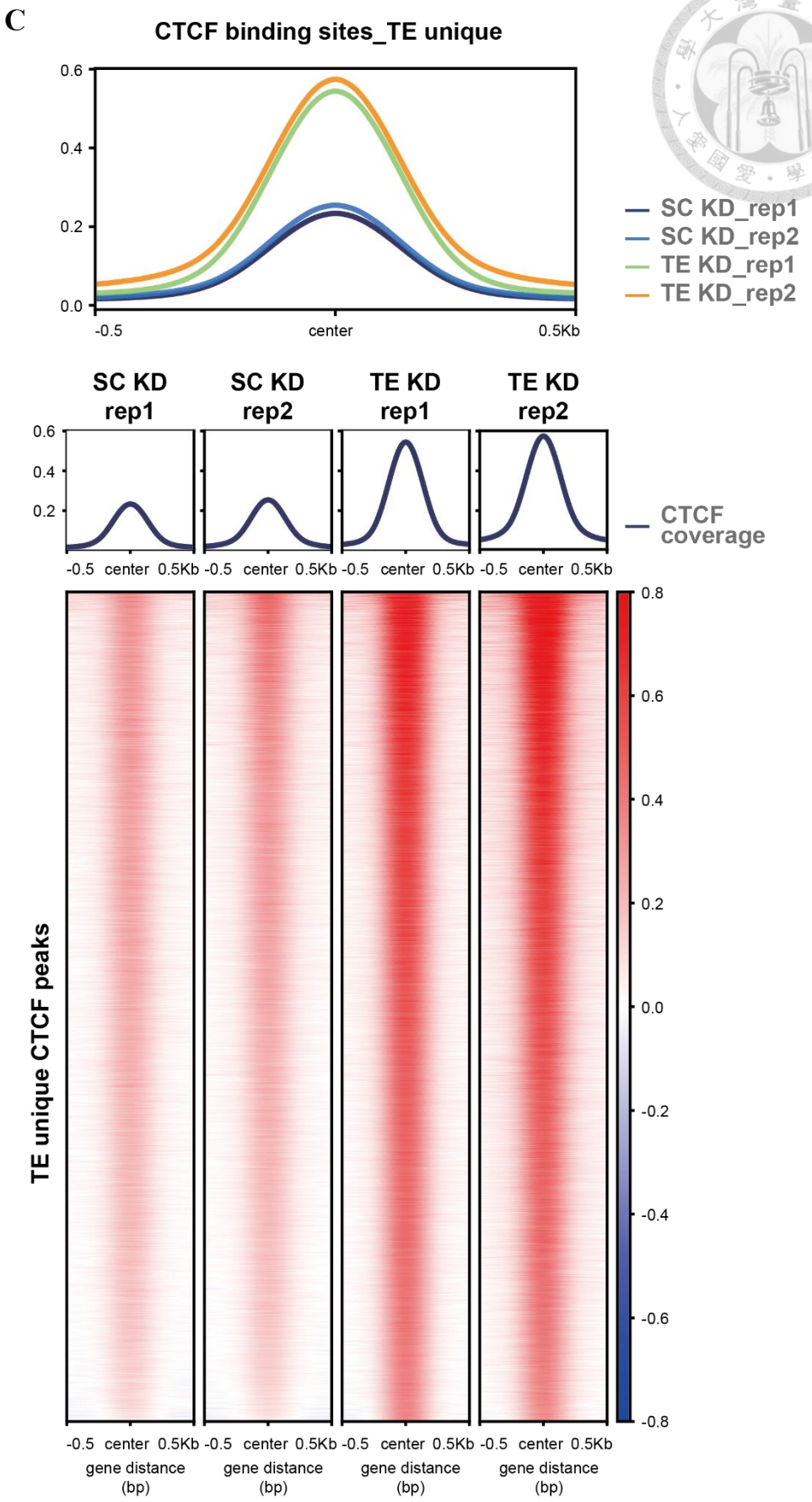
Figure 5.



doi:10.6342/NTU202504256



doi:10.6342/NTU202504256



doi:10.6342/NTU202504256

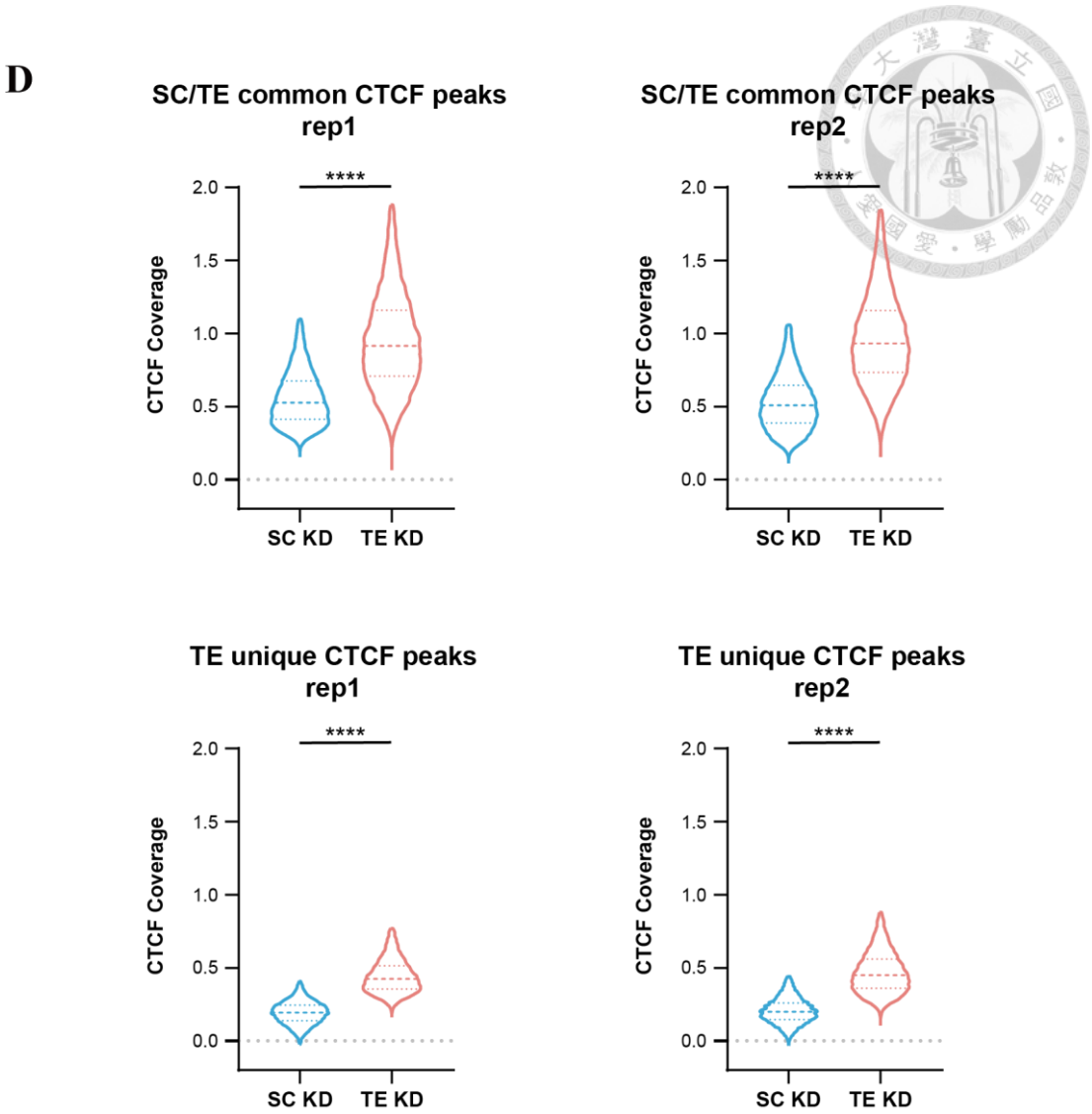
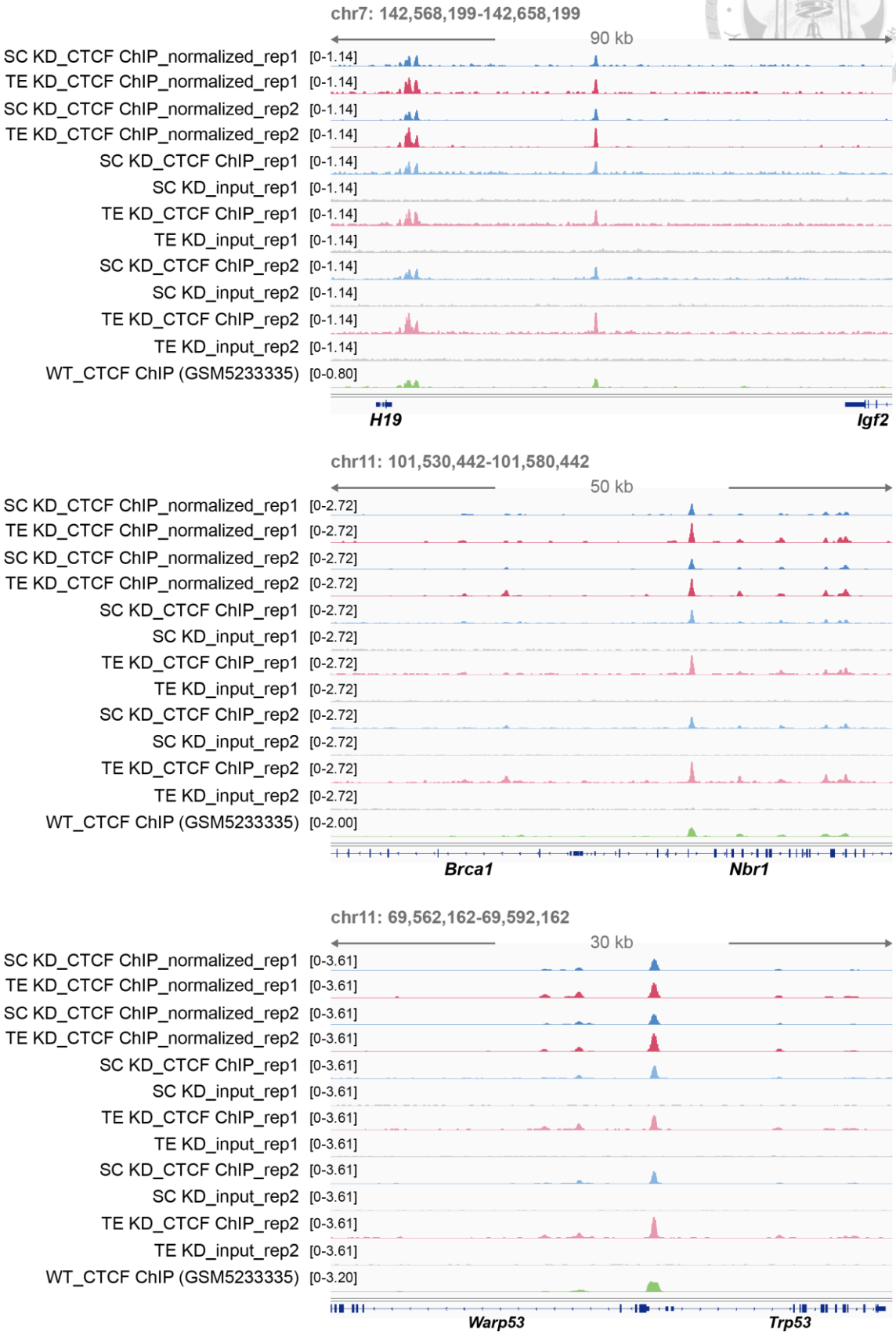


Figure 5. CTCF occupancy is increased in mostly common and unique binding sites upon TERRA depletion.

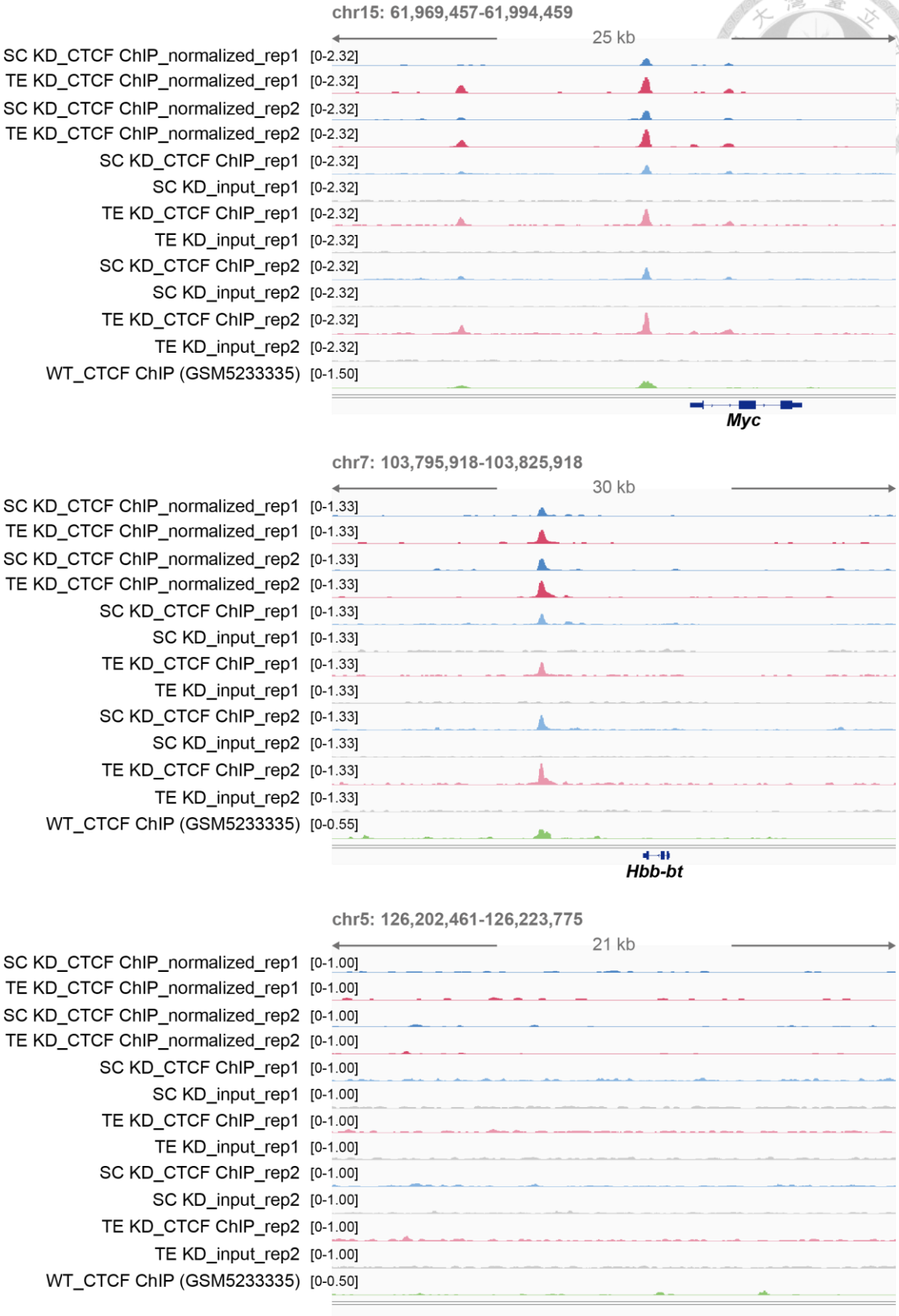
(A-C) Metaplots and heatmaps showing CTCF ChIP-seq signals in scramble (SC KD) and TERRA KD (TE KD) mESCs, centered around the summits of SC/TE common, SC unique, and TE unique CTCF peaks within ± 500 bp (see Materials and Methods 2.13). The classification of CTCF peaks is as defined in Figure 4A. Two replicates are shown. (D) Violin plots showing the average CTCF ChIP-seq coverage per CTCF peak in SC KD and TE KD groups, within SC/TE common and TE unique CTCF peaks (see Materials and Methods 2.13). Outliers were excluded using IQR-based filtering, with < 3% of peaks removed per group (see Materials and Methods 2.13). **** $p<0.0001$ by Mann-Whitney U test. Data from two replicates are shown separately.

doi:10.6342/NTU202504256

Figure 6.



doi:10.6342/NTU202504256



doi:10.6342/NTU202504256

Figure 6. Representative loci showing CTCF ChIP-seq signal changes upon TERRA depletion in mESCs.

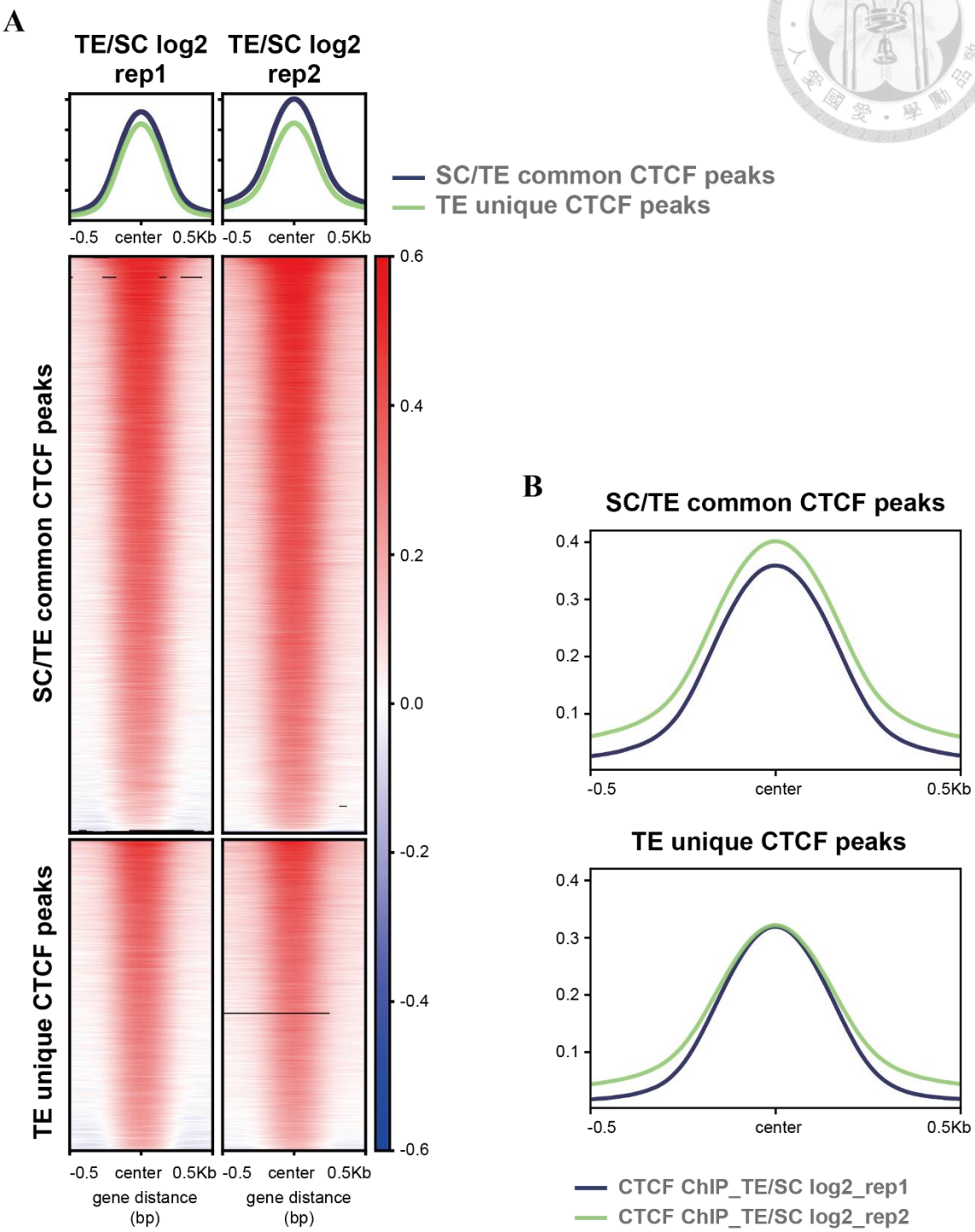
IGV snapshot visualizing CTCF ChIP-seq signals in scramble (SC KD) and TERRA KD (TE KD) groups across six selected genomic loci. Five CTCF-positive sites located near the *H19*, *Myc*, *Brca1*, *Hbb*, and *Trp53* genes are selected, along with one CTCF-negative control region on chromosome 5 corresponding to the 5nc primer used in CTCF ChIP validation. Tracks from top to bottom are:

1. Scramble KD mESCs input-normalized CTCF ChIP-seq signal, replicate 1;
2. TERRA KD mESCs input-normalized CTCF ChIP-seq signal, replicate 1;
3. Scramble KD mESCs input-normalized CTCF ChIP-seq signal, replicate 2;
4. TERRA KD mESCs input-normalized CTCF ChIP-seq signal, replicate 2;
5. Scramble KD mESCs raw CTCF ChIP-seq signal, replicate 1;
6. Scramble KD mESCs input signal, replicate 1;
7. TERRA KD mESCs raw CTCF ChIP-seq signal, replicate 1;
8. TERRA KD mESCs input signal, replicate 1;
9. Scramble KD mESCs raw CTCF ChIP-seq signal, replicate 2;
10. Scramble KD mESCs input signal, replicate 2;
11. TERRA KD mESCs raw CTCF ChIP-seq signal, replicate 2;
12. TERRA KD mESCs input signal, replicate 2;
13. Wild-type CTCF mESCs ChIP-seq input-normalized CTCF ChIP-seq signal [57].

Tracks 1-12 are scaled uniformly within groups, while track 13 is scaled individually.

doi:10.6342/NTU202504256

Figure 7.



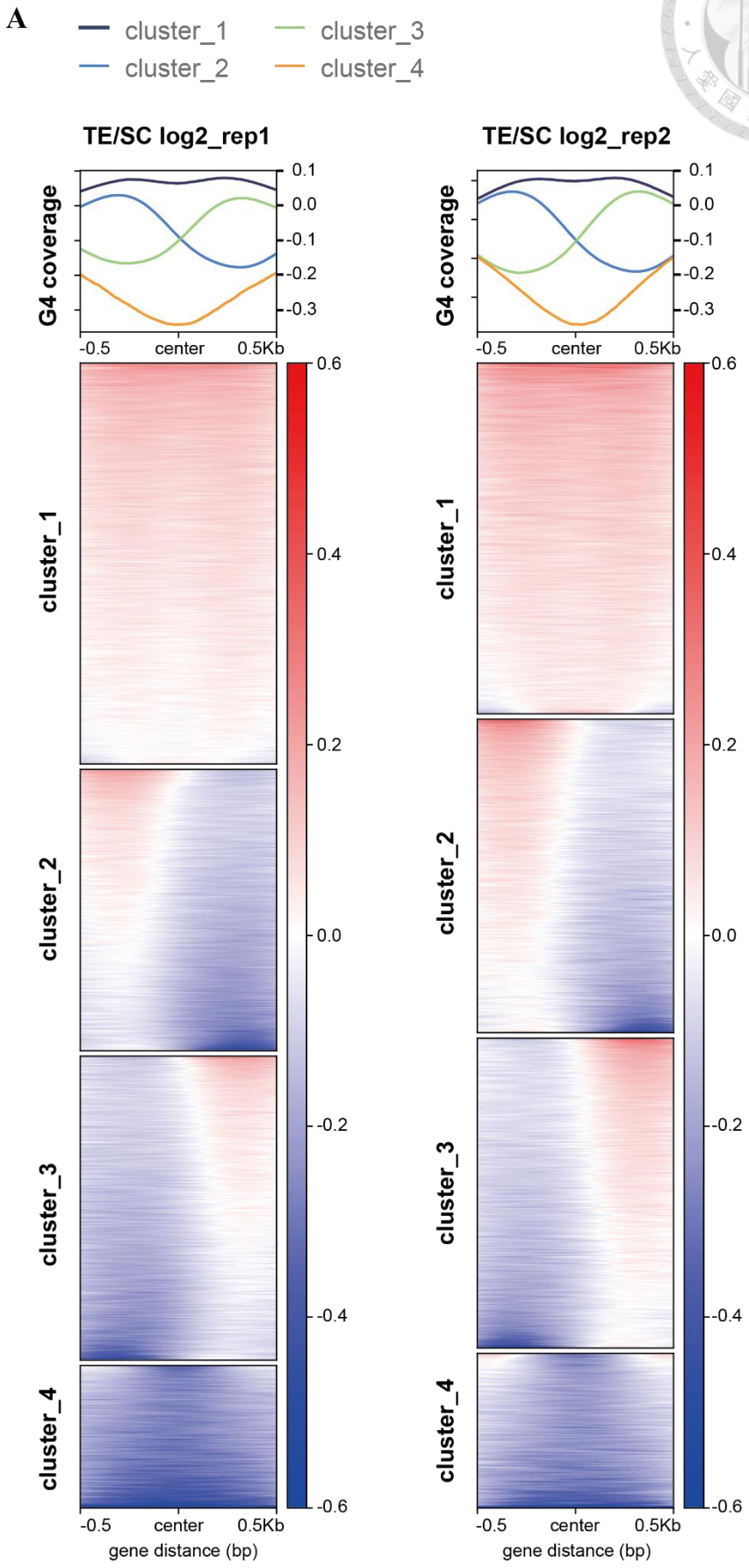
doi:10.6342/NTU202504256

Figure 7. CTCF occupancy at most common and unique CTCF peaks upon TERRA depletion.

(A-B) Metaplots and heatmaps showing the log2 fold change in CTCF ChIP-seq signal between TERRA and scramble KD groups, centered at the summits of SC/TE common and TE unique CTCF peaks within ±500bp (see Materials and Methods 2.13). The classification of SC/TE common and TE unique CTCF peaks was defined in Figure 4A. Two replicates are shown.

doi:10.6342/NTU202504256

Figure 8.



doi:10.6342/NTU202504256

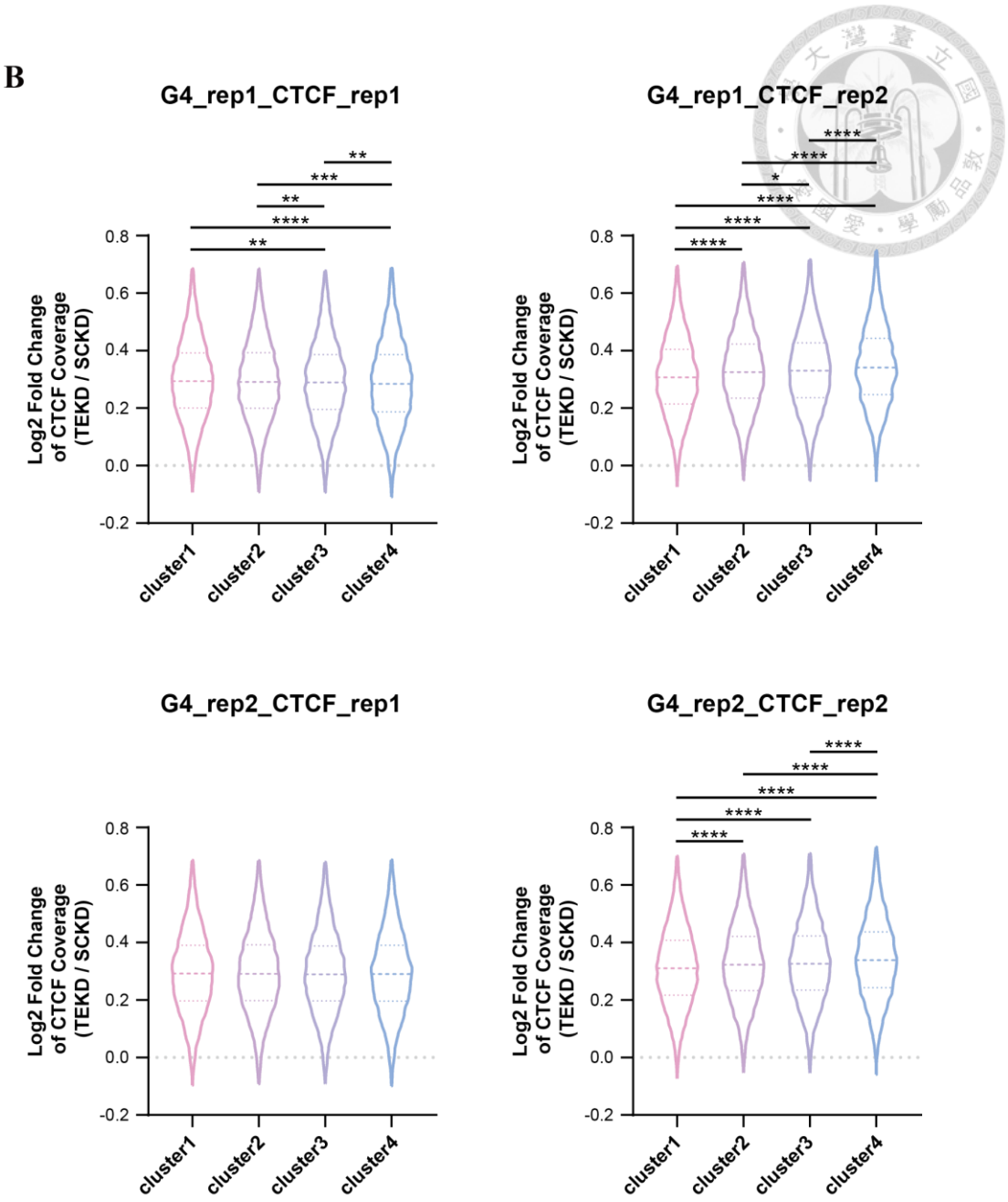


Figure 8. G4 reduction correlates with the greatest increase in CTCF binding.

(A) Metaplots and heatmaps showing the log2 fold change in G4 ChIP-seq signal between TERRA (TE KD) and scramble KD (SC KD) groups, centered at the summits of SC/TE common and TE unique CTCF peaks within ± 500 bp. G4 ChIP-seq data using TERRA KD mESCs were obtained from a previous published paper [21]. CTCF peaks were categorized into four clusters based on their log2 fold change of G4 signal using k-means clustering (see Materials and Methods 2.14). Two replicates of G4 ChIP-seq in scramble

doi:10.6342/NTU202504256

and TERRA KD mESCs are shown. (B) Violin plots showing the average log2 fold change in CTCF ChIP-seq signal between TERRA (TE KD) and scramble KD (SC KD) groups for each CTCF peak cluster defined in (A). Outliers were excluded based on IQR-based filtering, with < 5% of peaks removed per group. Statistical significance was assessed by pairwise Wilcoxon rank-sum test: **** $p \leqq 0.0001$; *** $0.0001 < p \leqq 0.001$; ** $0.001 < p \leqq 0.01$; * $0.01 < p \leqq 0.05$; n.s. $p > 0.05$ (n.s. not labeled). Each of the two replicate of both CTCF and G4 ChIP-seq in scramble and TERRA KD mESCs were analyzed and plotted individually.

doi:10.6342/NTU202504256

Figure 9.

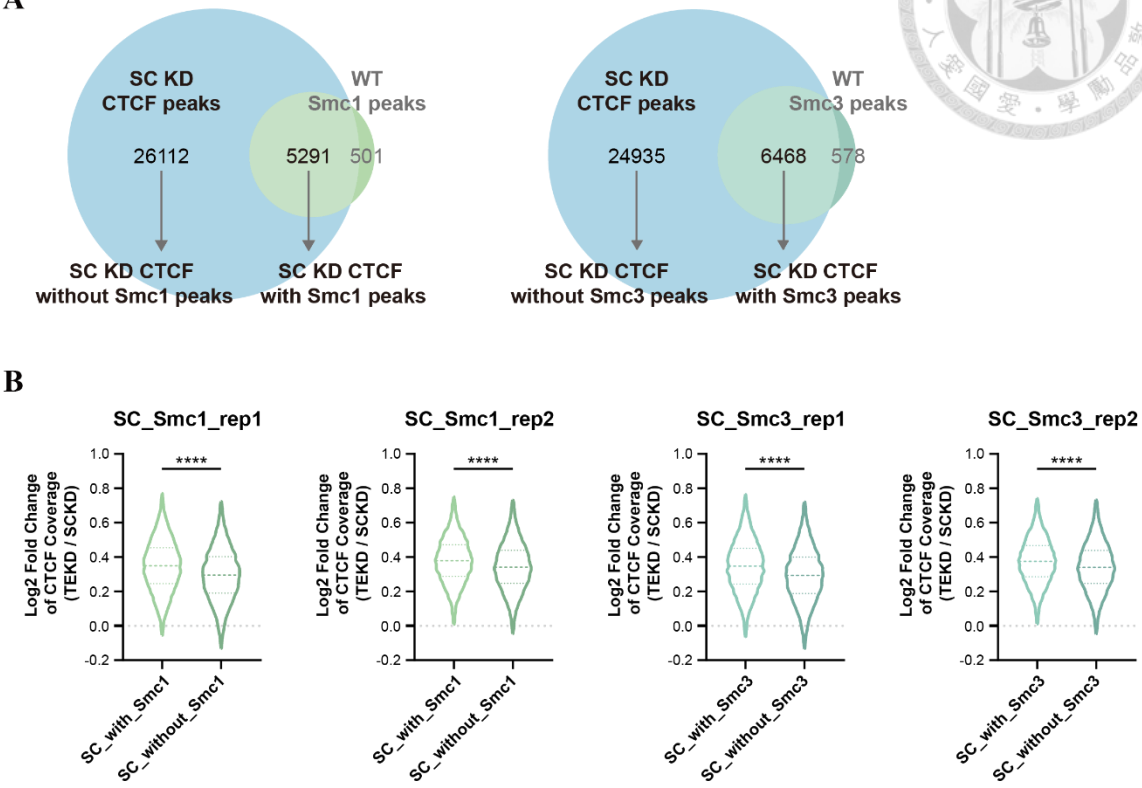


Figure 9. CTCF binding is more profound at cohesin-bound sites after TERRA depletion.

(A) Venn diagrams showing the overlap between CTCF peaks in scramble KD (SC KD) mESCs and Smc1 or Smc3 peaks identified in wild-type (WT) mESCs. The SC KD CTCF peaks (as defined in Figure 3D) were intersected with Smc1 or Smc3 peaks to classify them into two groups: SC KD CTCF peaks with or without overlapping cohesin binding. A total of 5,291 and 6,468 SC KD CTCF peaks overlapped with Smc1 and Smc3, respectively, while 26,112 and 24,935 peaks did not. (B) Violin plots showing the average log2 fold change in CTCF ChIP-seq signal between TERRA (TE KD) and scramble KD (SC KD) groups for SC KD CTCF peaks with or without overlapping Smc1 or Smc3 (see Materials and Methods 2.15). Outliers were excluded based on IQR-based filtering, with < 1.5% of peaks removed per group (see Materials and Methods 2.15). **** $p < 0.0001$ by Mann-Whitney U test. Two replicates are analyzed and plotted individually.

doi:10.6342/NTU202504256

Figure 10.

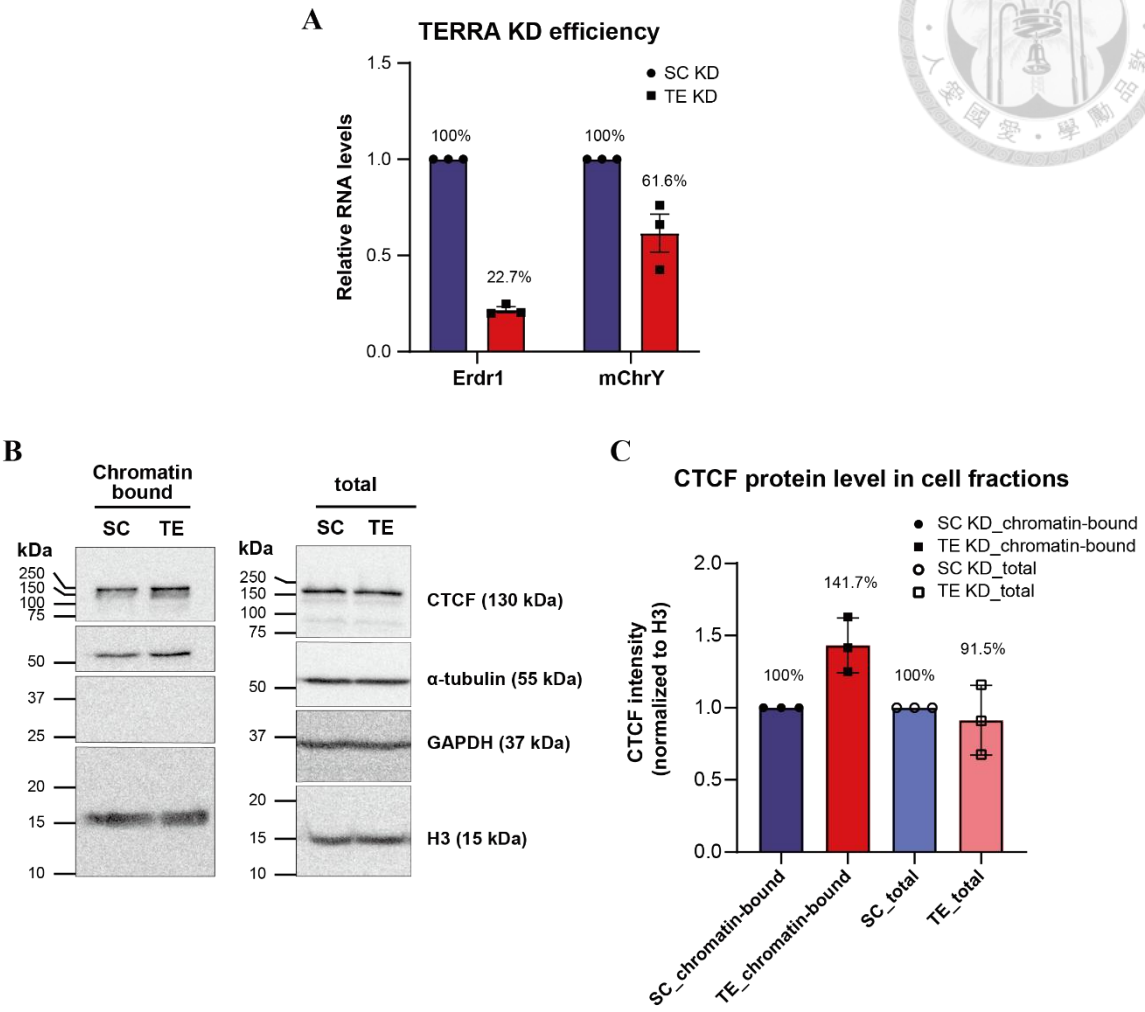
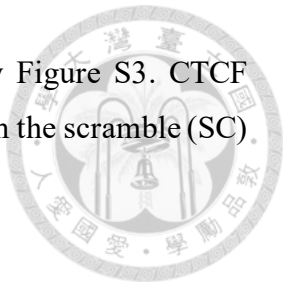


Figure 10. Chromatin-bound CTCF is increased upon TERRA depletion.

(A) RT-qPCR analysis result of TERRA expression in scramble (SC KD) and TERRA KD (TE KD) mESCs, using primers targeting *Erdr1* and the subtelomeric region on mouse chromosome Y. TERRA KD experiments were performed as described in Materials and Methods 2.7. Expression levels were normalized to *Gapdh* within each group, with the SC KD group set to 1 for normalization. Data from three replicates corresponding to cell fractionation samples are shown. (B) Representative western blot analysis of chromatin-bound and whole-cell lysate fractions from scramble (SC) and TERRA (TE) KD mESCs, showing the protein level of CTCF, α-tubulin, GAPDH, and histone H3 (see Materials and Methods 2.16, 2.17). Expected molecular weights of each protein are as follows: CTCF, 130 kDa; α-tubulin, 55 kDa; GAPDH, 37 kDa; and H3, 15 kDa. The results of western blotting from three replicates are shown in Supplementary Figure S3. (C) Quantification of CTCF band intensity in the chromatin-bound and whole-

doi:10.6342/NTU202504256

cell lysate fractions from three replicates shown in Supplementary Figure S3. CTCF levels were normalized to their corresponding H3 signals. Values from the scramble (SC) KD group were set to 1.



doi:10.6342/NTU202504256

Figure 11.

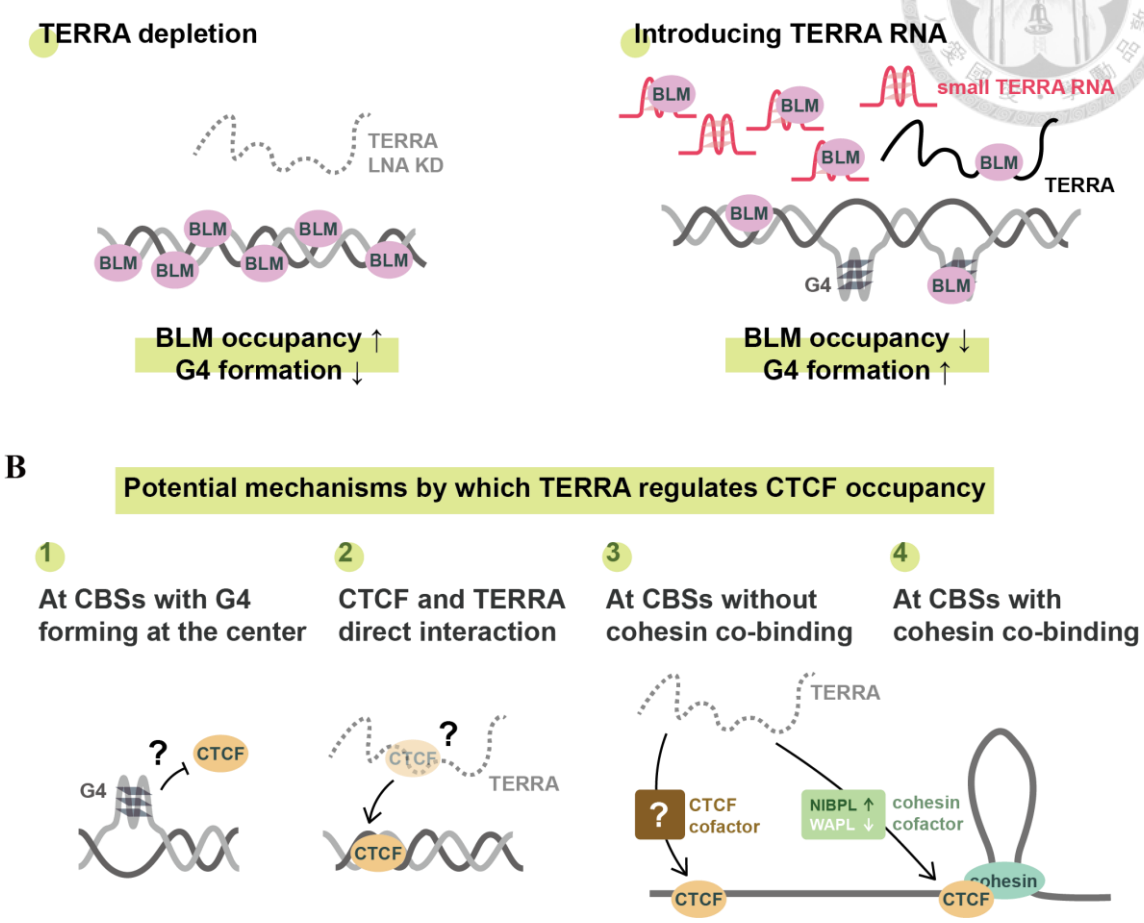


Figure 11. Proposed model for TERRA regulating DNA G4 formation, BLM occupancy, and CTCF binding across the genome.

(A) The potential model illustrating how TERRA may regulate BLM occupancy and DNA G4 formation at chromatin. Reduced TERRA expression may increase BLM occupancy, thereby promoting the reduction of DNA G4 structures. Otherwise, elevated TERRA expression levels may sequester BLM away from chromatin, leading to reduced BLM occupancy and accumulation of DNA G4 structures. (B) Proposed mechanisms by which TERRA may influence CTCF occupancy. Potential mode of regulation include: (i) G4-mediated inhibition of CTCF binding; (ii) direct interaction between TERRA and CTCF; (iii) modulation of CTCF cofactors; (iv) regulation of cohesin-associated factors.

doi:10.6342/NTU202504256

Table 1.

Source	K. Hirashima and H. Seimiya, 2015 [58]	I-Tien, Chiang's master thesis [20]	This study
Telomeric RNA sequence	r(UUAGGGUUAGGGUUAGGGUUAGGG)	r(UUAGGGUUAGGGUUAGGGUUAGGG)	Cy3-r(UUAGGGUUAGGGUUAGGGUUAGGG)
Ctrl RNA sequence	r(UUACCCUUACCCUUACCCUUACCC)	r(UUACCCUUACCCUUACCCUUACCC)	Cy3-r(AACGUACGCGGAAUACUUCGAGUU)
Other G-rich RNA sequence	AS1411 d(GGTGGTGGTGGTTGTGGTGGTGGTGG)		Cy3-r(GGGGCCGGGGCCGGGGCCGGGGCC)
Transfection reagent	Lipofectamine (RNA iMAX)	Nucleofector solution	Lipofectamine (RNA iMAX)
Transfection method	Reverse transfection	Nucleofection	Reverse transfection
RNA dosage	100 pmol for 8×10^4 cells	400 pmol for 3×10^6 cells	25 pmol for 4×10^4 cells
G4 folding before transfection	Yes	No	No
Transfection duration	72 hours	12 hours	36 hours
Cell line	PC-3, HBC4, MKN74	mESC	HeLa
Cell culture	3D culture	2D culture	2D culture
Validation	Circular dichroism spectra (RNA oligo G4 forming) Live-cell microscopy (transfection efficiency)	Northern blotting (transfection efficiency)	Live-cell microscopy (transfection efficiency)

doi:10.6342/NTU202504256

Table 1. Comparison of RNA transfection strategies using small RNA oligos containing telomeric repeat sequences.

Summary table comparing RNA transfection approaches utilizing small RNA oligos with telomeric repeat sequences from three sources: a previouly published study [58], a prior system established in our laboratory (I-Tien, Chiang's master thesis [20]), and the approach employed this study.

doi:10.6342/NTU202504256

Table 2.

Experimental group	Total reads number	UUAGGG×4 containing reads number	Percentage of UUAGGG×4 containing reads in total reads
CLIP-seq UV_rep1	7,129,439	734	0.01030%
CLIP-seq UV_rep2	11,708,571	564	0.00533%
CLIP-seq noUV_rep1	713,310	38	0.00482%
CLIP-seq noUV_rep2	1,121,858	241	0.02148%
RNA-seq SC_rep1	63,354,831	4,404	0.00695%
RNA-seq SC_rep2	28,273,992	2,965	0.01049%

Table 2. Analysis of CTCF binding to telomeric repeat-containing RNAs.

Quantification of telomeric repeat-containing reads from CTCF CLIP-seq data in wild-type (WT) mESCs [57] (see Materials and Methods 2.19). CLIP-seq without UV crosslinking and RNA-seq data were included as negative controls. RNA-seq data from scramble (SC) KD mESCs were obtained from a previous study [18]. For each dataset, the percentage of reads containing telomeric repeat sequences was calculated.

doi:10.6342/NTU202504256

Table 3.

A

Gene ID	baseMean	Log2Fold Change	lfcSE	stat	pvalue	padj
Brd2	14066.81	-0.43601	0.216838	-2.01077	0.04435	0.58946
Chd8	13918.8	-0.40598	0.203411	-1.9959	0.045945	0.602363
Ddx5	14903.62	-0.17143	0.203233	-0.84352	0.398939	0.999763
Ep300	10427.24	-0.15676	0.242206	-0.64723	0.517485	0.999763
H2ac20	1109.083	0.276762	0.318888	0.869119	0.384782	0.999763
H2az1	1034.134	-0.18148	0.324131	-0.56065	0.575038	0.999763
Hdlbp	20185.91	-0.01905	0.200429	-0.09503	0.924289	0.999763
Kdm5b	30251.71	-0.03396	0.245071	-0.13857	0.889792	0.999763
Lmna	3175.618	-0.11829	0.311491	-0.37967	0.704189	0.999763
Maz	8650.057	0.090943	0.206518	0.440363	0.659675	0.999763
Ncl	70827.7	-0.00621	0.288523	-0.02152	0.982829	0.999763
Npm1	16571.1	0.209816	0.290754	0.721712	0.470472	0.999763
Parp1	24452.87	-0.12801	0.226086	-0.5662	0.571261	0.999763
Polr2a	18012.5	-0.12731	0.20891	-0.60938	0.54227	0.999763
Pou5f1	18013.97	0.040434	0.263996	0.153167	0.878267	0.999763
Rad21	11815.48	-0.10714	0.212154	-0.505	0.613559	0.999763
Rad51	2047.827	-0.15729	0.215829	-0.72882	0.466111	0.999763
Rfx5	569.0624	0.026274	0.253543	0.103621	0.91747	0.999763
Rxra	1214.496	-0.01067	0.252477	-0.04226	0.966288	0.999763
Sin3a	9486.545	-0.4149	0.229529	-1.80769	0.070654	0.740335
Smad3	1673.22	-0.17781	0.266814	-0.66623	0.505267	0.999763
Smarca4	10697.41	-0.74445	0.211505	-3.51995	0.000432	0.025894
Smc1a	7034.42	0.153546	0.2724	0.563732	0.572937	0.999763
Smc3	6298.883	-0.1364	0.294805	-0.46274	0.643552	0.999763
Sp100	0					
Stag2	5027.202	0.032473	0.21683	0.149768	0.880947	0.999763
Suz12	9564.086	0.122259	0.267432	0.45719	0.647534	0.999763
Taf3	1436.778	0.091402	0.251114	0.364087	0.715793	0.999763

doi:10.6342/NTU202504256

Top2b	3882.814	0.107301	0.246981	0.434499	0.663926	0.999763
Ybx1	17565.1	-0.01669	0.262675	-0.06352	0.949349	0.999763
Yy1	4157.58	0.002719	0.233084	0.011668	0.99069	0.999763
Zbtb33	951.3092	-0.01115	0.254131	-0.04385	0.965027	0.999763
Znf143						

B

Gene ID	baseMean	Log2Fold Change	lfcSE	stat	pvalue	padj
Mau2	4707.843	-0.01874	0.20526	-0.0913	0.927258	0.999763
Nipbl	3463.105	0.705679	0.225256	3.133061	0.00173	0.074708
Pds5a	13783.11	-0.1173	0.230513	-0.50887	0.610845	0.999763
Pds5b	5080.107	-0.4777	0.203667	-0.23453	0.814575	0.999763
Rad21	11815.48	-0.10714	0.212154	-0.505	0.613559	0.999763
Smc1a	7034.42	0.153546	0.2724	0.563732	0.572937	0.999763
Smc1b	500.408	0.629145	0.297078	2.122125	0.033827	0.511277
Smc3	6298.883	-0.1364	0.294805	-0.46274	0.643552	0.999763
Stag1	5734.411	-0.20681	0.202781	-1.01987	0.307789	0.999763
Stag2	5027.202	0.032473	0.21683	0.149768	0.880947	0.999763
Wapl	9728.992	-0.46251	0.222062	-2.0829	0.03726	0.53924

Table 3. Lists of TERRA KD DEGs associated with CTCF and the cohesin complex.

(A) CTCF-associated cofactor genes selected from the TERRA KD differentially expressed gene (DEG) table. Candidate CTCF cofactors were selected based on overlap between four sources: a review article summarizing known CTCF interactors [63], the STRING database [64], the Integrated Interactions Database [65], and the systematic analysis from a previous study [66]. Proteins appearing in at least two of these four sources were defined as potential CTCF -associated cofactors, resulting in a total of 33 proteins for downstream analysis. DEGs were derived from RNA-seq data comparing scramble and TERRA KD mESCs obtained from a previous study [18] (see Materials and

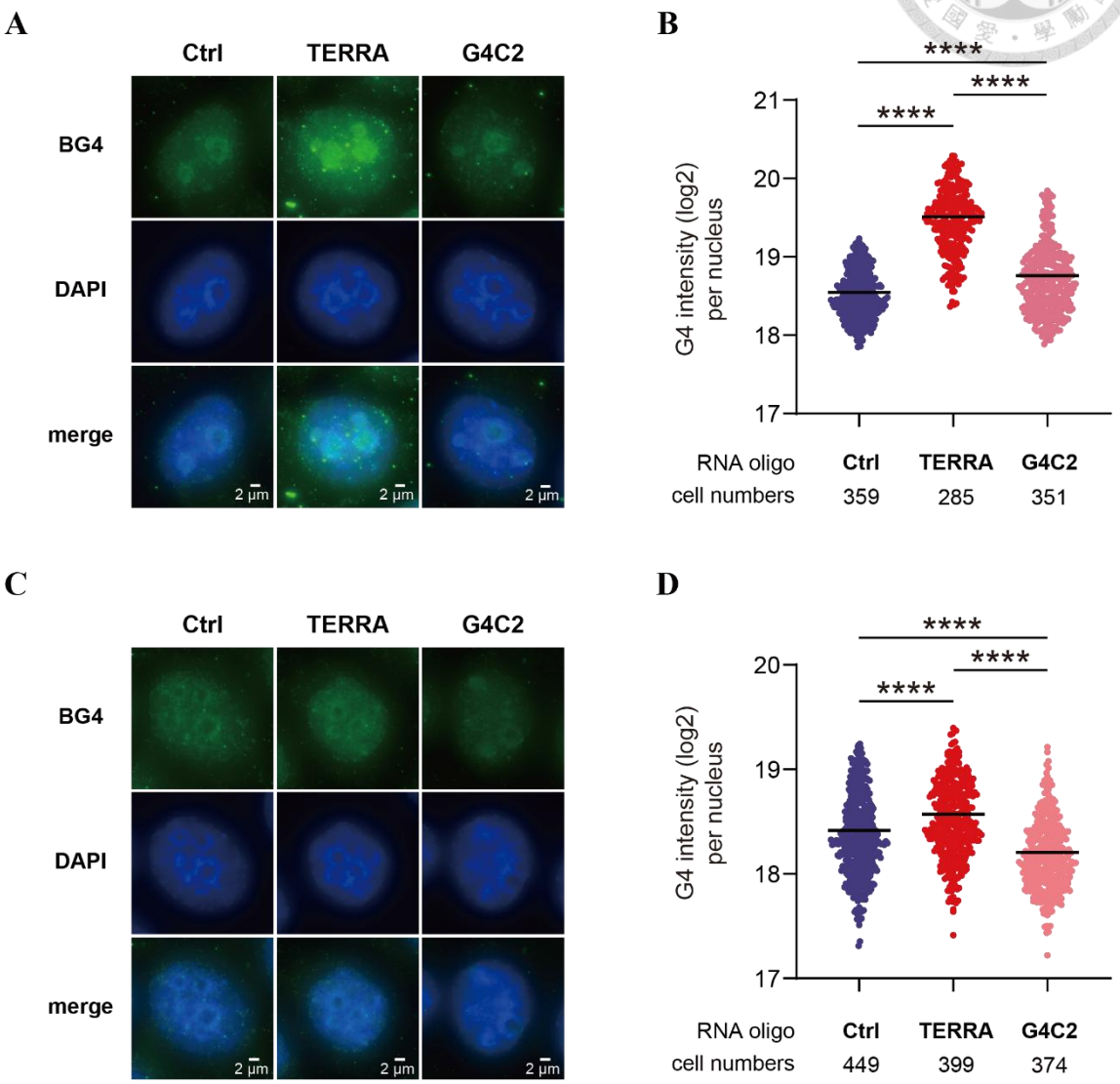
doi:10.6342/NTU202504256

Methods 2.20). Znf143 was not present in the DEG dataset. Genes with $p < 0.05$ and log2 fold change > 0 were defined upregulated, while those with $p < 0.05$ and log2 fold change < 0 were defined as downregulated. Upregulated and downregulated genes are highlighted in red and blue, respectively. (B) Cohesin-associated cofactor genes selected from TERRA KD DEG dataset. 11 cohesin-related proteins were selected from a previous review article [70] based on their known roles as cohesin complex components or regulators of cohesin occupancy. The same DEG dataset from (A) was used. Criteria and color labeling for upregulated and downregulated genes are identical to those described in (A).

doi:10.6342/NTU202504256

Chapter 6 Supplementary Figures and Tables

Supplementary figure S1.



Supplementary figure S1. The other two replicates for DNA G4 immunostaining related to Figure 1.

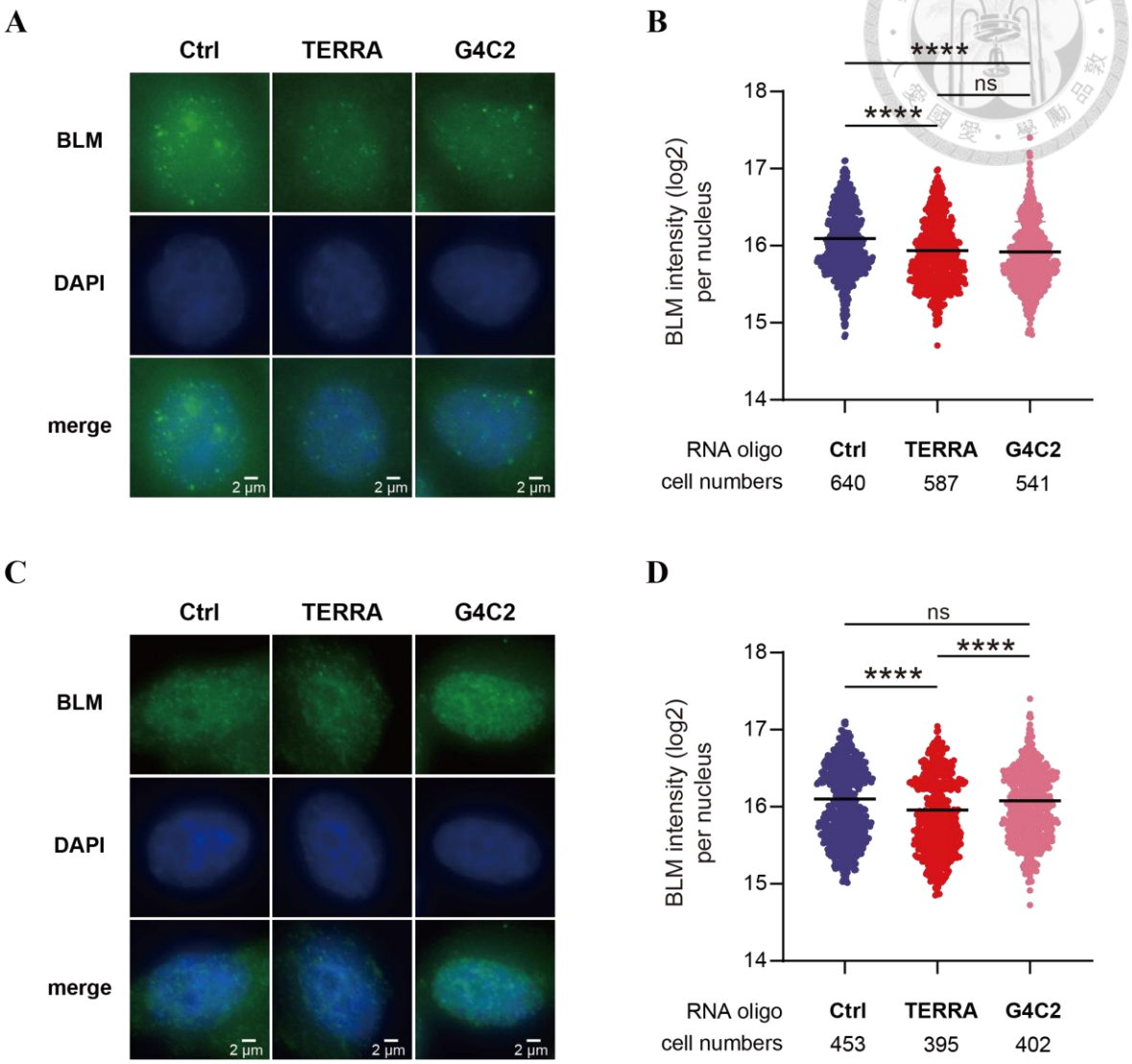
(A, C) Representative immunofluorescence microscopy images showing nuclear DNA G4 signals in HeLa cells, 36 hours post-transfection with the small RNA oligos (Ctrl RNA, TERRA RNA, and G4C2 RNA) (see Materials and Methods 2.4, 2.5). Prior to staining, cells were treated with RNase A to remove RNA. G4 structures were detected using the BG4 antibody. The FITC channel displays the G4 signal, and DAPI staining marks the nuclear regions. (B, D) Quantification of total nuclear G4 signal intensity from the immunofluorescence staining shown in (A) and (C), respectively, for Ctrl RNA, TERRA

doi:10.6342/NTU202504256

RNA, and G4C2 RNA transfection groups; at least 200 cells were analyzed per group. **** $p < 0.0001$ by Mann-Whitney U test.

doi:10.6342/NTU202504256

Supplementary figure S2.

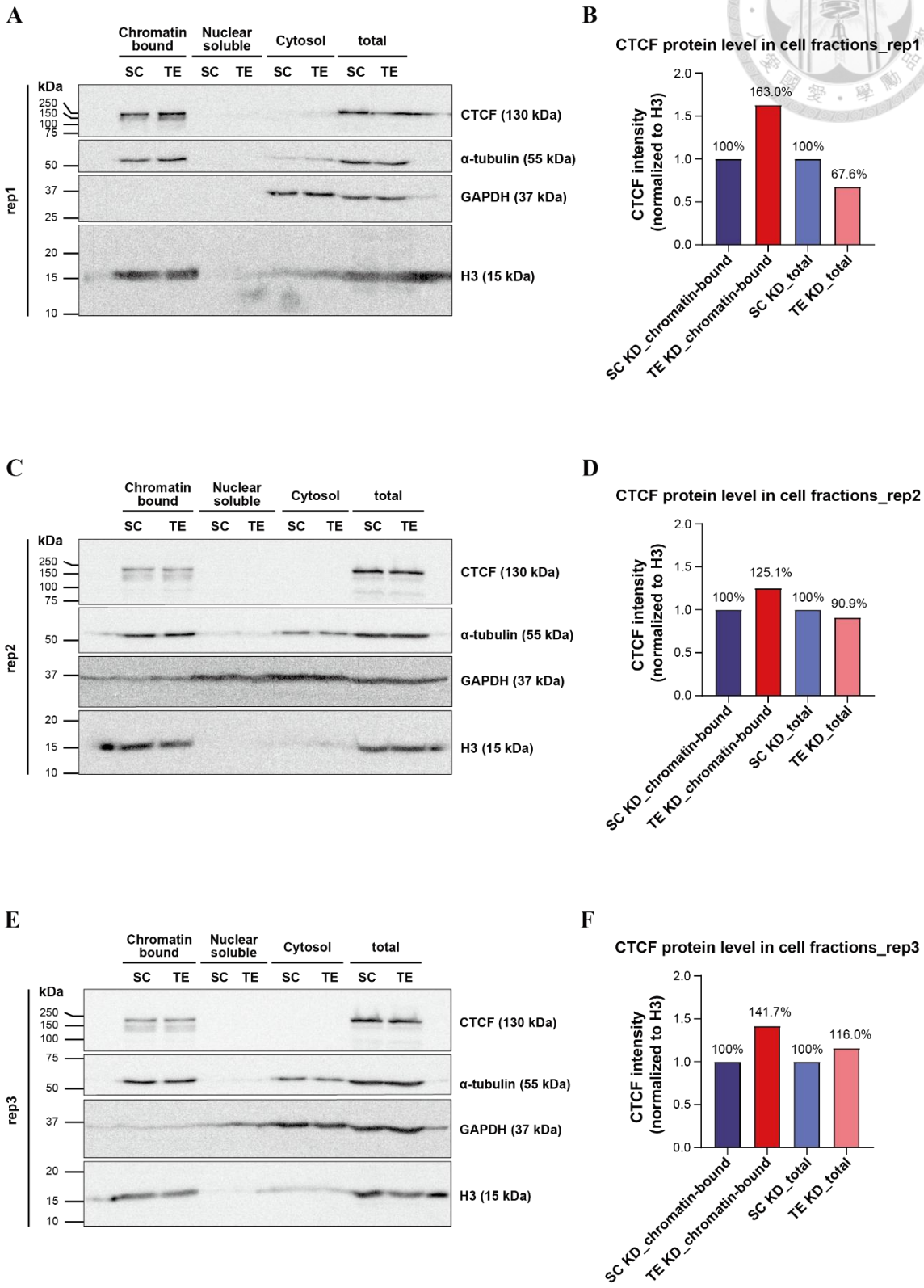


Supplementary figure S2. The other two replicates for BLM immunostaining related to Figure 2.

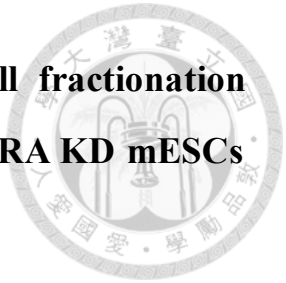
(A, C) Representative immunofluorescence microscopy images showing BLM signals in HeLa cells, 36 hours post-transfection with the small RNA oligos (Ctrl RNA, TERRA RNA, and G4C2 RNA) (see Materials and Methods 2.4, 2.5). The FITC channel displays the BLM signal, and DAPI staining marks the nuclear regions. (B, D) Quantification of total nuclear BLM signal intensity from the immunofluorescence staining shown in (A) and (C), respectively, for Ctrl RNA, TERRA RNA, and G4C2 RNA transfection groups; at least 200 cells were analyzed per group. **** $p < 0.0001$ by Mann-Whitney U test.

doi:10.6342/NTU202504256

Supplementary figure S3.



doi:10.6342/NTU202504256



Supplementary figure S3. The three replicates for cell fractionation followed by western blotting performed in SC and TERRA KD mESCs related to Figure 10.

(A, C, E) Results of western blot analysis for chromatin-bound, nuclear-soluble, cytosolic, and whole-cell lysate fractions from scramble (SC) and TERRA KD (TE) mESCs, across the three replicates (see Materials and Methods 2.16, 2.17). Protein levels of CTCF, α-tubulin, GAPDH, and histone H3 are shown. H3 and GAPDH were used as markers for chromatin-bound and cytosolic fractions, respectively. Expected molecular weights of each protein are as follows: CTCF, 130 kDa; α-tubulin, 55 kDa; GAPDH, 37 kDa; and H3, 15 kDa. (B, D, F) Quantification of CTCF band intensity in the chromatin-bound and whole-cell lysate fractions shown in (A), (C), and (E), respectively. CTCF signals were normalized to corresponding H3 signals, and values from the SC groups were set to 1.

doi:10.6342/NTU202504256

Supplementary Table S1. Small RNA oligos for transfecting in HeLa cells

RNA oligo	Sequence
Ctrl RNA	5’-Cy3-r(AACGUACGCGGAAUACUUCGAGUU)-3’
TERRA RNA	5’-Cy3-r(UUAGGGUUAGGGUUAGGGUUAGGG)-3’
G4C2 RNA	5’-Cy3-r(GGGGCCGGGGCCGGGGCCGGGGCC)-3’

Supplementary Table S2. LNA GapmeRs for TERRA knock down in mESCs

LNA	Sequence
Scramble ASO	5’-CACGTCTATACACCAC-3’
Antisense TERRA ASO	5’-TAACCCTAACCCTAAC-3’

Supplementary Table S3. Primers for qPCR

Primers for confirming TERRA knockdown efficiency		
Target	Forward (5’→3’)	Reverse (5’→3’)
GAPDH	CGTCCCGTAGACAAAATGGT	TTGATGGCAACAATCTCCAC
Erdr1	CACAGTGATGTCACCCACGA	GTGAGAATCGCTCCGTCCTG
mChrY	CCCTCTTCCCTTAGACACGC	CACACTCTCTGTCTGTGGACC

Primers for validating CTCF antibody used in Chromatin immunoprecipitation			
Target	Application	Forward (5’→3’)	Reverse (5’→3’)
h19	CTCF positive gene	GTCACTCAGGCAT AGCATTC	GTCTGCCGAGCAA TATGTAG
5nc	CTCF negative gene	TGTCCTGTCTGTCT CCCAGT	ATGTCCATGGCTC AGCAAGT

doi:10.6342/NTU202504256

Supplementary Table S4. Antibodies List

Antibodies for chromatin immunoprecipitation			
Target	Species	Source	Catalogue
IgG	Rabbit	Cell Signaling Technology	#2729
CTCF	Rabbit	Cell Signaling Technology	#2899

Primary and secondary antibodies and their dilution for Immunofluorescence				
Target	Species	Source	Catalogue	Dilution
BG4	Mouse	Absolute Antibody	#2899S	1:200
BLM	Mouse	Santa Cruz	#sc-365753	1:200
Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	Mouse	Invitrogen	#A-11029	1:500

Primary and secondary antibodies and their dilution for western blotting				
Target	Species	Source	Catalogue	Dilution
CTCF	Rabbit	Cell Signaling Technology	#2899S	1:1000
α -tubulin	Mouse	Cell Signaling Technology	#3873S	1:5000
GAPDH	Mouse	Cell Signaling Technology	#2118S	1:1000
H3	Rabbit	Abcam	#ab1761	1:7500
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, HRP	Mouse	Invitrogen	#A-11029	1:5000
Goat Anti-Rabbit IgG antibody (HRP)	Rabbit	Genetex	#GTX213110-01	1:5000

doi:10.6342/NTU202504256

Supplementary Table S5. Kits List

Reagent	Source	Catalogue	Application
Agilent High Sensitivity DNA Kit	Agilent Technologies	5067-4626	Quality control of ChIP-seq DNA library
FavorPrep™ GEL/PCR Purification Kit	FAVORGEN	FAGCK 001-1	DNA purification for ChIP
iQ™ SYBR® Green Supermix	Bio-rad	1708882	qPCR for DNA samples
Mouse Embryonic Stem Cell Nucleofector® Kit	Lonza	VPH-1001	TERRA KD in mESCs
NEBNext® ULtra™ II DNA Library Prep with Sample Purification Beads	New England Biolabs	E7103S	DNA library for ChIP
SuperScript™ IV Reverse Transcriptase	Thermo Fisher Scientific	18090200	RNA extraction for RT-qPCR
SuperSignal™ West Pico PLUS Chemiluminescent Substrate	Thermo Fisher Scientific	34580	ECL detection for western blot

Supplementary Table S6. Reagents List

Reagent	Source	Catalogue	Application
2-Mercaptoethanol	Thermo Fisher Scientific	21985023	Cell Culture
Acid-Phenol:Chloroform, pH 4.5 (with IAA, 125:24:1)	Thermo Fisher Scientific	AM9722-1QT	RNA extraction
Acrylamide/Bis Solution 29:1 (30%)	Bio-Rad	1610156	Western Blotting

doi:10.6342/NTU202504256

Adaptor for Illumina	New England Biolabs	E7337A	Library
Agarose	Bioman	PB1200	ChIP
APS	SERVA	17-1311-01	Western Blotting
Benzonase	Millipore	E1014-5KU	Cell Fractionation
Bromophenol Blue	Sigma-Aldrich	B8026	Western Blotting
Chloroform:Isoamyl alcohol 24:1	Sigma-Aldrich	C0549	RNA Extraction
DEPC	MERCK	SI-D5758-100ML	General
DMEM	Thermo Fisher Scientific	10569044	Cell Culture
DMSO	Scharlau	SU01551000	Cell Culture
dNTP Mix	Thermo Fisher Scientific	18427088	cDNA Synthesis
DPBS	Gibco™	14190235	Cell culture
Dynabeads™ Protein G for Immunoprecipitation	Thermo Fisher Scientific	10004D	ChIP
EDTA	AMRESCO	0105	ChIP, Cell Fractionation
EDTA (0.5M, pH 8.0)	盟基	IB3111	ChIP, Cell Fractionation
EGTA	Sigma-Aldrich	E3889	ChIP
Ethanol Dehydrate	Bioman	E23-500	Library
FBS	Thermo Fisher Scientific	26140079	Cell Culture
Formaldehyde (37%)	Sigma-Aldrich	F8775	ChIP
Gelatin	Sigma-Aldrich	G2500	Cell Culture
Glycerol	Bioshop	GLY001.500	Cell Fractionation, Western Blotting
Glycine	BioShop	GLN001.5	ChIP, Western Blotting
GlycoBlue™ Coprecipitant	Thermo Fisher Scientific	AM9516	RNA Extraction

doi:10.6342/NTU202504256

HEPES	Thermo Fisher Scientific	15630080	ChIP, Cell Fractionation
Isopropanol	J.T.Baker	9084-03	RNA Extraction
LIF	Merck	ESG1107	Cell Culture
Magnesium Chloride (MgCl)	Sigma-Aldrich	M0250	Cell Fractionation
MEM NEAA	Thermo Fisher Scientific	11140050	Cell Culture
Methyl Alcohol	Macron	3041-31	Western Blotting
Molecular Biology Grade Water (DNase and RNase-free)	Sartorius	01-869-1B	Library
MycoZap™ Plus-CL	Lonza	VZA-2011	Cell Culture
NEBNext Library Dilution Buffer	New England Biolabs	E7633A	Library
NEBNext Library Quant DNA Standard 2 (10 pM)	New England Biolabs	E6887AA	Library
NEBNext Library Quant DNA Standard 3 (1 pM)	New England Biolabs	E6888AA	Library
NEBNext Library Quant DNA Standard 4 (0.1 pM)	New England Biolabs	E6889AA	Library
NEBNext Library Quant DNA Standard 5 (0.01 pM)	New England Biolabs	E6890AA	Library
NEBNext Library Quant Master Mix	New England Biolabs	E7640AA	Library
NEBNext Library Quant Primer Mix	New England Biolabs	E7639AA	Library
NEBNext Ligation Enhancer	New England Biolabs	E7374AA	Library
NEBNext Sample Purification Beads	New England Biolabs	E7104L	Library

doi:10.6342/NTU202504256

NEBNext ULtra II End Prep Enzyme Mix	New England Biolabs	E7646AA	Library
NEBNext ULtra II End Prep Reaction Buffer	New England Biolabs	E7647AA	Library
NEBNext ULtra II Ligation Master Mix	New England Biolabs	E7648AA	Library
NEBNext ULtra II Q5 Master Mix	New England Biolabs	E7649AA	Library
NEBNext Universal PCR Primer for Illumina	New England Biolabs	S33102	Library
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1)	New England Biolabs	E7335L	Library
N-Lauroylsarcosine sodium salt	Sigma-Aldrich	L9150	ChIP
Nonfat Milk Powder	安佳		Western Blotting
NP-40	Bioman	NON505.100	ChIP, Cell Fractionation
Nuclease-Free Water (not DEPC-Treated)	Thermo Fisher Scientific	AM9938	Library
PBS	Bioman	PBS105000	General
PEN-STREP	Thermo Fisher Scientific	15140122	Cell Culture
PMSF	BioShop	PMS123	Cell Fractionation
Potassium chloride (KCl)	Sigma-Aldrich	P4504	Cell Fractionation
Precision Plus Protein Dual Color Standards	Bio-Rad	1610374	Western Blotting
Protease Inhibitor Cocktail (PIC)	Roche	04693132001	ChIP, Cell Fractionation
Proteinase K	Gproan Biotech	GPK003001	ChIP
PVDF membrane	Merck	IPVH85R	Western Blotting
QIAzol Lysis Reagent	Quagen	79306	RNA Extraction

doi:10.6342/NTU202504256

Random Primers	Thermo Fisher Scientific	48190011	cDNA Synthesis
RNase A	Bioshop	RNA888.100	ChIP
RNaseOUT™ Recombinant Ribonuclease Inhibitor	Thermo Fisher Scientific	10777019	cDNA Synthesis
Sodium Chloride (NaCl)	J.T.Baker	4058-01	ChIP, Cell Fractionation
Sodium Deoxycholate	Sigma-Aldrich	D6750	ChIP
Sodium Dodecyl Sulfate (SDS)	Bioshop	SDS-001	ChIP, Western Blotting
SuperScript™ IV Reverse Transcriptase	Thermo Fisher Scientific	18090200	cDNA Synthesis
Surcose	Bioshop	SUC507.5	Cell Fractionation
SYBR® Green I Nucleic Acid Gel Stain	Thermo Fisher Scientific	S7567	ChIP
SYBR™ Safe DNA Gel Stain	Thermo Fisher Scientific	S33102	Library
TAE	Bioman	TAE501000	ChIP
TBS (10x)	Omics.Bio	IB3052	Western Blotting
TE buffer	Sigma-Aldrich	93283	Library
TEMED	Bio-Rad	1610801	Western Blotting
TG-SDS (10x)	Bioman	TGS102000	Western Blotting
Tris	Bioman	TRS011.1	Western Blotting
Tris-HCl (1.0M, pH 6.8)	Bioman	TRS100568	Western Blotting
Tris-HCl (1.0M, pH 8.0)	Bioman	TRS100580	ChIP, Western Blotting
Triton X-100	Thermo Fisher Scientific	21568-0010	ChIP
Trypsin-EDTA (0.25%)	Thermo Fisher Scientific	25200072	Cell Culture
User Enzyme	New England Biolabs	E7428AA	Library

doi:10.6342/NTU202504256

Supplementary Table S7. NGS code list

Action	Tool_version	Parameter
Download data	sratoolkit_v3.0.3	prefetch --option-file [SRR_download_link] -O [output_path]
Transfer data into fastq format	sratoolkit_v3.0.3	fastq-dump [SRR_number.txt] -O [output_path] --split-files --gzip
Trim adaptor	TrimGalore_v0.6.3	<for paired reads> trim_galore --path_to_cutadapt [path_of_cutadapt_tool] --illumina --fastqc -q 30 --gzip --paired [sample_R1.fastq.gz] [sample_R2.fastq.gz] -o [output_path]
		<for single read> trim_galore --path_to_cutadapt [path_of_cutadapt_tool] --illumina --fastqc -q 30 --gzip [sample.fastq.gz] -o [output_path]
Filter reads containing telomeric repeats	BBTools_v37.62	<for paired reads> bbduk.sh in=[trimmed_R1_fq.gz] in2=[trimmed_R2_fq.gz] literal=TTAGGGTTAGGGTTAGGGTTAGGG outm=[output_TTAGGGx4_R1.fq] outm2=[output_TTAGGGx4_R2.fq] stats=[output_TTAGGGx4.stats.txt] k=21 hdist=2 edist=2 threads=14
		<for single reads> bbduk.sh in=[trimmed_R1_fq.gz] literal=TTAGGGTTAGGGTTAGGGTTAGGG outm=[output_TTAGGGx4_R1.fq] stats=[output_TTAGGGx4.stats.txt] k=21 hdist=2 edist=2 threads=14
	bowtie2_v2.4.2	<for paired reads>

doi:10.6342/NTU202504256

Remove rRNA and tRNA sequences		bowtie2 -p 14 -x [bw2_rRNAandtRNA_genome] -1 [sample_trimmed_R1.fq.gz] -2 [sample_trimmed_R2.fq.gz] --un-conc [output_name] -S [output_ rRNAandtRNA.sam]
		<for single read> bowtie2 -p 14 -x [bw2_rRNAandtRNA_genome] -U [sample_trimmed.fq.gz] --un-conc [output_name] -S [output_ rRNAandtRNA.sam]
Align sample reads onto reference genome		<for paired reads> bowtie2 -x [bw2_genome] -1 [sample_trimmed_R1.fq.gz] -2 [sample_trimmed_R2.fq.gz] -S [output.sam]
		<for single read> bowtie2 -x [bw2_genome] -U [sample_trimmed.fq.gz] -S [output.sam]
Align sample reads onto reference genome	STAR_v2.7.10a	<for paired reads> STAR --runThreadN 28 –genomeDir [STAR_genome] --runMode alignReads --readFilesCommand zcat --readFilesIn [input_trimmed_R1.fq.gz] [input_trimmed_R2.fq.gz] --outSAMprimaryFlag AllBestScore --outMultimapperOrder Random --outWigType wiggle --outSAMtype BAM SortedByCoordinate --outSAMattributes All --outFileNamePrefix [output_name]
		<for single reads> STAR --runThreadN 28 –genomeDir [STAR_genome] --runMode alignReads --

doi:10.6342/NTU202504256

		readFilesCommand zcat --readFilesIn [input_trimmed.fq.gz] --outSAMprimaryFlag AllBestScore --outMultimapperOrder Random --outWigType wiggle --outSAMtype BAM SortedByCoordinate --outSAMattributes All --outFileNamePrefix [output_name]
Count the reads number	Htseq_v0.13.5	htseq-count -f bam -m intersection-nonempty -i gene_id -s reverse -t exon [input_aligned.bam] [genome_index] > [output_HTcount.txt]
Convert SAM file to BAM format	SAMTOOLS_v1.13	samtools view --threads 14 -b [sample.sam] > [output.bam]
Sort the BAM file by genomic coordinates		samtools sort --threads 14 -O bam [output.bam] -o [output_sorted.bam] samtools flagstat --threads 14 [output_sorted.bam] > [output_sorted.flagstat.txt]
Remove PCR duplicates from the sorted BAM files		samtools rmdup [output_sorted.bam] [output_sorted.rmdup.bam] samtools flagstat --threads 14 [output_sorted.rmdup.bam] > [output_sorted.rmdup.flagstat.txt]
Index the deduplicated BAM file		samtools index -@ 14 [output_sorted.rmdup.bam] [output_sorted.rmdup.bam.bai]
Visualize for a bamfile and transfer into	deepTools_v3.3.1	bamCoverage --normalizeUsing CPM --binSize 20 --smoothLength 60 --centerReads --extendReads 150 -b [sample_sorted.rmdup.bam] -o

doi:10.6342/NTU202504256

bigwig format		[output_extend.bigwig] --outFileFormat bigwig
Calculate the difference between two bamfiles and transfer into bigwig format		bamCompare --operation subtract --bamfile1 [sample_ChIP_sorted.rmdup.bam] --bamfile2 [sample_input_sorted.rmdup.bam] --scaleFactorsMethod None --normalizeUsing CPM --smoothLength 250 --centerReads --outFileFormat bigwig -o [output_ChIP_subtract_input.bigwig]
Calculate the correlation between different ChIP data		multiBigwigSummary bins -b [samples_bigwig_files] --label [samples] -o [output_computeMatrix.npz] plotCorrelation -in [output_computeMatrix.npz] --corMethod pearson --removeOutliers --plotTitle [name_scatterplot] --whatToPlot scatterplot -o [output_scatterplot.pdf] plotCorrelation -in [output_computeMatrix.npz] --corMethod spearman --skipZeros --plotTitle [name_heatmap] --whatToPlot heatmap -o [output_heatmap.pdf] --colorMap 'bwr'
Compute the signal matrix over scaled genomic regions		computeMatrix scale-regions -S [samples_bigwig_files] -R [ref_bedfile] --skipZeros --regionBodyLength 5000 -b 3000 -a 3000 -p 14 -o [output_computematrix.gz]
Compute the signal matrix centered on		computeMatrix reference-point -S [samples_bigwig_files] -R [ref_bedfile] --skipZeros --referencePoint TSS -b 5000 -a 5000 -p 14 -o [output_computematrix.gz]

doi:10.6342/NTU202504256

a reference point		
Plot metagene profile		plotProfile -m [samples_computematrix.gz] -o [output_plotProfile.pdf] --perGroup
Plot heatmap		plotHeatmap -m [samples_computematrix.gz] -o [output_plotHeatmap.pdf] --outFileSortedRegions [output_plotHeatmap.bed] --colorMap 'bwr'
Call peaks	MACS_v2.2.9.1	<with input> macs2 callpeak -t [sample_CTCF_sorted.rmdup.bam] -c [sample_input_sorted.rmdup.bam] -f BAM -B -g mm -p 0.01 -n [output_name] --outdir [output_path] <without input> macs2 callpeak -t [sample_CTCF_sorted.rmdup.bam] -f BAM -B -g mm -p 0.01 -n [output_name] --outdir [output_path]
Intersect peaks between two bedfiles	BEDTOOLS_v2.31.1	<for common peaks> intersectBed -wa -a [sample_A.bed] -b [sample_B.bed] > [output_ABcommon.bed] <for unique peaks> intersectBed -v -a [sample_A.bed] -b [sample_B.bed] > [output_Aunique.bed]
Annotation	HOMER	annotatePeaks.pl [sample_peak.bed] mm10 -annStats [output_name.annstats] -go [output_name.go] -genomeOntology [output_name.genomeOntology] > [output_name_anno.txt]

doi:10.6342/NTU202504256

Chapter 7 References

[1] M. Gellert, M. N. Lipsett, and D. R. Davies, "Helix formation by guanylic acid," (in eng), *Proc Natl Acad Sci U S A,* vol. 48, no. 12, pp. 2013-8, Dec 15 1962, doi: 10.1073/pnas.48.12.2013.

[2] J. You *et al.*, "Effects of monovalent cations on folding kinetics of G-quadruplexes," (in eng), *Biosci Rep,* vol. 37, no. 4, Aug 31 2017, doi: 10.1042/bsr20170771.

[3] A. Joachimi, A. Benz, and J. S. Hartig, "A comparison of DNA and RNA quadruplex structures and stabilities," (in eng), *Bioorg Med Chem,* vol. 17, no. 19, pp. 6811-5, Oct 1 2009, doi: 10.1016/j.bmc.2009.08.043.

[4] D. H. Zhang, T. Fujimoto, S. Saxena, H. Q. Yu, D. Miyoshi, and N. Sugimoto, "Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors," (in eng), *Biochemistry,* vol. 49, no. 21, pp. 4554-63, Jun 1 2010, doi: 10.1021/bi1002822.

[5] R. Hänsel-Hertsch, J. Spiegel, G. Marsico, D. Tannahill, and S. Balasubramanian, "Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing," (in eng), *Nat Protoc,* vol. 13, no. 3, pp. 551-564, Mar 2018, doi: 10.1038/nprot.2017.150.

[6] G. Marsico *et al.*, "Whole genome experimental maps of DNA G-quadruplexes in multiple species," (in eng), *Nucleic Acids Res,* vol. 47, no. 8, pp. 3862-3874, May 7 2019, doi: 10.1093/nar/gkz179.

[7] M. Zhang *et al.*, "Mammalian CST averts replication failure by preventing G-quadruplex accumulation," (in eng), *Nucleic Acids Res,* vol. 47, no. 10, pp. 5243-5259, Jun 4 2019, doi: 10.1093/nar/gkz264.

[8] L. I. Jansson *et al.*, "Telomere DNA G-quadruplex folding within actively extending human telomerase," (in eng), *Proc Natl Acad Sci U S A,* vol. 116, no. 19, pp. 9350-9359, May 7 2019, doi: 10.1073/pnas.1814777116.

[9] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, and L. H. Hurley, "Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription," (in eng), *Proc Natl Acad Sci U S A,* vol. 99, no. 18, pp. 11593-8, Sep 3 2002, doi: 10.1073/pnas.182256799.

[10] J. A. Smestad and L. J. Maher, 3rd, "Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription," (in eng), *BMC Med Genet,* vol. 16, p. 91, Oct 8 2015, doi: 10.1186/s12881-015-0236-4.

[11] J. Spiegel, S. M. Cuesta, S. Adhikari, R. Hänsel-Hertsch, D. Tannahill, and S. Balasubramanian, "G-quadruplexes are transcription factor binding hubs in

doi:10.6342/NTU202504256

human chromatin," (in eng), *Genome Biol,* vol. 22, no. 1, p. 117, Apr 23 2021, doi: 10.1186/s13059-021-02324-z.

[12] J. Shen *et al.*, "Promoter G-quadruplex folding precedes transcription and is controlled by chromatin," (in eng), *Genome Biol,* vol. 22, no. 1, p. 143, May 7 2021, doi: 10.1186/s13059-021-02346-7.

[13] C. M. Azzalin, P. Reichenbach, L. Khoriauli, E. Giulotto, and J. Lingner, "Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends," (in eng), *Science,* vol. 318, no. 5851, pp. 798-801, Nov 2 2007, doi: 10.1126/science.1147182.

[14] S. Schoeftner and M. A. Blasco, "Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II," (in eng), *Nat Cell Biol,* vol. 10, no. 2, pp. 228-36, Feb 2008, doi: 10.1038/ncb1685.

[15] Y. Mei *et al.*, "TERRA G-quadruplex RNA interaction with TRF2 GAR domain is required for telomere integrity," (in eng), *Sci Rep,* vol. 11, no. 1, p. 3509, Feb 10 2021, doi: 10.1038/s41598-021-82406-x.

[16] M. Feretzaki, M. Pospisilova, R. Valador Fernandes, T. Lunardi, L. Krejci, and J. Lingner, "RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops," (in eng), *Nature,* vol. 587, no. 7833, pp. 303-308, Nov 2020, doi: 10.1038/s41586-020-2815-6.

[17] Z. Deng, J. Norseen, A. Wiedmer, H. Riethman, and P. M. Lieberman, "TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres," (in eng), *Mol Cell,* vol. 35, no. 4, pp. 403-13, Aug 28 2009, doi: 10.1016/j.molcel.2009.06.025.

[18] H. P. Chu *et al.*, "TERRA RNA Antagonizes ATRX and Protects Telomeres," (in eng), *Cell,* vol. 170, no. 1, pp. 86-101.e16, Jun 29 2017, doi: 10.1016/j.cell.2017.06.017.

[19] S. Redon, P. Reichenbach, and J. Lingner, "The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase," (in eng), *Nucleic Acids Res,* vol. 38, no. 17, pp. 5797-806, Sep 2010, doi: 10.1093/nar/gkq296.

[20] I.-T. Chiang, "TERRA acts as an epigenetic regulator to maintain DNA G-quadruplex formation in the genome," M.S. M.S., Institute of Molecular and Cellular Biology, National Taiwan University, Taipei, Taiwan, 2023.

[21] R. X. Tsai *et al.*, "TERRA regulates DNA G-quadruplex formation and ATRX recruitment to chromatin," (in eng), *Nucleic Acids Res,* vol. 50, no. 21, pp. 12217-12234, Nov 28 2022, doi: 10.1093/nar/gkac1114.

[22] M. DeJesus-Hernandez *et al.*, "Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS," (in

doi:10.6342/NTU202504256

eng), *Neuron,* vol. 72, no. 2, pp. 245-56, Oct 20 2011, doi: 10.1016/j.neuron.2011.09.011.

[23] A. E. Renton *et al.*, "A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD," (in eng), *Neuron,* vol. 72, no. 2, pp. 257-68, Oct 20 2011, doi: 10.1016/j.neuron.2011.09.010.

[24] P. E. Ash *et al.*, "Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS," (in eng), *Neuron,* vol. 77, no. 4, pp. 639-46, Feb 20 2013, doi: 10.1016/j.neuron.2013.02.004.

[25] F. Frottin, M. Pérez-Berlanga, F. U. Hartl, and M. S. Hipp, "Multiple pathways of toxicity induced by C9orf72 dipeptide repeat aggregates and G(4)C(2) RNA in a cellular model," (in eng), *Elife,* vol. 10, Jun 23 2021, doi: 10.7554/eLife.62718.

[26] B. Swinnen *et al.*, "A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism," (in eng), *Acta Neuropathol,* vol. 135, no. 3, pp. 427-443, Mar 2018, doi: 10.1007/s00401-017-1796-5.

[27] Y. B. Lee *et al.*, "Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic," (in eng), *Cell Rep,* vol. 5, no. 5, pp. 1178-86, Dec 12 2013, doi: 10.1016/j.celrep.2013.10.049.

[28] S. Rossi *et al.*, "Nuclear accumulation of mRNAs underlies G4C2-repeat-induced translational repression in a cellular model of C9orf72 ALS," (in eng), *J Cell Sci,* vol. 128, no. 9, pp. 1787-99, May 1 2015, doi: 10.1242/jcs.165332.

[29] Q. Wang, E. G. Conlon, J. L. Manley, and D. C. Rio, "Widespread intron retention impairs protein homeostasis in C9orf72 ALS brains," (in eng), *Genome Res,* vol. 30, no. 12, pp. 1705-1715, Dec 2020, doi: 10.1101/gr.265298.120.

[30] Z. F. Wang *et al.*, "The Hairpin Form of r(G(4)C(2))(exp) in c9ALS/FTD Is Repeat-Associated Non-ATG Translated and a Target for Bioactive Small Molecules," (in eng), *Cell Chem Biol,* vol. 26, no. 2, pp. 179-190.e12, Feb 21 2019, doi: 10.1016/j.chembiol.2018.10.018.

[31] P. Fratta *et al.*, "C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes," (in eng), *Sci Rep,* vol. 2, p. 1016, 2012, doi: 10.1038/srep01016.

[32] A. Ishiguro, J. Lu, D. Ozawa, Y. Nagai, and A. Ishihama, "ALS-linked FUS mutations dysregulate G-quadruplex-dependent liquid-liquid phase separation and liquid-to-solid transition," (in eng), *J Biol Chem,* vol. 297, no. 5, p. 101284, Nov 2021, doi: 10.1016/j.jbc.2021.101284.

[33] J. K. Karow, R. K. Chakraverty, and I. D. Hickson, "The Bloom's syndrome gene product is a 3'-5' DNA helicase," (in eng), *J Biol Chem,* vol. 272, no. 49, pp. 30611-4, Dec 5 1997, doi: 10.1074/jbc.272.49.30611.

doi:10.6342/NTU202504256

[34] S. Lee, A. R. Lee, K. S. Ryu, J. H. Lee, and C. J. Park, "NMR Investigation of the Interaction between the RecQ C-Terminal Domain of Human Bloom Syndrome Protein and G-Quadruplex DNA from the Human c-Myc Promoter," (in eng), *J Mol Biol,* vol. 431, no. 4, pp. 794-806, Feb 15 2019, doi: 10.1016/j.jmb.2019.01.010.

[35] J. B. Budhathoki, E. J. Stafford, J. G. Yodh, and H. Balci, "ATP-dependent G-quadruplex unfolding by Bloom helicase exhibits low processivity," (in eng), *Nucleic Acids Res,* vol. 43, no. 12, pp. 5961-70, Jul 13 2015, doi: 10.1093/nar/gkv531.

[36] R. Tippana, H. Hwang, P. L. Opresko, V. A. Bohr, and S. Myong, "Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases," (in eng), *Proc Natl Acad Sci U S A,* vol. 113, no. 30, pp. 8448-53, Jul 26 2016, doi: 10.1073/pnas.1603724113.

[37] H. Sun, J. K. Karow, I. D. Hickson, and N. Maizels, "The Bloom's syndrome helicase unwinds G4 DNA," (in eng), *J Biol Chem,* vol. 273, no. 42, pp. 27587-92, Oct 16 1998, doi: 10.1074/jbc.273.42.27587.

[38] W. C. Drosopoulos, S. T. Kosiyatrakul, and C. L. Schildkraut, "BLM helicase facilitates telomere replication during leading strand synthesis of telomeres," (in eng), *J Cell Biol,* vol. 210, no. 2, pp. 191-208, Jul 20 2015, doi: 10.1083/jcb.201410061.

[39] N. van Wietmarschen, S. Merzouk, N. Halsema, D. C. J. Spierings, V. Guryev, and P. M. Lansdorp, "BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes," (in eng), *Nat Commun,* vol. 9, no. 1, p. 271, Jan 18 2018, doi: 10.1038/s41467-017-02760-1.

[40] Y. M. Danino *et al.*, "BLM helicase protein negatively regulates stress granule formation through unwinding RNA G-quadruplex structures," (in eng), *Nucleic Acids Res,* vol. 51, no. 17, pp. 9369-9384, Sep 22 2023, doi: 10.1093/nar/gkad613.

[41] J. Yang *et al.*, "Structures of CTCF-DNA complexes including all 11 zinc fingers," (in eng), *Nucleic Acids Res,* vol. 51, no. 16, pp. 8447-8462, Sep 8 2023, doi: 10.1093/nar/gkad594.

[42] H. Nakahashi *et al.*, "A genome-wide map of CTCF multivalency redefines the CTCF code," (in eng), *Cell Rep,* vol. 3, no. 5, pp. 1678-1689, May 30 2013, doi: 10.1016/j.celrep.2013.04.024.

[43] A. C. Bell, A. G. West, and G. Felsenfeld, "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators," (in eng), *Cell,* vol. 98, no. 3, pp. 387-96, Aug 6 1999, doi: 10.1016/s0092-8674(00)81967-4.

doi:10.6342/NTU202504256

[44] J. R. Dixon *et al.*, "Topological domains in mammalian genomes identified by analysis of chromatin interactions," (in eng), *Nature,* vol. 485, no. 7398, pp. 376-80, Apr 11 2012, doi: 10.1038/nature11082.

[45] S. S. Rao *et al.*, "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping," (in eng), *Cell,* vol. 159, no. 7, pp. 1665-80, Dec 18 2014, doi: 10.1016/j.cell.2014.11.021.

[46] E. de Wit *et al.*, "CTCF Binding Polarity Determines Chromatin Looping," (in eng), *Mol Cell,* vol. 60, no. 4, pp. 676-84, Nov 19 2015, doi: 10.1016/j.molcel.2015.09.023.

[47] G. Wutz *et al.*, "Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins," (in eng), *Embo j,* vol. 36, no. 24, pp. 3573-3599, Dec 15 2017, doi: 10.15252/embj.201798004.

[48] I. F. Davidson *et al.*, "CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion," (in eng), *Nature,* vol. 616, no. 7958, pp. 822-827, Apr 2023, doi: 10.1038/s41586-023-05961-5.

[49] B. Xu *et al.*, "Acute depletion of CTCF rewires genome-wide chromatin accessibility," (in eng), *Genome Biol,* vol. 22, no. 1, p. 244, Aug 24 2021, doi: 10.1186/s13059-021-02466-0.

[50] P. T. Li *et al.*, "Expression of the human telomerase reverse transcriptase gene is modulated by quadruplex formation in its first exon due to DNA methylation," (in eng), *J Biol Chem,* vol. 292, no. 51, pp. 20859-20870, Dec 22 2017, doi: 10.1074/jbc.M117.808022.

[51] P. Wulfridge *et al.*, "G-quadruplexes associated with R-loops promote CTCF binding," (in eng), *Mol Cell,* vol. 83, no. 17, pp. 3064-3079.e5, Sep 7 2023, doi: 10.1016/j.molcel.2023.07.009.

[52] R. Saldaña-Meyer *et al.*, "CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53," (in eng), *Genes Dev,* vol. 28, no. 7, pp. 723-34, Apr 1 2014, doi: 10.1101/gad.236869.113.

[53] H. J. Oh *et al.*, "Jpx RNA regulates CTCF anchor site selection and formation of chromosome loops," (in eng), *Cell,* vol. 184, no. 25, pp. 6157-6173.e24, Dec 9 2021, doi: 10.1016/j.cell.2021.11.012.

[54] S. Kuang and L. Wang, "Deep Learning of Sequence Patterns for CCCTC-Binding Factor-Mediated Chromatin Loop Formation," (in eng), *J Comput Biol,* vol. 28, no. 2, pp. 133-145, Feb 2021, doi: 10.1089/cmb.2020.0225.

[55] C. Y. Wang, T. Jégu, H. P. Chu, H. J. Oh, and J. T. Lee, "SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the

doi:10.6342/NTU202504256

Inactive X," (in eng), *Cell,* vol. 174, no. 2, pp. 406-421.e25, Jul 12 2018, doi: 10.1016/j.cell.2018.05.007.

[56] G. Nagy *et al.*, "Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA," (in eng), *BMC Genomics,* vol. 17, no. 1, p. 637, Aug 15 2016, doi: 10.1186/s12864-016-2940-7.

[57] J. T. Kung *et al.*, "Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF," (in eng), *Mol Cell,* vol. 57, no. 2, pp. 361-75, Jan 22 2015, doi: 10.1016/j.molcel.2014.12.006.

[58] K. Hirashima and H. Seimiya, "Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo," (in eng), *Nucleic Acids Res,* vol. 43, no. 4, pp. 2022-32, Feb 27 2015, doi: 10.1093/nar/gkv063.

[59] M. M. Fay, P. J. Anderson, and P. Ivanov, "ALS/FTD-Associated C9ORF72 Repeat RNA Promotes Phase Transitions In Vitro and in Cells," (in eng), *Cell Rep,* vol. 21, no. 12, pp. 3573-3584, Dec 19 2017, doi: 10.1016/j.celrep.2017.11.093.

[60] R. N. Plasschaert *et al.*, "CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation," (in eng), *Nucleic Acids Res,* vol. 42, no. 2, pp. 774-89, Jan 2014, doi: 10.1093/nar/gkt910.

[61] A. Barral *et al.*, "SETDB1/NSD-dependent H3K9me3/H3K36me3 dual heterochromatin maintains gene expression profiles by bookmarking poised enhancers," (in eng), *Mol Cell,* vol. 82, no. 4, pp. 816-832.e12, Feb 17 2022, doi: 10.1016/j.molcel.2021.12.037.

[62] M. H. Kagey *et al.*, "Mediator and cohesin connect gene expression and chromatin architecture," (in eng), *Nature,* vol. 467, no. 7314, pp. 430-5, Sep 23 2010, doi: 10.1038/nature09380.

[63] A. Del Moral-Morales, M. Salgado-Albarrán, Y. Sánchez-Pérez, N. K. Wenke, J. Baumbach, and E. Soto-Reyes, "CTCF and Its Multi-Partner Network for Chromatin Regulation," (in eng), *Cells,* vol. 12, no. 10, May 10 2023, doi: 10.3390/cells12101357.

[64] D. Szklarczyk *et al.*, "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets," (in eng), *Nucleic Acids Res,* vol. 49, no. D1, pp. D605-d612, Jan 8 2021, doi: 10.1093/nar/gkaa1074.

[65] M. Kotlyar, C. Pastrello, Z. Ahmed, J. Chee, Z. Varyova, and I. Jurisica, "IID 2021: towards context-specific protein interaction analyses by increased coverage, enhanced annotation and enrichment analysis," (in eng), *Nucleic Acids Res,* vol. 50, no. D1, pp. D640-d647, Jan 7 2022, doi: 10.1093/nar/gkab1034.

doi:10.6342/NTU202504256

[66] G. Hu, X. Dong, S. Gong, Y. Song, A. P. Hutchins, and H. Yao, "Systematic screening of CTCF binding partners identifies that BHLHE40 regulates CTCF genome-wide distribution and long-range chromatin interactions," (in eng), *Nucleic Acids Res,* vol. 48, no. 17, pp. 9606-9620, Sep 25 2020, doi: 10.1093/nar/gkaa705.

[67] S. C. Hsu *et al.*, "The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries," (in eng), *Mol Cell,* vol. 66, no. 1, pp. 102-116.e7, Apr 6 2017, doi: 10.1016/j.molcel.2017.02.027.

[68] K. Ishihara, M. Oshimura, and M. Nakao, "CTCF-dependent chromatin insulator is linked to epigenetic remodeling," (in eng), *Mol Cell,* vol. 23, no. 5, pp. 733-42, Sep 1 2006, doi: 10.1016/j.molcel.2006.08.008.

[69] G. Ren *et al.*, "Acute depletion of BRG1 reveals its primary function as an activator of transcription," (in eng), *Nat Commun,* vol. 15, no. 1, p. 4561, May 29 2024, doi: 10.1038/s41467-024-48911-z.

[70] C. Hoencamp and B. D. Rowland, "Genome control by SMC complexes," (in eng), *Nat Rev Mol Cell Biol,* vol. 24, no. 9, pp. 633-650, Sep 2023, doi: 10.1038/s41580-023-00609-8.

[71] J. Minderjahn *et al.*, "Postmitotic differentiation of human monocytes requires cohesin-structured chromatin," (in eng), *Nat Commun,* vol. 13, no. 1, p. 4301, Jul 25 2022, doi: 10.1038/s41467-022-31892-2.

doi:10.6342/NTU202504256